

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE LETRAS

JÚLIA LUIZ SOSTRUZNIK DA SILVEIRA

**O USO DE CONJUNÇÕES EM PRODUÇÕES ESCRITAS NO EXAME
CELPE-BRAS: UM ESTUDO BASEADO EM CORPUS**

PORTO ALEGRE

2023

JÚLIA LUIZ SOSTRUZNIK DA SILVEIRA

**O USO DE CONJUNÇÕES EM PRODUÇÕES ESCRITAS NO EXAME
CELPE-BRAS: UM ESTUDO BASEADO EM CORPUS**

Trabalho de Conclusão de Curso de graduação apresentado ao Instituto de Letras da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Licenciada em Letras

Orientadora: Profa. Dra. Juliana Roquete Schoffen

Porto Alegre

2023

JÚLIA LUIZ SOSTRUZNIK DA SILVEIRA

**O USO DE CONJUNÇÕES EM PRODUÇÕES ESCRITAS NO EXAME
CELPE-BRAS: UM ESTUDO BASEADO EM CORPUS**

Trabalho de Conclusão de Curso de graduação apresentado ao Instituto de Letras da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Licenciada em Letras

Porto Alegre, 11 de setembro de 2023.

BANCA EXAMINADORA

Orientadora: Prof^a Dra. Juliana Roquele Schoffen
Universidade Federal do Rio Grande do Sul (UFRGS)

Marine Matte
PPGLetras/Universidade Federal do Rio Grande do Sul (UFRGS)

Prof^a. Dra. Elisa Marchioro Stumpf
Universidade Federal do Rio Grande do Sul (UFRGS)

AGRADECIMENTOS

À minha família que foi responsável por todo suporte emocional, familiar, educacional e profissional para que eu pudesse concluir esta etapa de ensino. Além de me ajudar, todos os dias, a construir minha trajetória.

À minha mãe que sempre acreditou na educação e me forneceu conselhos durante os desafios que me deparei logo nas primeiras entradas em sala de aulas. Ao meu pai, por todas as caronas e dicas para simplificar problemas enormes que eu mesma criei na minha cabeça.

Agradeço à Prof. Dra Juliana Schoffen, que é minha orientadora desde 2020, por acompanhar toda minha trajetória acadêmica, e, principalmente, por acreditar em mim. Obrigada por toda dedicação e conhecimento. Obrigada por ensinar a lidar com os futuros percalços da educação e também que estudar texto é estudar os gêneros discursivos.

Às minhas companheiras de graduação e parceiras de estágio, Helen e Júlia, por compartilharem comigo tantos momentos importantes da graduação, sendo eles enfrentar a papelada da UFRGS e despertar a criatividade para fazer os projetos de ensino mais incríveis possível.

Às minhas amigas as quais a UFRGS me presenteou. Obrigada por encararem a graduação ao meu lado e toparem todos os seminários intermináveis. A todos os amigos que sempre estiveram ao meu lado. À Júlia Santana por toda parceria e calma, por causa de todo pensamento positivo e “vai dar certo” que estamos nos formando juntas. À Helen, por todas as risadas e trabalhos de última hora que fizemos durante a graduação. À Manoela, por compartilhar as melhores referências de Orgulho e Preconceito que eu poderia pedir.

Ao Grupo Avalia, por ter ajudado a construir minha trajetória acadêmica com o singelo objetivo de dominar o mundo. Um agradecimento especial às colegas Isadora e Luiza que sempre estão presentes para publicar um trabalho à luz da Linguística de *Corpus*.

Às professoras Ana Bocorny, Elisa Marchioro Stumpf e Letícia Grubert dos Santos por todos os conselhos e materiais que ajudaram na composição dos cursos do CLA.

Aos professores do Instituto de Letras da UFRGS, pela dedicação em apresentar o clássico e o contemporâneo, e também, por contribuir para a formação de futuros professores que consigam transmitir gentileza e cuidado com os alunos.

À banca examinadora deste trabalho, composta por Elisa Marchioro Stumpf e Marine Matte. Obrigada por aceitarem o convite para avaliar e contribuir com a esta pesquisa.

À Universidade Federal do Rio Grande do Sul, por todas as oportunidades. Pela educação pública, gratuita e de qualidade.

RESUMO

Este trabalho tem por objetivo analisar o uso de conjunções em produções escritas do exame de Certificado de Proficiência em Língua Portuguesa (Celpe-Bras). Cushing (2022) destaca que pesquisas utilizando textos produzidos por falantes de segunda língua à luz da Linguística de Corpus permitem contribuir com a diferenciação e descrição de diferentes níveis de proficiência. Acrescenta-se também que estudos com *corpora* podem ter o potencial de aumentar a transparência e a consistência na avaliação de proficiência (Callies; Götz, 2015). Nesse contexto, o presente estudo visa explorar e examinar as conjunções utilizadas em produções-resposta às tarefas IV das edições de 2016/2 e 2017/1 do Celpe-Bras, avaliadas com as notas 5 e 2 (correspondente aos níveis Avançado Superior e Intermediário), utilizando o programa *Sketch Engine* (Kilgarrieff et al, 2004). Os *corpora* foram analisados usando as ferramentas disponíveis no *software*, como *Wordlist*, que permite escolher a classe gramatical desejada para as extrações, e também *Concordance*, que possibilita a leitura de trechos das produções textuais. Além disso, este estudo também analisou as ocorrências das conjunções de acordo com as categorias propostas por Azeredo (2011) - aditivas, adversativas, explicativas e conclusivas. Os resultados mostram que os textos avaliados como Avançado Superior (nota 5) utilizam um número maior de conjunções em relação aos avaliados como Intermediário (nota 2). Neste trabalho, foi também possível contabilizar as posições de uso das conjunções no interior das sentenças. Os resultados apresentados podem colaborar para o aumento de pesquisas sobre o Exame Celpe-Bras e de informações relacionadas aos níveis de proficiência certificados pelo exame. Além disso, o estudo tem potencial de incentivar o desenvolvimento de atividades pedagógicas que visem a auxiliar a prática escrita no ensino de PLA.

Palavras-Chaves: Conjunções; Linguística de Corpus; Celpe-Bras; Português como língua adicional

ABSTRACT

This study aims to analyze the use of conjunctions in written productions from the Celpe-Bras exam, the Certificate of Proficiency in the Portuguese Language. Cushing (2022) highlights that research using texts produced by second language speakers, in the context of Corpus Linguistics, can contribute to differentiating and describing various levels of proficiency. Additionally, it is noteworthy that corpus studies have the potential to enhance transparency and consistency in proficiency assessment (Callies; Götz, 2015). In this context, the present study seeks to explore and examine the conjunctions used in response productions to Task IV in the 2016/2 and 2017/1 editions of the Celpe-Bras exam, which were evaluated with scores of 5 (corresponding to Advanced Superior) and 2 (corresponding to Intermediate). To conduct this analysis, we utilized the Sketch Engine program (Kilgarriff et al., 2004). This software provides us with tools such as Wordlist, which allows the selection of the desired part of speech for extractions, and Concordance, which enables the examination of excerpts from the textual compositions. Furthermore, this study also categorized conjunction occurrences according to the classifications proposed by Azeredo (2011) - additive, adversative, explanatory, and conclusive. The results indicate that texts evaluated as Advanced Superior (score 5) employ a greater number of conjunctions compared to those assessed as Intermediate (score 2). Additionally, this research counted the positions of conjunctions within sentences. The results presented have the potential to contribute to an increased body of research on the Celpe-Bras Exam and provide valuable insights into the proficiency levels certified by the exam. Moreover, this study may inspire the development of pedagogical activities aimed at enhancing written practice in the teaching of Portuguese as an Additional Language (PLA).

Keywords: Conjunctions; Corpus Linguistics; Celpe-Bras; Portuguese as an Additional Language.

LISTA DE FIGURAS

Figura 1 - Número de Examinandos Anuais do Exame Celpe-Bras.....	14
Figura 2 - Material de insumo presente na Tarefa IV da edição 2016/2.....	32
Figura 3 - Enunciado e material de insumo presente na Tarefa IV da edição 2017/134	
Figura 4 - Dashboard do programa Sketch Engine.....	37
Figura 5 - Ferramenta Wordlist.....	37
Figura 6 - Calculadora de log-likelihood.....	39
Figura 7 - Ferramenta concordance a partir das extração da conjunção “e” na Tarefa IV da edição 2016/2 - Nota 5.....	40
Figura 8 - Acesso ao texto com a ferramenta concordance a partir das extração da conjunção “e” na Tarefa IV da edição 2016/2 - Nota 5.....	60

LISTA DE QUADROS

Quadro 1 - Certificações do Celpe-Bras.....	16
Quadro 2 - Tarefas da Parte Escrita.....	17
Quadro 3 - Conjunções Coordenativas.....	26
Quadro 4 - Conjunções subordinadas com exemplos.....	27
Quadro 6 - Corpora - Quantidade de textos analisados.....	35
Quadro 7 - Número de palavras por nota e tarefa.....	41
Quadro 8 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 5 da edição 2016/2.....	42
Quadro 9 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 2 da edição 2016/2.....	43
Quadro 10 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 5 da edição 2017/1.....	45
Quadro 11 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 2 na edição 2017/1.....	46
Quadro 12 - Total de conjunções presentes nos corpora.....	47
Quadro 13 - Comparação de frequências entre as conjunções presentes nas notas 5 e 2 da Tarefa IV da edição 2016/2.....	48
Quadro 14 - Comparação de frequências entre as conjunções presentes nas notas 5 e 2 da Tarefa IV da edição 2017/1.....	50
Quadro 15 - Comparação de frequências entre as notas 5 das edições 2016/2 e 2017/1.....	51
Quadro 16 - Comparação de frequências entre as notas 2 das edições 2016/2 e 2017/1.....	52
Quadro 17 - Categorização das conjunções - 2016/2.....	54
Quadro 18 - Categorização das conjunções - 2017/1.....	55
Quadro 19 - Tipos de conjunções.....	56
Quadro 20 - Comparativo do corpus 2016/2 T4, utilizando o teste estatístico do Log-Likelihood.....	58
Quadro 21 - Comparativo do corpus 2017/1 T4, utilizando o teste estatístico do	

Log-Likelihood.....	59
Quadro 22 - Comparando as posições das conjunções presentes na edição 2016/2..	
61	
Quadro 23 - Comparando as posições das conjunções presentes na edição 2017/1...	
62	

SUMÁRIO

1 INTRODUÇÃO.....	11
2 O EXAME CELPE-BRAS.....	13
2.1 CARACTERÍSTICAS DO EXAME.....	13
2.2 CONCEITO DE PROFICIÊNCIA.....	15
2.3 PARTE ORAL.....	16
2.4 PARTE ESCRITA.....	17
2.4.1 A AVALIAÇÃO DA PARTE ESCRITA.....	18
3 FUNDAMENTAÇÕES TEÓRICAS.....	19
3.1 LINGUÍSTICA DE CORPUS.....	20
3.1.1 ESTUDOS SOBRE O EXAME CELPE-BRAS QUE ANALISAM CORPUS.....	21
3.2 CONJUNÇÕES.....	25
3.2.1 PESQUISAS SOBRE CONJUNÇÕES EM PRODUÇÕES TEXTUAIS. 28	
4 METODOLOGIA.....	29
4.1 OBJETIVOS.....	29
4.2 CORPORA DE ESTUDO.....	30
4.2.1 AS TAREFAS SELECIONADAS.....	30
4.2.1.1 Tarefa IV da Edição 2016/2.....	31
4.2.1.2 Tarefa IV da Edição 2017/1.....	33
4.2.2 PRODUÇÕES ESCRITAS REALIZADAS PELOS EXAMINANDOS.....	35
4.3 O SOFTWARE SKETCH ENGINE.....	36
4.4 PROCEDIMENTOS PARA AS ANÁLISES.....	38
4.4.1 WORDLIST.....	38
4.4.2 LOG-LIKELIHOOD.....	38
4.4.3 CONCORDANCE.....	39
5 ANÁLISE DOS DADOS.....	40
5.1 CARACTERÍSTICAS DOS CORPORA.....	40

5.2 WORDLIST.....	41
5.2.1 CONJUNÇÕES.....	42
5.2.1.1 2016/2 Tarefa IV.....	42
5.2.1.2 2017/1 Tarefa IV.....	44
5.2.1.3 Comparando as notas.....	48
5.2.1.4 As categorias das conjunções.....	54
5.3 LOG-LIKELIHOOD.....	57
5.4 POSIÇÃO DAS CONJUNÇÕES.....	60
6 DISCUSSÃO DOS RESULTADOS.....	63
7 CONSIDERAÇÕES FINAIS.....	66
REFERÊNCIAS BIBLIOGRÁFICAS.....	68
APÊNDICE.....	73
APÊNDICE A – Quadro comparativo geral das edições.....	73

1 INTRODUÇÃO

O presente trabalho é fruto de mais de três anos de contato com as produções-resposta do exame Celpe-Bras, atrelado às leituras e às discussões propostas pelo Grupo Avalia. Após ingressar como bolsista de iniciação científica, sob coordenação da Prof. Dra. Juliana Schoffen, tive a oportunidade de digitar, revisar e analisar os textos produzidos pelos examinandos, além de ter contato com a área de avaliação. Dessa maneira, tive a oportunidade de me familiarizar com a área de avaliação e o ensino de Português como Língua Adicional (PLA), além de conhecer melhor o Exame, bem como o conceito de proficiência adotado, isto é, “a capacidade do aprendiz de usar adequadamente a língua para desempenhar ações no mundo” (Brasil, 2020a).

Uma das tarefas que realizei como bolsista de Iniciação Científica foi auxiliar na composição e na construção de um *corpus* de textos de examinandos do Celpe-Bras com cerca de 70 mil textos, atualmente ainda em processo de digitação, revisão e etiquetamento. No entanto, apesar do processo estar em andamento, já há diversos estudos que visam a caracterização dos níveis do Celpe-Bras utilizando esses dados. À luz disso, alguns dos estudos desenvolvidos foram qualitativos, analisando pequenas quantidades de textos, porém obtendo resultados significativos (Mendel, 2019; Sirianni, 2020). Após, pesquisas utilizando softwares de Linguística de Corpus começaram a ser desenvolvidas (Kunrath, 2019; Divino, 2021; Hanauer, 2021; 2022; 2023; Silveira, 2021; 2022) e, a partir delas, abriram-se portas ao possibilitar análises comparativas entre produções textuais de diferentes notas, utilizando metodologia de Linguística de Corpus.

Pesquisas que envolvam a língua portuguesa como língua adicional são muito importantes para a desenvolver esta área de ensino. Em Quevedo-Camargo e Scaramucci (2018), é destacado a necessidade de uma maior discussão que permita divulgar o ensino de PLA através de pesquisas disponíveis. Outrossim, é essencial que professores da área de língua adicional conheçam as avaliações e as possibilidades de trabalho que os resultados do exame Celpe-Bras permitem. Além disso, pesquisas podem contribuir com esclarecimentos sobre a elaboração e a avaliação do exame, capacitando os profissionais elaboradores a partir de maior contato com o construto do Celpe-Bras, além de detalhar os níveis de proficiência para chegar em uma melhor compreensão de como eles operam nas produções

escritas.

Nesse sentido, o objetivo deste trabalho é contribuir com informações sobre produções de PLA no contexto do Celpe-Bras, além de fornecer características sobre a capacidade linguística de diferentes níveis de proficiência. Para isso, nesta pesquisa, foram analisadas as características dos textos, a frequência no uso de conjunções, as categorias de conjunções presentes nos textos e o seu contexto de uso dentro das frases. Esses procedimentos permitem conhecer aspectos das produções textuais, e podem trazer contribuições pedagógicas a respeito de uso de conjunções, reforçando o efeito retroativo do exame Celpe-Bras.

Este trabalho foi dividido em sete capítulos, incluindo a introdução. O segundo capítulo aborda informações sobre o exame Celpe-Bras, como: construção da prova, o conceito de proficiência, a estrutura da aplicação e a avaliação da Parte Escrita. Em seguida, foram exploradas as fundamentações teóricas usadas para compor esta pesquisa, como conceitos da Linguística de Corpus e de estudos previamente realizados.

A metodologia também foi descrita, especialmente o conjunto de ferramentas utilizadas a partir do programa *Sketch Engine* (Kilgariff et al., 2004) e a também a calculadora *Log-Likelihood* que permite notar a significância dos dados. Ainda, o capítulo quatro traz os objetivos, as descrições sobre as tarefas utilizadas nos *corpora* e os procedimentos necessários para as extrações. No quinto capítulo, temos os resultados, e, no sexto, a discussão destes. Para concluir, são apresentadas as considerações finais e as referências bibliográficas.

2 O EXAME CELPE-BRAS

O Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras), inicialmente desenvolvido em 1993 pelo Ministério da Educação, foi elaborado para responder a uma demanda crescente de intercâmbios econômicos, culturais e científicos no Brasil, principalmente referentes a uma procura por cursos de graduação e pós-graduação (Schlatter et al 2009). A elaboração do exame por uma Comissão de professores da área de Português como Língua Adicional visava causar um efeito direcionador no ensino de PLA (Schlatter et al 2009).

2.1 CARACTERÍSTICAS DO EXAME

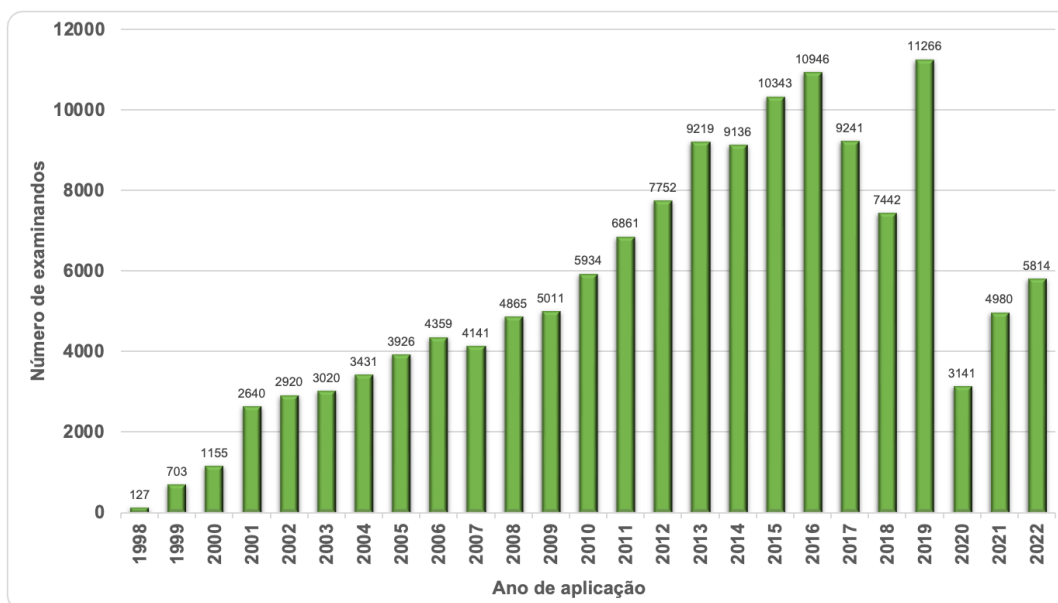
O exame, aplicado pela primeira vez em 1998, conta com duas edições anuais. Conforme Schlatter et. al (2009), o Celpe-Bras tem como principais públicos-alvos os indivíduos que não possuem a língua portuguesa como primeira língua e buscam ingressar em uma universidade brasileira, incluindo os estudantes contemplados pelos programas governamentais PEC-G e PEC-PG¹; profissionais estrangeiros que necessitam validar seus diplomas obtidos no exterior; indivíduos empregados em empresas multinacionais que precisam comprovar sua proficiência na língua portuguesa e também para fins de naturalização, de acordo com as diretrizes apresentadas na Portaria nº 623, de 13 de novembro de 2020 (Brasil, 2020b).

O exame Celpe-Bras tem um impacto significativo em várias áreas sociais, como no ensino, no currículo e na preparação de candidatos, conforme destacado por Schlatter et al (2005). Assim, de acordo Scaramucci e Quitzau (2018, p. 71), informações sobre “avaliação e a construção de ensino de PLA são impactadas também pelo teste” e é possível destacar que esse exame influenciou as atitudes das pessoas envolvidas e a elaboração de materiais didáticos para a área de PLA. Devido a esses impactos, o exame Celpe-Bras é considerado uma avaliação de alta relevância.

¹ Os Programas de Estudantes-Convênio de Graduação e Pós-Graduação (PEC-G e PEC-PG) oferecem oportunidades de formação superior a cidadãos de países em desenvolvimento com os quais o Brasil mantém acordos educacionais e culturais. (Brasil, 2020b.). Informações em: <https://www.gov.br/mre/pt-br/assuntos/cultura-e-educacao/temas-educacionais/programas-de-estudo-para-estrangeiros/pec-pg-pos-graduacao>

Na Figura 1, temos o número de examinandos atendidos anualmente pelo Celpe-Bras, o que é uma prova da crescente relevância social do exame.

Figura 1 - Número de Examinandos Anuais do Exame Celpe-Bras



Fonte: Adaptado de Schlatter et al (2009) e Damazo (2012) por Ellen Yurika Nagasawa e Isadora Dahmer Hanauer. FONTE: MEC e INEP. Disponível em <https://www.ufrgs.br/acervocelpebras/dados-celpe-bras/>

O crescente número de inscritos no exame exemplifica a importância da aplicação de duas edições anuais² e a necessidade da realização de mais edições. Além disso, com o aumento de possibilidades de uso da certificação, houve um crescimento no número de Postos Aplicadores cadastrados. Enquanto na primeira aplicação do exame, em 1998, 127 examinandos realizaram a prova em 8 Postos Aplicadores, a edição de Abril de 2023 contou com mais de 3.100 examinandos e mais de 130 postos aplicadores credenciados, no Brasil e no exterior.

Durante sua elaboração, o Celpe-Bras foi proposto para avaliar níveis de proficiência por meio de interações sociais. O exame tem como ideia avaliar as habilidades linguísticas nas produções feitas pelos examinandos, assim o Celpe-Bras é classificado como um exame de desempenho. Testes assim têm como características solicitar uso da língua, por meio de tarefas que se assemelham ao uso real e social da linguagem (Schoffen, 2009).

² No ano de 2020, 2021 e 2022, o exame Celpe-Bras teve somente uma aplicação por ano por motivos de restrições sanitárias relacionadas à pandemia de COVID-19.

2.2 CONCEITO DE PROFICIÊNCIA

No âmbito do exame Celpe-Bras, o conceito de proficiência é muito importante para compreender a funcionalidade do teste. Assim, no Documento-Base do exame, afirma-se: “Proficiência, nesse caso, implica ser capaz de engajar-se em diferentes situações de uso da língua portuguesa no mundo, mostrando adequação às demandas dos vários contextos.” (Brasil, 2020b, p. 30). Com isso, essa visão de proficiência leva em conta a prática de uso da língua, de forma situada e contextualizada, e não somente a comprovação de formas gramaticais. Em Scaramucci (2000), o conceito de proficiência dialoga com o exame, e ainda acrescenta sobre as variáveis que podem afetar o nível de proficiência como o contexto do examinando, a aprendizagem da língua, as características do local de prova e o próprio processo de aprendizado.

Dentro da Linguística Aplicada, o conceito de proficiência é bastante discutido devido a sua essencialidade para o aprendizado de uma língua adicional. Quevedo-Camargo (2019) e Scaramucci (2000) ressaltam em suas pesquisas as características que envolvem ser proficiente em uma língua, ao fim, chegando a um conceito comunicativo dinâmico, uma vez que o nível de proficiência dependeria da situação e propósito de uso da língua. Testar proficiência é, desse ponto de vista, avaliar a compreensão e produção a partir de um contexto solicitado, adequando as práticas linguísticas às demandas presentes nas tarefas (Schoffen, 2009).

Além disso, para o exame, não há a idealização do falante nativo, visto que, para o construto, é considerado principalmente a capacidade que o examinando tem de usar a língua portuguesa de uma forma social proposta, engajando-se adequadamente em ações no mundo com propósitos sociais. Ao defender essa ideia, o exame considera que a língua é indissociável da cultura, bem como do contexto social e histórico (Brasil, 2020a). Ademais, o Documento-Base traz algumas reflexões:

A proficiência conforme o construto do Exame, é sempre relativa, isto é, apresenta níveis definidos de acordo com as necessidades de uso futuro da língua. Em vez de uma única proficiência, 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, 2020a, p. 27).

A partir do conceito de proficiência apresentado, o exame Celpe-Bras concede quatro diferentes certificados aos examinandos, de acordo com as faixas

mostradas no Quadro 1, sendo que o certificado é correspondente à menor nota final obtida na parte oral ou na parte escrita:

Quadro 1 - Certificações do Celpe-Bras

NÍVEL	FAIXA DE NOTAS
Sem Certificação	De 0,00 a 1,99
Intermediário	De 2,00 a 2,75
Intermediário Superior	De 2,76 a 3,50
Avançado	De 3,51 a 4,25
Avançado Superior	De 4,26 a 5,00

Fonte: Brasil (2020a, p. 79)

Para compreender melhor o conceito de proficiência e as certificações, é necessário compreender que a prova é dividida em duas etapas: Parte Oral e Parte Escrita.

2.3 PARTE ORAL

A Parte Oral é realizada individualmente, durante uma interação de 20 minutos do examinando com dois avaliadores: o Avaliador Interlocutor (AI) e o Avaliador Observador (AO). Nesta etapa, os examinandos devem dialogar, a partir de perguntas guiadas pelo Avaliador-Interlocutor, com base em temas de interesse pessoal e em três Elementos Provocadores que tratam de temas da atualidade (Brasil, 2020). A avaliação desta etapa é feita em dois momentos, o AI, enquanto interage com o examinando, realiza uma avaliação holística, atribuindo nota de 0 (zero) a 5 (cinco) pontos correspondente ao desempenho geral. Já o AO utiliza uma grade avaliativa analítica, com notas de 0 a 5 para seis critérios diferentes: compreensão do fluxo da conversa, competência interacional, fluência na comunicação, adequação lexical, adequação gramatical e pronúncia. A nota final do exame consiste na média entre as avaliações feitas pelo AI e o AO.

2.4 PARTE ESCRITA

A Parte Escrita, foco deste trabalho, tem duração de três horas. Nesse período, os examinandos produzem textos em resposta a quatro tarefas, que integram avaliação de diferentes habilidades, conforme o Quadro 2 a seguir:

Quadro 2 - Tarefas da Parte Escrita

TAREFA	HABILIDADES	INSUMO
1	Compreensão oral e imagética e produção escrita	Vídeo
2	Compreensão oral e produção escrita	Áudio
3	Leitura e produção escrita	Texto escrito
4		

Fonte: (Brasil, 2020a, p.58)

Dessa forma, como organizado no Quadro 2, os examinandos devem escrever textos em resposta ao enunciado das tarefas que se encontram no caderno de questões, entregue aos examinandos durante o teste. O objetivo é cumprir um propósito explicitado, se adequar ao enunciador proposto e, também, selecionar recursos linguísticos adequados ao interlocutor indicado.

As tarefas I e II do exame avaliam a compreensão oral e a produção escrita. A Tarefa I tem um insumo multimodal de vídeo, que é exibido duas vezes para possibilitar que o examinando faça anotações no caderno de rascunho. Já a Tarefa II tem um insumo multimodal de áudio, que também é reproduzido duas vezes.

A Tarefa III e a Tarefa IV avaliam, de forma integrada, a compreensão de leitura e escrita, a partir de um texto de insumo presente no caderno de questões. Os materiais de insumo de todas as tarefas são autênticos e circulam na sociedade brasileira. Além disso, todos os enunciados propostos determinam que o examinando mobilize informações do insumo para escrever os textos de acordo com os interlocutores, propósitos e gêneros solicitados nos enunciados (Schoffen et al., 2018).

No trabalho feito por Schoffen et al. (2018), as pesquisadoras realizaram um estudo descritivo das tarefas da Parte Escrita do exame entre os anos de 1998 e 2017. Neste estudo, podemos ver que todas as tarefas apresentam as condições

para produção, assim, a tarefa especifica uma posição enunciativa para o autor do texto, um interlocutor para direcionar a escrita e um propósito de uma ação a ser realizada. Os materiais de insumo disponibilizados são utilizados para selecionar informações essenciais para a realização da tarefa (Schoffen et al, 2018).

No Documento-Base do Exame (Brasil, 2020b) também é destacado que, sobre a seleção de informações do material de apoio:

Considerando que, no Celpe-Bras, as habilidades são avaliadas de forma integrada, na Parte Escrita, avalia-se a compreensão para a produção, uma vez que os objetivos de compreensão para cada tarefa são determinados pelos objetivos de escrita. Isso significa que o texto a ser produzido na Parte Escrita 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, 2020b, p. 37).

Em Schoffen (2009), a autora reitera essas informações, partindo de que o examinando terá que, além de compreender o material, selecionar somente as informações adequadas ao seu propósito de escrita. A seleção de informações não relevantes para o texto também acarreta na demonstração do nível proficiência do examinando. Mendel (2019) também defende que o conhecimento pessoal do participante pode contribuir para realização dos propósitos dispostos, assim como a leitura do material de insumo.

2.4.1 A AVALIAÇÃO DA PARTE ESCRITA

A avaliação das produções procura corroborar com a definição de proficiência apresentada no construto do exame. Assim, o Documento-Base do exame afirma que a escolha pela avaliação holística na Parte Escrita busca operacionalizar o construto, uma vez que seria incompatível com a ideia de avaliar o conhecimento dos recursos linguísticos de forma independente do uso da língua (Brasil, 2020a). Isso significa que, durante a avaliação, os corretores devem considerar as eventualidades do enunciado e os ajustes da língua em cada tarefa, da mesma forma como acontece no uso cotidiano da língua nas práticas sociais.

O processo de correção começa com os postos aplicadores enviando as provas escritas para a empresa aplicadora contratada pelo Inep, responsável pela correção. Em segundo momento, após as provas serem desidentificadas, as tarefas

são separadas e agrupadas (Schoffen, 2009). Cada texto produzido pelo examinando é avaliado de forma independente por dois avaliadores que atribuem aos textos notas de zero a cinco, utilizando a mesma grade de correção; a nota final do texto é calculada através da média aritmética das notas dadas pelos dois avaliadores. Se as notas atribuídas diferem em mais de um ponto, é considerada discrepância entre os avaliadores, então o texto volta para a equipe para ser reavaliado por um terceiro corretor, que não é informado das notas anteriores.

Para atribuir a nota, os avaliadores seguem os parâmetros de avaliação holística que são utilizados em conjunto com uma resposta esperada ajustada para cada tarefa, a fim de reunir as características de enunciador, interlocutor, propósito e conteúdo informacional específicas a serem consideradas na avaliação (Brasil, 2020a). De acordo com o Documento-Base do Celpe-Bras (Brasil, 2020a), os avaliadores podem perceber se o enunciador e o interlocutor definidos pelo enunciado estão claramente explicitados, bem como o propósito cumprido, e devem levar em conta, ainda, como e quais informações do texto são recontextualizadas e organizadas. Os recursos linguísticos empregados, tanto lexicais quanto gramaticais, também são avaliados.

Por fim, para atribuir a nota final do examinando na Prova Escrita, é feita uma média aritmética das notas finais atribuídas às quatro tarefas. Para a certificação final, considera-se a nota mais baixa entre a Parte Escrita e a Parte Oral, visto que não são atribuídos certificados separadamente para cada prova.

3 FUNDAMENTAÇÕES TEÓRICAS

Nesta seção, serão apresentadas as fundamentações teóricas que embasaram este estudo. Inicialmente, será apresentada a Linguística de Corpus e os procedimentos de análise. Em seguida, um panorama dos estudos sobre PLA, a partir de pesquisas com uso de *corpus*, e, por fim, uma subseção sobre os conectivos que serão analisados neste trabalho.

3.1 LINGUÍSTICA DE CORPUS

A Linguística de Corpus (LC) é uma área de pesquisa que se dedica ao estudo de uma língua ou de uma variedade linguística valendo-se de evidências reais da língua em uso, extraídas de um *corpus*, isto é, um conjunto de textos que foram coletados pelo pesquisador especificamente para fins de pesquisa e armazenados em um computador (Berber Sardinha, 2004, 2011; Cushing, 2022; Granger, 2009; Huang, Yao, 2015; Sinclair, 1991). Em Cushing (2022), a autora afirma que: “uma das aplicações mais promissoras da LC é a descrição da aprendizagem de língua em diferentes níveis de proficiência.” (2022, p. 554, tradução própria)³. A composição de um *corpus*, de acordo com Kennedy (1998) é feita a partir de uma compilação de textos, cujo objetivo é o de descrever e perceber fatos em uma determinada língua ou algum aspecto pontual presente. Em uma das linhas de trabalho possíveis com a LC, o pesquisador analisa o *corpus* em busca de padrões na linguagem (Berber Sardinha, 2004) e ocorrências de traços linguísticos frequentes.

Dessa maneira, a LC pode auxiliar na análise dos diferentes níveis de aprendizagem de uma língua, e isso pode ser realizado a partir de produções feitas em exames de proficiência. Em Gablasova (2020), destaca-se que pesquisas sobre estudo de *corpora* são importantes para evidenciar características linguísticas típicas e particulares de domínio da língua por falantes de segundo idioma.

As pesquisas que utilizam a LC podem, inicialmente, utilizar uma metodologia quantitativa, como a identificação de frequência e a extração de recursos linguísticos em produções textuais, assim como explicado por Granger (2009). Após um período, pode surgir a necessidade de se utilizar também uma abordagem qualitativa, em que os resultados são contextualizados, o que possibilita reconhecer padrões de uso de linguagem entre textos de um mesmo *corpus*.

Com isso, estudos que conseguem unir metodologia quantitativa e qualitativa utilizando LC são capazes de analisar diferentes gêneros discursivos e registros, em aspectos lexicais, gramaticais e discursivos (Kennedy, 2001). Há, também, uma categorização para a metodologia usada a partir de *corpus*; segundo Callis e Gotz (2015), há as pesquisas baseadas em *corpus* e as pesquisas guiadas por *corpus*. A

³ No original: “One of the most promising applications of corpus linguistics is in describing learner language at different levels of proficiency.” (CUSHING, 2022, p. 554)

primeira refere-se a estudos que analisam padrões de variação e de uso, a partir de características e questionamentos pré-definidos. Já no segundo caso, as pesquisas objetivam a extração e a avaliação dos dados, levando a questão e a conclusão serem derivadas dos resultados.

3.1.1 ESTUDOS SOBRE O EXAME CELPE-BRAS QUE ANALISAM CORPUS

Pesquisas acerca do exame Celpe-Bras utilizando *corpus* foram feitas anteriormente, visando investigar potenciais formas de auxiliar os examinandos e professores de PLA com uma maior clareza de informações sobre as produções e os níveis certificados. Alguns trabalhos foram realizados por participantes do Grupo Avalia⁴, utilizando textos produzidos por participantes do exame ou produzidos por estudantes de cursos preparatórios que utilizaram tarefas aplicadas anteriormente, que contribuíram, inicialmente, no detalhamento dos níveis de proficiência avaliados no exame. Grande parte desses estudos realizaram pesquisas qualitativas e analisaram um número pequeno de textos.

Em Schoffen (2009), a pesquisa procurou analisar a validade do construto do Celpe-Bras, principalmente na Parte Escrita. A partir das análises realizadas, a autora utilizou 181 textos produzidos por examinandos em resposta às tarefas da edição 2006/1 do exame. As análises foram efetuadas a partir da leitura dos componentes dos enunciados, a fim de verificar se os contextos de produção e de recepção estavam de acordo com o solicitado no texto. Como resultado da pesquisa, foram elaborados novos parâmetros de avaliação, seguindo uma perspectiva bakhtiniana de linguagem, que são capazes de subsidiar uma avaliação de acordo com o construto do exame.

Gomes (2009) discute a complexidade nas tarefas de leitura e produção escrita do Celpe-Bras. A autora procurou analisar os enunciados e insumos de quatro tarefas e cerca de 144 textos-resposta produzidos por alunos do Programa de Português para Estrangeiros da UFRGS (PPE-UFRGS). Nesse trabalho, foi

⁴ Grupo de pesquisa que atua no Instituto de Letras da Universidade Federal do Rio Grande do Sul, registrado no CNPq (www.dgp.cnpq.br/dgp/espelhogrupo/8480692535262657), coordenado pela Professora Dra. Juliana Schoffen. O grupo realiza pesquisas referentes à avaliação de proficiência em uso da linguagem; avaliação para nivelamento em cursos de língua; níveis de proficiência em leitura e escrita; efeitos retroativos e impactos de instrumentos de avaliação.

concluído que a complexidade nas tarefas envolve uma relação entre a tarefa, o examinando e os procedimentos de avaliação, não sendo possível prever ou determinar aspectos antecipadamente. Dessa forma, enunciados que são suficientemente contextualizados, com informações adequadas sobre a relação de interlocução solicitada, auxiliam na compreensão dos examinandos e para que estes possam ser avaliados com mais coerência. Já Sirianni (2016) também compôs um *corpus* de análise a partir de textos produzidos por alunos do Curso Preparatório Celpe-Bras do PPE-UFRGS. Neste trabalho, foram utilizados 22 textos-resposta à tarefa IV da edição 2015/2, a partir dos quais a autora realizou a descrição dos níveis de proficiência demonstrado pelos estudantes, transformando os parâmetros gerais de avaliação do exame em parâmetros específicos para cada tarefa, com a finalidade de discutir a aplicabilidade de uma avaliação mais detalhada dos níveis de proficiência do Celpe-Bras.

Procurando compreender as recontextualizações feitas a partir das informações do insumo, Mendel (2019) analisou 236 produções escritas por examinandos em resposta a uma tarefa de leitura e escrita. A pesquisa consistiu em uma análise qualitativa das tarefas de leitura e escrita do exame, constatando que cada nível de certificação analisado contém recontextualizações de informações específicas do material de insumo e nos níveis mais avançados há uma maior mobilização de repertórios culturais externos. Além disso, a autora chegou à conclusão de que a autoria é essencial para atingir o nível Avançado Superior, uma vez que a mobilização de repertório está conectada principalmente ao gênero do discurso exigido pela tarefa. Em outro trabalho, Kunrath (2019), procurou distinguir os níveis Intermediário Superior e Avançado Superior, a partir da recontextualização de informações e do emprego de recursos linguístico-discursivos, levando em conta aspectos do material de insumo e dos propósitos interlocutivos das tarefas. Com um *corpus* formado por 50 textos e a utilização do *software Coh-Matrix*, foi possível propor uma progressão a partir da concretude e da coesão dos textos produzidos em cada nível.

Divino (2021) elaborou um estudo, utilizando o *software Sketch Engine* (Kilgarriff et al., 2004), para analisar dados de textos que tiraram a nota 5 e a nota 2, correspondentes ao Nível Avançado Superior e ao Nível Intermediário, respectivamente. Ainda, foi realizada uma comparação entre o *corpus* de estudo e um *corpus* de referência, a fim de comparar o uso de itens lexicais específicos

encontrados nas produções dos examinandos. Este trabalho analisou 865 textos com o auxílio do programa e buscou determinar as semelhanças e diferenças lexicais entre as produções dos examinandos e o texto de insumo, partindo da extração de palavras-chaves dos *corpora*. Os resultados mostraram que a maioria das *keywords*⁵ extraídas das produções de participantes estaria presente no texto de insumo ou no enunciado da tarefa. Ademais, o trabalho permitiu perceber que produções certificadas com a nota 5 apresentam maior adequação ao gênero discursivo solicitado, além da utilização de recursos linguísticos importantes que não constam no material de insumo ou no enunciado.

Hanauer (2023) utilizou os enunciados e os materiais de insumo de duas tarefas e um *corpus* de 400 textos-resposta a tarefas que solicitavam a produção do gênero discursivo carta/e-mail. As produções escritas selecionadas foram certificadas com a nota 5 e a nota 2, assim totalizando dois *subcorpora* de 200 textos cada. O objetivo do trabalho era contribuir para a caracterização dos níveis Intermediário e Avançado Superior pertencentes ao gênero carta/e-mail, por meio do uso do *software Sketch Engine*. Como resultado, a autora observou, a partir de extrações realizadas com o auxílio do programa, as diferenças na riqueza lexical entre os dois níveis, e também, que os textos que receberam a nota 5 mobilizam um maior número de recursos linguísticos para adequar as produções em relação a interlocução, suporte e propósito solicitados pelo enunciado. Ademais, também foi analisada a adequação ao gênero carta/e-mail feita pelos examinandos, destacando um aumento de saudações, despedidas e assinaturas em produções certificadas como Avançado Superior.

Ambos os trabalhos verificaram que as inadequações ortográficas influenciam nas análises, pois a adição ou a supressão de acentos ou sinais gráficos, ou de alguma letra, mesmo significando uma mesma palavra, é lida pelo programa como uma nova palavra, afetando na frequência dos dados. Isso ocorre porque os *softwares* não são capazes de processar uma mesma palavra englobando suas diferentes grafias. De acordo com Hanauer (2023), o maior número de inadequações foi encontrado na nota 2. Quando a autora realizou análises das *keywords* dos textos nota 5 e nota 2, percebeu-se que a presença de inadequações ortográficas pode influenciar fortemente no valor das frequências de

⁵ Ferramenta disponível no programa *Sketch Engine* que possibilita listar as palavras mais frequentes no *corpus* em relação a um *corpus* de referência.

palavras dos *corpora* formados por essas produções. A análise das *keywords* dos textos de nota 2 usando como referência os *corpora* de textos avaliados com nota 5 demonstra que aparecem muitas inadequações ortográficas, inclusive de acentuação.

Em um trabalho conjunto, Divino, Hanauer e Sostruznik (2021) realizaram uma análise dos textos avaliados como Avançado Superior de todas as tarefas de uma única edição do Celpe-Bras. Nestas produções, os resultados demonstraram a explicitação da relação de interlocução realizada pelos examinandos de maneira autoral, com a utilização de sinônimos para se referir ao interlocutor, como também, o uso de informações do material de insumo e informações adicionais ao longo dos textos. Em destaque para as tarefas de leitura e escrita, Sostruznik (2021) notou, no nível Avançado Superior, uma frequente utilização de sinônimos para recuperar informações do texto de insumo.

Em Sostruznik (2022), foi realizada uma pesquisa sobre a presença e categorização funcional de marcadores discursivos empregados nas produções textuais que receberam a nota máxima (nota 5). Nesta pesquisa, procurou-se extrair os marcadores discursivos mais frequentes utilizados nos textos, partindo de uma categorização funcional elaborada por Fraser (1999), a qual foi adaptada para a língua portuguesa. O *corpus* analisado é composto pelos textos que receberam a nota mais alta nas tarefas de leitura e escrita na edição 2015/2 do Celpe-Bras, totalizando 138 textos da Tarefa III e 237 textos da Tarefa IV. Com o programa *Sketch Engine*, foi realizada a extração com o auxílio da ferramenta *Wordlist*, selecionando somente as conjunções e os advérbios que eram usados para marcar uma opinião dentro das produções textuais. Como resultados das conjunções, notou-se uma menor frequência de conjunções adversativas comparadas com as demais, porém uma maior variedade de exemplares. Os resultados mostraram que os examinandos optaram mais pelos marcadores discursivos considerados “elaborativos” pela taxonomia criada por Fraser (1999). Esses marcadores discursivos elaborativos são usados principalmente para acrescentar informações essenciais ao texto. Além disso, o trabalho permitiu notar a maior ocorrência dos marcadores discursivos contrastivos e aditivos nos textos avaliados com a nota 5. Esta pesquisa possibilitou fazer uma análise inicial sobre o uso das conjunções pelos examinandos em suas produções textuais, incentivando este trabalho a aprofundar o uso dos conectivos em textos produzidos pelos falantes de PLA.

Inclusive, em Sostruznik (2022) foram usadas duas tarefas de gêneros discursivos distintos, o que pode ter afetado os resultados comparativos entre as tarefas, visto que a tarefa III solicitava que o examinando pedisse patrocínio de empresas para um projeto social e a tarefa IV propunha uma carta aberta direcionada à prefeitura, demandando medidas imediatas para salvar um patrimônio histórico. Por isso, as diferenças demandadas pelos enunciados das tarefas podem afetar no uso dos marcadores discursivos.

A pesquisa anterior motivou a construção deste trabalho, visto que, em Sostruznik (2022), buscou-se extrair os marcadores discursivos presentes em textos-resposta do Celpe-Bras, e durante o processo, analisou-se, conjuntamente, diferentes tipos de conjunções. Este trabalho pretende expandir essas análises, utilizando outras tarefas, dado que a escolha adequada de uma conjunção, em uma produção textual, auxilia no desenvolvimento dos argumentos durante a escrita de um texto. Assim, procurando descrever melhor suas frequências e usos nos textos, decidiu-se realizar este estudo com diferentes tarefas que fossem semelhantes quanto ao gênero discursivo e ao propósito solicitado.

3.2 CONJUNÇÕES

De acordo com Perini (2016), as conjunções (coordenativas e subordinativas) fazem parte do que chamamos de conectivos, que são palavras ou expressões que têm como funcionalidade unir ideias. Além das conjunções, as preposições também fazem parte dos conectivos; no entanto, não serão analisadas nesta pesquisa. Ainda, as conjunções podem estabelecer a ligação entre dois termos de uma oração, podendo transformar o sentido de uma sentença.

A partir disso, esses elementos são parte importante da estratégia retórica de uma produção textual, pois têm o poder de afetar a interpretação de uma sentença. O uso adequado das conjunções em produções textuais permite que o autor deixe claro seu posicionamento. A escolha de um conector auxilia na força de um argumento e também na compreensão do leitor. Koch (2011) defende que os elementos conectivos são formas importantes de mostrar a opinião do autor em relação ao conteúdo que expressa. Com isso, a coesão, por estabelecer relações de sentido, está entrelaçada com os usos dos conectivos, visto que, por meio desses, uma sentença se liga à outra que veio antes, mantendo o significado do discurso.

Ainda em Koch (2011), os conectivos permitem ligar as sentenças e estabelecer as relações de sentido escolhidas pelo autor, mantendo a coerência textual, de maneira que essa seja responsável pela continuidade semântica em um texto. A coerência não se trata somente de um trecho de uma escrita, ela faz parte de “uma complexa rede de fatores” (Koch, 2011, p.12).

Funcionando como conectivos, as conjunções têm como foco reunir duas ou mais orações em um mesmo enunciado (Bechara, 2010). Concomitantemente, essas unidades são tradicionalmente divididas em dois tipos: coordenativas e subordinativas.

Primeiro, as coordenativas, também chamadas de independentes, têm como objetivo juntar duas unidades de idêntica função sintática e podem ser classificadas em grupos, dialogando com o sentido que fornecem (Azeredo, 2011). Seguindo a classificação feita pelo autor, as conjunções coordenativas podem ser agrupadas em conectivos de:

1. adição (conjunção aditiva): expressa a união de dois fatos.
2. contraste (conjunção adversativa): expressa a ideia de relação contrastiva entre dois fatos, podendo realçar uma ideia em oposição a outra mencionada.
3. opção ou alternância (conjunção alternativa): expressa sentidos opostos ou semelhantes, porém um nega a aplicação do outro.
4. esclarecimento (conjunção explicativa e conclusiva): exprime a continuidade lógica de um raciocínio iniciado com a oração anterior ou inicia uma nova sentença ou parágrafo, dando sequência aos argumentos anteriores.

Quadro 3 - Conjunções Coordenativas

Categoria	Exemplos
Conjunção aditiva	e; nem;
Conjunção adversativa	mas; porém; senão;
Conjunção alternativa	ou (...ou); ora (...ora);
Conjunção explicativa	porque; pois;

Quadro elaborado pela autora a partir das conjunções propostas por Bechara (2010).

O segundo tipo são as conjunções subordinativas⁶, que são aquelas que ligam duas orações, em que uma determina ou completa o sentido da outra (Cunha & Cintra, 2013). Para exemplificar, os autores classificam essas conjunções em tipos funcionais:

1. Causais: iniciam uma oração subordinada denotadora de causa;
2. Concessivas: iniciam uma oração subordinada em que se afirma o contrário da oração principal;
3. Condicionais: iniciam uma oração subordinada que indica hipótese ou concessão para ser realizado ou não;
4. Finais: iniciam a oração subordinada, indicando finalização de uma ideia trazida na oração principal;
5. Temporais: iniciam orações que indicam uma circunstância de tempo;
6. Comparativa: inicia a oração subordinada comparando com a ideia apresentada na oração principal;

Quadro 4 - Conjunções subordinadas com exemplos.

Categoria	Conjunções	Exemplo
Causais	visto que, já que, dado que, porquanto, por isso que, uma vez que;	Tio Couto estava sombrio, já que aparecera um investigador da polícia perguntando por Gervásio. (É. Veríssimo, L S, 137.)
Concessivas	embora, conquanto, ainda que, apesar de;	Não saberei nunca escrever sobre ele, embora tenha tentado mais de uma vez. (F. Sabino, G, II, 76.)
Condicionais	se, caso, contanto que, salvo se;	Se aquele entrasse, também os outros poderiam entrar... (Branquinho da Fonseca, MS, 41.)
Finais	para que, a fim de que;	Recolheu a carta e a sobrecarta, para mostrá-las a Rubião, a fim de que ele visse bem que não era nada. (Machado de Assis, OC, 1,646.)
Temporais	quando, antes que, depois que, até que, logo que;	Quando tio Severino voltou da fazenda, trouxe para Luciana um periquito. (G. Ramos, Ins., 79.)

⁶ As conjunções subordinativas normalmente são compostas por mais de uma palavra, sendo assim, não será possível analisá-las neste trabalho, como será mencionado no próximo capítulo.

Comparativa	do que; qual; quanto; como; que nem;	“Ele comeu-a que nem confeitos”. (C. Castelo Branco, OS, 1,368.)
-------------	---	---

Quadro elaborado pela autora a partir das conjunções e exemplos trazidos por Cunha & Cintra, 2013.

Esta pesquisa pretende considerar a contexto de uso das conjunções no interior da frase, de modo a delimitar quais as que ocorrem preferencialmente em posição inicial e quais as que ocorrem preferencialmente em posição não inicial. Para Perini (2010), as conjunções são itens léxicos que devem ser colocados imediatamente antes de introduzir a outra oração, coordenada ou dependente. Entretanto, sabe-se que as conjunções podem variar de posição, apresentando-se no início de uma sentença em um deslocamento. Segundo Matos (2003, p. 539), as conjunções de coordenação ocupam tipicamente a posição inicial do membro coordenado, entre as duas orações; entretanto, podem deslocar-se para introduzir a primeira oração, sendo este considerado um recurso linguístico estilístico elaborado.

3.2.1 PESQUISAS SOBRE CONJUNÇÕES EM PRODUÇÕES TEXTUAIS

Trabalhos anteriores sobre conjunções em produções textuais utilizando Linguística de Corpus já foram realizados, porém, em grande maioria, sobre a aprendizagem da língua inglesa como língua adicional. A respeito da língua portuguesa, a maior parte das pesquisas investigam o uso de português como língua materna, que não é o foco deste trabalho.

A partir da língua inglesa como língua adicional, temos uma pesquisa realizada por Matte (2017) que procurou comparar o uso de conjunções por brasileiros e por falantes nativos que frequentavam universidades britânicas. Nesta pesquisa, a autora trouxe dois *corpora* com produções textuais que obtiveram notas elevadas. O programa Sketch Engine foi utilizado para a extração dos conectores, utilizando a ferramenta *Wordlist*, porém com o auxílio de uma “whitelist”, que seria uma lista prévia com conectores pré-definidos pela autora. Como resultados, percebeu-se que há uma diferença significativa nos usos dos conectores nos *corpora* escolhidos, bem como um uso excessivo de elementos que expressam ideias de adição por parte dos estudantes brasileiros. Com essa pesquisa, podemos notar a essencialidade de

estudos baseados em *corpus*, visto que com as evidências das produções textuais da língua é possível notar aspectos particulares de uso.

Em outro trabalho, desta vez com textos de PLA, Evers (2013) trabalhou com 177 produções textuais oriundas da edição de 2006/1 do exame Celpe-Bras. Utilizando a ferramenta Coh-Metrix, a autora tinha como um dos objetivos analisar o uso de conectivos. Como resultados, notou-se que textos classificados como Intermediário tendem a utilizar mais conectivos com relação ao conjunto de textos avaliados como nível Básico.

Ambos os trabalhos citados demonstram que a LC permite descrever aspectos da língua, comparar *corpora* e caracterizar algumas diferenças presentes entre os níveis de proficiência.

4 METODOLOGIA

Para dar seguimento aos estudos já feitos, serão utilizadas ferramentas de Linguística de Corpus para analisar textos produzidos por examinandos do Celpe-Bras, uma vez que a contribuição dos estudos com *corpora* pode oferecer aos pesquisadores a possibilidade de análise de dados autênticos e extração de elementos específicos a partir de um grande número de textos. Assim, neste capítulo, serão apresentados os objetivos deste trabalho, bem como os procedimentos adotados para atingir tais objetivos. Na sequência, são apresentados os *corpora* utilizados nesta pesquisa, o programa utilizado para a análise dos dados e, por fim, os procedimentos para as análises.

4.1 OBJETIVOS

Com o presente trabalho, busca-se contribuir com a construção de uma descrição mais detalhada dos níveis de proficiência avaliados no Celpe-Bras, utilizando ferramentas de Linguística de Corpus para analisar recursos coesivos presentes nos textos produzidos por examinandos em diferentes níveis de proficiência. Para isso, seguiu-se a pergunta norteadora: **Há diferença no uso de conjunções entre as produções dos examinandos certificados nos níveis Avançado Superior e Intermediário?**

A partir desta pergunta, buscou-se responder aos seguintes questionamentos:

1. Quais as conjunções mais frequentemente utilizadas nos *corpora* de cada nível?
2. Quais as classificações das conjunções coordenativas mais utilizadas nos *corpora* de estudo?
3. Quais as posições das conjunções utilizadas nas sentenças? Qual a diferença entre os contextos de uso em cada nível?

4.2 CORPORA DE ESTUDO

Para compor as análises, foram selecionados dois *corpora*, ambos compostos por produções escritas por examinandos em resposta às tarefas IV das edições 2016-2 e 2017-1 do exame Celpe-Bras. Essas produções escritas fazem parte de um *corpus* de produções escritas do Celpe-Bras compilado pelo Grupo Avalia⁷.

4.2.1 AS TAREFAS SELECIONADAS

Para cumprir os objetivos propostos, as tarefas selecionadas para esta pesquisa foram escolhidas por terem características semelhantes: ambas solicitam a produção de um texto do gênero Carta do Leitor e solicitam ao examinando que cumpra o propósito de “posicionar-se” diante de uma temática. Também, Schoffen, Sirianni e Kunrath (2020) notam uma estabilidade presente na tarefa IV nas edições do exame Celpe-Bras. As autoras afirmam que há uma incidência de repetições no gênero discursivo a ser escrito e no propósito comunicativo das tarefas solicitadas nos últimos anos de aplicação: o gênero do discurso que teve maior frequência é “Carta do Leitor”; já o propósito mais recorrente é “Posicionar-se”. De acordo com as autoras:

Percebe-se, assim, que a Tarefa IV pode ser considerada uma tarefa bastante estável no que diz respeito ao propósito comunicativo, priorizando a produção de textos nos quais os

⁷ No momento, o Grupo Avalia está compilando um *corpus* de textos produzidos por examinandos do Celpe-Bras de quatro edições do exame e avaliados com diferentes notas. O material foi cedido ao grupo pelo Inep, órgão brasileiro responsável pelo exame, em forma de imagens digitalizadas. Atualmente, o *corpus* está em fase final de digitação, revisão e etiquetagem (Schoffen et al, em preparação).

examinandos devam expressar opinião e argumentar. (Schoffen, Sirianni, Kunrath, 2020, p. 924).

Com isso, a escolha pelas tarefas IV das edições 2017/1 e 2016/2 está relacionada à estabilidade encontrada, o que as torna representativas das demais tarefas IV de outras edições.

4.2.1.1 Tarefa IV da Edição 2016/2

A tarefa IV da edição 2016/2 tem como título "Shopping em casa". Como material de insumo, há um texto adaptado do jornal *Folha de São Paulo*, escrito por Denise Fraga, que traz ao leitor o desafio de olhar as roupas e os objetos que temos em casa e, tentar trazer um novo olhar, ao invés de consumir impulsivamente. Na figura 2, é possível visualizar na íntegra o texto de insumo e o enunciado da tarefa.

Figura 2 - Material de insumo presente na Tarefa IV da edição 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 doída 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: Zdzisław Beksiński

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).

Fonte: Brasil (2016, p. 9)

A partir da leitura do insumo e do enunciado, os examinandos precisam produzir um texto para ser publicado na seção "Carta ao Leitor" do jornal *O Estado de São Paulo*, que traga um ponto de vista em relação às questões de consumo apresentadas pela autora. Para isso, os examinandos devem, além de

contextualizar informações relevantes do texto, trazer argumentos concordando ou discordando da opinião da autora.

4.2.1.2 Tarefa IV da Edição 2017/1

A tarefa IV da edição de 2017/1, denominada “Bibliotecas: metamorfose ou morte?”, traz um texto de insumo escrito por Cláudio Castro para a revista *Veja* sobre a situação atual das bibliotecas pelo Brasil e como elas estão se tornando desinteressantes e desatualizadas. O autor também oferece dicas de como rejuvenece-las e deixá-las atraentes para o público. Na figura 3, estão apresentados na íntegra o texto de insumo e o enunciado da tarefa:

Figura 3 - Enunciado e material de insumo presente na Tarefa IV da edição 2017/1

2017/1 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 4 | Bibliotecas Página 8

Você leu o artigo de opinião abaixo apresentado, e resolveu escrever uma carta para a seção "Leitor", respondendo à pergunta "Bibliotecas: metamorfose ou morte?". Em seu texto, você deverá posicionar-se sobre o assunto, concordando ou discordando dos argumentos apresentados pelo autor.

Bibliotecas: metamorfose ou morte?

Quando buscávamos um livro, a solução era óbvia: bastava ir à biblioteca. Mas rondam tempestades ameaçando essa respeitável instituição. Em poucos anos, caberão em um *notebook* todos os livros produzidos na história da humanidade (as estimativas flutuam entre 42 milhões e 130 milhões). Um pouco adiante, e enfia-se tudo em um celular.

Para que biblioteca? Periódicos científicos e muitas outras publicações migram para a sua versão digital, o mesmo acontecendo com os jornais. Diante do *www*, é risível o tamanho das bibliotecas em papel. A Wikipedia esmaga a mais ambiciosa enciclopédia tradicional. E para quê bibliotecário se o "Santo Google" acha tudo rapidinho?

Por 10 dólares ou pouco mais, a versão digital de praticamente todos os livros em inglês pode ser comprada na Amazon. Um minuto depois de um só clique, o livro está em nosso poder. É inevitável que o Brasil vá pelo mesmo caminho – apesar do atraso presente. E não há como impedir a digitalização pirata de livros populares.

Diante disso tudo, o que será das bibliotecas? São caras, e seu acervo no Brasil é inferior ao de muitos países. Pior, falta-nos o hábito de frequentá-las. Portanto, se definharem, sua falta não será notada.

As notícias sobre a morte das bibliotecas podem ser exageradas. De fato, seu lugar no futuro pode estar assegurado, desde que elas se transformem. Biblioteca careta e chata não sobreviverá. Como depósito de livros, está condenada.

É sintomático que algumas bibliotecas americanas tenham levado seus livros para depósitos, pois havia muitos usos mais nobres para o espaço. Eis a pista para a salvação: a biblioteca do futuro será um canivete suíço, fará tudo.

Se bem concebida, ela será um lugar aonde vamos sem pensar muito no que faremos lá. Vamos porque nos atrai, porque é bom estar nela. Para início de conversa, precisa ser supremamente formosa, confortável e atraente. A arquitetura externa tem de dar vontade de entrar. A interna, de ficar.

Seu ambiente será agradável como as Starbucks e os restaurantes chineses do Vale do Silício. Haverá abundância de jornais, revistas e livros de interesse geral. E, cada vez mais, vídeos. Livros desinteressantes, porém, doados por alguma viúva (três quartos dos nossos acervos são dessa origem), não trazem ninguém às bibliotecas.

De depósitos de livros, passam a oferecer quase tudo. Alguns espaços são silenciosos, para ler. Em outros, conversamos ou nos reunimos (com projetor de PowerPoint). Algumas poucas estão voltadas para a pesquisa, uma função essencial e cara. Mas, se a Amazon consegue entregar no dia seguinte os livros comprados, as bibliotecas também poderão. Títulos pouco procurados não precisam de mais de um exemplar, talvez no país inteiro. Basta um sistema para tomar emprestado, rapidamente, do acervo de outras bibliotecas.

Na nova biblioteca, salas e auditórios promovem conferências, concertos e exposições. Por que não jardins lindos, para os criativos filósofos? Ou espaços para meditar? No fundo, a biblioteca deve tornar-se um lugar de leitura, troca de ideias e interação criativa entre os frequentadores. Enfim, uma usina intelectual, contribuindo para o avanço do país. Naturalmente, quando bate a fome, lá comemos. E, afinal, um lugar onde se leem e se tomam livros emprestados por que não os vende também? Assunto e clientela são os mesmos das livrarias.

A fórmula salvadora já existe e é resumida pela celebrada arquiteta Maya Lin. Para ela, bibliotecas são os templos de hoje, espaços para reflexão, exploração intelectual e discussão de ideias. Mas engana-se quem pensa ser revolucionária tal visão. De fato, a primeira grande biblioteca que o mundo conheceu, a de Alexandria, tinha como ponto de partida uma arquitetura memorável, e sua concepção antecipa essa linha. Além dos livros, tinha jardins, exposições de arte, concertos e outras atividades culturais. No dizer de um contemporâneo, "era um lugar para curar a alma".

Ou seja, eis a receita para salvar nossas bibliotecas. Não é preciso inventar nada.

CASTRO, Cláudio de M. **Bibliotecas: metamorfose ou morte?** In: Revista *Veja*, 26 ago, 2015, p.24 (adaptado).

Fonte: Brasil (2017, p. 9)

Partindo dos argumentos trazidos pelo autor do texto, os examinandos deveriam concordar ou discordar da opinião do autor sobre a morte das bibliotecas. Para isso, os examinandos deveriam trazer elementos recontextualizados do material de insumo, como também argumentar com suas opiniões. Dessa forma, assim como na tarefa anterior, o texto solicitado nesta tarefa configura-se como uma "Carta ao leitor", e demanda um posicionamento.

4.2.2 PRODUÇÕES ESCRITAS REALIZADAS PELOS EXAMINANDOS

Nesta pesquisa, foram utilizados parte dos textos do *corpus* de produções escritas do exame Celpe-Bras (Schoffen et al, em preparação), assim, somente produções-resposta às tarefa IV das edições 2016/2 e 2017/1 foram analisadas. Para possibilitar a comparação entre os níveis de certificação, foram analisados cinquenta textos avaliados com nota 2 (Intermediário) e cinquenta textos avaliados com nota 5 (Avançado Superior) em cada uma das tarefas, resultando em um total de 200 textos analisados. A escolha desses níveis de proficiência se deu a partir do objetivo de realizar uma comparação entre o nível mais alto de certificação (Avançado Superior) e o nível mais baixo certificado pelo exame (Intermediário). No Quadro 6, são apresentadas as informações de composição dos *corpora*.

Quadro 6 - Corpora - Quantidade de textos analisados

CORPORA	NÚMERO DE TEXTOS
2016/2 T4 Nota 5	50
2016/2 T4 Nota 2	50
2017/1 T4 Nota 5	50
2017/1 T4 Nota 2	50
Total	200

Fonte: Elaborado pela autora.

A seleção dos textos foi determinada a partir daqueles que haviam recebido a mesma nota dos dois avaliadores em cada tarefa no processo de avaliação do Celpe-Bras. Dentre os textos com essas características, foram escolhidos randomicamente 50 textos de cada nota para compor esta pesquisa. Para randomizar os arquivos, foi utilizado o recurso "Lista Aleatória" do site *Invertexto*⁸.

Após a digitação dos textos a partir das imagens digitalizadas, foi realizada uma revisão por outra pessoa para garantir a fidelidade aos textos originais dos examinandos. Para assegurar a impossibilidade de identificação dos autores dos textos, nos textos que continham nomes próprios distintos dos que estão nos textos

⁸ Disponível em: <<https://www.invertexto.com/lista-aleatoria>>.

de insumo, esses nomes foram substituídos. A substituição escolhida foi “Bibiana Terra Cambará”⁹, respeitando sempre a quantidade de sobrenomes expressa no texto original.

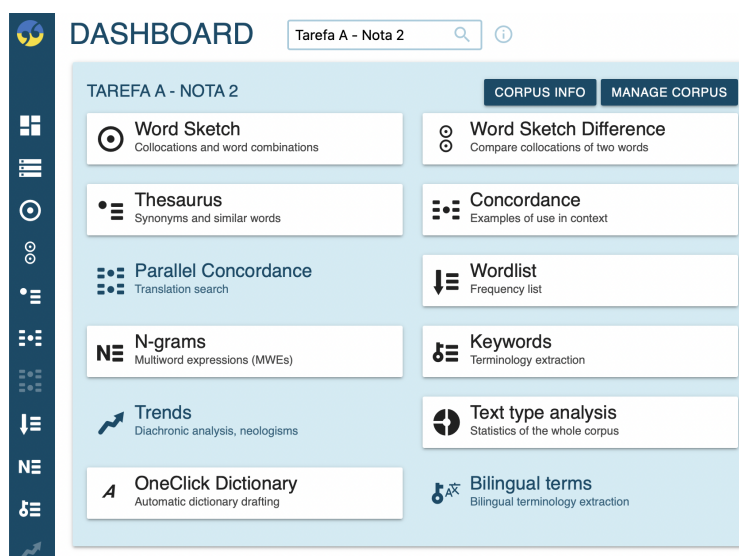
4.3 O SOFTWARE SKETCH ENGINE

Para realizar os procedimentos de análise dos *corpora* de estudo, foi utilizado o programa *Sketch Engine* (Kilgarriff et al., 2004), caracterizado por comportar um conjunto de ferramentas para análise de *corpora*. Além de sua versão gratuita disponível, há a possibilidade de adquirir uma licença para utilizar recursos pagos. Esse programa disponibiliza acesso a pesquisas de vários idiomas, incluindo o Português, sendo esse um diferencial, visto que muitos *softwares* ainda não estão habilitados para esta língua. Além disso, o *Sketch Engine* possibilita lidar com *corpora* de bilhões de palavras, apresentando resultados a partir de diversas funções do programa.

Através do site que hospeda o programa, o pesquisador consegue incluir *corpora* separadamente e pode utilizar as opções de *corpora* de referência de vários idiomas, por exemplo: *Portuguese Web 2018* (ptTenTen18) que contém textos escritos em língua portuguesa brasileira e europeia. Após adicionar os *corpora*, é possível extrair dados sobre o comportamento lexical e gramatical dos textos com as opções: palavras-chave (*keywords*), *n-grams*, e *collocations*. A tela inicial do programa, conhecida como *dashboard*, pode ser observada na Figura 4.

⁹ “Bibiana Terra Cambará” é personagem de uma da série de livros chamada *O Tempo e o Vento*, do autor gaúcho Érico Veríssimo. Essas obras, publicadas entre 1949 a 1962, são consideradas clássicos da literatura do estado e do país e, por isso, o Grupo Avalia escolheu realizar esta homenagem.

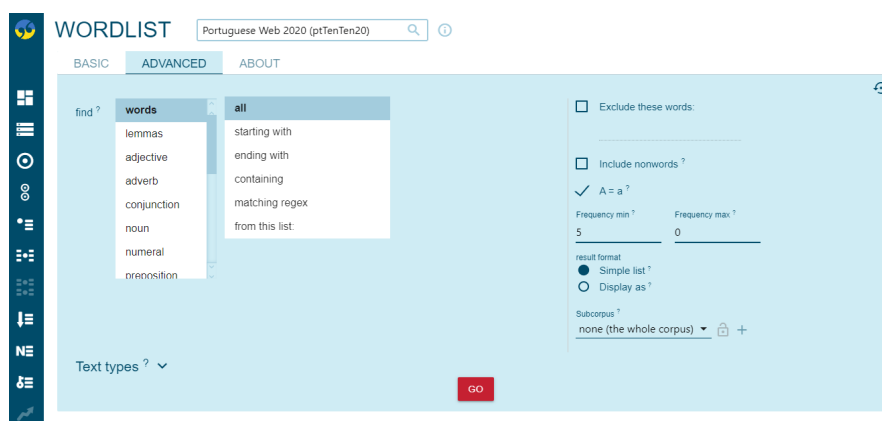
Figura 4 - Dashboard do programa Sketch Engine



Fonte: *Sketch Engine*

Para este trabalho, foi utilizada a opção *Wordlist* (lista de palavras), que permite produzir, a partir do *corpus* selecionado, uma lista de frequência de palavras. Com essa ferramenta, é possível escolher listas específicas de determinadas classes gramaticais, como adjetivos, advérbios, conjunções, preposições, substantivos, numerais, etc. Também é possível definir uma frequência mínima ou máxima para que as palavras da lista gerada apareçam no *corpus*. (Sketch Engine, s. d.). Entretanto, a ferramenta somente extrai uma única palavra por vez, gerando uma lista. Com isso, impedindo a verificação de conjunções compostas por mais de uma palavra, como alguns casos das conjunções subordinadas. Abaixo, na figura 5, podemos ver a ferramenta *Wordlist*:

Figura 5 - Ferramenta Wordlist



Fonte: *Sketch Engine*

4.4 PROCEDIMENTOS PARA AS ANÁLISES

O primeiro passo para efetuar a pesquisa foi a digitação, revisão e compilação dos *corpora*, como descrito anteriormente. Após, os textos foram convertidos no formato *txt* como demandado pelo *software*, para assim ser realizado o *upload* de cada *subcorpora*, individualmente, no *Sketch Engine*.

Com o uso deste programa, é possível obter o número de palavras total dos textos dentro de cada *subcorpora*, que ficam salvos no programa, e também utilizar as ferramentas disponibilizadas do modo que o pesquisador desejar. O *corpus* do presente trabalho totaliza 200 textos-respostas, sendo 50 de cada *subcorpora*, que foram produzidos por examinandos do Celpe-Bras, a partir de duas diferentes tarefas e que receberam notas diferentes (nota 5 e nota 2).

4.4.1 WORDLIST

Como mencionado anteriormente, o *Wordlist* permite que se extraia uma lista com as palavras mais frequentes, a partir da classe gramatical escolhida. Para a extração, basta clicar em "go" e é gerada a lista, nela é possível observar o número de vezes que a palavra aparece e o número de documentos nos quais aparece.

Neste trabalho, com a ferramenta *wordlist* foram extraídas listas usando-se o critério *conjunction*¹⁰.

4.4.2 LOG-LIKELIHOOD

Para compreender melhor os resultados de frequências extraídos de uma comparação entre *corpora*, de acordo com McEnery e Hardie (2012), é necessário utilizar os testes de significância estatística para avaliar se as diferenças são significativas. Esses testes permitem verificar a probabilidade de um resultado ser coincidência e descartar as chances de ser aleatório.

Nesta pesquisa, usamos o *log-likelihood* (LL) (figura 6), que mostra o desvio da frequência de uso de uma palavra em um *corpus* em comparação a outro. (Rayson, Garside, 2000). Para isso, Smith (2009) afirma que são necessárias quatro informações: o número total de palavras em cada *corpus* e o número de vezes que

¹⁰ A ferramenta *wordlist* permite somente a extração de uma única palavra por vez, assim conjunções compostas, como: *além disso*, *com isso*, *a partir disso*, *no entanto*, etc. não poderão ser analisadas integralmente neste trabalho.

a palavra específica aparece em cada *corpus* de estudo. Se o resultado for igual ou maior que 3,34, há cerca de 95% de chance de que essa variação não seja aleatória, sendo assim, estatisticamente significativa. Esse resultado permitiu apoiar as análises realizadas, principalmente para comparar os níveis de proficiência.

Figura 6 - Calculadora de log-likelihood

Log-likelihood and effect size calculator

To use this wizard, type in frequencies for one word and the corpus sizes and press the calculate button.

	Corpus 1	Corpus 2
Frequency of word	<input type="text"/>	<input type="text"/>
Corpus size	<input type="text"/>	<input type="text"/>

Notes:

1. Please enter plain numbers without commas (or other non-numeric characters) as they will confuse the calculator!
2. The LL wizard shows a plus or minus symbol before the log-likelihood value to indicate overuse or underuse respectively in corpus 1 relative to corpus 2.
3. The log-likelihood value itself is always a positive number. However, my script compares relative frequencies between the two corpora in order to insert an indicator for '+' overuse and '-' underuse of corpus 1 relative to corpus 2.

Fonte: Log-Likelihood calculator

4.4.3 CONCORDANCE

Para contabilizar os posição de uso das conjunções, foi utilizado a ferramenta *concordance*. A partir dela, podemos encontrar palavras, frases e expressões e exibir os resultados no contexto dentro nos textos compilados no *corpus*. A *concordance* permite ser utilizada com grandes e pequenos *corpora*, podendo encontrar todos os resultados para serem analisá-los e interpretá-los. Na figura 7, há o exemplo dos resultados encontrados pela *concordance* com a conjunção “e” extraído do *corpus* nota 5 da edição 2016/2.

Figura 7 - Ferramenta *concordance* a partir das extração da conjunção “e” na Tarefa IV da edição 2016/2 - Nota 5

	Left context	KWIC	Right context
1	doc#0 m onibus enquanto atravessa a cidade no meio do transito caótico.	</s><s>Hoje cada minuto é muito valioso	e a tecnologia muito acessível o que favorece o uso deixando de lado ir ate as bibliotecas.
2	doc#1 ser morta.	</s><s>para termos o melhor biblioteca que vai ajudar os futuros quadros a fazer leitura dos livros	e estudar melhor, precisamos de arranjar uma sala bem organizado e confortavel com o ar condicionado, com
3	doc#1 futuros quadros a fazer leitura dos livros e estudar melhor, precisamos de arranjar uma sala bem organizado		e confortavel com o ar condicionado, com a internet banda larga assim para poder fazer melhor as suas pesqu
4	doc#1 lo, com a internet banda larga assim para poder fazer melhor as suas pesquisas sobre net, com muita calma		e silêncio.
5	doc#1 /s><s>Ninguém não deve roubar os livros, si roubar vai ficar muita falta aos outros que querem ler.	</s><s>	E não deve fumar o cigarro la os que costumam fumar si botar o fumo la vai prejudicar a saúde de muitas gente
6	doc#2 : preferências, mentras exista o arte o dibujo o qualquer expressão das mãos sempre existirá libros impresos		e éstos não pousseem morrer porque a tecnologia avança, a tecnologia va em depença de aqueles libros impre
7	doc#3 ario.	</s><s>Tudo isso é muito mais simples que uma verdadeira biblioteca, basta estar conectado a internet	e temos todo o que queremos.
8	doc#3 aira biblioteca, basta estar conectado a internet e temos todo o que queremos.	</s><s>Porém, é mais seguro	e interactivo ir a biblioteca.
9	doc#3 o a internet e temos todo o que queremos.	</s><s>Porém, é mais seguro e interactivo ir a biblioteca.	E podemos ver uma transformação das bibliotecas.
10	doc#3 demos encontrar tudo em uma biblioteca.	</s><s>Vídeos, jornais, revistas, livros em varios idiomas.	E é muito melhor poder experimentar o livro antes de o comprar.
11	doc#4 quando quer livro de qualquer autor o Bibliotecário deve ordenar-lhe para ir procurar livro de aquele autor		e ler depois faz pesquisa, também os escritores devem usar uma política para usar mais Biblioteca do que bai
12	doc#5 cnologia não deve ser utilizado em lado caso, porque vai acabar com varias praticas.	</s><s>Ola meu nome e Bibiana eu sou um leitor de seu jornal	e eu gostei muito de seu publicação porque eu concordo com voce, é
13	doc#5 caso, porque vai acabar com varias praticas.	</s><s>Ola meu nome e Bibiana eu sou um leitor de seu jornal	e eu gostei muito de seu publicação porque eu concordo com voce, é eu acho que as noticias sobre a morte d
14	doc#5 1 pensar, vamos por que nos atraí, porque é bom estar nela.	</s><s>Ser supremamente formosa, confortavel	e atraente para inicio de conversa com um ambiente agradável como as Starbucks com abundância de jornais
15	doc#5 para inicio de conversa com um ambiente agradável como as Starbucks com abundância de jornais, revistas		e livros de um interesse geral. e tudo de forma tecnologica como em lpads, tablets, notebooks, a celular.
16	doc#5 1 ambiente agradável como as Starbucks com abundância de jornais, revistas e livros de um interesse geral.		e tudo de forma tecnologica como em lpads, tablets, notebooks, a celular.
17	doc#5 3>Basta um sistema para todos emprestado rapidamente por isso na nova biblioteca promover conferências		e exposições com espaços creativos para meditar as novas ideias e interação criativa entre os frequentadores.
18	doc#5 na nova biblioteca promover conferências e exposições com espaços creativos para meditar as novas ideias		e interação criativa entre os frequentadores.
19	doc#5 interação criativa entre os frequentadores.	</s><s>Na biblioteca do futuro se pode tomar livros emprestados	e comprar sem nenhum problema.

Fonte: Captura de tela. Disponível em: <<https://auth.sketchengine.eu/>>

A partir desta ferramenta, foram contabilizadas manualmente as posições das conjunções, possibilitando, em um próximo, contabilizar também a quantidade de palavras que antecedem ou sucedem a conjunção.

5 ANÁLISE DOS DADOS

Nesta seção será feita a análise dos dados obtidos através do *software Sketch Engine*, com informações comparativas entre as tarefas, os enunciados, os diferentes níveis de proficiência e o *corpus* de referência.

5.1 CARACTERÍSTICAS DOS CORPORA

Para facilitar a observação dos dados, decidiu-se trabalhar com o mesmo número de textos em todos os *subcorpora*. A análise inicial foi responsável por totalizar o número de palavras em cada *corpus*, informação obtida e calculada através dos dados presentes na página *corpus info*, que mostra os dados numéricos de cada *corpus* no *Sketch Engine*. No Quadro 7, temos as informações:

Quadro 7 - Número de palavras por nota e tarefa

TAREFA	NOTA	NÚMERO DE TEXTOS	NÚMERO DE PALAVRAS	NÚMERO MÉDIO DE PALAVRAS POR TEXTO
2016/2 IV	2	50	7.891	157,82
2016/2 IV	5	50	12.222	244,44
2017/1 IV	2	50	8.376	167,52
2017/1 IV	5	50	11.650	233

Fonte: Elaborado pela autora.

Ao incluir os *subcorpora* no programa, a primeira informação extraída é o número total de palavras dos *corpora* individualmente. Nota-se, ao compará-los, que nas duas tarefas analisadas os *corpora* de Nota 5 são mais extensos que os de Nota 2, demonstrando um certo padrão nos textos analisados e corroborando os estudos anteriores (Divino, 2021; Hanauer, 2023). Dessa maneira, percebe-se que os textos produzidos em resposta às duas tarefas e que receberam nota 2 são mais curtos que os textos produzidos em resposta às duas tarefas que receberam nota 5. Além disso, é possível perceber que os textos produzidos em resposta à tarefa 4 da edição 2017/1 têm menos palavras que os produzidos em resposta à tarefa 4 da edição 2016/2. Destaca-se que tanto na nota 2 quanto na nota 5 os textos são menores na edição 2017/1, o que pode indicar que essa tarefa induzia a produção de textos menos extensos. Isso pode ter acontecido em razão de o material de insumo da edição 2017/1 trazer uma pergunta explícita, o que pode ter induzido os candidatos a responderem de forma direta, produzindo, assim, textos mais curtos do que os encontrados na edição 2016/2.

5.2 WORDLIST

Nesta etapa, serão apresentados quadros com as conjunções extraídas do programa *Sketch Engine*, junto às frequências encontradas em cada *corpus*. Assim, como explicado anteriormente, a ferramenta *Wordlist* permite selecionar a classe gramatical desejada, neste caso as conjunções.

5.2.1 CONJUNÇÕES

Retomando as definições anteriores, as conjunções têm o papel de conectar as sentenças e podem ser categorizadas de acordo com o sentido que fornecem ao conectar dois termos oracionais. Abaixo, estão os quadros com as conjunções extraídas de acordo com as tarefas e os níveis de proficiência.

5.2.1.1 2016/2 Tarefa IV

A tarefa IV da edição 2016/2, para lembrar, solicitava ao examinando que escrevesse uma “Carta ao Leitor”, respondendo a pergunta da autora sobre o consumismo e as compras desnecessárias que podemos fazer. Além disso, o examinando teve que cumprir o propósito de “posicionar-se” diante dos comentários feitos na reportagem. Visando, abaixo, conhecer as conjunções encontradas, segue o Quadro 8 com as extrações feitas nos textos classificados como Nota 5 (Nível Avançado Superior).

Quadro 8 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 5 da edição 2016/2

2016/2 Tarefa IV Nota 5			
Conjunções	Ocorrências no <i>corpus</i>	Frequência em textos	Frequência relativa por documento
e	352	50	100%
que	218	49	98%
ou	67	33	66%
mas	61	32	64%
como	58	36	72%
porque	35	20	40%
nem	17	14	28%

se	16	12	24%
pois	15	10	20%
porém	13	13	26%
embora	5	5	10%
portanto	3	3	6%
contudo	3	2	4%
apesar	3	3	6%
enquanto	1	1	2%
Total	867	-	-

Fonte: Elaborado pela autora.

Para compor esta análise, extraiu-se as ocorrências do programa Skech Engine. Na segunda coluna “Ocorrências no *corpus*”, são apresentadas as ocorrências totais por *corpus*, mostrando o número total de vezes que a conjunção apareceu. Na terceira coluna, “Frequência em textos”, temos o número de textos do *corpus* em que aparece a conjunção selecionada para a extração. Por último, a coluna “Frequência relativa por documento” indica a porcentagem de textos pertencentes ao *corpus* que contém a conjunção pré-selecionada.

Como resultado desta extração, podemos perceber que há um maior número de usos de “e” nos textos em comparação com as demais conjunções, aparecendo em 100% das produções, além dessa conjunção ser utilizada em média sete vezes por texto, resultado que foi obtido a partir do cálculo do número total de conjunções dividido pelo total de textos.

Em seguida, apresentamos o Quadro 9, com as conjunções encontradas nos textos classificados como Nota 2 (Nível Intermediário) na tarefa 2016-2.

Quadro 9 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 2 da edição 2016/2

2016/2 Tarefa IV Nota 2			
Conjunções	Ocorrências	Frequência	Frequência

	no <i>corpus</i>	em textos	relativa por documento
e	191	48	96%
que	161	47	94%
mas	34	23	46%
ou	28	15	30%
como	22	19	38%
porque	14	13	26%
se	10	7	14%
nem	6	6	12%
apesar	4	3	6%
portanto	3	2	4%
pois	2	2	4%
embora	2	2	4%
porém	1	1	2%
Total:	478		

Fonte: Elaborado pela autora.

Nos textos avaliados com nota 2, destaca-se que a conjunção “e” não é utilizada em 100% dos textos, diferentemente do resultado anterior; existem 4 textos que não a utilizam. Podemos ver também que há uma queda no uso de todas as conjunções, quando comparamos com o *corpus* de textos nota 5; por exemplo, “mas” é usado em 64% nos textos que receberam a nota 5, porém em apenas 46% dos textos que receberam nota 2.

5.2.1.2 2017/1 Tarefa IV

A tarefa IV da edição 2017/1 tem como objetivo propor ao examinando a escrita de um texto do gênero “Carta do Leitor”, respondendo ao questionamento

feito em uma reportagem sobre o futuro das bibliotecas diante das novas tecnologias. O enunciado solicitava também que o texto apresentasse um posicionamento claro, concordando ou não com o autor do texto de insumo. Em seguida, o Quadro 10 apresenta as conjunções extraídas das produções que receberam nota 5 .

Quadro 10 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 5 da edição 2017/1

2017/1 Tarefa IV Nota 5			
Conjunções	Ocorrências no <i>corpus</i>	Frequência por texto	Frequência relativa por documento
e	343	50	100%
que	215	48	96%
ou	94	45	90%
como	71	37	74%
mas	38	29	58%
porque	19	14	28%
pois	12	10	20%
se	12	11	22%
porém	9	9	18%
nem	7	6	12%
portanto	7	6	12%
embora	5	5	10%
apesar	4	4	8%
caso	3	3	6%
contudo	2	2	4%

Total:	841		
--------	-----	--	--

Fonte: Elaborado pela autora.

No Quadro acima, notamos que a conjunção “e” aparece em 100% dos textos do *corpus*, cerca de 6 vezes por texto. Nesta edição, os examinandos avaliados com a nota 5 utilizaram mais vezes as conjunções “ou” e “como” quando comparamos com o *corpus* de nota 5 da edição 2016/2. A temática da tarefa pode ter envolvimento com a escolha desses conectivos, visto que o enunciado pedia ao examinando que falasse sobre a morte das bibliotecas ou a adaptação tecnológica que deveria ser feita. Além disso, o examinando deveria citar modos como as bibliotecas poderiam se renovar para novos públicos. Ainda, vale ressaltar que ambos os *corpora* de nota 5 tiveram as conjunções “e”, “que”, “ou”, “como” e “mas” como mais usadas, nesta ordem.

Abaixo, é possível verificar as frequências das conjunções presentes nos textos que receberam nota 2 nesta tarefa (Quadro 11).

Quadro 11 - Conjunções presentes nas produções escritas da Tarefa IV, avaliados com a nota 2 na edição 2017/1

2017/1 Tarefa IV Nota 2			
Conjunções	Ocorrências no <i>corpus</i>	Frequência em textos	Frequência relativa por documento
e	275	49	98%
que	128	42	84%
ou	60	37	74%
mas	42	27	54%
como	38	24	48%
porque	35	19	38%
porém	6	4	8%

pois	5	4	8%
portanto	4	3	6%
se	4	4	8%
nem	3	3	6%
embora	2	2	4%
pero	1	1	2%
caso	1	1	2%
apesar	1	1	2%
Total:	605	-	-

Fonte: Elaborado pela autora.

Assim como nos textos que receberam nota 2 na edição 2016/2, não houve a presença da conjunção “e” em 100% dos textos, mas essa continuou sendo a conjunção mais frequente em todos os *subcorpora*. Percebe-se, também, uma queda no uso da conjunção “como” quando comparamos com os textos que receberam nota 5 na mesma edição.

No Quadro 12, há uma comparação do número total de conjunções utilizadas em cada *subcorpora*.

Quadro 12 - Total de conjunções presentes nos corpora

		Frequência Absoluta	Frequência em relação ao número de palavras do subcorpora
2016/2	Nota 5	867	0.070
	Nota 2	478	0.060
2017/1	Nota 5	841	0.0721
	Nota 2	605	0.0722

Fonte: Elaborado pela autora.

Os dados acima indicam que os textos avaliados com nota 5 apresentam um número total de conjunções maior em relação aos textos avaliados com nota 2 nas

duas tarefas analisadas. Na edição 2016/2, a o *corpus* de nota 5 conta com 867 conjunções, enquanto o *corpus* de nota 2 conta com somente 478. Já na edição 2017/1 a diferença é menor, com 841 conjunções no *corpus* de nota 5 e 605 no *corpus* de nota 2. Sobre a edição 2017/1, percebemos que esse contraste não é tão expressivo como na edição anterior, sendo que os textos de nota 2 parecem apresentar mais conjunções do que na outra edição, visto que, calculado pelo número de palavras, eles utilizam quase o mesmo número de conjunções que os textos de nota 5. Além disso, destaca-se um maior uso de “mas” e “porque” pelos examinandos avaliados com a nota 2, quando comparado com os avaliados com a nota 5 nesta edição.

5.2.1.3 Comparando as notas

Nesta etapa, iremos comparar as conjunções encontradas nos *corpora* de estudo, verificando a frequência em cada nota de acordo com a edição. No Quadro 13, abaixo, temos as conjunções presentes nos textos avaliados em ambas as notas na edição 2016/2, ao lado de suas frequências extraídas anteriormente.

Quadro 13 - Comparação de frequências entre as conjunções presentes nas notas 5 e 2 da Tarefa IV da edição 2016/2.

Conjunções	2016/2 Tarefa IV - Nota 5		2016/2 Tarefa IV - Nota 2	
	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>
e	352	0.028	191	0.025
que	218	0.018	161	0.021
ou	67	0.0055	28	0.0035
mas	61	0.0049	34	0.0044
como	58	0.0047	22	0.0028

porque	35	0.0029	14	0.0018
nem	17	0.0014	6	0.0007
se	16	0.0013	10	0.0012
pois	15	0.0012	2	0.0002
porém	13	0.0010	1	0.0001
embora	5	0.0004	2	0.0002
portanto	3	0.0002	3	0.0003
contudo	3	0.0002	-	-
apesar	3	0.0002	4	0.0004
enquanto	1	0.00008	-	-

Fonte: Elaborado pela autora.

No Quadro acima, podemos notar, com as comparações entre as notas da edição 2016/2, que os textos avaliados com a nota 5 utilizam um número maior de conjunções do que os avaliados com a nota 2, lembrando que os textos também são mais extensos em relação ao número total de palavras. Ademais, os textos avaliados com nota 2 usam cerca de metade das conjunções quando comparamos com os textos avaliados com a nota 5, como por exemplo a conjunção “ou” que está presente 67 vezes nos textos de nota 5 e somente em 28 vezes nos textos de nota 2. Entretanto, isso não ocorre com todas as conjunções, visto que as demais contam um número baixo de uso. A comparação entre os dados permitiu perceber que no *corpus* de nota 2 não há a presença das conjunções “enquanto” nem “contudo”, somente no *corpus* de nota 5, mas aparecendo pouquíssimas vezes. Além disso, permitiu também notar que a conjunção “que”, percentualmente, é mais frequente na nota 2 do que na nota 5, indicando uma maior presença nos textos.

Já na edição 2017/1, temos abaixo o Quadro 13, com os dados extraídos de ambas as notas dos *corpora* da edição.

Quadro 14 - Comparação de frequências entre as conjunções presentes nas notas 5 e 2 da Tarefa IV da edição 2017/1.

Conjunções	2017/1 Tarefa IV - Nota 5		2017/1 Tarefa IV - Nota 2	
	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>
e	343	0.030	275	0.032
que	215	0.019	128	0.015
ou	94	0.0080	60	0.0071
mas	38	0.0033	42	0.0050
como	71	0.0060	38	0.0045
porque	19	0.0016	35	0.0041
nem	7	0.0006	3	0.0003
se	12	0.001	4	0.0004
pois	12	0.001	5	0.0005
porém	9	0.0007	6	0.00071
embora	5	0.0004	2	0.00023
portanto	7	0.0006	4	0.00047
contudo	2	0.0001	-	
apesar	4	0.0003	1	0.00011
enquanto	-	-	-	-
caso	3	0.0002	1	0.00011

Fonte: Elaborado pela autora.

Neste Quadro, podemos ver que os textos avaliados com a nota 2 apresentam um maior número de conjunções “e”, “mas” e “porque” em relação ao número de palavras do *corpus*, quando comparamos os textos avaliados com a nota 5. Entre os Quadros 13 e 14, diferencia-se a frequência de uso das conjunções, visto que na edição 2017/1, o *corpus* de nota 2 conta com um número maior de conjunções nos textos, já na edição 2016/2, a nota 5 conta com um percentual maior de uso de conjunções. Ademais, comparando os resultados das edições, destaca-se, especialmente, o conectivo “ou”, que teve um aumento de uso na edição 2017/1 em relação a edição 2016/2, enquanto a conjunção “porque” teve queda nas ocorrências.

No Quadro 15 abaixo, são comparados os *corpora* de notas 5 das duas edições.

Quadro 15 - Comparação de frequências entre as notas 5 das edições 2016/2 e 2017/1.

Conjunções	2016/2 Tarefa IV - Nota 5			2017/1 Tarefa IV - Nota 5		
	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência relativa por documento	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência relativa por documento
e	352	0.0289	100%	343	0.030	100%
que	218	0.0179	98%	215	0.019	96%
ou	67	0.0055	66%	94	0.0080	90%
mas	61	0.0049	64%	38	0.0033	58%
como	58	0.0047	72%	71	0.0060	74%
porque	35	0.0029	40%	19	0.0016	28%
nem	17	0.0014	28%	7	0.0006	12%
se	16	0.0013	24%	12	0.0010	22%
pois	15	0.0012	20%	12	0.0010	20%

porém	13	0.0010	26%	9	0.0007	18%
embora	5	0.0004	10%	5	0.0004	10%
portanto	3	0.0002	6%	7	0.0006	12%
contudo	3	0.0002	4%	2	0.0001	4%
apesar	3	0.0002	6%	4	0.0003	8%
caso	-	-	-	3	0.0002	6%
enquanto	1	0.00008	2%	-	-	-
Total:	867			841		

Fonte: Elaborado pela autora.

Em ambas as edições, as produções avaliadas com a nota 5 demonstram o uso de um número mais elevado de conjunções, quando comparamos com os textos avaliados com a nota 2. Há algumas conjunções com resultados expressivos em comum entre as duas tarefas, por exemplo: “e”, “que”, “como”, que normalmente são usadas para acrescentar ou explicar informações. Poderíamos constatar que os examinandos optam mais por essas conjunções em seus textos. Porém, ressalta-se também que as conjunções “mas” e “porque” contaram com um uso maior na edição de 2016/2. Contudo, verifica-se um uso maior de “ou” na edição de 2017/1, podendo ter conexão com o propósito da tarefa, que solicitava posicionamento do autor em relação a duas possibilidades. Vale destacar também que algumas conjunções foram utilizadas somente em uma das edições, visto que “caso” foi encontrada apenas na edição de 2017/1 e “enquanto” na edição de 2016/2. No entanto, tais conjunções aparecem poucas vezes.

Em seguida, apresentamos o Quadro comparativo referente ao *corpus* de textos avaliados com nota 2 nas edições analisadas.

Quadro 16 - Comparação de frequências entre as notas 2 das edições 2016/2 e 2017/1.

Conjunções	2016/2 Tarefa IV - Nota 2			2017/1 Tarefa IV - Nota 2		
	Frequência	Frequência	Frequência	Frequência	Frequência	Frequência

	Absoluta	em relação ao número de palavras do <i>subcorpora</i>	relativa por documento	Absoluta	em relação ao número de palavras do <i>subcorpora</i>	relativa por documento
e	191	0.025	96%	275	0.032	98%
que	161	0.021	94%	128	0.015	84%
ou	28	0.0035	30%	60	0.0071	74%
mas	34	0.0044	46%	42	0.0050	54%
como	22	0.0028	38%	38	0.0045	48%
porque	14	0.0018	26%	35	0.0041	38%
nem	6	0.0007	12%	3	0.00035	6%
se	10	0.0012	14%	4	0.00047	8%
pois	2	0.0002	4%	5	0.00059	8%
porém	1	0.0001	2%	6	0.00071	8,5%
embora	2	0.0002	4%	2	0.00023	4%
portanto	3	0.0003	4%	4	0.00047	6%
contudo	-	-	-	-	-	-
apesar	4	0.0004	6%	1	0.00011	2%
caso	-	-	-	1	0.00011	2%
Total:	478			605		

Fonte: Elaborado pela autora.

Os textos avaliados com nota 2 em ambas as edições apresentaram dados em comum em relação à frequência da conjunção “e”, a mais frequente em ambos os *subcorpora*. Entretanto, há um contraste de ocorrência notável entre as conjunções “e”, “ou” e “mas”, tendo sido mais usadas na edição 2017/1 quando comparamos com os dados da edição 2016/2. Percebe-se também que, além de serem mais

extensos em relação aos textos da edição 2016/2, os textos nota 2 da edição 2017/1 contam com um número maior de conjunções, destacando-se o uso de “mas” e “porque”.

Outrossim, no Apêndice A, temos como visualizar a comparação de ambas edições. A partir disso, percebe-se que a edição de 2017/1 tem resultados semelhantes na coluna de frequência em relação ao número de palavras do *subcorpora*, em oposição a edição 2016/2, em que os *subcorpora* de nível Avançado Superior e Intermediário têm resultado diferentes.

5.2.1.4 As categorias das conjunções

As conjunções são divididas em dois tipos pelas gramáticas: conjunções coordenativas e conjunções subordinativas. As conjunções coordenativas unem duas orações independentes. Já as conjunções subordinativas conectam orações dependentes, assim uma completa o sentido da outra. Este tipo de conjunção, como mencionado anteriormente, não será analisado detalhadamente neste estudo, somente contabilizaremos as conjunções subordinadas que são formadas por uma única palavra. Como citado anteriormente, as conjunções subordinadas, normalmente, são formadas por mais de uma palavra. Como a ferramenta utilizada neste estudo para extrair as listas não permite a extração de duas palavras juntas, sugere-se que essas conjunções sejam objeto de estudo de um próximo trabalho.

Neste trabalho, portanto, as conjunções coordenativas serão analisadas a partir de uma divisão feita por Azeredo (2011) na Gramática Houaiss, visto que este tipo aparece em maior número. No Quadro abaixo, são apresentadas as conjunções coordenativas dispostas em cinco classes: aditivas, adversativas, alternativas, conclusivas e explicativas. Para cada classe, são apresentados exemplos da lista de conjunções que as compõem. As conjunções apresentadas foram extraídas das produções dos examinandos a fim de contabilizar qual a categoria mais utilizada em cada nota e em cada edição analisada nesta pesquisa.

Quadro 17 - Categorização das conjunções - 2016/2

Categorias (Azeredo,2011)	Conjunções	Edição 2016/2 - Tarefa IV			
		Nota 5		Nota 2	
		Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>
Conjunção Aditiva	e, nem	369	0.030	197	0.032
Conjunção Adversativa	mas, porém, contudo	80	0.0061	39	0.0041
Conjunção Alternativa	ou	67	0.0055	28	0.0080
Conjunção Explicativa	porque, pois	50	0.0041	16	0.0026
Conjunção Conclusiva	portanto	3	0.0002	3	0.0006

Fonte: Elaborado pela autora.

Com o Quadro acima, podemos perceber que as conjunções explicativas apresentam um contraste em suas frequências, dado que, nos textos de nota 5, há 50 ocorrências, porém, nos de nota 2, encontram-se somente 16. Ainda, percebe-se que os examinandos utilizaram as conjunções aditivas em maior número em relação as demais conjunções tanto nos textos avaliados com nota 5 do que nos textos avaliados com nota 2. Observamos também a presença da conjunção “ou” que é superior na nota 2 do que na nota 5 desta edição.

Quadro 18 - Categorização das conjunções - 2017/1

Categorias (Azeredo, 2011)	Conjunções	Edição 2017/1 - Tarefa IV			
		Nota 5		Nota 2	
		Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>
Conjunção Aditiva	e, nem	350	0.032	278	0.033
Conjunção Adversativa	mas, porém	49	0.004	48	0.005
Conjunção Alternativa	ou	94	0.0030	60	0.0071
Conjunção Explicativa	porque, pois	31	0.002	40	0.004
Conjunção Conclusiva	portanto	7	0.0003	4	0.0004

Fonte: Elaborado pela autora.

Nos resultados acima apresentados, nesta edição, temos números parecidos em ambas as notas no que diz a respeito à quantidade de conjunções que marcam oposição, com 49 ocorrências na nota 5 e 48 na nota 2. As conjunções aditivas foram utilizadas com maior frequência, assim como na edição anterior. É possível destacar, também, as conjunções explicativas, que foram utilizadas 31 vezes no *corpus* de nota 5 e 40 vezes no *corpus* de nota 2, demonstrando um resultado diferente do encontrado nas demais conjunções e também nas notas da edição 2016/2.

Os dados dos Quadros 17 e 18 corroboram o estudo feito por Sostruznik (2022), que mostra maior frequência no uso de conjunções aditivas em textos das tarefas III e IV da edição 2015/2 quando comparadas com as demais categorias. Além disso, percebemos que, quanto aos tipos de conjunções (coordenativas e subordinativas), há, nos dados analisados, uma maior frequência de conjunções

coordenativas encontradas do que conjunções subordinativas¹¹. Somando os tipos de conjunções, temos como resultado o Quadro abaixo:

Quadro 19 - Tipos de conjunções

		Tipos de Conjunções			
		Coordenativas		Subordinativas	
		Frequência Absoluta	Frequência em relação ao número de palavras	Frequência Absoluta	Frequência em relação ao número de palavras
2016/2	Nota 5	569	0.046	298	0.0242
	Nota 2	283	0.035	195	0.0241
2017/1	Nota 5	531	0.045	310	0.026
	Nota 2	430	0.051	175	0.021

Fonte: Elaborado pela autora.

A partir da comparação acima, percebe-se a pouca frequência das conjunções subordinativas, já que percentualmente, têm a metade das ocorrências das conjunções coordenativas. Nos *corpora* de estudo, destacam-se as frequências em relação ao número de palavras de cada *subcorpora*, que mostram resultados semelhantes em questão ao número baixo de conjunções subordinativas utilizadas. Comparando somente os subcopora de nota 2, eles diferenciam-se nos resultados quanto às conjunções coordenativas, dado que os textos da edição 2016/2 apresentam menos conjunções do que os textos da edição 2017/1. Portanto, retomando que as conjunções subordinativas, muitas vezes, têm mais de uma palavra conectando as orações, ficando, assim, de fora da análise deste trabalho.

5.3 LOG-LIKELIHOOD

Quanto ao teste estatístico do *Log-Likelihood*, usado para comparar a frequência de palavras entre dois *corpora*, serão comparados os *corpora* de notas 5

¹¹ Entretanto, lembramos que o programa utilizado apenas extrai listas com as conjunções formadas por uma única palavra, assim, pode haver mais conjunções subordinativas que não foram extraídas.

e de notas 2, utilizando as frequências absolutas de uso de conjunções extraídas pelo *software Sketch Engine*.

Os resultados serão considerados significativos se forem iguais ou superiores ao número 3,34, mostrando que há 95% chances de não se tratar de um evento aleatório da amostra de textos. O resultado pode ser positivo (+) ou negativo (-), dependendo de qual *subcorpora* terá um maior número de ocorrências. Para padronizar o procedimento na calculadora, os dados que contiverem um sinal positivo (+) indicam que há mais ocorrências no *subcorpora* de nota 5. E se obtiverem um resultado negativo (-), há uma maior predominância no *subcorpora* nota de 2. Dessa forma, segue abaixo o Quadro 19 com as informações comparativas referentes à edição 2016/2.

Quadro 20 - Comparativo do *corpus* 2016/2 T4, utilizando o teste estatístico do Log-Likelihood.

Edição 2016/2			
Conjunção	Nota 5 (+)	Log-likelihood	Nota 2 (-)
	Frequência absoluta		Frequência absoluta
e	352	+3.80	191
que	218	-1.66	161
ou	67	+3.94	28
mas	61	+0.48	34
como	58	+4.84	22
porque	35	+2.44	14
nem	17	+1.76	6
se	16	+0.01	10
pois	15	+6.37	2
porém	13	+7.62	1
embora	5	+0.35	2

portanto	3	-0.28	3
apesar	3	-0.91	4

Fonte: Elaborado pela autora.

Como resultado da edição 2016/2, a partir do LL, pode-se afirmar que “e”, “ou”, “como”, “pois” e “porém” ocorreram significativamente mais no *corpus* de textos nota 5 do que no de textos nota 2, tendo um valor acima de 3,34. Quanto ao restante das conjunções, não há diferenças significativas na frequência de uso entre as duas notas. Em seguida, apresentamos a comparação, a partir do *Log-Likelihood*, da edição 2017/1.

Quadro 21 - *Comparativo do corpus 2017/1 T4, utilizando o teste estatístico do Log-Likelihood.*

Edição 2017/1			
Conjunção	Nota 5 (+)	Log-likelihood	Nota 2 (-)
	Frequência		Frequência
e	343	-1.80	275
que	215	+2.90	128
ou	94	+0.52	60
mas	38	-3.69	42
como	71	+2.21	38
porque	19	-11.55	35
nem	7	+0.60	3
se	12	+1.98	4
pois	12	+1.12	5
porém	9	+0.02	6
embora	5	+0.53	2
portanto	7	+0.14	4

apesar	4	+1.07	1
caso	3	+0.49	1

Fonte: Elaborado pela autora.

A partir do Quadro acima, percebe-se que “mas” e “porque” obtiveram resultados acima de 3,34 e ambos negativo, sendo mais frequente no *subcorpora* de nota 2 e, com isso, são considerados dados significantes pela calculadora.

Dessa forma, entende-se como significativa a maior presença de “mas” e “porque” no *corpus* de nota 2 do que no *corpus* de nota 5, o que pode indicar, talvez, que os textos mais proficientes tendem a utilizar outros recursos para explicar as informações, além de “mas” e “porque”; no entanto, haveria a necessidade de outros estudos para comprovar isso. Além disso, há a hipótese de que os textos de nível Avançado Superior possam variar as conjunções, utilizando outras com as mesmas funções.

5.4 POSIÇÃO DAS CONJUNÇÕES

De acordo com Perini (2011), as conjunções podem apresentar duas posições padrões: posição inicial ou não-inicial. Além disso, em Matos (2003), temos que as conjunções situadas em contexto de uso inicial podem demonstrar uso de um recurso estilístico mais elaborado nas produções escritas, dado que não afetam a compreensão, apenas se tratam de um deslocamento opcional.

Para analisar o contexto de uso das conjunções dentro do texto, é necessário utilizar a ferramenta *Concordance*, presente no *Sketch Engine*. Após a extração das conjunções a partir do *Wordlist*, é possível fazer a leitura e contagem das amostras de trechos de textos em que as conjunções aparecem. Na figura 8, há exemplificado a possibilidade de leitura do material e notar o uso dentro dos textos.

Figura 8 - Acesso ao texto com a ferramenta *concordance* a partir das extração da conjunção “e” na Tarefa IV da edição 2016/2 - Nota 5

Fonte: Captura de tela. Disponível em: <<https://auth.sketchengine.eu/>>

Para não estender demais este trabalho, apresentamos os dados já contabilizados nos Quadros 21 e 22, contendo apenas com as informações de posição das cinco conjunções mais frequentes em todos os *corpora*, que são “e”, “que”, “ou”, “como” e “mas”.

Quadro 22 - Comparando as posições das conjunções presentes na edição 2016/2.

2016/2 Tarefa IV				
Conjunções	Posição Inicial		Posição não inicial	
	Frequência absoluta		Frequência absoluta	
	Nota 5	Nota 2	Nota 5	Nota 2
e	12	5	340	186
que	1	1	217	160
ou	0	1	67	27
como	10	4	48	18
mas	9	8	52	26

Fonte: Elaborado pela autora.

Analisando o Quadro acima, podemos perceber que os textos que receberam nota 5 utilizam as conjunções em posição inicial de sentença com uma frequência maior que os que receberam nota 2; contudo, mesmo na nota 5 esse mecanismo é pouco utilizado. Além disso, os textos de ambas as notas assemelham-se no uso da conjunção “que” e “ou” utilizando-as em posição não inicial. Nos poucos casos de posição inicial, foi encontrada a conjunção “que” sendo utilizada como pronome interrogativo; e “ou” sendo utilizado para explicar uma definição ou argumento trazido na frase anterior. Podemos inferir que os examinandos, em ambos os níveis de proficiência, optam, nesta amostra de textos, por utilizar todas as conjunções em posição não-inicial, conectando a oração coordenada ou dependente.

Portanto, é possível dizer que, nos textos analisados que receberam nota 5 e nota 2, nesta tarefa, houve uma preferência pelo uso das conjunções como conectores de sentenças, assim como abordado por Perini (2001), que afirma que este é o uso mais comum das conjunções.

Agora, analisando a tarefa IV da edição 2017/1, foi elaborado um Quadro com as comparações de resultados encontrados na edição 2017/1 em ambas as notas.

Quadro 23 - Comparando as posições das conjunções presentes na edição 2017/1.

2017/1 Tarefa IV				
Conjunções	Posição Inicial		Posição não inicial	
	Frequência absoluta		Frequência absoluta	
	Nota 5	Nota 2	Nota 5	Nota 2
e	7	18	336	257
que	3	2	212	126
ou	0	2	94	58
como	8	3	63	35

mas	9	6	29	36
-----	---	---	----	----

Fonte: Elaborado pela autora.

No Quadro 22, podemos perceber que os textos avaliados com nota 5 na edição 2017/1 não utilizaram a conjunção “e” em posição inicial com tanta frequência quanto os textos avaliados com nota 2, embora ambos os resultados tragam um número muito pequeno de usos. Inclusive, ambas as notas têm um uso comum das conjunções “que” e “ou” em posições no interior das sentenças. Por conseguinte, podemos notar que, nesta edição, nos textos de nota 2, houve um aumento no uso do “e” em posição inicial em relação à mesma nota da edição anterior, analisada no Quadro 21, porém não nas demais conjunções, que foram utilizadas, em sua maioria, em posição não-inicial, apenas unindo as orações principais com as secundárias.

6 DISCUSSÃO DOS RESULTADOS

Os procedimentos de análise realizados foram ao encontro dos objetivos propostos anteriormente. Assim, este trabalho visou analisar as conjunções utilizadas pelos examinandos em dois níveis de proficiência em tarefas IV pertencentes a duas edições distintas do Celpe-Bras. As extrações de dados foram feitas a partir dos textos-resposta produzidos pelos examinandos e compilados em dois *corpora* para cada tarefa, formados por textos que receberam nota 5 e nota 2, a fim de caracterizar dois dos níveis de proficiência certificados pelo exame.

Ambas as tarefas IV escolhidas para a composição desta pesquisa solicitam a produção de textos do mesmo gênero discursivo e também têm o mesmo propósito estabelecido nos enunciados. Essas semelhanças auxiliam nas comparações, visto que, nas duas edições, era solicitado que os examinandos se posicionassem a favor ou contra os argumentos escritos pelos autores dos materiais de insumo.

O primeiro passo das extrações pelo programa *Sketch Engine* foi verificar a extensão dos *corpora*. O número total de palavras de cada *subcorpus* (cada nota) apresentou um contraste entre as notas, visto que os *subcorpora* dos textos avaliados com nota 5 apresentaram 12.222 palavras (edição 2016/2) e 11.650

palavras (edição 2017/1), enquanto os *subcorpora* de textos avaliados com nota 2 totalizaram 7.851 (edição 2016/2) e 8.376 (edição 2017/1) palavras, respectivamente. Trabalhos anteriores, como Divino (2021), Hanauer (2023) e Sostruznik (2022), obtiveram resultados semelhantes, analisando edições e tarefas diferentes. Assim, pode-se afirmar que os textos avaliados com nota 5 são, em média, mais extensos que os avaliados com nota 2, como pode ser observado no Quadro 7.

Posteriormente, foram extraídas as conjunções mais frequentes de cada *subcorpora* das edições analisadas. Primeiro, no Quadro 8, temos as conjunções mais frequentes nos textos de nota 5 e logo após nos textos de nota 2 no Quadro 9, ambas da edição 2016/2. Com a comparação dos *subcorpora*, percebe-se que eles tiveram como mais frequentes as mesmas conjunções: “e”, “que”, “ou”, “como” e “mas”. Esses resultados corroboram os resultados de Sostruznik (2022), visto que no *corpus* analisado nesse trabalho essas conjunções também ficaram nas primeiras posições como mais recorrentes.

A conjunção “e” destacou-se como a mais utilizada em todos os *subcorpora*. Com maior frequência de utilizações, podemos dizer que a maioria dos textos deste estudo avaliados com nota 5 e com nota 2 utilizam o conectivo “e” em suas frases. Já a conjunção “que”, embora seja menos frequente, ainda teve uma grande ocorrência em todos os *subcorpora*, demonstrando um bom uso, pelos examinandos, para interligar as sentenças durante a escrita.

A presença de inadequações ortográficas na grafia das palavras, como discutido anteriormente, pode influenciar a análise, principalmente, nos textos avaliados com Nota 2, visto que, de acordo com os parâmetros de avaliação, é possível haver desvios ortográficos e gramaticais, porém sem afetar a compreensão do texto. Essa variação na escrita das palavras poderia demonstrar um aumento no uso de alguns conectivos pelos examinandos avaliados com a nota 2, como variações de “pôrque”, “por que”, “más”, etc.. Entretanto, o programa *Sketch Engine* não consegue extrair diferentes ortografias de uma mesma palavra, o que impediu a realização dessa análise.

No que diz a respeito às conjunções mais frequentes, a edição 2017/1 teve resultados semelhantes à edição 2016-2, permanecendo “e”, “que”, “ou”, “como” e “mas” como mais usadas. Percentualmente, houve um aumento no uso de “que” e “ou” nos textos, e também, a conjunção “caso” teve ocorrência nas duas notas de

2017/1, não tendo aparecido na edição 2016/2. Destacando o uso de “ou” na edição, supõe-se que a tarefa pode ter relação com esse uso, dado que, além de solicitar ao examinando que concordasse ou discordasse da afirmação do autor do material de insumo, sugeria também que expressasse uma opinião sobre a morte ou a adaptação das bibliotecas. Assim, nos textos analisados, há uma presença grande da conjunção “ou” dando uma alternância de ideias.

Comparando as extrações feitas nas duas edições, temos resultados destacáveis quanto ao uso das conjunções “mas”, “ou” e “porque”. Na edição 2016/2, nota 5, as conjunções “mas” e “porque” têm um aumento de uso em relação aos números encontrados na nota 2 da mesma edição. Contudo, comparando com os resultados da edição de 2017/1, o oposto ocorre, já que no *corpus* de nota 2, há um aumento de frequência dessas mesmas conjunções.

Percebemos, também, que há uma diferença entre os tipos de conjunções utilizadas nas produções. As coordenativas tiveram um maior número de ocorrências do que as subordinadas em ambos *corpora* analisados. Porém, destaca-se, novamente, que a ferramenta *Wordlist* apenas extrai os conectivos formados por uma única palavra, sendo assim, é possível haver outras conjunções subordinativas que não foram analisadas. Com isso, há a necessidade de um estudo focado apenas nestes tipo de conectivos, utilizando outra ferramenta, que permita essa análise com mais palavras juntas. Por isso, foi realizada, neste trabalho, somente a análise das categorias de conjunções coordenativas.

Azeredo (2011) categorizou as conjunções coordenativas de acordo com o sentido que elas podem trazer em uma frase. Ao separar as conjunções extraídas pelo programa e categorizá-las, percebeu-se que há uma maior frequência de uso das conjunções aditivas. Esses dados corroboram as descobertas de Sostruznik (2022), que analisou tarefas diferentes e obteve os mesmos resultados quanto à frequência e variedade de exemplos desses tipos de conjunções. Na pesquisa, a autora mencionou também a pouca frequência de uso de conjunções explicativas e conclusivas, corroborando os dados encontrados neste estudo. Dessa maneira, isso demonstra que, neste *corpus* de estudo, os examinandos procuraram conectar as frases e adicionar informações por meio de conjunções aditivas.

Já a análise feita com o auxílio da calculadora *Log-Likelihood* apresentou significância em algumas conjunções nos resultados comparativos entre as notas das duas edições. No caso de 2016/2, destacou-se o uso das conjunções “e”, “ou”,

“como”, “pois” e “porém” que se mostraram mais relevantes nos textos avaliados com a nota 5, do que nas produções de nota 2. Na edição 2017/1, a calculadora apresentou significância apenas na conjunção “mas” e “porque” que são mais frequentes nos textos avaliados com nota 2 do que nos avaliados com nota 5. Podemos dizer, talvez, que os textos de nota 5 possam ter usado recursos linguísticos mais elaborados para fornecer explicações no texto, ou, utilizaram uma maior variedade de outros conectivos. Entretanto, isso só pode ser conferido, em outros estudos, com uma análise qualitativa.

No caso do contexto de uso das conjunções, percebeu-se que os textos de nota 5 da edição de 2016/2 contam com um número maior de conjunções em posição inicial de sentença, quando comparamos com os resultados da nota 2 da mesma edição. No entanto, na edição de 2017/1, o resultado se mostra contrário ao da anterior, uma vez que na nota 2 temos uma maior frequência de uso de conjunções em posição inicial, equiparando com a nota 5 desta edição. Entretanto, a posição não-inicial teve ocorrência majoritária geral no uso de todas as conjunções extraídas pelo programa. O uso de conjunções em posição inicial, de acordo com Matos (2003), demonstra um recurso estilístico de deslocamento, sendo opcional caso não afete o sentido da frase. Esse resultado demonstrou que as produções-resposta estudadas, tanto as que receberam nota 5 quanto as que receberam nota 2, optaram por utilizar conjunções no interior das frases, conectando duas orações, e não em posição inicial.

A partir do que foi mostrado o presente estudo buscou ampliar a pesquisa anterior de Sostruznik (2022) com um estudo mais abrangente do uso das conjunções, partindo de uma análise de *corpora* composto por produções escritas de examinandos do Celpe-Bras. Os resultados obtidos comprovaram que há diferença na quantidade de conjunções utilizadas nos diferentes níveis de proficiência avaliados pelo Celpe-Bras, visto que há uma maior frequência de uso de conjunções nos *corpora* de nota 5 em relação aos *corpora* de nota 2.

Assim como o aspecto analisado neste trabalho, compreendo que diversos aspectos dos textos podem ser analisados à luz da Linguística de Corpus, podendo contribuir para uma descrição mais ampla dos níveis de proficiência avaliados nas tarefas analisadas.

7 CONSIDERAÇÕES FINAIS

O presente trabalho, de forma geral, buscou contribuir para a caracterização dos níveis Intermediário e Avançado Superior em produções escritas de examinandos do Celpe-Bras. Com a utilização de produções-resposta de duas tarefas IV de edições distintas, 2016/2 e 2017/1, conseguiu-se analisar uma amostra de textos avaliados com a nota 5 (Avançado Superior) e com a nota 2 (Intermediário). Ambas as tarefas solicitavam a produção de textos do mesmo gênero discursivo (carta do leitor) e também com o mesmo propósito (posicionar-se), o que permitiu realizar comparações entre as conjunções usadas.

Através do programa *Sketch Engine*, foi possível realizar as devidas extrações. Além disso, o programa permite uma análise automatizada dos *corpora*, sendo, porém, necessária uma leitura mais detalhada para perceber os dados revelados pela ferramenta. Nesse caso, as extrações realizadas permitiram uma posterior análise comparativa entre as edições e as notas, que foram feitas em quadros para que seja possível visualizar com mais clareza possíveis justificativas para resultados esperados ou inesperados.

Os resultados obtidos ampliam o estudo anterior de Sostruznik (2022), que também analisou produções-resposta com uma metodologia baseada em *corpus*, tendo analisado conjunções utilizadas em diferentes tarefas e edições. Ademais, as frequências elevadas de conjunções aditivas corroboram os resultados encontrado no estudo anterior.

No que diz respeito às ocorrências, foi observado maior frequência de uso de conjunções nos *corpora* de nota 5 em comparação com os de nota 2. Para dar continuidade a esta pesquisa, outras notas poderão ser analisadas no futuro, buscando aumentar a comparação do uso das conjunções, visando caracterizar mais detalhadamente as nuances entre os textos de diferentes níveis, além de possibilitar uma descrição mais robusta das produções dos diferentes níveis avaliados na Parte Escrita do Celpe-Bras.

Entre os níveis de proficiência, percebe-se que as produções textuais avaliadas com a nota 5 demonstraram um maior uso de conjunções, sendo assim, pode-se afirmar que examinandos mais proficientes utilizaram mais conectivos em seus textos do que os que foram avaliados com nota 2.

O aumento de pesquisas utilizando ferramentas de Linguística de Corpus com textos em língua portuguesa demonstra a possibilidade de análises diversificadas e resultados empíricos para compreender melhor como se operacionalizam os parâmetros de avaliação dos testes de proficiência, neste caso o Celpe-Bras. As pesquisas de Divino (2021) e Hanauer (2023) são um exemplo disso, dado que as autoras procuraram descrever índices de riqueza lexical e as *keywords* presentes no material de insumo.

Este trabalho traz à tona a necessidade de pesquisas futuras que investiguem mais produções-resposta a outras tarefas e edições do Exame Celpe-Bras. Além da análise de um número maior de textos, seria interessante analisar outros recursos argumentativos de examinandos a fim de verificar seus usos nas produções textuais. Também seria necessário analisar *corpora* compostos pelas demais notas atribuídas aos textos, de modo a investigar as possíveis frequências presentes.

Por fim, instiga-se trabalhos futuros que trabalhem com um número maior de textos e também que utilize outras tarefas que solicitem outros gêneros discursivos. Além disso, é necessário divulgar informações a respeito de características e diferenciações dos níveis de proficiência certificados pelo Celpe-Bras.

REFERÊNCIAS BIBLIOGRÁFICAS

AZEREDO, José Carlos. **Gramática Houaiss da Língua Portuguesa**. 3ª ed. - São Paulo. Publifolha, SP, 2011.

BECHARA, Evanildo. **Moderna gramática portuguesa**. 37ª ed. - Rio de Janeiro: Nova Fronteira, 2009.

BRASIL. **Documento-base do exame Celpe-bras**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020a. 130 p. Disponível em: <https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/avaliacoes-e-exames-da-educacao-basica/documento-base-do-exame-celpe-bras>. Acesso em: 12 jun. 2023.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Caderno de Questões 2016/2: Certificado de Proficiência em Língua Portuguesa para Estrangeiros**. Brasília, 2016. 10 p. Disponível em: https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/PARTE-ESCRITA_CELPEBRAS_2016_II_Caderno-de-questoes.pdf. Acesso em: 12 ago. 2023.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Caderno de Questões 2017/1**: Certificado de Proficiência em Língua Portuguesa para Estrangeiros. Brasília, 2017. 10 p. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/Caderno-de-questoes-2017-1-compactado.pdf>. Acesso em: 10 ago. 2023.

BRASIL. Ministério da Educação. **PEC-G**. Disponível em: <http://portal.mec.gov.br/pec-g>. Acesso em: 24 ago. 2023.

BRASIL. **Portaria Nº 623, de 13 de Novembro de 2020**. Brasília, 2020b. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-623-de-13-de-novembro-de-2020-288547519>.

CALLIES, Marcus; GÖTZ, Sandra. **Learner corpora in language testing and assessment**. Studies In Corpus Linguistics, [S.L.], p. 1-10, 2015. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/scl.70.001int>.

CUNHA, Celso & CINTRA, Lindley. **Nova Gramática do Português Contemporâneo**. Rio de Janeiro: Nova Fronteira, 3a ed..2001.

CUSHING, Sara T. **Corpus linguistics and language testing**. In: FULCHER, Glenn; HARDING, Luke (ed.). The Routledge Handbook of Language Testing. 2. ed. Londres: Routledge, 2022. Cap. 32. p. 345-360.

DIVINO, Luiza Sarmento. **Índices lexicais de análise para a caracterização dos níveis intermediário e avançado superior no exame Celpe-Bras : uma pesquisa guiada por corpus**. 2021. 68 f. TCC (Graduação) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: <https://www.lume.ufrgs.br/handle/10183/235361>. Acesso em: 19 jun. 2023.

DIVINO, Luiza Sarmento; HANAUER, Isadora Dahmer; SOSTRUZNIK, Julia Luiz Sostruznik da. **O nível avançado superior no Celpe-Bras: uma análise de textos de examinandos**. In: I CONGRESSO DE PORTUGUÊS COMO LÍNGUA ESTRANGEIRA., 2021, Online. Caderno de Resumos do I Congresso de Português como Língua Estrangeira. Nova Iorque: 2021. p. 72. Disponível em: <https://www.lrc.columbia.edu/wp-content/uploads/2021/06/I-Congresso-PLC-Columbia-University.pdf>. Acesso em: 19 jun. 2023.

FRASER, Bruce. **What Are Discourse Markers?** In: Journal of Pragmatics, ed. 31, p. 931-952. Boston, 1999. Disponível em: [http://dx.doi.org/10.1016/S0378-2166\(98\)00101-5](http://dx.doi.org/10.1016/S0378-2166(98)00101-5). Acesso em: 28 de jul. 2023.

GOMES, Maíra da Silva. **A complexidade de tarefas de leitura e produção escrita no Exame Celpe-Bras**. 2009. 109 f. Dissertação (Mestrado) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <https://www.lume.ufrgs.br/handle/10183/16228>. Acesso em: 28 jun. 2023.

GRANGER, Sylviane. **The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation**. Studies In

Corpus Linguistics, [S.L.], p. 13-332, 2009. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/scl.33.04gra>. Disponível em: https://www.researchgate.net/publication/235942106_The_contribution_of_learner_corpora_to_second_language_acquisition_and_foreign_language_teaching_A_critical_evaluation. Acesso em: 20 jul. 2023.

GRANGER, Sylviane; WYNNE, Martin. **Optimising measures of lexical variation in EFL learner corpora**. In: J., Kirk. *Corpora Galore*. Amsterdam And Atlanta: Rodopi, 1999. p. 249-257. Disponível em: https://www.researchgate.net/publication/269693046_Optimising_measures_of_lexical_variation_in_EFL_learner_corpora. Acesso em: 20 jul. 2023.

HANAUER, Isadora Dahmer. **Caracterização do avançado superior em tarefa integrada de escrita e compreensão oral no Celpe-Bras**. XXXII Salão de Iniciação Científica, UFRGS. 2020. Disponível em: <https://www.lume.ufrgs.br/handle/10183/227306>. Acesso em: 20 jun. 2023.

HANAUER, Isadora Dahmer. **Caracterização dos Níveis Intermediário e Avançado Superior do exame Celpe-Bras em produções escritas por examinandos no gênero carta/e-mail: contribuições de uma análise guiada por corpus**. Trabalho de Conclusão de Curso (Graduação). UFRGS. 2023 Disponível em: . Acesso em: 20 jun. 2023.

KENNEDY, G.. *Corpus Linguistics. International Encyclopedia Of The Social & Behavioral Sciences*, [S.L.], p. 2816-2820, 2001. Elsevier. <http://dx.doi.org/10.1016/b0-08-043076-7/03056-4>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B0080430767030564>. Acesso em: 20 jul. 2023.

KILGARRIFF, Adam; RYCHLÝ, Pavel; SMRŽ, Pavel; TUGWELL, David. The Sketch Engine. In: **Proceedings of the XI EURALEX International Congress**. Université de Bretagne-Sud, p. 105–116, 2004. Disponível em: https://www.sketchengine.co.uk/wp-content/uploads/The_Sketch_Engine_2004.pdf. Acesso em 20 jul. 2023.

KOCH, Ingedore Villaça. **Ler e escrever: estratégias de produção textual**. 2. ed. São Paulo: Contexto, 2011.

KUNRATH, Simone Paula. **Os Descritores Gerais e a Progressão dos Níveis de Proficiência do Exame Celpe-Bras**. 2019. Tese (Doutorado em Linguística Aplicada) - Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul. Porto Alegre, 2019. Disponível em: <https://www.lume.ufrgs.br/handle/10183/239041>. Acesso em 26 jun. 2023.

MATOS, Gabriela. **Estruturas de coordenação**. In: Mateus, M. H. M. et al. (orgs.) *Gramática de língua portuguesa*. 5. ed. (pp. 549-592). Lisboa, 2003.

MCENERY, Tony; HARDIE, Andrew. **Statistics in corpus linguistics**. 2012. Disponível em: <http://corpora.lancs.ac.uk/clmtp/2-stat.php>. Acesso em: 21 jul. 2023.

MENDEL, Kaiane. **Proficiência e autoria na avaliação integrada de leitura e escrita do exame Celpe-Bras**. 2019. 185 f. Dissertação (Mestrado) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019. Disponível em: <https://www.lume.ufrgs.br/handle/10183/217789>. Acesso em: 21 jun. 2023.

PERINI, Mário A. **Gramática do português brasileiro**. São Paulo: Parábola, 2010.

PERINI, Mário A. **Gramática descritiva do português brasileiro**. Petrópolis, RJ: Vozes, 2016.

QUEVEDO-CAMARGO, Gladys. **Breve história da evolução do construto proficiência em línguas**. Em Aberto, [S.L.], v. 32, n. 104, p. 27, 18 jun. 2019. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. <http://dx.doi.org/10.24109/2176-6673.emaberto.32i104.4286>. Disponível em: <http://rbep.inep.gov.br/ojs3/index.php/emaberto/article/view/4192>. Acesso em: 22 jun. 2023.

QUEVEDO-CAMARGO, Gladys; SCARAMUCCI, Matilde Virgínia Ricardi. **O CONCEITO DE LETRAMENTO EM AVALIAÇÃO DE LÍNGUAS: origem de relevância para o contexto brasileiro**. Linguagem: Estudos e Pesquisas, [S.L.], v. 22, n. 1, p. 225-245, 16 ago. 2018. Universidade Federal de Goiás. <http://dx.doi.org/10.5216/lep.v22i1.54474>. Disponível em: https://www.researchgate.net/publication/327194920_O_CONCEITO_DE_LETRAMENTO_EM_AVALIACAO_DE_LINGUAS_ORIGEM_DE_RELEVANCIA_PARA_O_CONTEXTO_BRASILEIRO. Acesso em: 22 jun. 2023.

RAYSON, Paul. **Log-likelihood and effect size calculator**. 2008. Disponível em: <https://ucrel.lancs.ac.uk/llwizard.html>. Acesso em: 22 jul. 2023.

SARDINHA, Antonio Berber *et al.* **Corpus Brasileiro: coletânea on-line de um bilhão de palavras do Português Brasileiro Contemporâneo**. 2010. Disponível em: <https://www.pucsp.br/pesquisa-seleta-2011/projetos/081.php>. Acesso em: 30 jun. 2023.

SCARAMUCCI, Matilde Virginia Ricardi. **Proficiência em LE: considerações terminológicas e conceituais**. Trabalhos em Linguística Aplicada, Campinas, v. 36, n. 1, p. 11-22, dez. 2000. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/tla/article/view/8639310>. Acesso em: 24 jun. 2023.

SCARAMUCCI, Matilde Virginia Ricardi; QUITZAU, Luciana Amgarten. **Integração leitura/escrita na prova de redação do vestibular Unicamp**. Revista Examen, Brasília, v. 2, n. 2, p. 28, jun. 2018. Disponível em: <https://examen.emnuvens.com.br/rev/article/view/63>. Acesso em: 10 jun. 2023.

SCHLATTER, Margarete; SCARAMUCCI, Matilde V. R.; PRATI, Sílvia; ACUÑA, Leonor. **Celpe-Bras e Celu: impactos da construção de parâmetros comuns de avaliação de proficiência em português e em espanhol**. In: FONTANA, Mónica Graciela Zoppi. O português do Brasil como língua transnacional. Campinas: Editora Rg, 2009. p. 95-122.

SCHOFFEN *et al.* **Estudo descritivo das tarefas da parte escrita do exame Celpe-Bras**. Porto Alegre: Instituto de Letras - Ufrgs, 2018. 76 p. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/Schoffen-et-al-2018.pdf>. Acesso em: 20 jun. 2023.

SCHOFFEN, Juliana Roquele; SIRIANNI, Gabrielle; KUNRATH, Simone. **O perfil argumentativo da Tarefa IV do Exame Celpe-Bras**. Revista Abralín, v. 19, n. 3, p. 909-935, 2020.

SCHOFFEN, Juliana Roquele. **Gêneros do discurso e parâmetros de avaliação de proficiência em português como língua estrangeira no Exame Celpe-Bras**. 2009. 192 f. Tese (Doutorado) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <https://lume.ufrgs.br/handle/10183/16900>. Acesso em: 25 jun. 2023.

SOSTRUZNIK, Julia Luiz. **Marcadores Discursivos nas Tarefas de Leitura e Escrita no Nível Avançado Superior do Exame Celpe-Bras**. XXXIV Salão de Iniciação Científica, UFRGS.. 2022. Disponível em: <https://www.youtube.com/watch?v=7s4RWXj0JCc>. Acesso em: 12 ago. 2023.

SIRIANNI, Gabrielle Rodrigues. **Descrição dos níveis de proficiência em tarefa de leitura e escrita a partir de produções textuais de alunos do curso preparatório CELPE-BRAS**. 2016. 76 f. TCC (Graduação) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016. Disponível em: <https://www.lume.ufrgs.br/handle/10183/157769>. Acesso em: 20 jun. 2023.

SKETCH ENGINE. **User Guide**. Disponível em: <https://www.sketchengine.eu/guide/>. Acesso em: 03 ago. 2023.

SMITH, Nick. **Corpus Linguistics: a practical web-based course**. A Practical Web-based Course. 2009. Disponível em: https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/clc_top.htm. Acesso em: 12 ago. 2023.

WEIGLE, Sara Cushing. **Assessing Writing**. Cambridge: Cambridge University Press, 2002. 282 p. Acesso em: 26 jun. 2023.

APÊNDICE

APÊNDICE A – Quadro comparativo geral das edições

Conjunções	2016/2 Tarefa IV - Nota 5		2016/2 Tarefa IV - Nota 2		2017/1 Tarefa IV - Nota 5		2017/1 Tarefa IV - Nota 2	
	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>	Frequência Absoluta	Frequência em relação ao número de palavras do <i>subcorpora</i>
e	352	0.028	191	0.025	343	0.030	275	0.032
que	218	0.018	161	0.021	215	0.019	128	0.015
ou	67	0.0055	28	0.0035	94	0.0080	60	0.0071
mas	61	0.0049	34	0.0044	38	0.0033	42	0.0050
como	58	0.0047	22	0.0028	71	0.0060	38	0.0045
porque	35	0.0029	14	0.0018	19	0.0016	35	0.0041
nem	17	0.0014	6	0.0007	7	0.0006	3	0.0003
se	16	0.0013	10	0.0012	12	0.001	4	0.0004
pois	15	0.0012	2	0.0002	12	0.001	5	0.0005
porém	13	0.0010	1	0.0001	9	0.0007	6	0.00071
embora	5	0.0004	2	0.0002	5	0.0004	2	0.00023

portanto	3	0.0002	3	0.0003	7	0.0006	4	0.00047
contudo	3	0.0002	-	-	2	0.0001	-	
apesar	3	0.0002	4	0.0004	4	0.0003	1	0.00011
enquanto	1	0.00008	-	-	-	-	-	-
caso	-	-	-	-	3	0.0002	1	0.00011

Fonte: Elaborado pela autora