

LUANA APARECIDA NAZZI LARANJA

O USO DE AGRUPAMENTOS LEXICAIS EM ATIVIDADES DE
LÍNGUA INGLESA PARA O VESTIBULAR
UMA METODOLOGIA BASEADA EM CORPORA

São José do Rio Preto

2020

LUANA APARECIDA NAZZI LARANJA

**O USO DE AGRUPAMENTOS LEXICAIS EM ATIVIDADES DE
LÍNGUA INGLESA PARA O VESTIBULAR
UMA METODOLOGIA BASEADA EM *CORPORA***

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientadora: Prof^a. Dr^a. Paula Tavares Pinto

São José do Rio Preto

2020

L318u Laranja, Luana Aparecida Nazzi
O uso de agrupamentos lexicais em atividades de língua inglesa para o vestibular : uma metodologia baseada em corpora / Luana Aparecida Nazzi Laranja. -- São José do Rio Preto, 2020
148 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientadora: Paula Tavares Pinto

1. Linguística de corpus. 2. Agrupamentos lexicais. 3. Material didático. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

LUANA APARECIDA NAZZI LARANJA

**O USO DE AGRUPAMENTOS LEXICAIS EM ATIVIDADES DE
LÍNGUA INGLESA PARA O VESTIBULAR
UMA METODOLOGIA BASEADA EM CORPORA**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof^a. Dr^a. Paula Tavares Pinto
UNESP – São José do Rio Preto
Orientadora

Prof^a. Dr^a. Ana Eliza Pereira Bocorny
UFRGS – Universidade Federal do Rio Grande do Sul

Prof^a. Dr^a. Talita Serpa
UNESP – São José do Rio Preto

São José do Rio Preto
03 de abril de 2020

Aos meus pais, Paula e Dorival, por todo o apoio e
companheirismo.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por guiar o meu caminho a oportunidades como essa.

Aos meus pais, que, mesmo diante de tantas dificuldades, me apoiaram em todos os momentos, sendo o alicerce para todos os objetivos que tenho almejado alcançar.

À minha orientadora, Dra. Paula Tavares Pinto, que, com todo o cuidado e sabedoria transformou meu período no mestrado em algo mais leve, e me inspirou a continuar no mundo acadêmico.

Ao meu namorado e amigo, Luciano Franco, que me inspirou a entrar na vida acadêmica e tem sido luz ao me ajudar a permanecer nessa caminhada. Agradeço pelo carinho com que me ajudou a perseverar e não desistir diante das atribulações.

À Maria Franco, que, mesmo indiretamente, sempre me apoiou nessa jornada.

À minha tia Neuza, que, carinhosamente, disponibilizou sua casa para eu ficar durante o tempo necessário para realizar determinadas atividades durante o período do mestrado.

A todos os meus amigos que confiaram em mim, especialmente à Amanda Rossi e Pamela Suélen, que há tantos anos têm me abençoado com uma amizade sincera e construtiva. Agradeço pela confiança, paciência e carinho.

Ao Prof. Luis Carlos Caineli (*in memoriam*), que foi exemplo de profissional e amigo, meu muito obrigado cheio de saudades.

A todos os professores com os quais tive contato durante o mestrado, especialmente ao Prof. Dr. Eduardo Penhavel, Prof. Dra. Marize Mattos Dall’Aglio Hattnher e Profa. Dra. Adriane Orenha-Ottaiano, em que tive o imenso prazer de ter aula.

A todos os meus companheiros de estudo, especialmente ao William Danilo Garcia e ao Guilherme Souza, com os quais tive a alegria de compartilhar grandes experiências durante o mestrado.

Ao Colégio Coopere que, em nome da Profa. Dra. Elen Dias, que leciona a disciplina de língua inglesa, me disponibilizou espaço para poder aplicar as atividades desenvolvidas aqui. Agradeço muito pela oportunidade.

Ao Ibilce, que me possibilitou uma formação digna e espaço para compartilhar minhas ideias e trabalhos no meio acadêmico.

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

“A educação é sempre uma certa teoria do conhecimento posta em prática”
FREIRE, P. (2003, p. 40).

RESUMO

Tendo como base teórico-metodológica o ensino de língua inglesa (LI) para fins específicos (IFE) (DUDLEY-EVANS; ST. JOHN, 1998) em associação com a Linguística de *Corpus* (LC) (BERBER SARDINHA, 2004, 2009; O'KEEFFE, MCCARTHY, CARTER, 2007; VIANA, TAGNIN, 2011), o objetivo do presente trabalho é avaliar a aplicabilidade do levantamento de agrupamentos lexicais para o desenvolvimento de atividades de compreensão de escrita em LI a fim de contribuir para a identificação vocabular de notícias da atualidade. Para isso, foram compilados quatro *corpora* de textos jornalísticos autênticos com temática saúde, tecnologia, política e mundo, advindos dos sites CNN® e BBC® (coleta em 2018) com o intuito de desenvolver atividades voltadas para o vestibular da UNESP (Universidade Estadual Paulista). Por meio do *software Sketch Engine*® foram extraídas as palavras-chave mais frequentes de cada *corpus* e, concomitantemente, os agrupamentos lexicais dessas palavras para posterior desenvolvimento das atividades, de modo que os exercícios tenham como base as combinações lexicais mais frequentes das notícias compiladas. Em função disso, foi possível verificar que quando o conteúdo do *corpus* está em consonância com o objetivo dos alunos, é possível extrair informações relevantes para o desenvolvimento de materiais úteis para a sala de aula. Assim, os textos jornalísticos foram muito importantes para os exercícios elaborados por meio dos *corpora*, visto que esse gênero textual é constante nos vestibulares prestados pelos alunos. A aplicação das atividades foi feita em uma escola particular de ensino regular situada no interior de São Paulo. Com o intuito de analisar o efeito das atividades com base em *corpora*, um questionário foi entregue aos alunos com perguntas acerca dos exercícios e sua validade no contexto de um vestibulando. A análise dos questionários apresentou um resultado positivo, visto que uma parte significativa dos estudantes respondeu que as atividades apresentadas antes da leitura dos textos os auxiliaram na compreensão dos mesmos. Cada exercício proporcionou ao aluno o contato com um vocabulário, por meio de palavras-chave e agrupamentos lexicais do próprio texto e do *corpus* compilado, e, ainda, despertou no aluno a consciência de uma leitura feita por meio de estratégias, como o uso e reconhecimento de agrupamentos lexicais.

Palavras-chave: Linguística de *Corpus*. Inglês para fins específicos. Compreensão escrita. Vestibular.

ABSTRACT

Based on the theoretical and methodological approach of English for Specific Purposes (ESP) (DUDLEY-EVANS; ST. JOHN, 1998) along with the use of Corpus Linguistics (CL) BERBER SARDINHA, 2004, 2009; O'KEEFFE, MCCARTHY, CARTER, 2007; VIANA, TAGNIN, 2011), the aim of this research is to evaluate the applicability of lexical bundles for the development of reading comprehension activities in English in order to contribute to the lexical comprehension of current news. To this end, four corpora of authentic journalistic texts on health, technology, politics and world, from the CNN® and BBC® websites, were compiled (collected in 2018) in order to develop activities aimed at the entrance exam of UNESP (São Paulo State University). We extracted the most frequent keywords from each corpus with the aid of the Sketch Engine® software and, concomitantly, the lexical bundles originated from these words for further development of teaching activities, so that the exercises are based on the most frequent lexical combinations of the compiled news. As a result, it was possible to verify that when the content of the corpus is in line with the students' objective, it is possible to extract relevant information for the development of useful materials for the classroom. Thus, the journalistic texts were very important for the exercises elaborated through the corpora, since this textual genre is constant in the entrance exams applied to the students. The activities were administered to students of a private regular school located in the northwest of the state of São Paulo. In order to analyze the effect of activities based on corpora, a questionnaire was given to the students with questions about the exercises and their relevance to the context of an entrance examination. The analysis of the questionnaires showed a positive result, since a significant part of the students answered that the activities presented before the reading of the texts helped them to understand the questions. Each exercise provided the student with the contact with a vocabulary, mainly by the new vocabulary approached, through the keywords and lexical bundles from the text and the compiled corpus, and also raised the student's awareness of reading based upon strategies, such as the use and recognition of lexical bundles.

Keywords: *Corpus Linguistics. English for specific purposes. Written comprehension. Entrance exam.*

LISTA DE ILUSTRAÇÕES

IMAGEM 1 - PROCEDIMENTO DE PALAVRAS-CHAVE	51
IMAGEM 2: EXEMPLO DE LINHAS DE CONCORDÂNCIAS.....	64
IMAGEM 3 - SKETCH ENGINE – DASHBOARD.....	70
IMAGEM 4 - SKELL-WEBSITE.....	72
IMAGEM 5 - SKELL – FERRAMENTA <i>WORD SKETCH</i>	73
IMAGEM 6 - SKELL – FERRAMENTA <i>EXAMPLES</i>	74
IMAGEM 7 - SKELL – FERRAMENTA <i>SIMILAR WORDS</i>	75
IMAGEM 8 - LISTA DE AGRUPAMENTOS LEXICAIS DO <i>CORPUS TECH</i>	81
IMAGEM 9 - SEQUÊNCIA DE ATIVIDADES COM AGRUPAMENTOS LEXICAIS	84
IMAGEM 10 - SELEÇÃO DE TEXTOS DO <i>CORPUS HEALTH</i>	88
IMAGEM 11 - SELEÇÃO DE TEXTOS DO <i>CORPUS POLITICS</i>	89
IMAGEM 12 - SELEÇÃO DE TEXTOS DO <i>CORPUS WORLD</i>	90
IMAGEM 13 - SELEÇÃO DE TEXTOS DO <i>CORPUS TECH</i>	90
IMAGEM 14 - SEQUÊNCIA DE SELEÇÃO DO TEXTO E DOS AGRUPAMENTOS LEXICAIS	91
IMAGEM 15 – EXEMPLO DE ATIVIDADES DO <i>CORPUS HEALTH</i> #1	92
IMAGEM 16 - EXEMPLO DE ATIVIDADES DO <i>CORPUS HEALTH</i> #2	93
IMAGEM 17 - EXEMPLO DE ATIVIDADE DO <i>CORPUS POLITICS</i> #1.....	94
IMAGEM 18 - FERRAMENTA <i>WORD SKETCH</i>	95
IMAGEM 19 – EXEMPLO DE ATIVIDADE DO <i>CORPUS WORLD</i> #1	96
IMAGEM 20 – EXEMPLO DE ATIVIDADE DO <i>CORPUS TECH</i> #1	97

LISTAS DE QUADROS

QUADRO 1 – FATORES IMPORTANTES NA ESCOLHA DE MATERIAL DE LEITURA AUTÊNTICO.....	39
QUADRO 2 - CONTEÚDO E PROPÓSITO DOS <i>CORPORA</i>	76
QUADRO 3 - ORGANIZAÇÃO DOS INSTRUMENTOS DE PESQUISA	77
QUADRO 4 - LISTA DOS CRITÉRIOS OBRIGATÓRIOS PARA A PREPARAÇÃO DE ATIVIDADES.....	78
QUADRO 5 – LISTA DOS CRITÉRIOS NÃO OBRIGATÓRIOS PARA A PREPARAÇÃO DE ATIVIDADES.....	79

LISTA DE TABELAS

TABELA 1 - AGRUPAMENTOS LEXICAIS DO <i>CORPUS TECH</i>	55
TABELA 2 - NÚMEROS REFERENTES AO <i>CORPUS</i> ENTENTEN.....	70
TABELA 3 - ITENS E FORMAS DOS <i>CORPORA</i> COMPILADOS.....	76
TABELA 4 - ITENS E FORMAS DOS <i>CORPORA</i> COMPILADOS.....	87
TABELA 5 – PERGUNTA #1 O QUE VOCÊ ACHOU MAIS INTERESSANTE NA AULA DE HOJE?	101
TABELA 6 - PERGUNTA #2 – QUAL A DIFICULDADE DAS ATIVIDADES?.....	102
TABELA 7 – PERGUNTA #3 AS ATIVIDADES DE VOCABULÁRIO TE AJUDARAM A COMPREENDER O TEXTO?	103
TABELA 8 - PERGUNTA #1 DO QUESTIONÁRIO FINAL	104
TABELA 9 - PERGUNTA #2 DO QUESTIONÁRIO FINAL	105
TABELA 10 – PERGUNTA 3 DO QUESTIONÁRIO FINAL.....	105
TABELA 11 - PERGUNTA 3 DO QUESTIONÁRIO FINAL	105

LISTAS DE ABREVIATURAS E SIGLAS

BNCC	Base Nacional Comum Curricular
CE	Compreensão Escrita
CO	Critérios Obrigatórios
DIY	<i>Do-It-Yourself</i>
IFE	Inglês para Fins Específicos
IFG	Inglês para Fins Gerais
LC	Linguística de <i>Corpus</i>
LE	Língua Estrangeira
LI	Língua Inglesa
LinFE	Línguas para Fins Específicos
NO	Não obrigatórios
PCNs	Parâmetros Curriculares Nacionais
SkELL	<i>Sketch Engine for Language Learning</i>
TCLE	Termo de Consentimento Livre e Esclarecido
TALE	Termo de Assentimento Livre e Esclarecido
UNESP	Universidade Estadual Paulista

SUMÁRIO

1. INTRODUÇÃO	15
2. SUBSÍDIOS TEÓRICOS	21
2.1 A importância da abordagem IFE	21
2.2 O processo de compreensão escrita	24
2.2.1 O ensino de compreensão escrita no contexto IFE	30
2.2.2 Autenticidade nos textos de compreensão escrita	38
2.3 A Linguística de Corpus na pesquisa	41
2.3.1 O uso de LC no ensino de LI	45
2.3.2 IFE e o ensino de vocabulário sob a perspectiva da LC	48
2.3.2.1 A importância das palavras-chave	50
2.3.2.2 O uso de agrupamentos lexicais em atividades de CE	52
2.4 Desenvolvimento dos corpora	57
2.4.1 Tamanho dos corpora (small corpora)	58
2.4.2. Corpus reciclável: características e utilidade na pesquisa	61
2.4.3 O ensino por meio de dados e o foco em atividades do tipo hands-off	63
3. MATERIAIS E MÉTODOS.....	69
3.1. <i>Sketch Engine</i> – características do concordanciador e ferramentas utilizadas...69	
3.1.1 SkELL – principais características e utilidade na pesquisa	71
3.2 Os <i>corpora</i> de estudo – características	75
3.3 Critérios obrigatórios e não obrigatórios na preparação de atividades didáticas baseadas em <i>corpora</i>	78
3.4 Elaboração das atividades	80
3.4.1 Seção de <i>warm-up</i>	81
3.4.2 Seção de vocabulário	83
3.4.3 Seção de atividades de compreensão escrita	83
3.5 Aplicação das atividades – contexto	85
4. ANÁLISE DOS RESULTADOS	86
4.1 Análise de variação lexical – FI	86
4.2 Elaboração das atividades por meio da análise lexical de cada corpus	87
4.2.1 Atividade de Warm-up	87
4.2.2 Atividade de Vocabulário	91
4.2.3 Exercícios de compreensão escrita	98

4.3 Análise dos questionários.....	100
4.3.1 Análise dos questionários semanais	100
4.3.2 Análise do questionário final	104
5. CONCLUSÕES E ENCAMINHAMENTOS	107
5.1 Limitações do estudo	107
5.2. Conclusões e Encaminhamentos	108
REFERÊNCIAS	111
APÊNDICES	119
APÊNDICE A – Atividade do corpus World	120
APÊNDICE B – Atividade do corpus Health	124
APÊNDICE C – Atividade do corpus Tech.....	132
APÊNDICE D – Atividade do corpus Politics.....	136
APÊNDICE E – Chavicidade relativa do corpus World.....	144
APÊNDICE F – Chavicidade relativa do corpus Health	145
APÊNDICE G – Chavicidade relativa do corpus Tech	146
APÊNDICE H – Chavicidade relativa do corpus Politics	147

1. INTRODUÇÃO

Considerado o idioma estrangeiro de maior difusão no mundo, bem como a língua franca (CRYSTAL, 2003; SEIDLHOFER, 2005; JENKINS, 2009), o inglês é um pré-requisito essencial em qualquer área profissional ou acadêmica. Tendo essa grande relevância, a língua inglesa (LI) passou a ser parte integrante das disciplinas que compõem os vestibulares brasileiros – exames que possibilitam o aluno adentrar uma universidade pública ou privada no país.

Os estudantes, especialmente do ensino médio, buscam aprimorar a compreensão escrita (CE) de textos na LI, visto que essa ainda é a única habilidade exigida pelos exames vestibulares, bem como no vestibular da UNESP (Universidade Estadual Paulista) – exame utilizado como referência para essa pesquisa – sendo justificada pelo fato de que a leitura bibliográfica em inglês tem prioridade no contexto acadêmico, quando comparada às outras habilidades linguísticas (SCARAMUCCI, 1999).

Sendo assim, muitos alunos procuram por cursinhos, ou até mesmo aulas particulares, com o intuito de aperfeiçoar a CE em LI e aprender técnicas que os ajudem a interpretar melhor os textos no momento do exame. Levando em consideração o contexto específico de um vestibulando – sua necessidade de aprender a ler em inglês, bem como a velocidade com que precisa realizar essa prática e as técnicas que precisa utilizar – a abordagem teórico-metodológica Inglês para Fins Específicos (IFE) auxilia tanto professor quanto aluno em busca de um ensino-aprendizado mais efetivo em sala de aula, visto que o que está sendo evidenciado é a necessidade do estudante.

A IFE é uma das ramificações da abordagem LinFE – Línguas para Fins Específicos – que foca nos objetivos que o aprendiz necessita atingir em situações reais de uso da LI, e, a partir da Segunda Guerra Mundial, essa abordagem começou a se expandir, tornando o inglês uma língua internacional que tinha o intuito de promover uma melhor comunicação entre povos de diferentes locais. Dessa maneira, as pessoas passaram a demonstrar um enorme interesse e, conseqüentemente, uma maior procura pelo aprendizado dessa língua, não mais por prestígio, mas sim por interesse no comércio e tecnologia internacional (HUTCHINSON, WATERS, 1987).

Assim sendo, o aprendizado passou a surgir de interesses particulares, e houve, então, a necessidade de cursos voltados para pessoas com objetivos já determinados, fossem eles acadêmicos, profissionais ou pessoais, fazendo a IFE ganhar uma abrangência cada vez maior.

Nesse contexto, a Linguística de Corpus (LC) vem acrescentar ao ensino de línguas no sentido em que pode possibilitar, bem como facilitar, o desenvolvimento de atividades propostas em um curso com abordagem IFE, visto que, como afirma Bernardini (2004, p. 21) “os professores de IFE estavam entre os primeiros a apreciar o potencial pedagógico do trabalho com *corpus*¹”.

Berber Sardinha (2011, p. 302) assegura que “ao contrário de outros meios de ensinar, com a LC é necessário ter dados, concordâncias [...] o professor/linguista de *corpus* não pode depender apenas de si mesmo, de sua oratória e do quadro negro, para dar aula”.

Há diversas vantagens em se utilizar LC no ensino de línguas estrangeiras (LE), como por exemplo, o uso de materiais autênticos, textos selecionados criteriosamente e a observação da língua em uso. Dessa forma, a LC tem muito a contribuir no desenvolvimento de materiais de línguas, assim como para as atividades voltadas para essa pesquisa, visto que as aulas ministradas focam os objetivos dos vestibulandos.

Dessa forma, as atividades elaboradas na presente pesquisa tiveram como base a primeira fase do vestibular da UNESP – que é realizado pela Fundação Vunesp. Segundo o artigo 12 do manual do candidato do vestibular de 2018 (bem como no de 2019), as provas são desenvolvidas de acordo com os Parâmetros Curriculares Nacionais (PCNs) do Ensino Médio em conjunto com o Currículo do Estado de São Paulo, portanto, esses documentos serão citados ao longo da pesquisa. Entretanto, o atual documento que rege o ensino no Brasil é a BNCC – Base Nacional Comum Curricular, assim, esse último também será utilizado como subsídio para o presente trabalho. É importante mencionar que a BNCC foi homologada pelo MEC em dezembro de 2017 e todas as instituições escolares do Brasil tiveram de implementar essa base no ensino até o final de 2019.

Nesse sentido, de acordo com o manual do candidato (2018, 2019, p.34) do vestibular em questão, “o Concurso Vestibular consiste na seleção e classificação de candidatos à matrícula inicial nos Cursos de Graduação da Universidade”. O objetivo é a eleição de candidatos que:

- a) articulem ideias de modo coerente;
- b) compreendam ideias, relacionando-as;
- c) se expressem com clareza;
- d) conheçam o conteúdo do currículo da Educação Básica do Estado de São Paulo.

¹ ESP teachers were among the first to appreciate the pedagogic potential of corpus work (BERNARDINI, 2004, p. 21).

II - integrar os objetivos da Universidade àqueles desenvolvidos pelo Sistema de Ensino Fundamental e Médio;

III - dar condições para o desenvolvimento de potencialidades e aptidões do estudante nas áreas específicas da Universidade.

O vestibular de 2018 foi realizado em duas fases, sendo que a primeira dispunha de conteúdo de conhecimentos gerais enquanto a segunda de conhecimentos específicos e uma redação. A disciplina de LI está inclusa nas duas provas, apesar de a presente pesquisa focar apenas na primeira. Essa é composta de 90 questões objetivas no total e inclui 30 perguntas de cada área, sendo elas: Linguagens, Códigos e suas tecnologias (contendo LI); Ciências Humanas e suas tecnologias e Ciências da Natureza, Matemática e suas tecnologias.

Já na segunda prova, além da redação, o aluno encontra conteúdo relativo às áreas de: Linguagens, Códigos e suas tecnologias (em que se encontra a LI novamente); Ciências Humanas e suas tecnologias e Ciências da Natureza, Matemática e suas tecnologias, sendo 36 questões discursivas – 12 de cada área específica (FUNDAÇÃO VUNESP, 2018, p. 24-35).

Como pode-se notar no exame em questão, a LI é contemplada nas duas fases da prova, ou seja, ainda que o aluno tenha passado pela primeira fase desse vestibular, ele ainda encontrará conteúdo dessa disciplina na segunda fase, fato que pode despertar nos alunos mais dedicação no estudo da LI. Vale ressaltar que, apesar da importância da segunda fase da prova, esta pesquisa foca apenas na primeira, como mencionado anteriormente, devido à estrutura da prova, que é composta por questões múltipla escolha e que, por esse motivo, não adentra ao ensino de técnicas de produção escrita – habilidade exigida na segunda fase da prova.

Embora atualmente a relevância do aprendizado dessa língua seja mais intensificado, nem sempre foi dessa maneira. Segundo Bartholomeu (2002), o *status* do inglês, na época, passava por um conflito, ou seja, havia valorização desta língua pela perspectiva do mercado de trabalho, todavia, por outro lado, em sala de aula, essa disciplina não era valorizada, visto que, como a própria autora afirma em sua pesquisa, o inglês era oferecido em apenas algumas escolas particulares e públicas em uma aula de 45 minutos no terceiro ano do ensino médio.

De acordo com o PCN (BRASIL, 2000) – Parâmetros Curriculares Nacionais para o ensino médio – base para os vestibulares brasileiros,

No âmbito da LDB, as Línguas Estrangeiras Modernas recuperam, de alguma forma, a importância que durante muito tempo lhes foi negada. Consideradas, muitas vezes e de maneira injustificada, como disciplina pouco relevante, elas adquirem, agora, a configuração de disciplina tão importante como qualquer outra do currículo, do ponto de vista da formação do indivíduo. (BRASIL, 2000, p. 25).

Nesse contexto mencionado pelo PCN (BRASIL, 2000), é possível verificar a não contextualização dos materiais utilizados em sala de aula. Em outras palavras, ainda quando a LI era considerada pouco relevante para a realidade do aluno, a disciplina de quaisquer línguas estrangeiras pautava-se, segundo o próprio PCN (BRASIL, 2000), em estruturas gramaticais da língua, sendo que o conteúdo era transmitido de maneira descontextualizada, “desvinculada da realidade” (BRASIL, 2000, p. 26).

Em contraste com essa realidade, em 2006, no documento PCN+ (BRASIL, 2006), que diz respeito às Orientações Nacionais Complementares aos Parâmetros Curriculares Nacionais, é afirmado que “o objetivo primordial do professor de língua estrangeira deve ser o de tornar possível a seu aluno atribuir e produzir significados, meta última do ato de linguagem” (BRASIL, 2006, p.93).

Já o atual documento que conduz a educação brasileira, o BNCC, “prioriza o foco da função social e política do inglês e, nesse sentido, passa a tratá-la em seu status de língua franca (BRASIL, 2018, p. 241). Não adentrando muito profundamente no conceito de língua franca, o intuito da disciplina, é fazer com que essa ideia de língua estrangeira, como “inglês dos EUA” ou “inglês da Inglaterra”, seja desmitificado para que os alunos se identifiquem dentro de uma língua que já é considerada de comunicação global, e seja ferramenta de interação social e empoderamento. Trata-se, portanto de “possibilitar aos estudantes cooperar e compartilhar informações e conhecimentos por meio da língua inglesa, como também agir e posicionar-se criticamente na sociedade, em âmbito local e global” (BRASIL, 2018, p. 485).

No que diz respeito à autenticidade dos textos abordados em LI, é sugerido pelo PCN+ na seção “Texto” do documento, que se trabalhe com a “análise de textos de diferentes gêneros (*slogans*, quadrinhos, poemas, notícias de jornal, anúncios publicitários, textos de manuais de instrução, entre outros)” (BRASIL, 2006, p. 96). Essa informação vem enfatizada no BNCC, que afirma que a leitura aborda práticas linguísticas, “especialmente sob o foco da construção de significados, com base na compreensão e interpretação dos gêneros escritos em LI, que circulam nos diversos campos e esferas da sociedade” (BRASIL, 2018, p. 243).

Pensando acerca da abordagem de gêneros, a presente pesquisa utiliza apenas o gênero textual “notícia de jornal” para a compilação dos *corpora* e posterior elaboração das atividades, visto que não seria possível abordar todos os gêneros que circulam no meio social.

Esta pesquisa se justifica, dessa maneira, pela relevância de se abordar a LI no contexto de exames de vestibular no Brasil, visto que, contrastando o cenário de quinze anos atrás e atualmente, pode-se perceber que o inglês passou a ser necessário nos exames de vestibular nas

redes de universidades públicas e privadas, sendo que, em algumas, como é o caso da Unesp, a LI é exigida até mesmo na segunda fase do exame.

Também é importante mencionar que os textos jornalísticos são parte do que é exigido dos alunos no vestibular. Assim sendo, faz-se relevante desenvolver meios para que os estudantes aprimorem seus estudos na LI, a fim de alcançarem sucesso no vestibular por meio de materiais que utilizam a língua em seu contexto autêntico.

Ainda que muitos trabalhos já tenham sido desenvolvidos com o foco na CE em LI (CELANI, 1983, 2005; RAMOS, 2004; TOMITCH, 2009), a presente pesquisa justifica-se pelo fato de que poucos são os que se voltam para a elaboração de atividades com base no levantamento lexical por meio de *corpora*.

Desse modo, atividades que envolvem textos jornalísticos na LI foram elaboradas a fim de extrair informações lexicais consideradas estatisticamente relevantes de acordo com os *corpora* compilados. É importante esclarecer que, esta pesquisa não tem como objetivo principal levantar um vocabulário importante no que diz respeito à LI de modo geral. O que se pretende aqui é levar aos alunos textos que abordam tópicos contemporâneos para que eles aprendam, conseqüentemente, o léxico de notícias atuais por meio dos temas *mundo, tecnologia, saúde e política*. Assim, se muito está sendo falado acerca de “guerra comercial” no assunto política, então espera-se abordar o léxico *trade war* na sala de aula, que se trata de um vocábulo específico desse contexto.

Sabendo, portanto, que a CE é a habilidade exigida nos exames que abrem portas para o ensino superior e devido à relevância de se ter um curso voltado para as necessidades específicas de cada aluno, o presente trabalho tem como **objetivo geral** avaliar a aplicabilidade do levantamento de agrupamentos lexicais no desenvolvimento de atividades de CE em LI. Os **objetivos específicos** são:

1. Compilar quatro *corpora* recicláveis (DIY²) de textos jornalísticos autênticos advindos dos sites CNN® e BBC®;
2. Desenvolver atividades de CE com o uso dos agrupamentos lexicais extraídos a partir dos *corpora* compilados na busca pela identificação lexical de temas atuais;
3. Trabalhar técnicas de CE a partir dos agrupamentos lexicais extraídos a fim de auxiliar os alunos no momento de prestar o exame.

Em função disso, é esperado responder às seguintes perguntas de pesquisa:

² Mais informações acerca desse tipo de *corpus* serão dadas no subcapítulo 2.4.2, na página 61.

1. Em que medida o uso de *corpora* criados a partir de textos jornalísticos autênticos pode auxiliar no levantamento de agrupamentos lexicais na preparação de alunos para a prova de LI do vestibular?

2. Em que medida as atividades baseadas em agrupamentos lexicais podem desenvolver o léxico dos alunos e serem relevantes para o contexto dos vestibulandos?

As atividades de CE foram aplicadas no período de quatro aulas de LI no segundo e terceiro anos do ensino médio de uma escola particular de ensino regular situada em uma cidade do interior de São Paulo, e, podem ser replicadas por quaisquer outros professores IFE – público-alvo deste trabalho, mesmo que, para isso, tenham de reformular e adaptar os exercícios ao seu contexto de trabalho.

Assim sendo, a presente pesquisa está dividida da seguinte forma: O Capítulo 2 – Subsídios Teóricos – é voltado para os subsídios teóricos que embasam toda esta pesquisa, sendo que os principais temas abordados são: a abordagem IFE, a LC e o ensino de técnicas de leitura. Cada subcapítulo esclarece determinados pontos que são aparatos teóricos para o desenvolvimento do presente estudo.

O Capítulo 3 – Material e métodos – destaca o material e a metodologia desta pesquisa, sendo dividido em subcapítulos que explicitam tanto os materiais utilizados, bem como os procedimentos metodológicos adotados para a realização do presente estudo. O Capítulo 4 – Análise – é dedicado às verificações feitas para o desenvolvimento de cada exercício, assim como também foram analisados dois questionários aplicados em sala de aula, o questionário semanal e o final, que auxiliaram na compreensão dos dados desta pesquisa. Finalmente, o Capítulo 5 – Conclusões e encaminhamentos – faz referência às conclusões e desdobramentos futuros do presente estudo e é seguido pelas referências, os apêndices, que contam com as atividades desenvolvidas nesta pesquisa, e com as listas dos agrupamentos lexicais extraídos em cada *corpus*.

2. SUBSÍDIOS TEÓRICOS

Na presente pesquisa, duas principais correntes teórico-metodológicas embasam todo o estudo aqui elaborado: a abordagem IFE e a LC. Acredita-se que, com a união de uma vertente teórica que se volta para as necessidades específicas dos alunos e outra que trabalha com a língua em seu contexto de uso, tanto aluno quanto professor podem se beneficiar na prática ao desenvolver o material didático e ao aplicá-lo em sala de aula.

Desse modo, neste capítulo serão abordados os subsídios teóricos que serviram como aparato para o desenvolvimento da presente investigação. Para uma melhor organização, dividiu-se este capítulo da seguinte forma: o primeiro subcapítulo, 2.1, discorre sobre a importância da abordagem IFE e sobre como se encontra no presente estudo. O subcapítulo 2.2 trata do processo de compreensão escrita, o qual engloba determinadas estratégias de leitura que foram abordadas em sala de aula com os alunos. O subcapítulo seguinte está ligado aos anteriores ao mostrar como se dá o processo de compreensão escrita dentro do contexto IFE, conectando-se ao subcapítulo 2.2.2 que discorre a respeito da autenticidade nos textos escolhidos para se trabalhar em sala de aula. Em seguida, explica-se como a LC auxilia nas pesquisas linguísticas, especialmente no que concerne o seu uso no ensino de LI – assunto tratado no subcapítulo seguinte.

É possível, também, encontrar, neste capítulo, como a abordagem IFE e o ensino de vocabulário se constroem sob a perspectiva da LC, que engloba a importância das palavras-chave, bem como o uso de agrupamentos lexicais nas atividades de CE. Além disso, explicita-se como funciona o processo de desenvolvimento dos *corpora* e os seus tamanhos. Por fim, um ponto muito importante para a presente pesquisa é o desenvolvimento de um *corpus* reciclável; neste penúltimo subcapítulo, serão abordadas as suas características e a sua utilidade na pesquisa, enquanto a última sessão deste capítulo explica como se dá o ensino por meio de dados e o foco em atividades do tipo “*hands-off*”.

2.1 A importância da abordagem IFE

Diferentemente de cursos gerais de línguas que abordam a língua em contextos comuns de uso, em que o aluno aprende a língua de modo que o ensino se aplica às suas quatro habilidades, e, comumente, vai do nível A1 até o C1-C2, tendo uma abrangência mais geral, a abordagem IFE se volta para o objetivo de cada estudante, cuja consciência linguística tem

implicação direta na aprendizagem da língua. Conforme afirmado por Hutchinson e Waters (1987),

Se professor e aluno sabem o porquê estão estudando o idioma, essa consciência terá uma influência positiva ao selecionar conteúdos que serão pertinentes à aprendizagem, e quais habilidades serão exploradas em sala de aula. (HUTCHINSON, WATERS, 1987, p. 53).

Em consonância com os autores citados, Holmes (1981) e Dudley-Evans e St. John (1998) mencionam algumas vantagens da LinFE (e, conseqüentemente, IFE) em relação aos cursos gerais de línguas:

1. *alunos aprendem o que realmente precisam de acordo com suas necessidades;*
2. *alunos tornam-se mais motivados, pois percebem que estão em um curso feito sob medida (tailored course) e que suas necessidades são o foco norteador das aulas;*
3. *professores têm liberdade para preparar aulas específicas a alunos específicos; isso reflete a consciência que o professor possui sobre a importância das necessidades de cada aluno.*

Levando-se em conta as características citadas acima no contexto desta pesquisa, os vestibulandos irão estudar conteúdo e habilidade que realmente estão inseridos em seu contexto, ou seja, a CE – que é exigida nos vestibulares atuais – bem como textos jornalísticos, que também se mostram presentes nos exames que esses alunos prestam. Por esse motivo, acredita-se que os estudantes se motivam mais em sala de aula, uma vez que entendem a importância do que está sendo abordado, excluindo a famosa pergunta que muitos alunos se fazem ao entrar em contato com determinado tópico de uma disciplina: “Por que eu estou estudando isso?”.

Faz-se importante, ainda, mostrar a relevância das técnicas de leitura e de como se aborda a LI, mesmo depois que esses alunos já tiverem passado no vestibular, pois, provavelmente, eles vão ter de lidar com textos acadêmicos em inglês na graduação.

Em relação ao terceiro tópico, é essencial que o professor esteja ciente dos objetivos de seus alunos. No caso desta pesquisa, o foco é o vestibular, o qual foi considerado como norte para a elaboração das atividades, ou seja, questionou-se “o que é exigido nesse tipo de exame?”. A partir de então, o desenrolar dos exercícios se deu por meio do que os *corpora* ofereciam ao professor-pesquisador.

Quando o professor está ciente das necessidades de seus alunos, é necessário saber *como* abordar o conteúdo que eles precisam aprender e é nesse sentido que a abordagem IFE está relacionada à sala de aula. Assim, pode-se dizer que a IFE teve início no Brasil com o *Projeto*

de *Inglês Instrumental* (antigo nome) em 1978 e foi coordenado pela Profa. Dra. Maria A. Celani (PUC - SP). A oficialização do projeto só ocorreu em 1980, e a habilidade em foco era a CE de alunos universitários que necessitavam compreender textos acadêmicos em inglês.

Posteriormente, o *Projeto de Inglês Instrumental* transformou-se em *Programa de Línguas para Fins Específicos – LINFE* - (RAMOS, 2012), sob a coordenação da Professora Dra. Rosinda de Castro Guerra Ramos. Ramos (2005) apresenta sete mitos que foram inventados acerca dessa abordagem com o decorrer do tempo no Brasil e dois são aqui destacados: instrumental é leitura e instrumental é inglês técnico (RAMOS, 2005: 116-118).

Com base nos mitos apontados por Ramos (2005), destaca-se, em primeiro lugar, o mito de que essa abordagem só se desenvolve por meio da habilidade de CE, ou seja, a abordagem IFE é voltada somente para a leitura. Apesar de a presente pesquisa voltar-se para essa habilidade, já que o público-alvo é composto de vestibulandos, é importante saber que vários trabalhos já foram desenvolvidos com foco em outras habilidades linguísticas (ALVES, PINTO, 2018; SANTOS, 2018; SILVA, 2018).

Além disso, outro ponto a ser considerado é de que essa abordagem não trabalha apenas com a LI em seu nível técnico. Em outras palavras, não se trabalha IFE apenas com aqueles alunos que desejam aprender a linguagem técnica de seu campo de estudo; caso esse fato fosse verídico, a abordagem não seria empregada nesta pesquisa, visto que os textos trabalhados com os alunos em sala de aula são de linguagem geral, ou seja, notícias jornalísticas que abordam diferentes temas. O que se torna específico nesta pesquisa são o público-alvo, o gênero e os temas abordados.

Nesse contexto de ensino-aprendizagem de IFE, é importante levar em consideração como se dá o processo tanto de ensinar quanto de aprender. Para Hutchinson e Waters (1987, p. 59),

Uma abordagem centrada no aluno para a aprendizagem e o ensino considera a aprendizagem como a construção ativa de significado e o ensino como o ato de guiar, montar e facilitar o aprendizado. Esta abordagem considera o conhecimento como sendo um processo em constante mudança, construído com base na experiência prévia dos alunos. (HUTCHINSON, WATERS, 1987, p. 59, tradução nossa³).

³ A learner-centered approach to learning and teaching sees learning as the active construction of meaning, and teaching as the act of guiding, scaffolding and facilitating learning. This approach considers knowledge as being an ever-changing process, which is built upon the learners' prior experience (HUTCHINSON, WATERS, 1987, p. 59)

Assim, o professor deixa de ser o centro detentor do saber e passa a ser facilitador no processo de ensino, guiando o aluno na solidificação do próprio conhecimento por meio do *background* já existente do estudante, isto é, aquilo que o aluno já conhece em relação à língua. É dessa maneira que, na presente pesquisa, os alunos entram em contato com o inglês, ativando, especialmente, o que eles já têm de informação e acrescentando o que há de novo sendo abordado nas atividades, como as técnicas de leitura, que serão discutidas mais adiante, bem como o vocabulário apresentado durante os exercícios.

Nesse contexto, conforme mencionado por Dutra e Silero (2010, p. 911), “a Linguística de *Corpus* pode colaborar com a Linguística Aplicada”, visto que o uso de *corpus* em sala de aula fornece descrições da língua em uso, o que se torna uma excelente ferramenta no ensino de línguas, em especial, para contextos específicos, e que pode levantar textos apropriados a serem selecionados como materiais pertinentes para a aula.

Em função das vantagens acima mencionadas, é esperado que o professor transite com relativa segurança por temas diversificados e específicos, sendo que tanto professor quanto aluno aproveitam melhor o tempo em sala de aula, visto que, como mencionado pelos autores acima, os alunos aprendem o que condiz com suas necessidades e, pelo mesmo motivo, tornam-se mais motivados. Isso faz com que as aulas sejam mais interessantes e a relação entre professor e aluno se solidifique de forma mais colaborativa.

Visto que a habilidade abordada neste estudo é a CE, o próximo subcapítulo discorre sobre a leitura e, posteriormente, sobre como ela se dá no âmbito IFE.

2.2 O processo de compreensão escrita

A partir do projeto nacional Ensino de Inglês Instrumental em Universidades Brasileiras (CELANI, 1984), foi possível constatar que alunos envolvidos nesse projeto apresentavam dificuldades não apenas na interação com o texto em LI, como também na compreensão textual em sua língua materna.

Levando esse fato em consideração, é evidente que a lacuna encontrada no ensino de leitura é algo que deva ser abordado com seriedade desde a formação inicial dos alunos (alfabetização). Entretanto, o objetivo desta pesquisa não é debater tal situação, mas apresentar algumas técnicas de leitura em LI e como elas se refletem no desenvolvimento de materiais, focando na CE da língua.

De acordo com os PCN+ (BRASIL, 2006),

O substrato sobre o qual se apoia a aquisição dessas competências constitui-se no domínio de técnicas de leitura – tais como *skimming*, *scanning*, *prediction* – bem como na percepção e na identificação de índices de interpretação textual (gráficos, tabelas, datas, números, itemização, títulos e subtítulos, além de elementos de estilo e gênero) (BRASIL, 2006, p. 97).

Já o documento BNCC reforça a ideia apresentada acima, afirmando que

As práticas de leitura em inglês promovem, por exemplo, o desenvolvimento de estratégias de reconhecimento textual (o uso de pistas verbais e não verbais para formulação de hipóteses e inferências) e de investigação sobre as formas pelas quais os contextos de produção favorecem processos de significação e reflexão crítica/problematização dos temas tratados (BRASIL, 2018, p. 244).

As técnicas *skimming*, *scanning*, *prediction* e identificação/interpretação de títulos e subtítulos serão abordadas aqui; porém, primeiramente, é importante apresentar os modelos de interação do leitor com o texto descritos por Kato (1985), Farrell (2003) e Carvalho (2005). Tais modelos são apresentados pela teoria cognitivista como: *bottom-up* (modelo ascendente), *top-down*, (modelo descendente) e modelo interativo, que se dá pela interação entre os dois primeiros – *bottom-up* e *top-down*.

De modo sucinto, explica-se o modelo *top-down* como o exercício que instiga o leitor a buscar seus próprios conhecimentos e experiências prévias, a fim de compreender o texto e, então, continuar sua leitura enquanto o texto confirma as suas expectativas. Já no modelo *bottom-up*, encontra-se o inverso desse processo: o leitor lê as palavras e as sentenças presentes no texto e olha para a sua organização sem fazer qualquer relação com o seu conhecimento e/ou experiência anteriores; assim, o leitor é capaz de construir o significado do que foi escrito no texto a partir do envolvimento do léxico com a estrutura sintática. Farrell (2003) esclarece que a oposição entre os dois primeiros modelos se dá pelo fato de que, na abordagem *top-down*, o leitor trabalha a partir do significado para o texto; em contrapartida, no *bottom-up*, o leitor parte do texto para o significado, focando na palavra e no nível da frase.

Farrell (2003) explica o processo que acontece no modelo interativo, em que a CE é feita “por um leitor maduro e fluente” (FARELL, 2003, p. 4), enquanto ilustra a interação entre os modelos *bottom-up* e *top-down* em quatro passos que podem ser seguidos por esses leitores na abordagem interativa. Os passos são:

01. Formular hipóteses ou fazer previsões: o leitor inicia o texto prevendo o assunto que será abordado, inferindo, por meio de seu *background* (conhecimento prévio), se o texto será interessante para o prosseguimento da leitura;

02. Experimentar o texto: o leitor não utiliza muito de seu tempo com a gramaticalidade do texto, mas seleciona palavras que se encaixam no momento da leitura;

03. Continuar o ciclo: como em um ciclo, o leitor sempre continua examinando o texto e deduzindo seu sentido, ignorando aquilo que não é de seu entendimento até esse tipo de problema se tornar recorrente e perceber que sua previsão estava equivocada;

04. Sendo a compreensão impossível, o leitor formula outra hipótese e reinicia o ciclo de previsão.

Faz-se relevante mencionar que, durante as aulas ministradas, bem como no desenvolvimento do material, os quatro passos foram levados em consideração. O primeiro exercício de cada atividade aborda o assunto de forma geral, com perguntas do tipo “o que você já ouviu falar a respeito disso?”, ou então, “a partir do título do texto, você consegue inferir o que será tratado em seu conteúdo?” Dessa forma, como defende Farrell (2003), é possível inferir o assunto antes de iniciar a leitura em si, ativando seu conhecimento prévio.

O segundo passo, que se refere a experimentar o texto, se deu em forma de exercício envolvendo as palavras-chave do texto. Assim, toda atividade trazia cinco dos agrupamentos lexicais mais frequentes de determinado texto e era pedido para que os alunos identificassem palavras cognatas e/ou conhecidas, para posterior identificação do conteúdo textual.

Na terceira parte, temos a continuação do ciclo, passo que os alunos seguiam a cada atividade, já que eram instruídos a identificar as palavras conhecidas a fim de compreender, de forma geral, o conteúdo do texto. Geralmente, quando não sabiam muitas palavras ou não entendiam determinado parágrafo, discutia-se, então, o parágrafo com a turma, fato que poderia levar os alunos ao quarto passo: reformular outra hipótese. Todavia, devido à discussão das palavras-chave logo no início do texto e, também, ao levantamento de conhecimento prévio antes da leitura, o quarto passo não foi completamente desenvolvido em sala de aula.

É válido mencionar que, apesar de Farrell (2003) discutir a respeito de um leitor “maduro e fluente”, não era esperado que os alunos alvo desta pesquisa possuísem tais características. Entretanto, foram trabalhados os quatro passos descritos pelo autor em aula a fim de auxiliar os estudantes a aprimorarem a sua leitura. Os quatro passos da abordagem interativa buscam exatamente levar o aprendiz a alcançar esse nível de CE para que se torne um leitor mais independente, isto é, para que consiga lidar com o texto em LI sem a ajuda de um profissional ou de quaisquer outros meios de auxílio, que, muitas vezes, não são permitidos no momento do vestibular, como o dicionário ou a internet, por exemplo.

Desse modo, partindo da ideia de que a inferência é parte essencial para a compreensão textual, abordaremos essa técnica também à luz dos estudos de Kleiman (1985). A autora

explica o uso da inferência em textos autênticos, como foi feito no presente trabalho. Para a estudiosa, o leitor, muitas vezes, possui um conhecimento lexical limitado e é por esse motivo que uma estratégia como a inferência se faz necessária para a leitura de textos em uma língua estrangeira. Kleiman defende a ideia de que “se o conhecimento linguístico do aluno é limitado, ele passará a usar outros tipos de conhecimentos, de caráter textual, discursivo, enciclopédico, a fim de suprir as limitações” (KLEIMAN, 1985, p. 73).

A inferência é abordada de forma contextualizada nas atividades desenvolvidas nesta pesquisa, para que os alunos se sintam mais seguros em fazer determinados exercícios, ou seja, muitas vezes, quando determinadas palavras são inseridas em uma atividade de maneira descontextualizada, pode-se gerar algum tipo de confusão de significado, por exemplo. Os agrupamentos lexicais trabalhados com os alunos serão, posteriormente, apresentados dentro do texto em que eles farão a leitura, de modo que a inferência seja útil para a análise textual, uma vez que o aluno já terá tido contato com o agrupamento e tentará identificá-lo dentro do contexto inserido.

Schmitt (2000) afirma que fazer uso da inferência a fim de compreender um texto é essencial no que tange ao aprendizado lexical, já que nem sempre, como mencionado, o aluno terá conhecimento de todas as palavras presentes no texto (ou a maioria delas). Corroborando a ideia do autor, nesta pesquisa a inferência é muito presente no momento em que os participantes fazem os exercícios, pois, como é sabido, os vestibulares tratam de assuntos diversos que, muito provavelmente, trarão palavras novas para os alunos, sendo assim, é primordial que eles conheçam e apliquem técnicas de leitura.

Segundo Solé (1988), as estratégias de leitura são muito necessárias para o desenvolvimento de uma leitura proficiente. Dentre as várias técnicas existentes para auxiliar na CE de textos em LE, Brown (2001) menciona duas muito relevantes: *skimming* e *scanning* – citadas nos PCN+ (BRASIL, 2006). A primeira faz referência ao ato de olhar rapidamente pelo texto, a fim de encontrar características-chave que mostrem o que o texto aborda de modo geral. Pode-se encontrar o uso dessa técnica em leituras eventuais de jornais, por exemplo, em que o leitor busca rapidamente por notícias que são de seu interesse. A segunda estratégia refere-se à busca mais detalhada de informações específicas no texto, como palavras-chave, por exemplo, e, para o autor, essas técnicas que introduzem um tema são de grande valia no processo de compreensão textual.

Santos (2011, p. 39) sugere que, para utilizar o *skimming* em um texto, é necessário iniciar a leitura “projetando sua atenção para títulos, imagens e outras informações em destaque”, desse modo é importante se atentar não apenas ao conteúdo textual em si (corpo do

texto), mas também a informações “extras” que o texto traz, já que elas também são consideradas pontos-chave na interpretação textual.

Ao ler o título do texto e notar qual o assunto que ele tratará, o aluno deve fazer uma conexão com o que há de visual na página, isto é, com as imagens que ajudam a esclarecer o que está sendo tratado, colaborando para a interpretação textual. Juntando, por exemplo, o conhecimento prévio que o leitor tem acerca de determinado assunto, as imagens que esse texto contém e as palavras cognatas encontradas no título e/ou no corpo do texto e aquelas que são familiares, o aluno dificilmente se sentirá tão perdido a ponto de dizer que não entende nada acerca daquele texto.

Ainda de acordo com Santos (2011, p. 34), o título do texto às vezes “utiliza uma fonte diferente, para dar maior destaque” e o uso do negrito mostra que “certas partes do texto são mais importantes do que outras [...] os trechos em negrito contêm as informações centrais”. A autora, ainda, explica que, com o impacto da internet no nosso meio, há ainda mais “o uso intenso de recursos visuais em seus textos e que comunica não apenas através de palavras, mas também através de cores, fontes tipográficas, fotos, ilustrações diversas, layout, entre outros” (SANTOS, 2011, p. 34).

A autora Santos (2011) também enfoca a importância de estabelecer objetivos durante a leitura, isto é, ter clareza do “porquê” se está lendo e, a partir de então, estabelecer as técnicas para a compreensão do texto, já que, como a própria autora defende, “podemos e devemos ler de formas diferentes” (SANTOS, 2011, p. 38). Assim, fazendo ligação com o que foi dito acerca das técnicas *skimming* e *scanning*, o aluno provavelmente irá se deparar, no vestibular, com perguntas do tipo: “qual a ideia principal do texto?” ou “qual a ideia do primeiro parágrafo?” Esse é o momento de se estabelecer os objetivos e pensar qual a melhor técnica a ser usada.

A estudiosa Françoise Grellet (1981) corrobora a ideia de Santos ao afirmar que os alunos não vão conseguir ler de maneira eficiente caso não adaptem a velocidade de leitura e técnica ao seu objetivo no momento da leitura. Ler textos diversificados de uma mesma maneira faz com que os alunos percam tempo e deixem de lembrar pontos importantes pelo simples fato de estarem absorvendo informações desnecessárias (GRELLET, 1981, p. 17).

Em relação à segunda técnica – *scanning* – Santos (2011) afirma que essa estratégia é utilizada quando se busca por informações mais específicas dentro de um texto. Dessa maneira, a autora explica que leituras como a de um quadro de horário no aeroporto, informações nutricionais de um alimento e, até mesmo, a programação da TV na decisão de qual programa assistir são exemplos de uso do *scanning*, uma vez que, nesses casos, o que se torna relevante não é toda a informação textual, mas aquela pequena parte que nos interessa. No caso do quadro

de horário do aeroporto, por exemplo, pouco importam as informações acerca dos outros voos; o que eu busco, ao examinar esse quadro, é a informação do meu voo, assim, eu vou ler o quadro buscando palavras específicas, como o número do voo ou o nome do destino.

Geralmente, em questões de múltipla escolha, verdadeiro ou falso, entre outras presentes em testes de inglês, como o vestibular, faz-se muito relevante que o leitor saiba aplicar técnicas como *scanning*, pois, desde a pergunta, há palavras-chave das quais o estudante fará uso para buscar no corpo do texto a resposta de que precisa. Em outras palavras, se a pergunta é algo do tipo “*What is the biggest city of Brazil, according to João da Silva?*”, é necessário que o aluno busque no texto, utilizando *scanning*, palavras como *biggest city*, *Brazil* e até mesmo *João da Silva*, visto que essas unidades vão levar o aluno às partes do texto em que provavelmente terá a resposta correta.

Voltando-se para a técnica *prediction*, que diz respeito às previsões que podem (e devem) ser feitas antes da leitura textual, Santos (2011) mostra que essa estratégia estabelece certa conexão entre o texto e o leitor, pois “essa conexão ajudará o leitor a construir sentidos para o que lê” (SANTOS, 2011, p. 63). A autora, ainda, afirma que o relacionamento com o texto faz o leitor ativar o conhecimento prévio que já possui em relação a determinado tópico e a antecipar o conteúdo textual.

É possível afirmar que, em um vestibular, a utilização dessa estratégia também é de grande relevância e pode fazer diferença na leitura do texto. Muitos podem dizer que fazer previsões acerca de um texto é algo que se dá naturalmente, ou seja, ao ler o título, qualquer leitor ativa seu conhecimento prévio e tenta prever o que o texto traz em seu conteúdo. Todavia, isso nem sempre acontece de modo eficaz; o aluno pode até ativar seu conhecimento prévio, porém de nada irá adiantar se ele insistir em gastar o resto de seu tempo lendo e tentando traduzir palavra por palavra, por exemplo. Em outras palavras, do que adianta o estudante conhecer o assunto que está sendo abordado no texto, sendo que isso não vai fazer “diferença” no seu ponto de vista e no seu processo de leitura?

Como afirma Santos (2011), quando o leitor lê o texto fazendo previsões e ativando seu conhecimento acerca daquele tópico, ele

Estabeleceu uma postura diante do texto que lhe deu objetivos concretos para sua leitura: confirmar (ou não) suas previsões; refinar (ou não) tais previsões. Independentemente de acertar ou errar tais previsões, sua conexão com o texto fez [...] estabelecer um diálogo com ele. As previsões impedem que o texto seja maior que você, ‘transmitindo’ informações que você vai absorver passivamente; ao prever o assunto do texto, você está se posicionando

ativamente na leitura e tornando-se agente da leitura que está por iniciar (SANTOS, 2011, p. 63).

Como é possível observar, é preciso pensar em como essas previsões serão utilizadas, elas estando corretas ou não, já que, como afirma a autora, o importante é estabelecer uma relação íntima com o texto. Na presente pesquisa, essa estratégia foi utilizada explicitamente em perguntas para incitar o aluno a ativar o seu *background* acerca do assunto dos textos abordados e, também, em questões de *warm up*, a fim de abrir uma discussão entre os alunos e trocar ideias e experiências.

Levando em consideração que um vestibulando não dispõe de muito tempo para realizar as questões da prova, é importante que as técnicas mencionadas acima sejam ensinadas aos alunos a fim de que eles as apliquem quando forem realizar os exercícios de compreensão textual, já que, tanto inferência, como o *skimming* e o *scanning* auxiliarão os alunos no que diz respeito à introdução ao assunto do texto, bem como o entendimento de determinadas questões, além de não gastar tempo com informações que, muitas vezes, não são necessárias para responder as perguntas da prova.

Utilizar todos esses recursos oferecidos por um texto faz com o que leitor tenha mais segurança e sucesso no momento da sua leitura. Santos (2011, p. 35) ainda enfatiza que “é importante notar que precisamos aprender a usar esses recursos como fontes de informações” e que “esses dois aspectos (o visual e o verbal) contribuem em conjunto para a produção de sentido em um texto, e um leitor competente precisa atentar para ambas as formas de comunicação” (SANTOS, 2011, p. 35).

As técnicas de CE aqui explicitadas serão voltadas para um público específico, que como já se sabe, são os alunos do ensino médio (segundo e terceiro ano) que pretendem prestar o exame vestibular. Assim, no próximo subcapítulo, será discorrido acerca do ensino de leitura em um contexto IFE.

2.2.1 O ensino de compreensão escrita no contexto IFE

A habilidade de CE é vista como uma das habilidades mais próximas dos alunos, isto é, aquela que os alunos mais têm afinidade, sentem menos dificuldade. Isso se deve ao fato, talvez, de que, como afirma Moore (1983, p. 120, tradução nossa), no contexto IFE, “o aluno raramente vem ‘cego’ para o campo da leitura⁴”. Em outras palavras, o estudante de inglês, ainda que

⁴ *The student rarely comes ‘blind’ to the business of reading* (MOORE, 1983, p.120).

tenha nível básico, normalmente já teve algum contato com a leitura na LI, mesmo que de modo superficial, visto que, no Brasil, essa é uma das habilidades mais exigidas da população.

Pode-se notar que, na área acadêmica, cursos de graduação e, principalmente, de pós-graduação, exigem do estudante pelo menos a leitura (de nível intermediário) de textos acadêmicos, tornando-se um pré-requisito para o ingresso nesses cursos, nos quais o aluno terá de lidar com leituras, especialmente na LI, visto que, grande parte das pesquisas, de quaisquer áreas, são realizadas nessa língua. Prova dessa necessidade, são cursos presenciais e/ou online que abordam apenas a habilidade de CE para fins acadêmicos, como, por exemplo, o curso “Inglês para Fins Específicos: Leitura de Textos Acadêmicos – Online” disponibilizado pela PUC - Pontifícia Universidade Católica de São Paulo.

Faz-se importante, assim, destacar como as pesquisas têm focado na habilidade de CE dentro da abordagem IFE, corroborando o que Jordan (1997, p. 143, tradução nossa) afirma: “Quando os alunos leem, é para um propósito⁵”, e esse é o “ponta pé” inicial para a leitura. Desse modo, como os professores/pesquisadores podem desenvolver essa habilidade no aluno por meio de um curso/aula voltado especialmente para as suas necessidades?

De acordo com Jordan (1997, p. 51), os alunos tendem a afirmar em autoavaliações, questionários e pesquisas que a CE é a habilidade que eles têm menos dificuldade e isso também pode ser provado, nesta pesquisa, por meio de uma pergunta feita no primeiro dia de aula, em que os estudantes respondem ter mais intimidade com a leitura em relação às outras três habilidades – compreensão oral, produção oral e produção escrita. Além do mais, o autor menciona que esse fato não quer dizer que não haja problema algum quando diz respeito à abordagem da CE. Os alunos dizem ter afinidade com a leitura em LI, porém, muitas vezes, ela não é eficaz o suficiente ou alguns problemas são encontrados nesse processo. Dessa maneira, como é possível ajudar os alunos a desenvolverem a leitura de modo mais efetivo por meio da abordagem IFE?

O que se pretende mostrar aqui é que, de um modo geral, a visão acerca do texto tem mudado e, como Johns e Davies (1983) afirmam, “onde quer que o propósito subjacente à aprendizagem de uma língua estrangeira seja claro para o aluno e/ou para o seu professor, o uso de textos nessa língua imediatamente assume importância central⁶” (JOHNS; DAVIES, 1983, p. 2).

⁵ *When students read, it is for a purpose* (JORDAN, 1997, p. 143).

⁶ *Wherever the purpose underlying the learning of a foreign language is clear to the learner and/or to his teacher, the use of texts in that language immediately assumes central importance* (JOHNS, DAVIES, 1983, p. 2)

Novamente, é perceptível que o objetivo e as necessidades do estudante (bem como a clareza que o professor tem dessas necessidades) se tornam essenciais no contexto da abordagem IFE, uma vez que, caso essa percepção não seja precisa, ambos (aluno e professor) falham no processo de ensino-aprendizagem da língua – no caso, o inglês – visto que o professor não estará transmitindo ao aluno o que ele necessita saber, e, conseqüentemente, o aluno não aprende o que necessita.

Desse modo, os autores propõem, por meio dos acrônimos TALO (*text as linguistic object*) – texto como objeto linguístico – e TAVI (*text as a vehicle of information*) – texto como veículo de informação – uma nova perspectiva no que diz respeito ao texto utilizado em sala de aula. Para os estudiosos, o primeiro método foca na estrutura sintática da língua em um determinado texto e pode ser utilizado para o aprendizado de novo vocabulário. Nesse método, o professor seleciona um texto em que haja exatamente o léxico (ou o tópico gramatical) que ele deseja ensinar ao aluno e, por isso, esse método define o texto como um objeto linguístico.

Por outro lado, o segundo método, o mais defendido pelos autores, representa, como eles próprios afirmam, uma junção de novos desenvolvimentos na área tanto de inglês como primeira língua quanto segunda língua, sendo o texto um veículo de informação e não apenas transmissor de características linguísticas. Desse modo, ainda segundo Johns e Davies (1987), as abordagens TALO e TAVI podem ser comparadas em cinco aspectos:

1. *Princípios subjacentes à seleção de textos;*
2. *Atividades preparatórias para a leitura de textos;*
3. *Trabalhar com texto;*
4. *O tipo de interação de ensino/aprendizagem envolvida;*
5. *Atividades de acompanhamento para a leitura de texto.*

A respeito da escolha de textos, levando em conta o método TALO, Johns e Davies (1983) asseguram que o conteúdo deveria apresentar, como dito, o vocabulário e a estrutura gramatical a ser ensinada pelo professor em determinada aula e, a fim de os encaixar a essas características, os textos selecionados podem ser adaptados ou até mesmo reescritos. Aqui, o assunto abordado no texto, ou seja, seu conteúdo informativo, fica em segundo plano. Outro ponto importante desse método é que os textos são selecionados contendo temas gerais e não específicos (especializados), tornando-se mais “compreensível” para ambos (professor e aluno).

No que diz respeito à seleção de textos referente ao método TAVI, essa seleção dependerá do objetivo do aluno, se é imediato ou diferido. Os textos selecionados precisam ter valor para o estudante. Valor, aqui, é explicado por Johns e Davies (1983, p. 4) como algo que,

nos textos, “pode ser avaliado na medida em que eles correspondem às necessidades da especialização dos alunos: o que os alunos *precisam* saber⁷”.

No ponto de vista TAVI, o conteúdo dos textos deve ir ao encontro das necessidades do estudante, ou seja, não há sentido em abordar um tópico que o aluno não irá se deparar no contexto em que vivencia. O conteúdo (mensagem do texto) importa mais do que a estrutura gramatical e/ou o lexical contida em si. Entretanto, Johns e Davies (1983) explicam que léxico e sintaxe também são levados em consideração no método TAVI, mas não são o ponto mais relevante. Assim, o objetivo desse método é “fornecer aos alunos a experiência de lidar com uma gama completa de textos autênticos que eles provavelmente irão encontrar em seus estudos”⁸ (JOHNS, DAVIES, 1983, p. 4).

É importante mencionar que, no presente estudo, não há como selecionar textos especializados para os alunos, visto que, no contexto em que eles estão inseridos – sala de aula no ensino médio – eles se deparam com textos de temas diversos, assim, há uma grande diferença em selecionar textos para esses alunos do que para aqueles que estudam uma área já específica, como Biologia, por exemplo. Professores IFE que pretendem selecionar textos com o intuito de preparar material para alunos de Biologia irão, com certeza, escolher textos dessa área. Porém, visto que os vestibulandos se deparam com textos diversos tanto nas aulas quanto no momento de prestar o exame, foi necessário fazer uma seleção diferente para esse contexto. O uso real dos vestibulandos é exatamente prestar um exame, é para isso que eles serão preparados, esse é o uso que eles terão “fora da sala”.

Por esse motivo foram compilados os quatro corpora dos temas saúde, política, tecnologia e mundo. Reitera-se que o objetivo desta pesquisa não foi trabalhar com agrupamentos de especialidade como na área de saúde, por exemplo, mas sim abordar textos sobretudo atuais, a fim de fazer com que os alunos o lessem sob uma perspectiva diferente – não lendo e traduzindo palavra por palavra, mas observando e buscando encontrar agrupamentos que, no texto em questão, os ajudasse a compreender melhor o contexto. Ou seja, é importante que os vestibulandos se atualizem tanto no que diz respeito às notícias (textos jornalísticos), quanto em relação ao seu modo de leitura.

Dessa maneira, no que concerne à seleção textual, há presente, nesta pesquisa, o uso do método TAVI. Isso se justifica pelo fato de as atividades fornecerem experiência de textos

⁷ *May be measured by the extent to which they match the needs of the students' specialization: what the students need to know* (JOHNS, DAVIES, 1983, p. 4).

⁸ *To provide students with the experience of dealing with the full range of authentic texts they are likely to encounter in their studies* (JOHNS, DAVIES, 1983, p. 4).

autênticos para os alunos, uma vez que eles vão lidar com isso na vida real e no vestibular. Esses textos não são modificados, tampouco reescritos; eles são selecionados de acordo com o que o *corpus* oferece ao professor/pesquisador, a fim de abordar um léxico e/ou um ponto gramatical específico, e a mensagem do texto faz-se relevante, pois há questões que exigem do aluno buscar pelo entendimento do texto, como “qual a ideia geral do texto?”.

Em relação às atividades preparatórias para a compreensão textual, o segundo tópico que diferencia os dois métodos abordados, Johns e Davies (1983) afirmam que, geralmente, não há atividades preparatórias para a compreensão textual no método TALO e que isso faz com que aluno e professor percam a essência da leitura, qual seja: “decifrar o que o texto significa⁹” (JOHNS, DAVIES, 1983, p. 9). Por outro lado, levando em consideração o método TAVI, as atividades preparatórias são muito relevantes para a pré-leitura do texto e, segundo os autores:

Elas atuam como localizadores de direção, despertando interesse e estabelecendo o propósito para o qual o texto deve ser lido, o tipo de informação que pode ser encontrado nele e o valor que a informação pode ter para o aluno (tradução nossa)¹⁰ (JOHNS, DAVIES, 1983, p. 9).

No que diz respeito às atividades preparatórias, Johns e Davies (1983) dão algumas opções de exercícios que podem ser utilizados nesse momento da aula, como a inversão da ordem das atividades (colocar os exercícios de compreensão antes da leitura do texto), um quiz de pré-exercícios acerca do assunto do texto, diagramas relacionados ao texto, bem como uma atividade mais livre para os alunos. Esse último exercício foi utilizado nesta pesquisa, em que, nas atividades entregues aos alunos, foi nomeado de *warm up*. Esse exercício é dito como mais livre, pois se dá por meio de uma discussão do assunto entre os alunos. Inicialmente, pede-se para que os estudantes leiam o título do texto em questão e abra uma discussão acerca do que pode ser abordado nesse texto.

Além desse tipo de exercício que utiliza o título do texto, também foram discutidas outras questões já trazidas pelo próprio professor-pesquisador. Quando se abordou o tema “mundo”, por exemplo, o assunto principal do texto era um desastre natural; foi, então, aberta uma discussão acerca desse tópico, com perguntas do tipo: “Quais desastres naturais você já ouviu falar/conhece?” e “Quais fatores podem ocasionar um desastre natural?”.

⁹ *The puzzling out of what the text means* (JOHNS, DAVIES, 1983, p. 9).

¹⁰ *They act as Direction-finders, awakening interest and establishing the purpose for which the text is to be read, the sort of information that may be found in it, and the value that information may have for the student* (JOHNS, DAVIES, 1983, p. 9).

Dessa forma, acredita-se que, como asseguram Johns e Davies (1983), esses pré-exercícios de CE preparam os alunos para a leitura, por isso os autores os nomeiam *preparatory activities* – atividades preparatórias. Conforme os próprios estudiosos, essas atividades dão direção ao aluno, despertam seu interesse pelo assunto que é abordado no texto, além de instigar o tipo de informação a ser encontrada na leitura. Assim, pode-se afirmar que, nesse sentido, a presente pesquisa faz uso do método TAVI.

O trabalho com o texto é o terceiro tópico que difere os dois métodos. Segundo Johns e Davies (1983), quando o texto é trabalhado de acordo com o TAVI, o aluno focaliza não detalhes do texto, mas a ideia textual de um modo geral, além de utilizar o que já sabe em vez de informações desconhecidas. No entanto, quando utilizado o TALO, o foco se dá no ensino linguístico em oposição à informação. Ademais, os estudantes focalizam naquilo que *não* é de seu conhecimento em vez do que já é conhecido, e os detalhes são destacados em contraposição com o entendimento geral do texto. De modo geral, pode-se dizer que as questões são de nível mais linguístico e as respostas devem ser exatas, ou seja, não há espaço para discussões acerca do *background* do aluno.

É importante explicar o fator “linguístico” discutido aqui. Como mencionado, foram abordados, em todas as atividades, exercícios de *warm up* como preparatórios para a posterior compreensão textual. Nesse momento, os alunos tinham espaço para discutir acerca de seu conhecimento a respeito do tema das atividades, opinar e dividir suas vivências em relação àquele tema em específico. Entretanto, como é sabido, o vestibular é um exame que exige do candidato não apenas conhecimento do assunto geral do texto, mas também dados linguísticos a seu respeito, como o sentido de determinadas palavras ou pontos gramaticais, sendo, desse modo, inviável não abordar tópicos acerca da linguagem nas atividades.

Um ponto muito relevante mencionado por Johns e Davies (1983) em seu trabalho foi a utilização de estratégias de leitura por alunos “bem-sucedidos” no que concerne à compreensão textual em contradição àqueles que não eram bem-sucedidos na leitura. Para isso, os autores mencionam a pesquisa de Hosenfeld (1977 *apud* JOHNS, DAVIES, 1983), a qual assegura que esses dois tipos de leitores utilizam estratégias diferentes no momento da leitura. Para a pesquisadora, alunos bem-sucedidos tendiam a tentar compreender o texto de um modo geral, não focando, tampouco parando para traduzir palavras desconhecidas no texto. Em contraste a esses alunos, aqueles que falhavam mais na compreensão textual escolhiam “travar” a leitura em busca dessas palavras desconhecidas, fator que tende fazer com que os alunos esqueçam do significado do texto com um todo.

Por outro lado, o estudante bem-sucedido buscava pular as palavras que não eram de seu conhecimento, a fim de entendê-las por meio do contexto em que estavam inseridas. Assim, o uso do dicionário era mais raro, pois eles desenvolviam sua habilidade de inferência.

Novamente, faz-se necessário enfatizar que, apesar dos pontos linguísticos terem sido parte essencial nos exercícios de CE desenvolvidos nesta pesquisa, estratégias de leitura como a inferência também foram trabalhadas no decorrer das aulas.

No que diz respeito ao quarto tópico que diferencia os dois métodos, Johns e Davies (1983) afirmam que:

TALO está associado a dois estilos de ensino: um monólogo de um professor, os alunos tomando notas; ou um diálogo em que o professor faz as perguntas, os alunos respondem, e o professor avalia e comenta as respostas (ou seja, dá feedback) [...] Com o método TAVI, por outro lado, nosso objetivo é fazer com que os alunos trabalhem no texto em grupos de dois ou três dentro do quadro claro das atividades descritas acima [...] dá-lhes uma oportunidade de comunicação real na sala de aula [...] a língua-alvo pode ser introduzida gradualmente [...] o argumento mais importante para o trabalho em grupo é que ele pode fornecer um modelo para o auto estudo dos alunos fora da sala de aula, onde é provável que ocorra a maior parte de sua aprendizagem real (JOHNS, DAVIES, 1983, p. 12, tradução nossa)¹¹.

Voltando às duas propostas para a presente pesquisa, pode-se dizer que o tipo de interação entre os alunos se encaixa no segundo método no que diz respeito à interação com a sala toda durante o exercício de *warm up*, em que todos os alunos tinham a oportunidade de opinar e falar um pouco do que sabiam a respeito de determinado tópico. As outras atividades – as preparatórias, como mencionado anteriormente – foram desenvolvidas em pares, ou seja, o aluno não precisava fazer a atividade sozinho; pares eram formados para eles trocarem ideias no que concerne ao vocabulário que estava sendo utilizado.

Em relação aos exercícios de CE ao final de cada atividade, TALO foi o método mais utilizado, visto que os alunos respondiam às questões e o professor-pesquisador dava o *feedback*.

¹¹TALO is associated with two styles of teaching: a monologue by a teacher, the students taking notes; or a dialogue in which the teacher asks the questions, the students reply, and the teacher assesses and comments on their responses (i.e. give feedback) [...] With TAVI method, on the other hand, our aim is to get the students to work on the text in groups of two or three within the clear framework of the activities described above [...] it gives them an opportunity for real communication in the classroom [...] the target language can be gradually introduced [...] the most important argument for groupwork is that it can provide a model for the students' individual self-study outside the classroom, where most of their real learning is likely to take place (JOHNS, DAVIES, 1983, p. 12).

No que diz respeito às atividades de acompanhamento para a leitura de texto – quinto e último tópico – os métodos TALO e TAVI diferenciam-se exatamente pelo tipo de exercício envolvido depois dos exercícios preparatórios. O fato é que o primeiro método se preocupa com questões de compreensão textual, sejam elas gramaticais e/ou lexicais; já com o método TAVI, os estudantes se deparam com o uso real da língua, e os autores Johns e Davies (1983) classificam TAVI em quatro tópicos – transferência de informação, aplicação ou explicação de informação, extensão da informação e aplicação de técnicas.

Voltando-se novamente para o presente trabalho, pode-se assegurar que ambos os métodos são utilizados, porém, de modo e em momentos diferentes. TALO é certamente usado no desenvolvimento e, posteriormente, na resolução dos exercícios, pois eles se enquadram em um método que aborda a compreensão textual no que concerne, principalmente, ao vocabulário abordado no texto, já que, nas atividades preparatórias, o foco se dá no léxico apresentado como relevante pelo *corpus* de estudo. Por outro lado, no método TAVI, Johns e Davies (1983) explicam, no tópico “a aplicação de técnicas”, que é muito relevante a aplicação das técnicas ensinadas em sala de aula a outros textos de escolha dos próprios alunos. Os estudiosos ainda acrescentam que:

Isso é importante por três razões principais. [...] É essencial que eles vejam como podem aplicar as técnicas aprendidas ali aos textos extraídos de sua própria disciplina. Em segundo lugar, este aspecto do trabalho de acompanhamento será a indicação mais importante para o professor de quão bem-sucedido foi o trabalho em sala de aula [...] isto é, que as técnicas podem ser usadas, e usadas imediatamente, numa vasta gama de textos aparentemente muito diferentes. [...] Em terceiro lugar, é provável que tenha um efeito muito importante no aumento da autoconfiança dos alunos e de sua capacidade de trabalhar de forma independente (JOHNS, DAVIES, 1983, p. 14, tradução nossa)¹².

Ademais, pode-se dizer que TAVI também foi utilizado, pois é extremamente relevante deixar clara a importância de se “replicar” as técnicas ensinadas em sala de aula. Um dos pontos destacados durante a aplicação das atividades aos alunos foi exatamente o fato de se aplicar, o que foi ensinado, em outros textos. Em outras palavras, deve-se atentar às técnicas de leitura, pois elas são reaplicáveis em quaisquer outros textos em inglês e/ou outra língua.

¹²*This is important for three main reasons. [...] it is essential that they should see how they can apply the techniques learnt there to texts drawn from their own discipline. Secondly, this aspect of the follow-up work will be the most important indication to the teacher of how successful the classroom work has been [...] that is to say, that the techniques can be used, and used immediately, on a wide range of apparently very different texts. [...] Thirdly, it is likely to have a very important effect in increasing the students' self-confidence and their ability to work independently (JOHNS, DAVIES, 1983, p. 14).*

Assim sendo, esta pesquisa propõe a junção dos dois métodos para casos em que a compreensão do texto de modo geral é importante (como deve ser), mas também quando é necessária a abordagem de tópicos linguísticos – lexicais e/ou gramaticais, como é o caso do presente trabalho.

Em um curso IFE para vestibular, os alunos têm de entender o sentido do texto como um todo e utilizar estratégias de leitura para esse fim, visto que há perguntas de compreensão geral. Todavia, esses alunos também necessitam entender a parte mais linguística do texto, pois isso também é exigido deles no momento de prestar o exame. Deve-se, portanto, entender como cada característica dos dois métodos se encaixam em cada pesquisa de modo específico.

O próximo subcapítulo enfatiza a importância de um dos pontos tratados aqui: a autenticidade nos textos de CE e sua relevância em sala de aula.

2.2.2 Autenticidade nos textos de compreensão escrita

A autenticidade é um termo que gera muita discussão e divide opiniões no meio acadêmico, especialmente no que diz respeito ao desenvolvimento e ao uso de material didático. Enquanto alguns autores (WIDDOWSON, 1990; HARMER, 1994; NUTTALL, 1996; CARTER, NUNAN, 2001) defendem a ideia da autenticidade no contexto educacional, outros estudiosos (MARTINEZ, 2002; KILICKAYA, 2004) citam, na verdade, desvantagens em se utilizar materiais autênticos. No entanto, muitas vezes, o que acontece é a desmotivação do professor ao preparar a aula, devido ao tempo consumido para preparação e difícil vocabulário e estrutura encontrados em determinados textos, bem como à desmotivação por parte do aluno ao se deparar com um material de difícil entendimento.

Por outro lado, o linguista Sinclair defende sua forte opinião ao afirmar que “não se deve permitir que as dificuldades apoiem a noção absurda de que exemplos inventados podem realmente representar a linguagem melhor do que os reais”¹³ (SINCLAIR, 1991, p. 5). Além disso, o autor ainda acrescenta que “atualmente, há sinais de um reconhecimento crescente de que o estudo abrangente da linguagem deve ser baseado em evidências textuais. Não se estuda toda a botânica fazendo flores artificiais”¹⁴ (SINCLAIR, 1991, p. 6).

¹³ *The difficulties should not be allowed to support the absurd notion that invented examples can actually represent the language better than real ones* (SINCLAIR, 1991, p. 5).

¹⁴ *Currently there are signs of a growing recognition that the comprehensive study of language must be based on textual evidence. One does not study all of botany by making artificial flowers* (SINCLAIR, 1991, p. 6).

Materiais autênticos e a própria autenticidade são definidos por diferentes autores, ainda que compartilhem da mesma opinião. Para Martinez (2002), o material autêntico é aquele que foi preparado por nativos e não tem fins educacionais. Os autores Carter e Nunan (2001, p. 68) corroboram essa ideia ao afirmarem que materiais autênticos são aqueles que contêm “textos comuns não produzidos especificamente para o ensino de línguas”¹⁵. O uso de *corpora* faz da autenticidade um dos pontos cruciais no desenvolvimento de materiais, “uma vez que a LC pode fornecer descrições da língua real em uso, essa informação pode então ser usada para modelar e desenvolver materiais para ensino de línguas”¹⁶ (REPPEN, 2010, p. 4).

Harmer (1994), por sua vez, defende a ideia de que há vários efeitos positivos em adotar materiais autênticos no ensino de línguas, como ajudar os alunos a produzirem uma linguagem melhor e a terem uma aquisição mais rápida, além de se tornarem mais confiantes para lidar com situações da realidade. Na CE, especialmente, Berardo (2006, p. 62) explica que o uso da linguagem não autêntica torna os textos “muito diferentes de qualquer coisa que o aluno vai encontrar no mundo real e, muitas vezes, eles não refletem como a língua é realmente usada”¹⁷. O autor acrescenta, ainda, que esses textos podem ser úteis no ensino de estruturas, mas, quando se trata de melhorar as habilidades de CE, textos não autênticos não são uma boa escolha, exatamente por não permitirem uma leitura natural.

É interessante mencionar que, assim como Berardo (2006, p. 62) esclarece, “os materiais autênticos permitem que os alunos interajam com língua e conteúdo reais ao invés da forma. Os alunos sentem que estão aprendendo a língua-alvo como ela é usada fora da sala de aula”¹⁸. Ainda de acordo com Berardo (2006, p. 63), há quatro importantes fatores no que diz respeito a escolha de materiais autênticos para leitura, como se pode verificar no quadro abaixo:

Quadro 1 – Fatores importantes na escolha de material de leitura autêntico

Fatores importantes na escolha de material de leitura autêntico ¹⁹	
Adequação do conteúdo	<i>O texto interessa ao aluno?</i>
	<i>É relevante para as necessidades do aluno?</i>

¹⁵ Ordinary texts not produced specifically for language teaching purposes (CARTER; NUNAN, 2001, p. 68).

¹⁶ Since corpus linguistics can provide descriptions of actual language use, this information can then be used to shape and develop language-teaching materials (REPPEN, 2010, p. 4).

¹⁷ The artificial nature of the language and structures used, make them very unlike anything that the learner will encounter in the real world and very often they do not reflect how the language is really used (BERARDO, 2006, p. 62).

¹⁸ Authentic materials enable learners to interact with the real language and content rather than the form. Learners feel that they are learning a target language as it is used outside the classroom (BERARDO, 2006, p. 62).

¹⁹ Acerca da tabela acima serão feitos comentários no Capítulo Materiais e Metodologia.

	<i>Representa o tipo de material que o aluno usará fora da sala de aula?</i>
Explorabilidade	<i>O texto pode ser explorado para fins de ensino?</i>
	<i>Para que finalidade o texto deve ser explorado?</i>
	<i>Quais habilidades / estratégias podem ser desenvolvidas explorando o texto?</i>
Legibilidade	<i>O texto é muito fácil / difícil para o aluno?</i>
	<i>É estruturalmente muito exigente / complexo?</i>
	<i>Quanto vocabulário novo contém? Isso é relevante?</i>
Apresentação	<i>“parece” autêntico?</i>
	<i>É “atraente”?</i>
	<i>Isso chama a atenção do aluno?</i>

Fonte: Berardo (2006, p. 63)

Como visto, várias perguntas são feitas antes da escolha do material e, por mais simples que pareça, muitas vezes os professores não param para analisar cada tópico desse, assim, o texto até pode abranger o vocabulário que o profissional deseja abordar em sala de aula, mas será que esse texto é adequado para os alunos, ou seja, será que eles teriam algum tipo de interesse por essa leitura? Ou, ainda, será que o texto não é fácil/difícil demais? Desse modo, o quadro acima será analisado mais detalhadamente dentro desta pesquisa no Capítulo 4.

Carter e Nunan (2001) compartilham do mesmo pensamento dos autores supracitados ao afirmarem que o aprendizado é facilitado por meio do uso da língua em seu meio autêntico e enfatizam o uso de *corpora* como mediadores entre o ensino-aprendizado do inglês e sua autenticidade: “O papel dos *corpora* de uso autêntico da língua é importante para entender como a linguagem é usada em contextos autênticos, e tais *corpora* também podem ser usados como recursos didáticos e no desenvolvimento de materiais didáticos”²⁰ (CARTER; NUNAN, 2001, p. 215).

O fato de se utilizar *corpora* no ensino também requer do professor que alguns requisitos sejam seguidos para que a atividade não fique descontextualizada. Como foi mostrado acima,

²⁰ *The role of corpora of authentic language use is important in understanding how language is used in authentic contexts, and such corpora can also be used as teaching resources and in the development of teaching materials* (CARTER; NUNAN, 2001, p. 215)

há fatores que são importantes em uma atividade de CE. Para evitar problemas desse tipo, O'Keeffe, McCarthy e Carter (2007) afirmam que:

Em nossa experiência, *corpora*, falados e escritos, realmente contêm muitos textos que são obscuros e culturalmente opacos, mas eles também contêm inúmeros textos que são transparentes, facilmente contextualizados e interpretáveis por qualquer ser humano maduro. É simplesmente uma questão de como alguém cuidadosamente seleciona o material, quem é o público-alvo e o que eles querem e esperam de um curso de idiomas (O'KEEFFE; MCCARTHY; CARTER, 2007, p. 27, tradução nossa²¹).

Dessa maneira, desde que os exercícios sejam selecionados atenciosamente, contextualizados e estejam seguindo os critérios estabelecidos para o desenvolvimento de atividades com *corpora* (a tabela de critérios está disponível no subcapítulo 3.3), a LC muito tem a contribuir no campo do ensino-aprendizagem de línguas, especialmente no que tange ao ensino de leitura com textos autênticos.

Para melhor entender os fundamentos da LC e como ela tem atuado no campo do desenvolvimento de materiais, os subcapítulos a seguir irão discorrer sobre essa abordagem teórico metodológica e sobre como ela é utilizada na presente pesquisa.

2.3 A Linguística de *Corpus* na pesquisa

A LC é um campo de pesquisa que se ocupa da “coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” (BERBER SARDINHA, 2004, p. 03). Sua base teórica está pautada no estudo empírico da língua, isto é, em evidências do uso real da língua. De acordo com Viana (2011, p. 26), “é necessário basear o estudo em um alicerce empírico no qual os resultados advêm da observação de exemplos reais”.

Os dados linguísticos que, antes, eram pautados em especulações, passaram a ter uma nova perspectiva ao perceberem que a língua poderia ser analisada por meio do estudo de seu uso, isto é, com ela era realmente produzida pelos seus usuários, focando na probabilidade e não mais na possibilidade. Segundo Sinclair (1991, p. 4)

²¹ In our experience, *corpora*, both spoken and written, do indeed contain many texts that are obscure and culturally opaque, but they also contain numerous texts that are transparent, easily contextualised and interpretable by any mature human being. It is simply a matter of how carefully one selects the material, who the end-users are and what they want and expect from a language programme (O'KEEFFE; MCCARTHY; CARTER, 2007, p. 26).

De fato, o contraste exposto entre as impressões de detalhes da linguagem observadas pelas pessoas e as evidências compiladas objetivamente a partir de textos é enorme e sistemático. Isso nos leva a supor que a intuição humana sobre a linguagem é altamente específica, e de modo algum um bom guia para o que realmente acontece quando as mesmas pessoas realmente usam a linguagem (SINCLAIR, 1991, p. 4, tradução nossa²²).

Como observado pelo pesquisador acima, os dados linguísticos fornecidos por meio do *corpus* são tidos como confiáveis e precisos devido a sua característica empírica, sistemática e estatística que um computador/concordanciador é capaz de oferecer. Segundo os pensamentos de McCarthy e O’Keffee (2010, p. 5), “foi a revolução em hardware e software nas décadas de 1980 e 1990 que realmente permitiu a linguística de corpus como a conhecemos emergir”²³.

A partir desse fato que, de certo modo, revolucionou o estudo linguístico e fez surgir um campo tão relevante para a linguística, McCarthy e O’Keffee (2010, p. 12) também afirmam que

Em suma, pode-se argumentar que a LC é uma disciplina saudável e vibrante no âmbito geral do estudo da língua. Suas origens eram não computacionais, mas sua explosão e expansão nos campos da linguística descritiva e aplicada são devidas principalmente à revolução da informação do final do século XX, uma revolução que continua e da qual LC, sem dúvida, continuará se beneficiando (MCCARTHY, O’KEFFEE, 2010, p. 12, tradução nossa ²⁴).

Assim, a LC se tornou grande influenciadora na pesquisa linguística, sendo “uma das áreas de pesquisa de linguagem mais ativas nos últimos anos” (BERBER SARDINHA, 2000, p. 324). Vários tipos de *corpora* passaram a emergir de acordo com as necessidades dos pesquisadores, como *corpus* geral, monitor, especializado, reciclável, de aprendiz, paralelo, entre outros que serviram como base teórico-metodológica para diversas pesquisas em diferentes áreas que buscavam, de alguma forma, um alicerce empírico em seus estudos.

Não tendo a intenção de explorar todos os tipos de *corpora* nesta pesquisa, apenas dois serão abordados e explicados brevemente, devido a sua presença neste trabalho, que são o

²² Indeed, the contrast exposed between the impressions of language detail noted by people, and the evidence compiled objectively from texts is huge and systematic. It leads one to suppose that human intuition about language is highly specific, and not all a good guide to what actually happens when the same people actually use the language (SINCLAIR, 1991, p. 4).

²³ It was the revolution in hardware and software in the 1980s and 1990s which really allowed corpus linguistics as we know it to emerge (MCCARTHY; O’KEFFEE, 2010, p. 5).

²⁴ All in all, CL can be argued to be a healthy, vibrant discipline within the general umbrella of language study. Its origins were non-computational but its explosion and expansion in the fields of descriptive and applied linguistics are due mainly to the information revolution of the late twentieth century, a revolution which continues, and from which CL will undoubtedly continue to benefit (MCCARTHY; O’KEFFEE, 2010, p. 12).

corpus geral e o *corpus* especializado, ou de estudo. O estudioso Sinclair (1991) afirma que, para ser considerado como geral, o *corpus* precisa ter uma boa seleção de exemplos de determinado idioma e servir para diferentes propósitos. Já o *corpus* especializado está em busca de um objetivo mais específico de pesquisa. Gavioli (2005) explica que o especializado costumava ter entre 50.000 e 100.000 palavras dependendo do propósito do pesquisador e, corroborando o pensamento de Sinclair, a autora diferencia os dois *corpora*, esclarecendo que

Os *corpora* “especializados” são corpora projetados com a finalidade de criar uma amostra de linguagem especializada, coletando textos de conteúdo semelhante (por exemplo, ciência, medicina, negócios, filosofia) ou de tipo ou gênero de texto semelhante (por exemplo, documentos de pesquisa, cartas, livros capítulos) ou ambos (por exemplo, artigos de pesquisa médica ou palestras científicas) ou mesmo textos de outros tipos de categorias especializadas, como a linguagem dos jornais ou a linguagem acadêmica. Os corpora “gerais” são normalmente corpora projetados com o objetivo de representar um idioma “inteiro” (por exemplo, inglês) ou uma variedade geográfica (por exemplo, inglês britânico) (GAVIOLI, 2005, p. 8, tradução nossa²⁵).

Tendo em vista as características de cada *corpus* apresentado acima, é possível afirmar que a presente pesquisa faz uso do *corpus* especializado no que concerne a sua finalidade, ou seja, cada *corpus* compilado para esta pesquisa teve um propósito específico que não era o de representar uma determinada língua, mas de analisar as palavras-chave de cada especialidade – os tópicos mundo, tecnologia, saúde e política – e os agrupamentos lexicais de cada *corpus*. Como Gavioli (2005) explica, *corpora* especializados não são caracterizados dessa maneira apenas por representarem tipos ou tópicos textuais específicos, “mas também porque foram criados para um interesse específico de ensino e/ou aprendizado” (GAVIOLI, 2005, p. 6).

É importante mencionar que, além de especializado, cada um dos quatro *corpora* compilados nesta pesquisa caracteriza-se, também, por ser reciclável, ou seja, por ter um objetivo específico: o pesquisador atualiza o seu *corpus* de acordo com a necessidade de sua pesquisa. O presente estudo necessita de um *corpus* desse tipo, pois, tratando-se de vestibular, os alunos que terão contato com as atividades precisam de conteúdo atual. Cada *corpus* abrange textos jornalísticos atuais do ano correspondente ao vestibular que os alunos prestarão o exame,

²⁵ *Specialized corpora are corpora designed for the purpose of creating a sample of specialized language either by collecting texts of similar content (e.g. science, medicine, business, philosophy) or of similar text-type or genre (e.g. research papers, letters, book chapters) or both (e.g. medical research articles or science lectures), or even texts from other types of specialized categories, such as newspaper language or academic language. “General” corpora are normally corpora designed with the aim of representing a “whole” language (e.g. English) or a geographical variety (e.g. British English) (GAVIOLI, 2005, p.8).*

porém, esses *corpora* estarão desatualizados para os anos seguintes, sendo necessário adicionar ou, até mesmo, retirar textos do *corpus* em questão. Considera-se, dessa forma, reciclável. Mais informações acerca desse tipo de *corpus* serão fornecidas no subcapítulo 2.4.2.

Levando em consideração todas as características que até aqui descreveram parte do que a LC tem sido na pesquisa linguística, pode-se assegurar que ao longo dos anos a LC passou a atuar em diversas áreas, dentre elas, no processamento de linguagem natural, no campo da lexicografia, da tradução, do ensino de línguas, entre outros (BERBER SARDINHA, 2004). Como dito, esse último é o foco desta pesquisa: a utilização de *corpus* no ensino de línguas, que também se encaixa na produção de materiais. Mais acerca desse assunto será discorrido no próximo subcapítulo.

Tendo o *corpus* definido segundo o objetivo do pesquisador, é necessário saber como a pesquisa será guiada, ou seja, dirigida ou baseada em *corpus*. De acordo com Biber (2012)

Os estudos de *corpus* usaram duas abordagens principais de pesquisa: 'baseada em *corpus*' e 'dirigida por *corpus*'. A pesquisa baseada em *corpus* assume a validade de formas e estruturas linguísticas derivadas da teoria linguística. O objetivo principal da pesquisa é analisar os padrões sistemáticos de variação e uso desses recursos linguísticos predefinidos. A pesquisa dirigida por *corpus* é mais indutiva, de modo que as construções linguísticas emergem da análise de um *corpus* (BIBER, 2012, p. 162, tradução nossa²⁶).

O autor explica, desse modo, que se determinada pesquisa for baseada em *corpus*, isso significa que o pesquisador já possui uma hipótese e/ou teoria, e ela será confirmada ou não pela análise do *corpus*, por meio dos dados que são oferecidos ao observá-lo. Por outro lado, se o trabalho está sendo dirigido por *corpus*, o pesquisador desenvolve uma hipótese/teoria por meio daquilo que é observado nele, isto é, a hipótese emerge dos dados apresentados pelo próprio *corpus*.

Tendo isso em vista, é possível afirmar que a presente pesquisa se encaixa na segunda abordagem, a dirigida por *corpus*, visto que não há uma hipótese do que será encontrado nos *corpora* de estudo; é somente a partir da análise que os agrupamentos são extraídos e os exercícios são elaborados de acordo com o que cada *corpus* oferece. Prova disso é a diversidade de exercícios nos diferentes temas de atividades, ou seja, os exercícios desenvolvidos com o léxico encontrado no *corpus Politics*, por exemplo, não são os mesmos criados para a atividade

²⁶ *Corpus studies have used two major research approaches: 'corpus-based' and 'corpus-driven'. Corpus-based research assumes the validity of linguistic forms and structures derived from linguistic theory. The primary goal of research is to analyse the systematic patterns of variation and use for those pre-defined linguistic features. Corpus-driven research is more inductive, so that the linguistic constructs themselves emerge from analysis of a corpus* (BIBER, 2012, p. 162).

do tema *Health*, já que cada *corpus* apresentou uma gama diferente de agrupamentos lexicais e, por esse motivo, as atividades desenvolvidas foram diferentes.

No próximo subcapítulo será abordado como se dá a LC no campo de ensino-aprendizagem de línguas, mais especificamente de LI, discorrendo principalmente acerca das suas vantagens nessa área que pode beneficiar tanto o aluno quanto o professor

2.3.1 O uso de LC no ensino de LI

É sabido que professores de línguas estrangeiras procuram sempre ferramentas que os auxiliem na preparação de aulas, com o intuito de aperfeiçoar, cada vez mais, o processo de ensino-aprendizagem, oferecendo aos aprendizes aulas que atinjam seus objetivos de maneira eficaz.

Para muitos professores, um dos fatores que implica muito trabalho quando utilizada a abordagem IFE é a necessidade de se organizarem muito previamente, para se dedicarem à preparação e à seleção de materiais específicos utilizados em suas aulas, pois cada aluno possui um objetivo diferente.

Dessa maneira, pesquisadores da área de Linguística Aplicada passaram a utilizar *corpora* eletrônicos como fornecedores de contribuições importantes para o desenvolvimento de suas pesquisas, não apenas na elaboração de materiais, mas na pesquisa linguística de modo geral, visto que a LC

Ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica -se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BERBER SARDINHA, 2004, p. 3).

Segundo Berber Sardinha (2004), a utilização de *corpora* no ensino se divide em quatro áreas principais: descrição da linguagem nativa, descrição da linguagem do aprendiz, transposição de metodologias de pesquisa acadêmica para a sala de aula e desenvolvimento de materiais de ensino, currículos e abordagens. Complementando isso, Viana (2011) também aborda o ensino de línguas por meio de *corpora*, afirmando que há duas maneiras de se obter mais informações a respeito de uma língua e de seu funcionamento.

O primeiro modo diz respeito à intuição, ou seja, julga-se quando uma regra ou determinado uso da língua é adequado ou não, obtendo, dessa maneira, uma ideia do que é/seria a língua em si, o que, segundo o autor, não corresponde exatamente ao seu uso real. Esse tipo

de situação pode, muitas vezes, deixar o professor em uma situação um tanto delicada, já que, para se ter certeza se determinados usos da língua estariam apropriados ou não em certas ocasiões, eles teriam de consultar um falante nativo.

A segunda maneira de se estudar uma língua é por meio da observação de dados autênticos, ou seja, como ela é realmente utilizada em um contexto que não se preocupa em utilizar conhecimentos intuitivos na comunicação, mas, ao contrário disso, tem-se a língua em sua espontaneidade. Nesse caso, o estudo é baseado em informações empíricas, por meio do uso da língua em interações que ocorrem naturalmente entre falantes nativos (VIANA, 2011).

Baseando-se na segunda abordagem da língua citada por Viana (2011), Berber Sardinha (2011) assegura que o professor já não deve depender apenas de meios tradicionais como lousa, giz e oratória para conduzir sua aula, uma vez que o mundo moderno oferece ferramentas computacionais e tecnológicas que auxiliam na criação de materiais, principalmente quando os objetivos são mais específicos, já que, assim, o professor pode pesquisar exatamente a característica linguística que deseja. Em outras palavras, se, por exemplo, o objetivo é ensinar *present perfect* para o aluno, é possível buscar em um *corpus* de inglês geral ocorrências que apresentam essa característica e o mesmo vale para o ensino do léxico.

Sinclair (1991; 2004), Carter, McCarthy (2001) e Bernardini (2004) compartilham a ideia de que a LC representa uma grande mudança no que diz respeito à educação em geral, tanto em relação aos métodos de ensino quanto ao papel do professor em sala de aula, além de fazer um importante intermédio entre teoria e prática.

Nesse contexto, Bernardini (2004) afirma que

Essa mudança de ênfase das rotinas de aprendizado dedutivo para indutivo tem efeitos abrangentes sobre: (a) o professor, que se torna coordenador da pesquisa ou facilitador; (b) o aluno, que aprende a aprender através de exercícios que envolvem a observação e interpretação de padrões de uso; (c) o papel das gramáticas pedagógicas, cujo nível de abstração muitas vezes contraria sua eficácia (BERNARDINI, 2004, p. 16-17, tradução nossa²⁷).

Tendo efeito sobre esses três pilares: professor, aluno e pedagogia, a LC se alia de modo eficaz ao contexto de ensino-aprendizagem, visto que, como explicitado pela autora, o professor passa a ser um facilitador que não somente transmite o conteúdo ao aluno, mas o faz envolver no seu próprio aprendizado por meio da observação do funcionamento da língua e não apenas

²⁷ This shift of emphasis from deductive to inductive learning routines has wide-ranging effects on: (a) the teacher, who becomes a coordinator of research, or facilitator; (b) the learner, who learns how to learn through exercises that involve the observation and interpretation of patterns of use; (c) the role of pedagogic grammars, whose level of abstraction often works against their effectiveness (BERNARDINI, 2004, p. 16-17)

por memorização gramatical e/ou lexical. É assim que, conseqüentemente, o papel pedagógico no ensino também sofre uma mudança, em que o aprendizado da língua se dá pela associação do significado e da forma em um processo indutivo (SINCLAIR, 1991; BERNARDINI, 2004).

Tendo em vista o processo de ensino-aprendizagem como uma descoberta do aluno em relação à língua, a LC se torna um meio efetivo e adaptável nesse contexto, já que, além de oferecer dados autênticos da língua, o professor pode adaptar o material às necessidades de seus alunos “graças à riqueza de dados e às infinitas possibilidades oferecidas pelos programas de software que são cada vez mais projetados com os alunos em mente”²⁸ (BERNARDINI, 2004, p. 22).

No que concerne ao uso de *corpora* em sala de aula, autores têm comprovado por meio de suas pesquisas que, muito do conteúdo abordado em sala de aula, não abrange unidades/frases que são muito frequentes na língua (HOLMES, 1988; TSUI, 2004; O’KEEFFE; MCCARTHY, 2007; DUTRA; BERBER SARDINHA, 2014). Seguindo essa linha de pensamento, Tsui (2004, p. 40) defende sua ideia, afirmando que

Concentrando-se em palavras com alta frequência de ocorrência e concentrando-se no usual e não no excepcional, os professores podem ajudar os alunos a adquirir o idioma de forma mais eficiente. [...] os resultados da análise de *corpus* podem ser usados como base para selecionar e sequenciar o conteúdo linguístico, bem como para determinar ênfases relativas (TSUI, 2004, p. 40, tradução nossa²⁹).

Como visto, a utilização de *corpora* no ensino pode colaborar no que diz respeito ao uso de conteúdo gramático e lexical que são frequentes na língua e que, desse modo, pode ajudar os alunos a se comunicarem e/ou compreenderem a língua de modo mais eficaz, inclusive ao economizar tempo focalizando em conteúdo que realmente importa ao aluno e não em frases que não são frequentes em sua área, e, muitas vezes, nem aparecem em determinado campo de estudo. Mesmo quando se trata da língua geral e não específica, há ocorrências que são muito mais frequentes que outras, o que seria muito mais interessante trabalhar com o aluno, em vez de dar mais relevância a conteúdos de baixa frequência.

Os estudos de Holmes (1988 *apud* TSUI, 2004) se voltam para a comparação de livros didáticos com o conteúdo extraído por *corpora*, e o autor encontra dados que comprovam a

²⁸ *thanks to the richness of the data and the endless possibilities offered by software programs which are more and more often designed with learners in mind* (BERNARDINI, 2004, p. 22).

²⁹ *By focusing on words which have a high frequency of occurrence and by concentrating on the usual rather than the exceptional, teachers can help learners acquire the language more efficiently [...] The findings of corpus analysis can be used as a basis for selecting and sequencing linguistic content, as well as for determining relative emphases* (TSUI, 2004, p. 40).

diferença de frequência em relação ao conteúdo abordado em cada situação. Os estudos levam em consideração verbos que expressam modalidade, como *appear, believe, doubt* e *suppose*, e unidades, como *possibility, tendency, likelihood*, bem como alguns advérbios do tipo *probably, of course, perhaps*, que são parte de um vocabulário muito utilizado na LI, todavia é dada pouca atenção em sala de aula.

O mesmo é apresentado na pesquisa de Ljung (1991, *apud* TSUI, 2004) que, ao comparar livros didáticos de IFE com o *corpus* Cobuild, observou que 20% das cem palavras mais frequentes dos textos dos alunos não constavam no *corpus* utilizado na comparação.

Fatos como esses mostram que o tempo dos alunos e do professor é possivelmente "desperdiçado" com informações que poderiam ser mais relevantes para o ensino-aprendizagem de línguas. Acerca desse assunto, Berber Sardinha (2004) explica que:

Não devemos sobrecarregar os alunos com vastas quantidades de informação sintática ["gramática"] de um lado e informação lexical ["vocabulário"] de outro, que eles terão de juntar de acordo com princípios não disponíveis naturalmente para eles na condição de falantes não-nativos. (BERBER SARDINHA, 2004, p. 261).

É dessa forma que a LC pode contribuir para a criação de materiais desta pesquisa, visto que as aulas são voltadas para objetivos específicos de alunos que irão prestar o vestibular, não sendo necessário expandir o ensino da língua inglesa de modo com que o aluno tenha de começar a estudar no nível básico até alcançar o avançado, fazendo que, assim, haja um melhor aproveitamento do conteúdo e do tempo disponíveis em sala de aula. A seguir será discutido o ensino lexical no contexto IFE sob a luz dos estudos com *corpora*.

2.3.2 IFE e o ensino de vocabulário sob a perspectiva da LC

Sabe-se que o vocabulário é parte fundamental para a compreensão de um texto, esteja ele em qualquer língua. Muitos alunos, quando iniciam um curso de LI, querem se expressar em inglês, mas sentem dificuldade em “encontrar” as palavras, ou seja, sentem uma “lacuna” tanto na produção da língua – fala e escrita – quanto na compreensão – leitura e audição.

É desse modo que, abordar o léxico da língua antes de uma leitura, é um fator fundamental para o entendimento do texto que será lido. Todavia, é importante mencionar que não é possível, e até inviável, abordar todas as palavras contidas em um texto, visto que, como elucidado anteriormente, o aluno não deve “travar” sua leitura para ficar procurando ou traduzindo todas as palavras que não consegue entender no texto em questão. Faz-se necessário

que, nessas situações, o aluno tente inferir o sentido de determinadas palavras por meio do contexto em que estão inseridas.

Sabendo disso, e voltando para a grande relevância que o léxico tem na leitura textual, Hirvela (2013, p. 84) afirma que “o vocabulário ocupa um lugar importante na leitura tanto na língua nativa quanto na segunda língua/língua estrangeira, visto que as palavras na página são o ponto de partida para a leitura”³⁰.

Nesse sentido, cabe ao professor pensar em formas de ensinar vocabulário aos seus alunos e fazer desse vocabulário parte importante para a compreensão textual. De acordo com Marshall e Gilmour (1993, p. 69), exercícios de pré-leitura são muito relevantes para uma posterior leitura do texto, além disso, deve-se evitar listas de palavras científicas, e “ensinar os contextos e relações estruturais dentro das quais as palavras têm significado”³¹ (MARSHALL, GILMOUR, 1993, p. 75).

Pautadas nos pensamentos dos autores supracitados, as atividades da presente pesquisa contam com exercícios de pré-leitura e exercícios contextualizados. Por meio do *corpus*, é possível extrair não apenas os agrupamentos lexicais como o contexto em que estão inseridos. Assim, os alunos foram instigados a observar os padrões apresentados, muitas vezes pelo contato com as linhas de concordância, e a inferir o funcionamento e o significado de determinados agrupamentos, pois, como afirma Frankenberg-Garcia *et al.* (2019), “quando você percebe, você aprende”.

Para esse tipo de tarefa, o contexto se torna um fator muito importante, já que, sem ele, não seria possível inferir padrões e sentidos das frases/palavras. E, segundo Coxhead (2013, p. 118), “estudos baseados em corpus permitem investigações em larga escala de palavras em contexto”³².

Voltando-se para a abordagem IFE, Hirvela afirma que “o tratamento de vocabulário nos cursos de IFE pode, de certa forma, ser mais desafiador do que nos cursos de inglês geral”³³ (HIRVELA, 2013, p. 84). De acordo com o autor, isso se dá especialmente devido ao vocabulário especializado que costuma ser abordado nesse contexto. Assim, se o curso é voltado para alunos de Direito, por exemplo, é claro que há a tendência de se focalizar o léxico

³⁰ *Vocabulary occupies an important place in both native language and second/ foreign language reading scholarship, since the words on the page are the starting point for reading* (HIRVELA, 2013, p. 84)

³¹ *Teach the contexts and structural relations within which the words have meaning* (MARSHALL, GILMOUR, 1993, p. 75).

³² *Corpus - based studies allow for larger - scale investigations of words in context* (COXHEAD, 2013, p. 118).

³³ *The treatment of vocabulary in ESP courses may in some ways be more challenging than in general purpose English courses* (HIRVELA, 2013, p. 84)

abordado em textos dessa especialidade. Pode-se dizer que, de uma área para outra, é o vocabulário que as difere, pois a estrutura da língua continua a mesma.

Nesse contexto, Hirvela (2013, p. 84) assegura que “o foco muito mais comum que surgiu desde o início dos anos 80, pelo menos no que diz respeito ao vocabulário, tem sido recursos e insights colhidos no campo da LC”³⁴.

O *corpus* pode ser incorporado no ensino de vocabulário, pois, por meio dele, é possível extrair informação de textos autênticos da língua, e, além disso, essas informações (palavras-chave, agrupamentos lexicais, linhas de concordância) são estatisticamente analisadas pelo concordanciador, disponibilizando dados que, na maioria das vezes, não são notados a olhar nu. Em outras palavras, quando o(a) professor(a)/pesquisador(a) observa um texto em busca de palavras-chave, por exemplo, ele(a) vai selecioná-las por intuição, pois não há uma certeza, estatisticamente falando, de que aquelas palavras são chaves no texto em questão.

Por outro lado, quando o(a) profissional utiliza *corpus* em sua pesquisa/material, o concordanciador faz esse trabalho estatístico por ele, inclusive com informações adicionais, como os textos em que essas informações são mais encontradas, quais textos não têm certos agrupamentos, qual a porcentagem dessa ocorrência, entre outras informações que podem auxiliar e aprimorar a pesquisa/trabalho do(a) profissional.

Como afirma Frankenberg-Garcia *et al.* (2019), “usando *corpora*, você não está perguntando para uma pessoa, mas para várias pessoas juntas”. Além disso, Coxhead (2013, p. 118) explica que “estudos com *corpus* têm contribuído muito para nossa questão de identificar e entender mais acerca de vocabulário especializado”³⁵.

Todavia, é de escolha do(a) professor(a)/pesquisador(a), como abordar esse vocabulário, ou seja, dentre tantas opções – listas de palavras, linhas de concordância, palavras-chave, agrupamentos lexicais – faz-se necessário que o(a) profissional proponha a melhor alternativa para sua pesquisa/material. Para o presente estudo, foram abordadas as palavras-chave e os agrupamentos lexicais de cada *corpus*, assim, explicações mais detalhadas dessa escolha serão apresentadas nos próximos subcapítulos.

2.3.2.1 A importância das palavras-chave

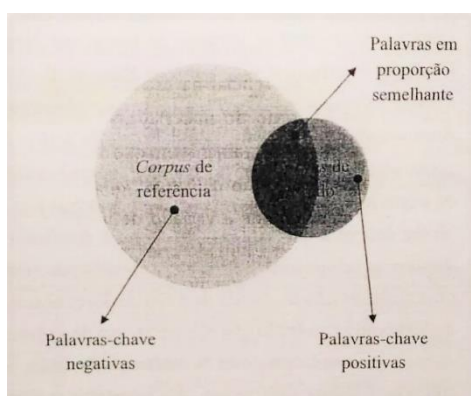
³⁴ The much more common focus that has emerged since the early 1980s, at least with respect to vocabulary, has been resources and insights gleaned from the field of corpus linguistics (HIRVELA, 2013, p. 84).

³⁵ Corpus studies have contributed a great deal to our quest to identify and understand more about specialized vocabulary (COXHEAD, 2013, p. 118).

Como mencionado, a presente pesquisa objetiva extrair as palavras-chave (*Keywords*) de um *corpus* específico, ou seja, daquele compilado por meio de textos jornalísticos em comparação a um *corpus* de referência de inglês geral, o *English TenTen*³⁶, que tem de ser necessariamente maior do que o *corpus* compilado pelo pesquisador. Berber Sardinha (2009, p. 193) afirma que as palavras-chave são “aquelas cujas frequências são diferentes, de modo estaticamente significativo, das frequências encontradas em um *corpus* de referência”. O autor também explica que essas palavras-chave não têm de ser necessariamente uma palavra individual, como por exemplo “café”, mas também, o que chamamos de *clusters*, ou seja, agrupamentos lexicais, como “café com leite”.

As palavras-chave podem ser classificadas entre positivas e negativas. As primeiras se referem àquelas cuja frequência é mais alta no *corpus* compilado pelo pesquisador do que no *corpus* de referência; elas representarão a língua de maneira geral. As negativas são exatamente o contrário: há uma frequência de palavras-chave muito mais relevante no *corpus* de referência em relação ao de estudo. Essa relação pode ser visualizada mais claramente por meio da figura abaixo, que mostra o procedimento de geração de palavras-chave, o qual inclui aquelas que aparecem em ambos os *corpora*, ou seja, são proporcionalmente semelhantes, assim como também inclui as palavras-chave positivas e as negativas, sendo que, em concordanciadores em geral, apenas as duas últimas são enaltecidas na tela, visto que o objetivo é exatamente explicitar a diferença entre os *corpora* em questão (VIANA, 2011, p. 61).

Imagem 1 - Procedimento de palavras-chave



Fonte: Viana (2011, p. 61)

Embora muito pareça que as palavras-chave ocorram unicamente em um ou outro *corpus*, Viana (2011) esclarece que é possível encontrá-las tanto na coletânea do pesquisador

³⁶ Mais informações serão dadas a respeito desse *corpus* no subcapítulo 2.1, p.69-70.

quanto na de inglês geral, sendo que o ponto principal dessa comparação é onde as palavras aparecem mais frequentemente em relação aos dois *corpora*.

Berber Sardinha (2009) afirma que não existem palavras-chave que sejam absolutas, isto é, que sejam chave independente do *corpus* de referência. O autor assegura que o *corpus* que o pesquisador escolhe para ser a referência tem influência direta na busca pelas palavras-chave, sendo assim, utilizando *corpora* de referência distintos, as palavras-chave também mudarão. O autor enfatiza que o *corpus* compilado pelo analista não deve estar contido no de referência, visto que a comparação estaria sendo realizada entre o próprio *corpus* de estudo, deixando de apresentar palavras que, comparadas a outro *corpus* de referência, seriam chaves, todavia, deixam de ser quando autocomparadas.

Tomemos como exemplo a expressão *generative design* encontrada no *corpus* de estudo *Tech* da presente pesquisa. Há a frequência relativa de 160.245 no *corpus* de estudo, enquanto a mesma palavra tem a frequência relativa de 0.013 no *corpus* de referência. Nesse caso, a chavicidade relativa de *generative design* é de 159.33, visto que, por meio da comparação entre os *corpora*, ela se ressalta no *corpus* da pesquisa, tendo maior chance de ser significativa para o estudo e conseqüentemente para a elaboração das atividades.

Berber Sardinha (2009) chama a atenção para a escolha do *corpus* de referência que, como já mencionado, muito pode alterar na pesquisa em si. O autor afirma que um *corpus* de características genéricas semelhantes ao *corpus* de estudo tende a “filtrar”, eliminar os elementos genéricos (relativos a um mesmo gênero) em comum, resultando em uma lista de palavras-chave que não inclui esses elementos. Alguns traços linguísticos que podem vir a ser filtrados são, entre outros, marcadores discursivos privilegiados, escolhas lexicais típicas e formas verbais flexionadas em comum.

Todavia, como afirmado por Berber Sardinha, a palavra-chave não é apenas uma unidade; ela pode ser mais de uma, o que chamamos de agrupamento lexical, como já mencionado. Assim, esse será o tópico do próximo subcapítulo, que explica, especialmente, o uso desses agrupamentos em exercícios de CE em LI.

2.3.2.2 O uso de agrupamentos lexicais em atividades de CE

De acordo com os estudos voltados para a LC, é possível verificar que essa abordagem teórico-metodológica não considera a língua em suas possibilidades de ocorrência, como era constatado nos estudos gerativistas. Assim, não se é estudado, aqui, quais as possíveis combinações que aparecem na língua. Ao contrário, estuda-se, por meio da análise de *corpora*,

ocorrências geradas por meio da língua em seu uso real, sendo o foco da análise o funcionamento da língua e as prováveis combinações que o falante faz ao se utilizar tanto da língua escrita quanto da falada.

Berber Sardinha (2009) afirma que o caráter probabilístico é baseado nas associações não aleatórias entre as palavras, ou seja, elas combinam entre si com graus de probabilidade diversos. De acordo com Greaves e Warren, esses agrupamentos ocorrem quando

Se encontram associados a uma frequência suficiente para excluir a possibilidade de serem co-ocorrências casuais. Isso foi levado um passo adiante e, com base em abundantes evidências de *corpus*, a linguística de *corpus* concluiu que as palavras geralmente têm uma preferência pelo que elas combinam (GREAVES, WARREN, 2010, p. 212, tradução nossa³⁷).

Sinclair (1991) busca explicar o funcionamento da língua em uso por meio de dois princípios: o da livre escolha e o idiomático. O primeiro diz respeito à crença de que quando um usuário produz um texto, seja ele oral ou escrito, ele envolve escolhas complexas em sua produção, já que decide utilizar determinada palavra em vez de outra, assim, o falante está sempre à mercê de escolhas de palavras individuais para construir seu texto. Entretanto, ainda que essa escolha seja livre e individual, todo falante deve seguir as regras gramaticais postuladas pela língua em questão.

No segundo princípio, o usuário da língua a utiliza em blocos lexicais sem ter o esforço de escolha por palavras soltas como aparece no primeiro caso. Em outras palavras, o sujeito busca mentalmente as combinações mais frequentes entre as unidades das quais fará uso em sua produção e essas são repetições de padrões já reconhecidos na língua e pelo usuário.

Esses padrões podem ser divididos em áreas diferentes e se associarem a termos que são empregados na área da Fraseologia, como *colocação* no âmbito lexical, ou *coligação*, voltada para elementos gramaticais da análise de padrões. Sinclair (1991) acredita em uma linguagem que une colocação e coligação. As duas esferas são, assim, trabalhadas em conjunto, produzindo agrupamentos significativos, sem que uma seja superior à outra; elas são complementares.

Padrões linguísticos são também observados pela perspectiva semântica, onde encontra-se a *prosódia semântica* ou a *preferência semântica*, em que se trabalha o valor afetivo das colocações ou o significado das mesmas, respectivamente. Desse modo, a LC pode atuar em

³⁷ *They are found to be associated with sufficient frequency to exclude the possibility that they are chance co-occurrences. This has been taken one step further and, based on abundant corpus evidence, corpus linguistics have concluded that words often have a preference for what they combine with* (GREAVES, WARREN, 2010, p. 212).

diversos campos com intuitos diferentes e objetos que também divergem entre si, apesar de se complementarem muitas vezes.

No entanto, não tendo a finalidade de destrinchar e analisar as áreas acima, que, sem dúvida são fontes ricas de informação acerca da língua e de seu funcionamento, o presente projeto foca em apenas um tipo de colocação oferecida pela LC, denominada *cluster*, *chunk*, *lexical bundle* ou simplesmente *agrupamento lexical*. Assim, os agrupamentos lexicais são "sequências de ocorrência regular de duas ou mais palavras que parecem possuir significados ou funções unitárias"³⁸ (O'KEEFFE, MCCARTHY, CARTER, 2007, p. 46) ou ainda, "são palavras que se seguem mais frequentemente do que se espera acontecer por acaso, ajudando a moldar significados de texto e a contribuir para o nosso senso de distinção em um registro"³⁹ (HYLAND, 2008, p. 42).

Um exemplo de agrupamento lexical encontrado no *corpus* do tema "saúde" da presente pesquisa é *vegan diet*, que em português significa "dieta vegana" - um conjunto de duas palavras que coocorre frequentemente e que mostra características significativas acerca do *corpus* compilado. Essas características contribuem de maneira significativa para as atividades desenvolvidas, visto que os alunos tomam consciência das ocorrências mais frequentes de determinado tema do *corpus*, bem como uma perspectiva diferente acerca de como analisar/compreender um texto, não levando em consideração apenas as palavras soltas, mas também os agrupamentos que nele se encontram. Assim como afirma Berber Sardinha (2009, p. 12), os agrupamentos são "bastante salientes devido à sua rigidez e assim se tornam bons padrões para se ensinar em aula de inglês, pois são facilmente notados".

Os estudiosos O'Keeffe, McCarthy e Carter (2007) explicam que os agrupamentos são, muitas vezes, tão frequentes quanto palavras "soltas" quando não, mais encontrados no vocabulário em geral. A LC apresenta, portanto, o segundo princípio defendido por Sinclair (1991, p. 201), o idiomático, pois "agrupamentos lexicais estão prontos para o uso em qualquer momento e não precisam ser reunidos toda vez que são utilizados"⁴⁰. Os autores Greaves e Warren (2010) afirmam que Sinclair colocou o estudo de multi-palavras no centro da LC por meio da ênfase dada ao estudo do significado. Abaixo, pode-se verificar uma lista de agrupamentos lexicais de um dos *corpora* compilados nesta pesquisa, o *Tech*.

³⁸ *Regularly occurring strings of two or more words which seem to possess unitary meanings or functions* (O'KEEFFE, MCCARTHY, CARTER, 2007, p. 46).

³⁹ *Essentially, these are words which follow each other more frequently than expected by chance, helping to shape text meanings and contribute to our sense of distinctiveness in a register* (HYLAND, 2008, p. 42).

⁴⁰ *Chunks are ready for use at any moment and do not need re-assembling every time they are used* (SINCLAIR, 1991, p. 201).

Tabela 1 - Agrupamentos lexicais do *corpus Tech*

Agrupamento	<i>Corpus</i> de estudo	<i>Corpus</i> de referência
<i>generative design</i>	160.245	0.013
<i>artificial intelligence</i>	364.193	1.304
<i>fake news</i>	160.245	0.088
<i>5g service</i>	101.974	0
<i>facial recognition</i>	131.109	0.337
<i>coal plant</i>	116.542	0.223

Fonte: Dados da presente pesquisa.

No que tange ao contexto de ensino-aprendizagem de LE, esses agrupamentos lexicais, como os exibidos acima, apresentam grande utilidade, visto que, como asseguram O'Keeffe, McCarthy, Carter (2007), o contexto educacional era baseado em listas de palavras que derivavam de textos escritos e/ou da intuição daqueles que organizavam os materiais pedagógicos, portanto, não era possível ter confiança absoluta nesses materiais, pois não se sabia se, de fato, a língua era utilizada da forma como estava sendo apresentada aos alunos. Com o avanço das pesquisas e com o auxílio da LC, não somente textos escritos são utilizados, mas também a língua falada é reportada e pode ser ensinada em sala de aula.

Berber Sardinha (2009, p. 12-13) explica que há diversos motivos pelos quais os padrões devem ser ensinados quando se tratando de L2, já que “os padrões são um tipo de ‘unidade pré-fabricada’ da língua, que parece residir na memória em forma ‘bruta’, isto é, sem estar decomposto em unidades individuais”. Sinclair (1991) já defendia essa ideia quando afirmou que o acesso aos padrões da língua permite que o processamento linguístico seja mais rápido e esse fator sustenta uma produção e compreensão mais fluente da língua, já que não é necessário se atentar a palavras individuais do texto.

Corroborando os pensamentos de Sinclair (1991), Berber Sardinha (2009) afirma que:

Normalmente, os aprendizes de língua estrangeira tendem a ver a linguagem com um conjunto de palavras individuais sustentadas por regras gramaticais; desse modo, a compreensão e a produção são, em muitos casos, com base na junção de palavras individuais. (BERBER SARDINHA, 2009, p. 13).

Para o autor, trabalhar com os alunos utilizando *corpora* pode melhorar situações como a descrita acima, visto que os aprendizes estarão em contato com unidades “pré-fabricadas” e não soltas, o que dificultaria a compreensão da LE.

Não apenas para o nível básico, mas considerando que o presente trabalho se volta para um público específico, é de extrema relevância o levantamento de agrupamentos lexicais dos textos compilados para os alunos que prestam vestibulares, uma vez que, a partir deles, um tema será enfatizado e o aluno terá a possibilidade de compreender melhor o que é pedido. Ademais,

é importante citar que “um *corpus* também pode fornecer mais informações a respeito de uma forma principal e enfatiza padrões léxico-gramaticais e semânticos associados à forma”⁴¹ (O'KEEFFE, MCCARTHY, CARTER, 2007, p. 100).

Por meio da leitura, é possível constatar que os agrupamentos lexicais são muito válidos na compreensão de um texto, visto que fazem parte de seu vocabulário e, assim, podem auxiliar os alunos no entendimento tanto do tema geral de determinado texto quanto de sentenças mais específicas. Os padrões lexicais também podem remeter à parte gramatical, ou seja, à estrutura que o texto apresenta, o que, novamente, vai ajudar o aluno a compreender melhor a mensagem que está sendo transmitida, principalmente por estar inserida dentro de um contexto, como é proposto no material desenvolvido para o presente estudo.

O uso de agrupamentos lexicais permite que os aprendizes utilizem a língua de modo mais automático, o que facilita a compreensão da linguagem em seu modo oral ou escrito assim como na produção da língua. Os estudiosos Girard e Sionis (2004, p. 1) analisam, em sua pesquisa, o uso de *formulaic speech* (FS), isto é, discurso formulado, que leva em conta as unidades de multi-palavras (agrupamentos lexicais) na produção oral. Os autores levam em consideração a função psicolinguística do FS, pois ela “pode ser uma estratégia de pré-planejamento e uma maneira de o aluno poupar esforços no processamento e, assim, se concentrar no aprendizado do idioma”.

Em outras palavras, os autores afirmam que o uso de agrupamentos lexicais “libera” determinada carga de processamento cognitivo, tendo em vista focar em outros elementos que apresentam outras funções no processo de produção linguística, como a organização do discurso, por exemplo, e isso evidencia a menor quantidade de esforço que um falante terá no momento de produção oral.

Wood (2010), que também estuda a mesma área de pesquisa que Girard e Sionis (2004) voltando-se para a produção oral na L2, afirma que algumas pesquisas têm indicado que um dos pontos chave para a fluência está nas unidades do discurso formulado e cadeias de *multi-words* (palavras-múltiplas ou agrupamentos lexicais).

Segundo o autor, “esses agrupamentos incluem entre outras categorias, colocações de duas palavras como *good time*, ou *first step*, verbos compostos como *run into*, ou *come across*”⁴² (WOOD, 2010, p. 2). Percebe-se que a ideia é a mesma trazida nesta pesquisa: o

⁴¹ A corpus can also provide more information about a key form and underline lexico-grammatical and semantic patterns associated with the form (O'KEEFFE, MCCARTHY, CARTER, 2007, p. 100).

⁴² These multiword units include, among other categories, 2-word collocations such as *good time*, or *first step*, phrasal verbs such as *run into*, or *come across* (WOOD, 2010, p. 2).

trabalho com agrupamentos lexicais (e não palavras-soltas apenas) voltado para o desenvolvimento de uma habilidade linguística. No caso dos autores citados, há uma preocupação com a produção oral e como esses conjuntos de palavras podem facilitar o desenvolvimento da fala na língua estudada.

No que diz respeito a esta pesquisa, o objetivo se volta para a CE, porém o trabalho com agrupamentos de palavras se torna praticamente o mesmo. A questão principal é mostrar como um bloco de palavras pode ser tão benéfico (ou até mais) do que trabalhar com palavras soltas e/ou descontextualizadas. Assim, Wood (2010) defende que

Dado que as sequências formuladas são unidades com várias palavras que parecem ser tratadas cognitivamente como palavras únicas, o conhecimento de sequências de fórmula pode ser automatizado como as palavras únicas são. Isso pode permitir que a expressão ocorra de forma fluente sob as restrições tempo que o discurso da vida real implica (WOOD, 2010, p. 2, tradução nossa⁴³).

Percebe-se, desse modo, que o mesmo há de ocorrer com a leitura textual quando o aluno se der conta de que certas palavras são combinações tão frequentes que possivelmente serão vistas mais de uma ou duas vezes por eles em textos de determinado assunto. Assim, as atividades buscam mostrar aos alunos determinados agrupamentos dos tópicos *mundo*, *saúde*, *tecnologia* e *política*, a fim de que eles não apenas aprendam esses agrupamentos, mas passem a olhar a língua por uma perspectiva diferente, em que combinações facilitam e agilizam o entendimento do texto, seja ele oral ou escrito.

Quando alunos têm acesso à agrupamentos lexicais que se mostram formas prontas do uso da língua, o esforço que fariam para identificar as palavras ou temas dos textos deve recair sobre outros elementos textuais que os levarão a uma compreensão mais efetiva.

2.4 Desenvolvimento dos *corpora*

Uma das questões essenciais quando se trabalha com *corpora* é por onde se deve começar e como se dará a construção do *corpus* de estudo. De acordo com Reppen (2010, p. 31), “ter uma pergunta claramente articulada é o primeiro passo essencial na construção de um

⁴³ Given that formulaic sequences are multiword units which appear to be dealt with cognitively as single words, knowledge of formulaic sequences may be automatized as single words are. This could allow expression to occur fluently under the constraints of time which real-life speech entails (WOOD, 2010, p. 2).

corpus, já que isso guiará no design do *corpus*”⁴⁴. Corroborando a afirmação da autora, Nelson (2010, p. 53) afirma que o pesquisador precisa se perguntar “o que eu quero alcançar criando o *corpus* e para que ele será usado?”⁴⁵

Dessa forma, a presente pesquisa e, concomitantemente, a construção dos *corpora*, basearam-se em duas perguntas: **1.** Em que medida o uso de *corpora* criados a partir de textos jornalísticos autênticos pode auxiliar no levantamento de agrupamentos lexicais na preparação de alunos para a prova de LI do vestibular?

2. Em que medida as atividades baseadas em agrupamentos lexicais podem desenvolver o léxico dos alunos e serem relevantes para o contexto dos vestibulandos?

Assim, por mais que os 4 *corpora* desta pesquisa tenham temáticas diferentes, foram construídos de igual maneira, já que tiveram a mesma utilidade: o desenvolvimento de atividades de compreensão escrita para o vestibular. Em todos os *corpora* foram feitas análises para descobrir quais agrupamentos lexicais encontrados seriam relevantes para a elaboração das atividades.

Depois de o pesquisador esclarecer seu objetivo de pesquisa, outros pontos ainda são muito relevantes a serem discutidos, e um deles diz respeito ao tamanho do *corpus*. Por esse motivo, o próximo subcapítulo tratará desse assunto, especialmente no que concerne aos *corpora* pequenos (*small corpora*).

2.4.1 Tamanho dos *corpora* (*small corpora*)

Muito já se foi discutido a respeito da representatividade de um *corpus* (BIBER, 1993; TRIBBLE, 1997; SINCLAIR, 2002; BERBER SARDINHA, 2004). Esse tópico está estritamente ligado ao tamanho do *corpus*, que objetiva, muitas vezes, representar dada língua ou parte dela. A orientação dada por Sinclair (1991, p. 18) é de que “um *corpus* deveria ser o maior possível e continuar crescendo”⁴⁶. Isso se dá, pois, como explica Berber Sardinha (2004, p. 23), “algumas palavras têm frequência de ocorrência muito rara e, para que haja probabilidade de ocorrerem no *corpus*, é necessário incorporar uma quantidade grande de palavras”, assim as palavras de menor frequência têm maior chance de estarem representadas no *corpus*.

⁴⁴ Having a clearly articulated question is an essential first step in corpus construction, since this will guide the design of the corpus (REPPEN, 2010 p. 31).

⁴⁵ What do you want to achieve by creating the corpus and what will it be used for? (NELSON, 2010 p. 53).

⁴⁶ A corpus should be as large as possible and keep on growing (SINCLAIR, 1991, p. 18)

Ainda segundo Berber Sardinha (2004, p. 24), a representatividade de um *corpus* leva o pesquisador a duas questões fundamentais: “representativo do que e para quem?”. Cada *corpus* deve ser representativo de uma parte específica da língua (quando não for de referência), e o pesquisador deve estar atento aos textos que são compilados para que seu estudo atinja o objetivo final da pesquisa. Assim, se o objetivo de determinada pesquisa é analisar o dialeto “caipira” do interior de São Paulo, não faz sentido o pesquisador coletar dados em outro Estado ou na capital.

Voltando ao presente estudo, para as perguntas feitas acima, pode-se afirmar que os *corpora* desta pesquisa representam parte da linguagem utilizada em artigos jornalísticos dos sites BBC® e CNN®, mais especificamente acerca dos temas saúde, tecnologia, política e mundo. E, no que concerne à segunda pergunta, os *corpora* são voltados para o público estudantil, especialmente os segundo e terceiro anos do ensino médio, já que a presente pesquisa se volta para atividades de vestibular.

Entretanto *corpora* como os desenvolvidos nesta pesquisa não têm de ser necessariamente muito grandes, pois o estudo é voltado para um público bem específico e não se objetiva representar a LI de modo geral. Assim, muitos autores defendem a utilidade e a importância do uso de *corpora* pequenos em pesquisas (TRIBBLE, 1997; GAVIOLI, 2005; KOESTER, 2010; REPPEN, 2010; NELSON, 2010). Segundo Reppen (2010, p. 32):

Para a maioria das questões que são pesquisadas por *corpus*, a questão do tamanho é resolvida por dois fatores: representatividade (eu coletei textos suficientes (palavras) para representar com precisão o tipo de linguagem sob investigação?) E praticabilidade (restrições de tempo). (REPPEN, 2010, p.32, tradução nossa⁴⁷).

A autora também afirma que o tamanho do *corpus* depende da investigação de determinado pesquisador, ou seja, ele deve se perguntar se os textos compilados serão representativos para responder sua pergunta de pesquisa, como já foi esclarecido acima.

De acordo com Koester (2010) há várias vantagens em se trabalhar com um *corpus* pequeno, como por exemplo:

Todas as ocorrências, e não apenas uma amostra aleatória de itens de alta frequência podem ser examinadas. [...] eles (*corpora* pequenos) permitem um elo muito mais estreito entre o *corpus* e os contextos em que os textos do

⁴⁷ For most questions that are pursued by corpus researches, the question of size is resolved by two factors: representativeness (have I collected enough texts (words) to accurately represent the type of language under investigation?) and practicality (time constraints) (REPPEN, 2010, p.32).

corpus foram produzidos. [...] Com um corpus pequeno, o compilador geralmente também é o analista e, portanto, geralmente tem um alto grau de familiaridade com o contexto (KOESTER, 2010 p. 66-67, tradução nossa⁴⁸).

Desse modo, pode-se notar que há uma divergência entre o *corpus* que tem a necessidade de ser grande no que tange a quantidade de palavras para que seja representativo de determinada língua e o *corpus* de estudo que caracteriza determinada pesquisa, envolvendo não apenas a língua em si, mas também o objetivo do pesquisador. Sinclair (2001, p. 11) esclarece que “agora os *corpora* 'pequenos' e 'grandes' são vistos em contraste um com o outro. Então a diferença deve ser metodológica, porque não pode ser apenas tamanho”⁴⁹. (SINCLAIR, 2001 p. 11).

Uma diferença metodológica pode ser, por exemplo, como se dá a compilação desse *corpus*. Assim como será explicado no próximo subcapítulo, o pesquisador que escolhe trabalhar com um *corpus* reciclável tem a opção de ir “abastecendo” esse *corpus*, ou seja, adicionando mais textos que forem úteis para sua pesquisa, bem como retirando textos desnecessários ou inúteis para a continuação do estudo.

É possível afirmar, desse modo, que essa pode ser considerada uma diferença no método de construção de *corpora*. O *corpus* de referência de determinada língua, por exemplo, precisa ter o máximo de textos possíveis dessa língua para que possa ser representativo e abranger todos os seus dialetos. Já no especializado, ou de estudo, como é o desta pesquisa, leva-se em conta o objetivo do estudo, como já mencionado.

Comparando esses dois tipos de *corpora*, Tribble (1997, p. 112) afirma que “se quisermos investigar o léxico de um domínio específico (ex: saúde), um micro *corpus* especializado pode ser mais eficiente que um maior, de língua geral”⁵⁰. Gavioli (2005, p. 53) corrobora a afirmação do autor ao assegurar que é “necessário um conjunto de textos muito menor e especializado para provar, por exemplo, trabalhos de pesquisa médica”⁵¹. Assim, caso um pesquisador esteja em busca, por exemplo, das colocações mais comuns na área da enfermagem, ele dificilmente encontrará informações precisas no *corpus* geral da língua, mas,

⁴⁸ All occurrences, and not just a random sample, of high frequency items can be examined. [...] they (small corpora) allow a much closer link between the corpus and the contexts in which the texts in the corpus were produced. [...] With a small corpus, the corpus compiler is often also the analyst, and therefore usually has a high degree of familiarity with the context (KOESTER, 2010 p. 66-67).

⁴⁹ Now ‘small’ and ‘large’ corpora are seen as in contrast with each other. So the difference must be methodological, because it cannot be just size (SINCLAIR, 2001, p.11).

⁵⁰ if one wishes to investigate the lexis of a particular content domain (e.g. health) a specialist micro-corpus can be more often be useful than a much larger general corpus (TRIBBLE, 1997, p. 112).

⁵¹ a much smaller and specialized set of texts was needed to sample, for instance, medical research papers (GAVIOLI, 2005, p.53).

se, por outro lado, esse pesquisador compilar seu próprio *corpus*, contendo artigos científicos da área desejada, sua pesquisa será mais eficiente e obterá sucesso com mais facilidade e precisão.

Além disso, Tribble (1997) esclarece que *corpora* pequenos podem ser “um recurso útil no projeto de ensino-aprendizagem de línguas”⁵² (TRIBBLE, 1997, p. 111). Compartilhando da mesma ideia, Nelson (2010, p. 55) afirma que “a necessidade de *corpora* menores - olhando para áreas específicas da língua - tem sido reconhecida, especialmente em relação ao ensino e uso na sala de aula”⁵³.

É dessa maneira que o presente estudo une as funções e a utilidade de *corpora* pequenos por meio do desenvolvimento de *corpora* recicláveis para a elaboração de atividades no contexto educacional, trabalhando com os quatro temas: saúde, tecnologia, mundo e política, já que não se fazia necessário e nem era o objetivo desta pesquisa compilar um grande número de artigos que não estivessem representando as notícias atuais na LI.

Em outras palavras, a compilação de notícias que foram publicadas anteriormente aos 3 últimos meses de 2018 seria incoerente, pois fugiria aos objetivos da pesquisa, uma vez que os dados encontrados nos *corpora* foram utilizados para o desenvolvimento das atividades que foram aplicadas logo no primeiro bimestre de 2019.

O próximo subcapítulo discorrerá acerca das características de um *corpus* reciclável e como ele, mesmo sendo pequeno, pode ser útil na presente pesquisa.

2.4.2. *Corpus* reciclável: características e utilidade na pesquisa

O *corpus* reciclável (ou DIY - *do-it-yourself*) consiste em um *corpus* criado pelo próprio pesquisador e de acordo com seus interesses e necessidades (BOWKER, 2002; ZANETTIN, 2002; TAGNIN, 2007). Zanettin (2002) elenca algumas características de um *corpus* dessa natureza, entre elas, duas são muito relevantes para este estudo: é uma coleção de documentos (textos) de páginas da Internet e é um *corpus* aberto, ou seja, mais materiais podem ser adicionados após a compilação, além de algumas vantagens para a área da Tradução, mas que não serão abordadas nesse trabalho.

Charles (2012) corrobora essas ideias e afirma que a utilização de um *corpus* reciclável pode ser de grande valia levando em consideração a multidisciplinariedade e

⁵² *a useful resource in the language learning / teaching project* (TRIBBLE, 1997, p. 106).

⁵³ *The need for smaller corpora – looking at specific areas of the language – has been recognized, especially in relation to teaching and use in the language classroom* (NELSON, 2010, p. 55).

reutilização/modificação do *corpus* compilado, além de outras diversas razões para se utilizar um *corpus* desse caráter. A autora, em seu artigo *Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building* (2012), menciona uma pesquisa feita por ela com alunos universitários de 12 diferentes áreas. Cada um tinha uma necessidade divergente relacionada a seu próprio campo de pesquisa, sendo impossível, segundo a estudiosa, que um único *corpus* correspondesse às necessidades de todos eles.

Dessa maneira, ensinou-se aos alunos a como compilar o seu próprio *corpus*, a fim de alcançarem seus objetivos voltando-se para suas pesquisas. Charles (2012) afirma que um importante motivo para se promover a construção de *corpus* reciclável é que este pode aumentar a autonomia do aluno de diferentes maneiras. Primeiramente, é o aluno quem decide quais textos compilar e quais não. Assim, não apenas eles têm o conhecimento específico do seu conteúdo, mas também têm o controle sob os recursos que eles utilizam para aprender. E os alunos podem mudar seu *corpus* a qualquer momento, deletando ou adicionando arquivos quando necessário (CHARLES, 2012, p. 94).

Uma pesquisa semelhante é realizada por Pinto (2018) ao trabalhar com alunos de três áreas diferentes, humanas, exatas e biológicas. O estudo se dá por meio da compilação de *corpora* recicláveis de cada ciência citada, em que os próprios alunos compilam resumos científicos e/ou resumos de congressos internacionais de sua área a fim de analisar as características encontradas nesse tipo de escrita, como o léxico utilizado para introduzir partes específicas de um resumo científico, a terminologia de cada área e a estrutura retórica utilizada nos textos compilados. Segundo a autora, esses tipos de características encontradas por meio do *corpus* auxiliam na melhora da escrita acadêmica.

Trazendo o projeto descrito acima para a realidade da presente pesquisa, pode-se notar que a utilização de um *corpus* reciclável é de grande utilidade, devido a seu caráter variante e autônomo, além de não ser necessário o uso da internet para utilizar o material compilado; uma vez construído, estará sempre disponível para o pesquisador (CHARLES, 2012).

Os *corpora* compilados neste trabalho são de textos atuais e de diferentes assuntos, portanto, é imprescindível que eles sejam sempre atualizados, visto que o material utilizado para as atividades em sala de aula em determinado ano não será tão útil para o ano seguinte e isso faz com que a aula perca a credibilidade, considerando que o material é voltado para vestibulandos, ou seja, um material atual se faz mais do que necessário.

Nesse sentido, o *corpus* reciclável pode ser muito válido para pesquisadores de diversas áreas, inclusive para professores em geral, sejam eles atuantes em cursinho, escola de idiomas ou, até mesmo, do ensino regular, uma vez que, com esse tipo de *corpus*, o pesquisador é quem

avalia o que se faz necessário ensinar naquele determinado momento, levando em consideração seu contexto profissional.

Visto que esta pesquisa se volta ao contexto educacional, principalmente no que diz respeito à elaboração de materiais didáticos, a seguir entenderemos mais acerca da característica *hands-off* de uma atividade baseada em *corpus*.

2.4.3 O ensino por meio de dados e o foco em atividades do tipo *hands-off*

A presente pesquisa, como visto até o momento, introduz o uso de *corpora* no contexto de ensino-aprendizagem de línguas, mais especificamente a LI. Para tanto, atividades do tipo “*hands-off*” foram desenvolvidas e inseridas em sala de aula a fim de que os alunos tivessem contato com o contexto autêntico da LI por meio de *corpora*, mesmo que sem notar, explicitamente, a presença dessa abordagem em seus exercícios. Dessa forma, faz-se necessário esclarecer do que se tratam atividades “*hands-off*” e como elas são utilizadas em sala de aula.

A abordagem DDL – *data-driven learning* – ou seja, aprendizagem por meio de dados, foi inicialmente introduzida pelo estudioso Johns (1988) que explica que, nesse contexto de aprender por meio de dados,

A tarefa do aluno é ‘descobrir’ a língua estrangeira, e que a tarefa do professor de línguas é fornecer um contexto em que o aluno possa desenvolver estratégias de descoberta - estratégias por meio das quais ele ou ela possa ‘aprender a aprender’ (JOHNS, 1988, p. 1, tradução nossa⁵⁴).

O autor centraliza o aluno no contexto de ensino-aprendizagem de uma língua, ou seja, não mais passivo, mas ativo; o aluno passa a ser o próprio descobridor de como essa língua funciona, e o professor passa a ser mediador desse processo de aprendizagem, aquele que ensina o aluno a aprender, como afirma Johns (1988). Desse modo, o aluno passa a ser um “pesquisador” da língua “cuja aprendizagem precisa ser dirigida por dados linguísticos”, de onde vem o nome da abordagem (JOHNS, 1988, p. 2).

Pensando ainda no aluno como integrante ativo de sua aprendizagem, o intuito da abordagem DDL é fazer com que o aprendiz desenvolva certa habilidade para enxergar padrões por meio dos dados apresentados a ele, assim supõe-se que o aluno faça generalizações

⁵⁴ *The task of the learner is to ‘discover’ the foreign language, and that the task of the language teacher is to provide a context in which the learner can develop strategies for discovery - strategies through which he or she can ‘learn how to learn’ (JOHNS, 1988, p. 1).*

(JOHNS, 1988, p. 2). Por exemplo, é mostrado ao aluno linhas de concordância⁵⁵ em que se destaca o padrão em que o professor quer chamar sua atenção, como mostra a figura abaixo para o agrupamento *life expectancy*. Por meio dessas linhas, os alunos observam, por exemplo, em que contexto esse agrupamento aparece.

Imagem 2: Exemplo de linhas de concordâncias

Details	Left context	KWIC	Right context
1 doc#26 • Childh...	urnal of Medicine that predicted a significant shortening of life	expectancy	in the US by mid-century due to the effect of obesity on longevi
2 doc#26 • Childh...	ears sooner than expected. </s><s> In 2015 and in 2016, life	expectancy	decreased for the first time since the Civil War in the United St
3 doc#33 • Life e...	l promoting positive health outcomes," she said. </s> <s> Life	expectancy	drops in the US and the UK, rises in Australia, a new study find
4 doc#33 • Life e...	JK and the United States -- most countries saw declines in life	expectancy	between 2014 and 2015, according to the study, published We
5 doc#33 • Life e...	the only countries in the study that showed an increase in life	expectancy	across all years for both men and women. </s><s> The drop in
6 doc#33 • Life e...	ote. </s><s> Outside the United States, declines in overall life	expectancy	were focused among people 65 and older, with the rise in deatl
7 doc#33 • Life e...	rvious system disorders contributed in part to driving down life	expectancy	among those 65 and older. </s><s> In countries where environ
8 doc#33 • Life e...	young adults in the US In the United States, the decline in life	expectancy	was concentrated among a younger cohort of the population: tl
9 doc#33 • Life e...	en the country's already lackluster performance regarding life	expectancy	among high-income countries, according to the study. </s><s>
10 doc#33 • Life e...	ording to the study. </s><s> "The USA now has the lowest life	expectancy	levels among high income developed countries, and American
11 doc#33 • Life e...	ia, Japan, Denmark and Norway -- saw a continued rise in life	expectancy	across the study years. </s><s> The average life expectancy fi
12 doc#33 • Life e...	expectancy across the study years. </s><s> The average life	expectancy	for women in Australia, Japan, Denmark and Norway in 2016 v
13 doc#33 • Life e...	7.17, 82.79 and 84.17, respectively. </s><s> For men, the life	expectancy	was 81.49, 81.01, 78.95 and 80.61, respectively. </s><s> The
14 doc#33 • Life e...	r why these four countries were able to retain increases in life	expectancy	over time. </s><s> Continued decline the US and UK Most cou
15 doc#33 • Life e...	nen rebounded between 2015 and 2016, showing gains in life	expectancy	to make up for the loss. </s><s> But the US and the UK were €
16 doc#33 • Life e...	re exceptions to this trend, continuing to show declines in life	expectancy	in more recent years. </s><s> Social inequality, poverty and th

Fonte: *Corpus Health* (presente pesquisa)

Faz-se relevante esclarecer que nem todos os exercícios propostos nesta pesquisa têm linhas de concordância, ou seja, fez-se o uso de *corpora* no desenvolvimento dos exercícios, entretanto foram utilizadas diferentes ferramentas para isso, como será visto posteriormente. Também é muito importante insistir na ideia de que “o uso de *corpus* não propõe substituir as metodologias de ensino existentes, mas a enriquecê-las e aprimorá-las” (GABRIELATOS, 2005, p. 15).

De qualquer modo, independente da ferramenta utilizada nas atividades – linhas de concordância, palavras-chave, agrupamentos lexicais, entre outros – um ponto de extrema importância em atividades que envolvem *corpora* é a autonomia desenvolvida no aluno. Em outras palavras, o estudante deverá ser sempre preparado para aprender a buscar informações acerca da língua por meio de dados autênticos que lhe permitam compreender o funcionamento

⁵⁵ “A concordância é uma listagem das ocorrências de um item específico, dispostas de tal modo que a palavras de busca (aquela que se tem interesse em investigar) aparece centralizada na página (ou tela do computador). A palavra de busca é acompanhada do seu contexto original, isto é, das palavras que ocorreram junto com ela no corpus” (BERBER SARDINHA, 2004, p.188)

do idioma, fator muito importante no contexto dos vestibulandos, já que todos os textos que são apresentados nos vestibulares são retirados de seu contexto real de uso.

Para isso, também há mudanças no que diz respeito ao papel do professor em sala de aula. Segundo Chambers (2010), é função do(a) professor(a) auxiliar os alunos a interpretar os dados da língua e analisar os resultados encontrados por eles. Todavia, essa mediação exige do professor “novas habilidades [...], incluindo conhecimento acerca de quais *corpora* estão disponíveis, consulta ao *corpus* a habilidades de análises, e a capacidade para decidir como melhor apresentar os dados aos alunos”⁵⁶ (CHAMBERS, 2010, p. 355).

É desse modo que a autora explica que, além da aquisição de novas habilidades, o(a) professor(a) também é levado a encarar a língua e o seu ensino de modo descritivo e não prescritivo, já que, no contexto de LC, a língua é analisada em seu ambiente autêntico; descreve-se o que é encontrado nos *corpora* como o funcionamento real da língua em uso. Assim, o funcionamento da língua não é prescrito, mas descrito por meio da observação de dados e isso requer um novo olhar do professor em relação à própria língua bem como ao seu ensino.

Tendo em vista todo esse processo de mudança tanto no papel do professor em sala de aula quanto do aluno, faz-se necessário discorrer a respeito das atividades que foram desenvolvidas por meio dos dados encontrados nos *corpora* compilados.

Primeiramente, quando se fala em DDL, dois tipos de atividades podem ser propostas: *hands-on* e *hands-off*. A principal diferença entre as duas está no meio em que o(a) professor(a) utiliza para passar os exercícios, ou seja, via computador – *hands-on* – ou via papel impresso – *hands-off*. O segundo tipo de atividade foi o utilizado nesta pesquisa em razão do restrito acesso aos computadores da escola, visto que outros professores poderiam utilizar no mesmo horário para atividades de outras disciplinas, além de que, para a utilização de atividades *hands-on*, os alunos precisariam de, no mínimo, um treinamento acerca de como utilizar as ferramentas que um *corpus* pode proporcionar, o que demandaria mais tempo, impossibilitando a conclusão das aulas.

Além disso, Boulton (2012) discorre acerca de alguns pontos positivos em se trabalhar com atividades *hands-off*, utilizadas por ele como atividades *paper-based*:

O uso de materiais preparados permite que o professor decida os pontos a serem abordados, conduza as pesquisas com antecedência, selecione e classifique pequenas quantidades de dados adequadamente e apresente-as em atividades controladas, adaptadas às necessidades e habilidades dos alunos. O meio de impressão pode usar tipos de atividades familiares (preenchimento de

⁵⁶ *New skills [...], including knowledge about what corpora are available, corpus consultation and analysis skills, and the capacity to decide how best to present the data to the learners* (CHAMBERS, 2010 p. 355).

lacunas, correspondência, etc.), apresentando assim um objetivo mais realista a curto prazo e pode, portanto, ser mais motivador do que o trabalho prático, bem como mais apropriado para alguns estilos de aprendizado. Os materiais *paper-based* podem ser mais eficientes, já que muito do ruído de fundo é eliminado antecipadamente, portanto, deve ser possível usá-los imediatamente com pouco ou nenhum treinamento prévio (BOULTON, 2012, p. 153, tradução nossa⁵⁷).

Um dos pontos fundamentais pela escolha de atividades *hands-off* também se deve ao fato de que a intenção era utilizar a LC de modo “camuflado” nos exercícios, ou seja, os alunos tiveram acesso a dados por meio de *corpora*, mas em nenhum momento esses dados pareciam fazer parte de algo novo e difícil de ser estudado. Em outras palavras,

Pode haver então um argumento contrário à crença comum, de apresentar a DDL não como uma nova técnica radical, mas como prática comum ao lado de outras atividades e materiais comuns [...]. A própria integração deve facilitar as coisas para alunos e professores, forçando-nos como pesquisadores a encontrar maneiras de reduzir alguns dos aspectos mais radicais, eliminar o excesso de bagagem em vez de continuamente adicionar novos recursos, bem como ajudar a identificar DDL como prática comum. Tudo isso precisa ser feito sem perder as vantagens da DDL, e sem que ela se torne completamente invisível ou se derreta no plano de fundo (BOULTON, 2009, p. 4, tradução nossa⁵⁸).

Desse modo, é importante mencionar aos alunos que os exercícios foram feitos com base em dados autênticos e que o computador foi mediador para essa prática, porém os exercícios em si não precisam utilizar algumas nomenclaturas que não fazem parte da realidade do aluno. Por exemplo, em vez de escrever “frases retiradas do *corpus* X”, pode-se optar apenas por “frases”, “sentenças”, entre outros.

Dessa maneira, o aluno enxerga o exercício como uma prática corriqueira, e a LC está tendo suas contribuições no ensino-aprendizagem de línguas ao mostrar, por exemplo, sua

⁵⁷ *Use of prepared materials allows the teacher to decide the points to cover, conduct the searches in advance, select and sort small amounts of data appropriately, and present them in controlled activities tailored to the learners' needs and abilities. The print medium can use familiar activity types (gap-fills, matching, etc.), thus presenting a more realistic goal in the short term, and may therefore be more motivating than hands-on work, as well as more appropriate for some learning styles. Paper-based materials can be more efficient as much of the background 'noise' is weeded out in advance, so it should be possible to use them immediately with little if any prior training* (BOULTON, 2012, p. 153).

⁵⁸ *There might then be an argument, contrary to common belief, for presenting DDL not as a radical new technique, but as ordinary practice alongside other ordinary activities and materials [...] The integration itself should make things easier for learners and teachers by forcing us as researchers to find ways to reduce some of the more radical aspects, to eliminate excess baggage rather than continually seeking to add new features and should also by association help to identify DDL with ordinary practice. All of this needs to be done without losing the advantages of DDL, and without it becoming completely invisible or melting into the background* (BOULTON, 2009, p. 4).

utilidade em atividades de línguas, aproximando o aluno de dados autênticos, sem deixá-lo desconfortável com o “novo”.

A presente pesquisa trabalha, principalmente, com o uso de *corpora* para a exploração de vocabulário, a fim de aprimorar a compreensão de textos escritos na LI. Nesse sentido, Gabrielatos (2005, p. 13) afirma que “as amostras de *corpus* também se prestam para trabalhar em habilidades de leitura e, em particular, para desenvolver estratégias para inferir o significado de léxico desconhecido no texto”⁵⁹.

Além disso, a autora também afirma que,

Um texto que conterà apenas algumas instâncias do item lexical, geralmente vai demonstrar seu significado e utilizá-lo em um contexto e poderá não fornecer pistas suficientes para inferir o seu significado. As amostras de *corpus*, por outro lado, contêm um grande número de exemplos que demonstram significado e uso em diversos contextos e oferecem inúmeras pistas (GABRIELATOS, 2005, p. 13, tradução nossa⁶⁰).

Esses tipos de atividades poderiam ter sido desenvolvidas e adaptadas como *hands-on*, entretanto, como Boulton (2012, p. 156) explica, as atividades *hands-off* “representam uma opção que pode ser aplicada de diferentes maneiras, de acordo com as condições locais”⁶¹. Isso significa que, ainda que o local onde as atividades serão aplicadas não permita acesso ao computador e/ou à internet, o(a) professor(a) tem outra opção: utilizar *corpus* em atividades impressas, o que, muitas vezes, pode até ser mais positivo, dependendo do contexto e do público que participará das aulas.

É importante mencionar que o objetivo da escolha de se desenvolver atividades *hands-off* ao invés de *hands-on* não está na desvalorização da segunda opção, muito pelo contrário: acredita-se que exercícios do tipo *hands-on* podem ser muito eficazes em diversos casos, inclusive na divulgação de atividades por meio de LC, independentemente de como são abordadas. Porém, como Boulton (2009) esclarece, o acesso a computador/internet pode ser uma barreira e

⁵⁹ *Corpus samples also lend themselves to work on reading skills, and, in particular, to developing strategies for inferring the meaning of unknown lexis in the text* (GABRIELATOS, 2005, p. 13).

⁶⁰ *A text will contain only a few instances of the lexical item, will usually demonstrate its meaning and use in one context, and may not provide sufficient clues for inferring meaning. Corpus samples, on the other hand, contain a large number of examples which demonstrate meaning and use in diverse contexts and offer a wealth of clues* (GABRIELATOS, 2005, p. 13).

⁶¹ *they represent one option which may be applied in different ways according to local conditions* (BOULTON, 2012, p. 165).

Talvez o principal argumento para introduzir o DDL por meio de materiais impressos é que ele elimina essa barreira e, portanto, tem o potencial de abri-la para um público mais amplo. Por sua vez, esperamos que este seja o cenário para o trabalho posterior em computadores, à medida que os aprendizes obtenham o conhecimento e as habilidades necessárias para um trabalho mais autônomo (BOULTON, 2009, p. 09, tradução nossa⁶²).

É de grande importância, pois, que, de qualquer modo, o aluno tenha acesso a dados autênticos da língua e que ele esteja inserido em um processo de autonomia em relação ao seu aprendizado. Para isso, o concordanciador utilizado nesta pesquisa – *Sketch Engine*® - oferece várias ferramentas que podem auxiliar o(a) professor(a) na elaboração de exercícios e, por essa razão, o próximo capítulo abordará um pouco mais acerca dessa plataforma, apresentando suas principais características e utilidade nas pesquisas linguísticas.

⁶² *Perhaps the main argument for introducing DDL via printed materials is that it cuts out this barrier and thus has the potential to open it up to a wider audience. This in turn will hopefully set the scene for later work on computers as learners gain the knowledge and skills necessary for more autonomous work* (BOULTON, 2009, p. 09).

3. MATERIAIS E MÉTODOS

Neste capítulo serão discutidos os materiais e procedimentos metodológicos que sustentaram a pesquisa por meio do desenvolvimento das atividades de CE voltadas para o contexto de um vestibulando, que tiveram como base os agrupamentos lexicais extraídos dos *corpora* de estudo. Destaca-se, nesta seção, detalhes acerca dos *softwares* utilizados para análise e elaboração das atividades, as características dos *corpora* da pesquisa, como estes serviram para a extração dos agrupamentos lexicais, os critérios utilizados no desenvolvimento das atividades, bem como a elaboração das mesmas.

3.1. *Sketch Engine* – características do concordanciador e ferramentas utilizadas

Para a compilação dos *corpora* desta pesquisa, bem como a sua análise e extração das informações necessárias para a elaboração das atividades, foi utilizada a plataforma do *Sketch Engine*® que conta com algoritmos que analisam diversos textos autênticos que são observados para diversos fins, como pesquisas na área da linguística, lexicografia, tradução, terminologia bem como historicistas, professores e estudantes, isso porque consegue “identificar instantaneamente o que é típico na linguagem e o que é raro, incomum ou emergente” (KILGARRIFF *et al.*, 2014). Segundo informações do próprio website, o *Sketch Engine* é composto por 500 *corpora* em mais de 90 idiomas, sendo que cada um deles possui até 30 milhões de palavras.

Como já mencionado, esta pesquisa faz uso de palavras-chave e agrupamentos lexicais, e para a extração dessas informações faz-se necessário que o(a) pesquisador(a) compare o seu *corpus* de estudo a um *corpus* de referência, ou seja, da língua geral, que neste caso é o inglês. Para que essa comparação seja realizada, alguns concordanciadores exigem que esse *corpus* de referência seja baixado e depois inserido no concordanciador para a comparação, pois o mesmo não possui em sua base um *corpus* de referência “fixo”, ou seja, que já acompanha o concordanciador.

Diferentemente do que foi explicado acima, o *Sketch Engine* já possui em sua plataforma vários *corpora* de referência, que, como explicado, possui milhões de palavras cada, sendo considerado representativo da língua. No caso do inglês, o *corpus* de referência utilizado foi o *English Web 2015 (EngTenTen)*, que possui mais de 15 milhões de palavras na LI a partir de textos coletados da internet. Abaixo pode-se visualizar os números referentes a este *corpus*,

contendo, de cima para baixo, os itens (número total de palavras), as formas (número de palavras não repetidas), posteriormente, a quantidade de sentenças, parágrafos e o número de documentos.

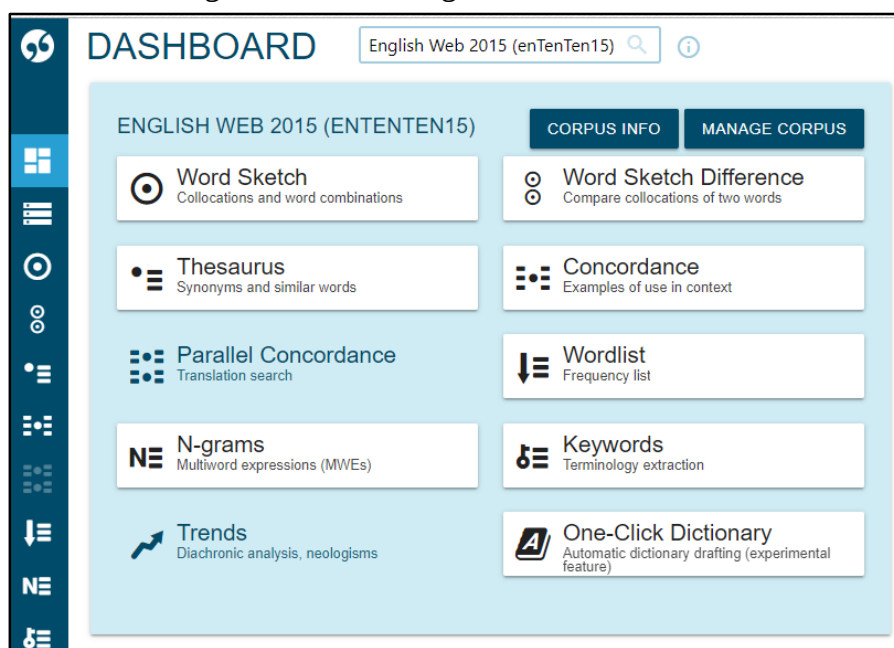
Tabela 2 - Números referentes ao *corpus* enTenTen

Unidade	Número
Itens	18,389,070,604
Formas	15,703,895,409
Sentenças	856,428,595
Parágrafos	351,102,999
Documentos	37,810,139

Fonte: Sketch Engine - enTenTen

Abaixo pode-se visualizar a imagem que faz referência ao website do *Sketch Engine*. Como é possível notar, há diversos itens que são relevantes para uma pesquisa linguística, uma vez que geram lista de palavras, palavras-chave, apresentam palavras sinônimas, linhas de concordância, entre outras ferramentas que podem auxiliar pesquisas voltadas para o estudo da linguagem. Entretanto, no contexto desta pesquisa, foram utilizadas as seguintes: *Keywords* e *Concordance*.

Imagem 3 - Sketch Engine – Dashboard



Fonte: Sketch Engine (KILGARRIFF, *et al*, 2014)

As *Keywords* se referem à lista de palavras-chave, que foram mencionadas anteriormente no subcapítulo 2.3.2.1, e têm grande importância no desenvolvimento dos exercícios. Já a ferramenta *Concordance* apresenta exemplos da língua em seu contexto de uso, por meio de linhas de concordância, que já foram introduzidas no subcapítulo 2.4.3. Assim, ao fazer uma busca por determinada unidade, o pesquisador encontra vários excertos que mostram como essa palavra tem sido utilizada no contexto real de uso, e pode utilizar essas linhas com o intuito de voltar a atenção do aluno para algum ponto gramatical e/ou lexical relevante para a aula.

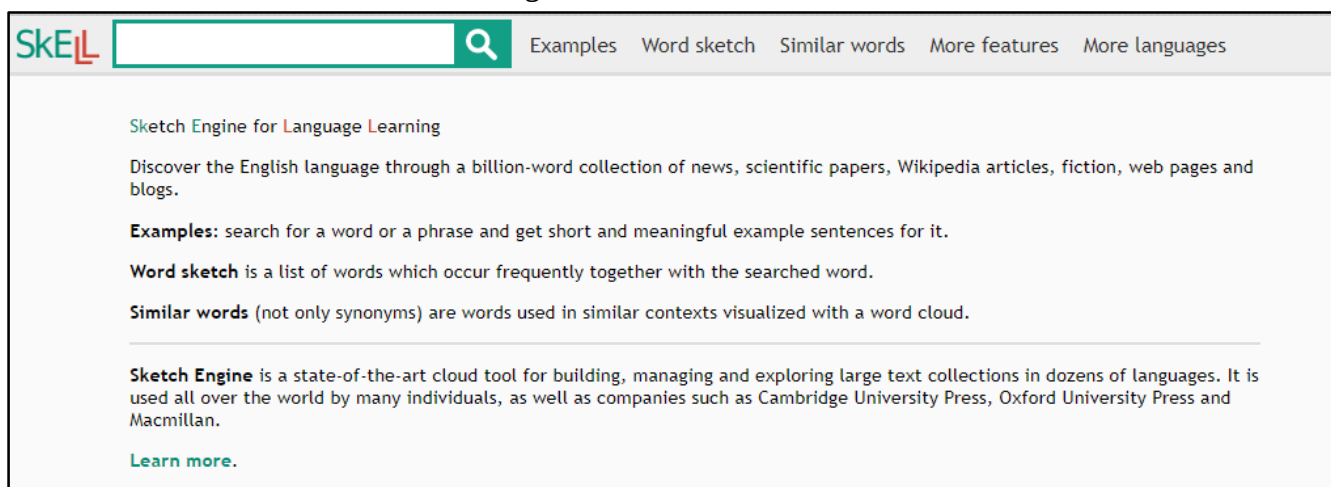
A plataforma do *Sketch Engine* sugere que professores e alunos utilizem o *SkELL* – *Sketch Engine for Language Learning* (KILGARRIFF, et al, 2014). Desse modo, na próxima seção essa plataforma será mais detalhadamente apresentada e, também será explicado como essa teve sua relevância na presente pesquisa.

3.1.1 SkELL – principais características e utilidade na pesquisa

O *SkELL* é uma vertente do *Sketch Engine* voltada para o ensino-aprendizagem de línguas devido a maior “simplicidade” dos exemplos e linhas de concordância oferecidos na tela. Dessa forma, é possível rebater as críticas feitas por se utilizar linguagem autêntica no ensino (MARTINEZ, 2002; KILICKAYA, 2004), em que o argumento estava na dificuldade do entendimento de determinado vocabulário e/ou estrutura presente em certos textos. Ou seja, ao se preparar atividades com linhas de concordância, por exemplo, era necessário que o aluno apresentasse um nível de inglês mais avançado para compreender o que estava escrito.

Assim, o *SkELL* possui várias abas que oferecem diferentes ferramentas para serem utilizadas nesse contexto entre professor e aluno. A imagem abaixo apresenta o website do *SkELL*.

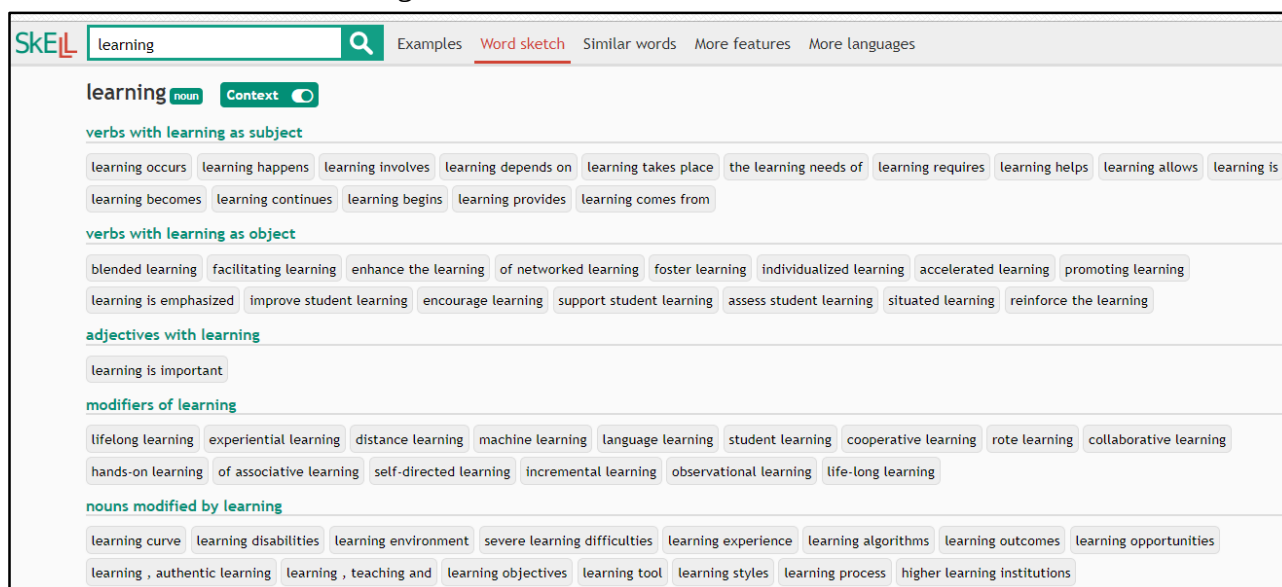
Imagem 4 - SkELL-website



Fonte: SkELL (KILGARRIFF, *et al*, 2014)

Não tendo a intenção de explorar detalhadamente todas as possibilidades oferecidas pela plataforma, serão explicitadas apenas aquelas que foram utilizadas na presente pesquisa – *Word Sketch*, *Examples* e *Similar Words*. Posteriormente, todas elas serão abordadas novamente na metodologia, com o intuito de mostrar como cada uma foi utilizada na prática e exemplificar os tipos de exercícios elaborados a partir de suas funções.

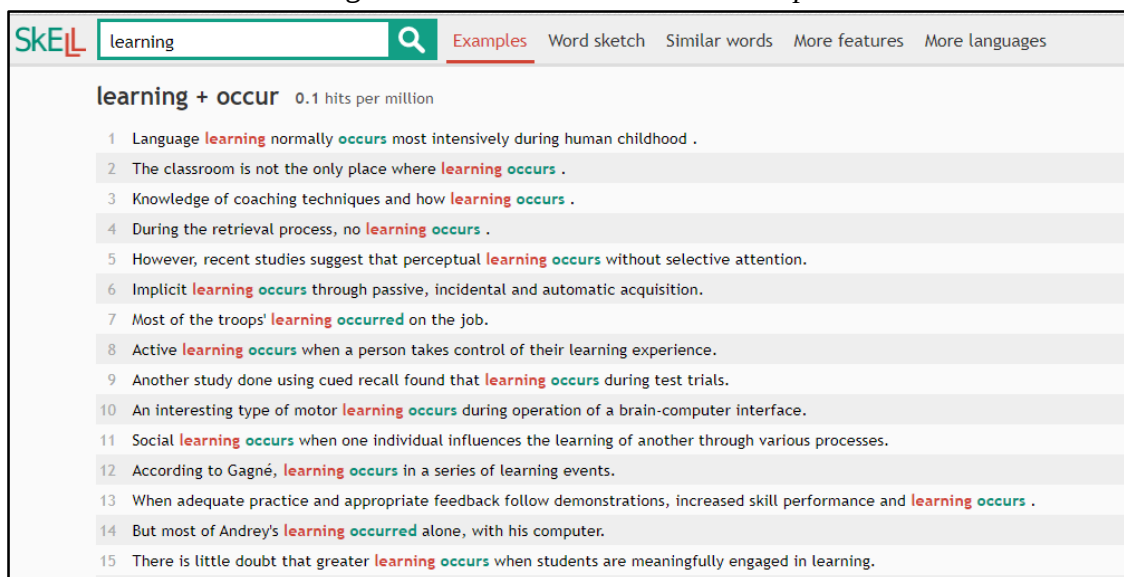
Voltando-se às ferramentas *Word sketch*, *Examples* e *Similar words*, a primeira diz respeito à uma lista de palavras oferecida pelo *corpus* que dispõe de uma coleção de bilhões de palavras contidas em notícias, artigos científicos, artigos da Wikipédia, ficção, páginas da web e blogs, com detalhes suficientes para facilitar o trabalho do(a) professor(a). Para exemplificar seu uso, foi utilizada a palavra *learning*, como podemos notar na imagem abaixo:

Imagem 5 - SkELL – Ferramenta *Word sketch*

Fonte: SkELL (KILGARRIFF, *et al*, 2014)

Como visto na imagem, há várias maneiras de se utilizar essa ferramenta. O *word sketch* separa a palavra de busca, que no caso é *learning*, em suas diferentes formas, ou seja, há a unidade como sujeito em uma frase, como objeto, os adjetivos mais usados com *learning*, os modificadores dessa palavra, entre outros. Assim, é possível que o(a) professor(a) prepare sua atividade de acordo com o que é necessário em determinada aula. Desse modo, se o objetivo da aula é abordar diferentes adjetivos, essa ferramenta pode facilitar muito o trabalho do(a) profissional, já que ele(a) pode ter a certeza de que esses elementos são classificados pela sua frequência de uso em conjunto com determinada palavra que o(a) professor(a) escolher abordar em sala.

Clicando em cima de alguma opção, como *learning occurs*, por exemplo, aparece a imagem abaixo, que leva à aba “*examples*”:

Imagem 6 - SkELL – ferramenta *Examples*

Fonte: SkELL (KILGARRIFF, *et al*, 2014)

Os exemplos apresentados na tela são diferentes dos que uma lista de concordância oferece, pois esta mostra parte de um texto, ou seja, o(a) leitor(a) não tem, muitas vezes, a visualização do início e do fim da frase, mas sim a parte em que se encontra a unidade pesquisada. Caso o pesquisador queira ver o contexto todo em que a palavra está inserida, é necessário clicar sobre ela para que isso ocorra. Já por meio do “*examples*” no SkELL, é visto apenas a frase em que se encontra a palavra de pesquisa. Isso facilita novamente o trabalho do(a) professor(a), pois facilmente este selecionará exemplos para os exercícios que deseja desenvolver.

Já em “*Similar words*”, é possível encontrar palavras sinônimas àquela pesquisada pelo(a) professor(a). Ainda tendo como exemplo a unidade *learning*, pode-se verificar como se dá essa ferramenta por meio da imagem abaixo:

Tabela 3 - Itens e formas dos *corpora* compilados

<i>Corpus</i>	<i>Formas</i>	<i>Itens</i>
<i>Health</i>	53,653	63,330
<i>Politics</i>	55.861	64,811
<i>Tech</i>	58,716	68,645
<i>World</i>	54,930	64,300

Fonte: Dados dos *corpora* da presente pesquisa

A primeira coluna apresenta os *corpora* de estudo, sendo os quatro temas abordados nesta pesquisa. Na segunda coluna, temos o número de formas (*types*), ou seja, o número de palavras diferentes no *corpus*, não contando as repetidas, enquanto a terceira coluna mostra o número de itens (*tokens*), o total de palavras, contando as que se repetem.

Outro ponto muito importante em relação aos *corpora*, é sua variação lexical, ou seja, quanto o *corpus* varia no que diz respeito seu vocabulário, o quanto há de repetição de vocábulos em sua composição. Para isso, é feita uma conta estatística que apresenta esses números, a que chamamos de razão forma/item (FI). Segundo Berber Sardinha (2004, p. 94) essa razão “indica a riqueza lexical do texto. Quanto maior o seu valor, mais palavras diferentes o texto conterá. Em contraposição, um valor baixo indicará um número alto de repetições, o que pode indicar um texto menos rico”. A razão FI será abordada mais detalhadamente no Capítulo 5.

A tabela abaixo, baseada nos critérios de Berber Sardinha (2004) para a construção de um *corpus*, apresenta o conteúdo e o propósito dos *corpora* da presente pesquisa.

Quadro 2 - Conteúdo e propósito dos *corpora*

MODO	Escritos, os <i>corpora</i> são compostos de textos eletrônicos.
TEMPO	Contemporâneo.
SELEÇÃO	Compostos por textos escritos, sendo eles jornalísticos e autênticos, retirados dos sites BBC® e CNN®.
CONTEÚDO	O conteúdo do <i>corpus</i> é de caráter especializado, os textos são dos temas <i>saúde</i> , <i>política</i> , <i>tecnologia</i> e <i>mundo</i> .

AUTORIA Os autores são falantes nativos ou não-nativos.

FINALIDADE Os *corpora* foram utilizados para o desenvolvimento de atividades de CE em LI.

Fonte: Dados da presente pesquisa

Acerca do Quadro 2, podemos dizer que o critério “tempo” é muito importante em quaisquer pesquisas, porém, voltando-se para esta pesquisa, a contemporaneidade é um ponto chave, visto que, por fazer uso de *corpora* recicláveis, o intuito da presente pesquisa é estar sempre atualizando os textos que serão compilados, para, consequentemente abordar novos vocábulos durante as aulas.

Após a apresentação dos critérios de compilação de um *corpus*, discutiremos quais foram os instrumentos da pesquisa utilizados neste estudo, que foram organizados da seguinte forma:

Quadro 3 - Organização dos instrumentos de pesquisa

OBJETIVOS	INSTRUMENTOS
Coleta de dados	- Compilação dos <i>corpora</i>. Textos retirados dos sites BBC® e CNN®. - Aplicação de um questionário ao final de cada aula, bem como ao final da pesquisa em sala de aula.
Análise de dados	- Uso do software <i>Sketch Engine</i>® para análise dos <i>corpora</i> – extração das palavras-chave e agrupamentos lexicais. - Análise dos questionários semanais e final.
Desenvolvimento das atividades	- As atividades foram desenvolvidas por meio dos agrupamentos lexicais que são chave no contexto do tema abordado.
Aplicação das atividades	- A aplicação das atividades se deu em uma escola particular de ensino regular situada em uma cidade do interior de São Paulo. As aulas foram ministradas durante o primeiro semestre de 2019. O público são alunos do segundo e terceiro anos do ensino médio.
Apresentação dos resultados	- Apresentação dos dados obtidos pela análise lexical e pela análise dos questionários.

Fonte: Dados da presente pesquisa

Acima, o quadro apresenta a organização referente aos instrumentos de pesquisa, desde a coleta de dados até a apresentação dos resultados. Em relação ao desenvolvimento das atividades, terceiro item do quadro, será detalhadamente explicado no subcapítulo – “Elaboração das atividades”. A seguir serão apresentados os critérios obrigatórios e não obrigatórios a serem adotados em uma pesquisa com base em *corpus*.

3.3 Critérios obrigatórios e não obrigatórios na preparação de atividades didáticas baseadas em *corpora*

As atividades propostas pela presente pesquisa baseiam-se nos critérios obrigatórios (CO) para atividades com base *corpora* (DELFINO, 2016) como exhibe o quadro abaixo:

Quadro 4 - Lista dos critérios obrigatórios para a preparação de atividades

- | |
|---|
| O1 – O exercício faz uso de <i>corpus</i> ;
O2 – O exercício precisa ter enunciados claros;
O3 – O exercício tem como foco principal o padrão léxico-gramatical;
O4 – O exercício é ético;
O5 – O exercício é replicável;
O6 – O exercício é motivador;
O7 – O exercício não simplifica a língua usada nos textos/concordâncias/lista de palavras, etc;
O8 – O exercício deve apresentar nível de dificuldade adequado;
O9 – O exercício contém conteúdo relevante para o aluno e para a construção do conhecimento em inglês;
O10 – O professor é facilitador e não distribuidor de conhecimento;
O11 – O aluno é descobridor, pesquisador e não recipiente de conhecimento. |
|---|

Fonte: DELFINO (2016, p. 56)

No que diz respeito especialmente o CO 5, o exercício é replicável na medida em que outro professor pode reutilizar as atividades aqui propostas ou simplesmente seguir a sequência sugerida no Quadro 6. Dessa forma, pensando que um outro profissional queira abordar exercícios de CE baseados em um assunto que não está em nenhuma das atividades aqui elaboradas, ele pode criar seu próprio *corpus* de determinado tema e seguir os três passos que neste estudo também foram seguidos.

No que concerne o CO 7, é importante ressaltar que, em nenhum momento, a linguagem retirada do *corpus* é simplificada no sentido em que a pesquisadora a modifica com a intenção de facilitar o acesso dos alunos à língua, visto que até os textos são abordados em sua forma autêntica. Todavia, um ponto relevante a ser discutido aqui é a utilização do *SkELL*®. Essa

plataforma não faz mudanças na língua, faz uso da autenticidade, porém é voltada exatamente para o aprendizado de línguas pela facilidade de entendimento das frases exibidas. Assim, buscou-se utilizar algo que os alunos não se sentissem “perdidos” nas atividades devido a linguagem difícil, mas em nenhum momento foram feitas alterações na língua, e sim o uso de uma ferramenta voltada para esse contexto.

A próxima tabela apresenta os critérios não obrigatórios (NO), ou seja, as características de atividades baseadas em *corpora*, porém que não precisam ser seguidas pelo professor/pesquisador no desenvolvimento de seus exercícios. É importante ressaltar que, desta tabela, apenas os passos 3, 6, 7, 13 e 17 não são seguidos.

Quadro 5 – Lista dos critérios não obrigatórios para a preparação de atividades

NO1 – O exercício é participativo;
NO2 – O exercício é colaborativo;
NO3 – O exercício exige pouco tempo de preparação por parte do professor;
NO4 – O exercício trabalha com o conceito de padronização;
NO5 – O exercício lida com o conceito de frequência;
NO6 – O exercício trabalha com o conceito de variação;
NO7 – O exercício trabalha com o conceito de variedade textual;
NO8 – O exercício incorpora mídias diferentes;
NO9 – O exercício incorpora linhas de concordância;
NO10 – O exercício incorpora textos;
NO11 – O exercício incorpora lista de frequência de palavras;
NO12 – O exercício incorpora lista de palavras-chave;
NO13 – O exercício faz com que os alunos trabalhem diretamente com <i>corpora</i> ;
NO14 – O exercício incorpora diagramas e diferentes formas de visualização;
NO15 – O exercício é de fácil adaptação;
NO16 – O exercício desenvolve a autonomia;
NO17 – O exercício ensina e não testa.

Fonte: DELFINO (2016, p. 59)

O critério NO 3 não foi seguido, visto que a preparação das atividades exigiu muito tempo da pesquisadora, já que não foi utilizado um *corpus* já existente, mas foram compilados 4 *corpora* de temas diferentes para o desenvolvimento das atividades. É fato que, nem todos os professores/pesquisadores necessitariam seguir esse procedimento, dependendo do tipo de exercício, um *corpus* de referência seria suficiente, e, ainda que o profissional optasse por compilar um *corpus*, a compilação seria mais trabalhosa no seu início, já que, se tratando de um *corpus* reciclável, textos poderiam ser adicionados de acordo com a necessidade dos alunos.

No que diz respeito os critérios NO 6 e 7, estes não condizem com os objetivos da presente pesquisa, já que visam trabalhar com variações da língua, enquanto espera-se, aqui, analisar o vocabulário mais frequente de determinados temas de artigos jornalísticos.

O critério NO 13 está relacionado ao uso direto de *corpora* pelos alunos, porém isso não acontece em sala de aula se for pensar que esse uso direto diz respeito à atividade *hands-on*, ou seja, aquela em que os participantes acessam o *corpus* pelo computador para buscar informações acerca da língua. Nesta pesquisa, essas informações já são selecionadas anteriormente e dadas aos estudantes, para que estes busquem informações da língua nos próprios exercícios.

O último critério não foi utilizado neste estudo, pois, acredita-se que, quando atividades são desenvolvidas e inicialmente aplicadas em determinado contexto, elas estão automaticamente testando o conhecimento do aluno, já que este pode ou não conhecer o conteúdo apresentado. Analisando por uma perspectiva construtiva, o que não funcionar como esperado, ou até mesmo não tiver certa aprovação dos participantes, pode sofrer alterações para a melhora dos exercícios. Assim sendo, além de ensinar, as atividades também testam, principalmente no que diz respeito a exames como o vestibular.

3.4 Elaboração das atividades

Como mencionado, para a elaboração das atividades foram utilizadas as ferramentas palavras-chave (*keywords*) e agrupamentos lexicais (*clusters*) a partir dos *corpora* de estudo compilados para o trabalho. Com estas ferramentas são geradas, automaticamente, duas listas de palavras-chave – uma referente às palavras soltas (apenas uma unidade lexical) e outra com os agrupamentos lexicais mais frequentes, como visto na imagem abaixo. Na observação dessa lista de blocos de palavras, foram escolhidos os agrupamentos que apresentavam chavicidade relevante, ou seja, que tinham mais probabilidade de serem importantes para o contexto dos textos. Foram selecionados os agrupamentos de duas ou três unidades. Esta lista foi a que serviu como base para a criação das atividades.

Imagem 8 - Lista de agrupamentos lexicais do *corpus Tech*

</

Fonte: Sketch Engine (KILGARRIFF, *et al*, 2014)

Faz-se relevante mencionar que foram escolhidos até 10 agrupamentos lexicais para o desenvolvimento das atividades, os quais evitou-se repetição de palavras, como por exemplo, *life expectancy*, que aparecia em outros agrupamentos como em *lackcluster performance regarding life expectancy* e *regarding life expectancy*, assim quando aconteceu esse tipo de repetição, apenas um dos agrupamentos foi considerado.

Após a seleção dos agrupamentos lexicais, as atividades foram organizadas com a mesma estrutura de exercícios, em uma sequência de três seções diferentes: (1) *Warm-up*, (2) exercícios de vocabulário com uso dos agrupamentos lexicais, e, finalmente, (3) os exercícios de CE, os quais possuem um dos textos compilados no *corpus* referente a determinado assunto. Cada seção dessa segue dois critérios que, necessariamente, deve ser levado em conta no desenvolvimento de cada exercício que será detalhado a seguir.

3.4.1 Seção de *warm-up*

Dois critérios selecionados para serem utilizados na seção de *warm-up* são: o uso de palavras-chave e a discussão acerca do tema abordado. Primeiramente, o *warm-up* tem o objetivo de incitar os alunos a discutirem o que eles já sabem do assunto que será abordado nos exercícios, assim, eles buscam na memória, e em seu *background*, ou seja, seu conhecimento prévio, o que já ouviram falar a respeito desse tópico. Acredita-se que esse passo pode ser facilitador no que tange a compreensão do texto, e, também das atividades, já que, ainda que o aluno não compreenda determinada palavra solta, por exemplo, ele pode inferir seu significado por meio do que ele já conhece do assunto.

Voltando-se estritamente para o contexto desta pesquisa, a seção de *warm-up* é constituída de perguntas em português, ao passo que alguns outros exercícios presentes ao longo da atividade possuem questões na LI. Essa escolha se deu devido ao fato de que este estudo não tem o objetivo de avaliar a produção oral dos alunos na LI, desse modo, para dar início a uma discussão, especialmente para levantar o que os alunos já sabiam acerca do assunto abordado, não houve a necessidade de ter uma comunicação em inglês. Além disso, a produção oral não é uma habilidade exigida em vestibulares, assim, é importante esclarecer que toda a aula foi ministrada em português, visto que os alunos apresentavam níveis variados de LI, levando em consideração tanto a produção quanto a compreensão oral.

Ainda que as imagens das atividades não sejam um fator considerado parte do *warm-up* elas ajudam os alunos no desenvolvimento da discussão e, muitas vezes, sua compreensão acerca do texto. Dessa maneira, voltamos ao ponto de que, apesar de não serem obrigatórias nesta pesquisa, as imagens também têm um papel importante na leitura de um texto.

Um fator muito importante a ser mencionado, e que será mais detalhadamente explicitado no subcapítulo 4.2, é que, para a atividade de *warm-up*, o texto selecionado para os exercícios de CE foi utilizado isoladamente do *corpus* para a extração dos agrupamentos lexicais chave. Ou seja, os agrupamentos abordados no início do *warm-up*, são retirados de um único texto compilado (que faz parte do *corpus* da pesquisa), que é o que será abordado nos exercícios de CE. As outras atividades de vocabulário levam em consideração o *corpus* do tema em geral.

A técnica de inferência foi utilizada durante toda a leitura, mas principal e explicitamente no *warm-up*, isso porque logo no primeiro exercício é pedido para que os alunos tentem inferir o assunto do texto por meio das palavras-chave, todavia não somente aqui é utilizada essa técnica. Em exercícios de vocabulário, também é pedido para que os alunos infiram o sentido de determinadas palavras por meio do contexto em que é dado a eles, como será visto a seguir na seção de vocabulário.

Também é na primeira seção – *warm-up* – que os alunos colocam em prática a estratégia de identificação/familiarização lexical. Quando feito o primeiro exercício, a professora pesquisadora pedia aos alunos que identificassem as palavras cognatas e as familiares, para notarem que muito do vocabulário apresentado nas palavras-chave era de conhecimento deles e estava presente no texto em que eles leriam em seguida. Durante toda a leitura essa estratégia foi enfatizada – destacar palavras que eram familiares aos alunos e inferirem o que estava sendo dito. As próximas técnicas serão discutidas no próximo subcapítulo em que os exercícios lexicais são destacados.

3.4.2 Seção de vocabulário

A utilização de agrupamentos lexicais e o do contexto de uso desses agrupamentos foram considerados obrigatórios nessa seção ao empregar o vocabulário extraído dos *corpora* na elaboração dos exercícios.

Assim, a segunda parte da sequência – vocabulário – é seguida em todas as atividades, todavia cada assunto aborda essa seção diferentemente. Em outras palavras, o desenvolvimento dos exercícios de vocabulário se deu de acordo com os agrupamentos lexicais escolhidos, isto é, cada agrupamento levou à elaboração de um tipo de atividade diferente, que utilizou aparatos que melhor se encaixavam para o ensino de determinado léxico. Dependendo do que o *corpus* apresentava como relevante, foram desenvolvidos exercícios relativos a esse léxico de maneira que enfatizasse as palavras-chave do *corpus*, bem como seus agrupamentos lexicais mais frequentes e o uso do contexto dessas unidades.

3.4.3 Seção de atividades de compreensão escrita

Após as atividades que abordavam vocabulário, os alunos encontraram exercícios de CE com textos retirados do próprio *corpus* de estudo referente ao tema que estava sendo abordado. Esses textos não foram escolhidos de forma aleatória, para a sua seleção foram seguidos os seguintes passos:

1. **Verificação** de qual agrupamento lexical tem maior frequência relativa no *corpus* de estudo.
2. **Seleção** do texto que tenha maior concentração do agrupamento escolhido.
3. **Compilação** do texto selecionado.
4. **Elaboração** dos exercícios de CE com base no texto selecionado.

Após a seleção dos textos, foram desenvolvidos os exercícios de CE que tiveram como base os vestibulares da UNESP – Universidade Estadual Paulista – dos últimos três anos, 2016, 2017 e 2018. Assim sendo, as atividades propostas nesse estudo tiveram questões de múltipla escolha em português e em inglês, como é exigido nos exames da universidade citada.

Buscou-se, nesta pesquisa, trabalhar apenas com exercícios lexicais, ou seja, por mais que, de alguma forma, a gramática esteja inclusa nas questões em geral, esse não foi o foco das atividades. Dessa forma, os exercícios contêm, geralmente, perguntas acerca da compreensão

de determinados parágrafos, da opinião de determinada pessoa dentro do texto e palavras sinônimas.

Durante a aplicação das atividades, eram enfatizadas algumas técnicas de leitura para que os alunos tivessem uma compreensão textual facilitada e, de certa forma, mais rápida, visto que nos vestibulares é exigido que os alunos busquem formas de resolver os exercícios rapidamente a fim de terminar a prova toda dentro do tempo estipulado. Uma das técnicas mais ressaltadas foi ler o texto se atentando aos agrupamentos lexicais e não somente às palavras soltas, além de fazer, primeiramente, a leitura das perguntas para, posteriormente, se dirigir ao texto a fim de encontrar a resposta adequada.

Apesar do uso das estratégias acima mencionadas, dois critérios também foram utilizados para a seção de atividades de CE: o uso da técnica *skimming* e da técnica *scanning*, assim, eram necessárias questões de cunho geral, bem como de entendimento mais específico, como acontece nos vestibulares em geral.

Como já foi discutido, o uso de *corpora* auxilia no desenvolvimento da habilidade de leitura, especialmente no que diz respeito ao léxico de um texto. Todavia, faz-se necessário observar o *corpus* e buscar por palavras que sejam relevantes aos alunos e ao texto que eles irão ler.

Para melhor visualização da sequência dos exercícios, foi desenvolvido um quadro a fim de ser seguido em todas as atividades, que se deu da seguinte forma:

Imagem 9 - Sequência de atividades com agrupamentos lexicais



Fonte: Sequência elaborada pela autora

A seguir será explicitado o contexto em que se deu a aplicação das atividades aqui elaboradas e será abordado em que turmas especificamente foi trabalhado o conteúdo dessa pesquisa.

3.5 Aplicação das atividades – contexto

As atividades desenvolvidas nesta pesquisa foram aplicadas em uma escola particular de ensino regular situada no interior do Estado de São Paulo. Como mencionado, foram selecionadas apenas as turmas de segundo e terceiro anos do ensino médio, devido o maior interesse e, conseqüentemente, maior desempenho por parte dos alunos em relação aos exames de vestibular.

Ambas as turmas estudavam no período matutino, sendo que as aulas de LI ocupadas para a aplicação desta pesquisa ocorria às terças-feiras no período das 7h15 até 8h55 no terceiro ano, e, no mesmo dia, porém das 8h55 às 10h50 no segundo ano.

No total, foram 24 alunos participantes. O nível de inglês de cada aluno variava muito, visto que, alguns estudantes já haviam frequentado ou ainda frequentavam cursos de idiomas (inglês) fora da escola de ensino regular. Por outro lado, outros alunos foram identificados com um nível mais baixo de inglês pelo fato de só ter contato com a língua nas aulas oferecidas na escola em si – fato que, segundo a professora de LI da escola, impossibilitava as aulas serem regidas em inglês.

A multiplicidade de nível dos alunos pôde ser também identificada por um questionário em que foi entregue no primeiro dia da pesquisa em sala de aula. Foi perguntado para cada aluno qual era o seu nível de proficiência em LI. Faz-se necessário esclarecer que, essa análise aplicada aos alunos não tinha a intenção de avaliar, por meio de questões de LI, o nível de cada um, mas sim, identificar o nível por meio da opinião dos próprios estudantes. Um teste de proficiência não foi aplicado aos alunos anteriormente por motivos de tempo disponibilizado para a pesquisadora realizar sua pesquisa na escola.

4. ANÁLISE DOS RESULTADOS

Neste capítulo será discorrido acerca das análises feitas durante todo o processo de elaboração e aplicação das atividades e dos questionários semanais e final entregues aos alunos. Assim sendo, a divisão desta seção dá-se da seguinte forma: O primeiro subcapítulo explica o conceito de FI e como se dá a variação lexical de cada *corpus* de estudo do presente estudo, enquanto, em seguida, o subcapítulo 4.2 discorre a respeito do processo de elaboração das atividades por meio da análise lexical de cada *corpus*. O último subcapítulo apresenta a análise dos questionários entregues aos alunos, tanto os semanais, entregues ao final de cada atividade, bem como o final, que foi respondido pelos alunos quando a aplicação de todas as atividades havia sido feita.

4.1 Análise de variação lexical – FI

Na análise FI (forma/item), é apresentada a razão entre a forma e o item de cada *corpus*, a divisão do primeiro pelo segundo apresenta o número de variação lexical do *corpus*. Segundo Thomas (2005, p. 01), “Um FI alto indica um alto grau de variação lexical, enquanto um FI baixo indica o oposto. O intervalo varia entre um 0 teórico (repetição infinita de uma única forma) e 1 (a não repetição completa encontrada em uma concordância)”. Este número mostra que quanto maior o resultado, mais variação de vocabulário foi utilizada pelo autor do texto, ao passo que, se este valor for baixo, o autor repetiu mais os vocábulos e não variou tanto, o que pode evidenciar um texto menos rico lexicalmente (BERBER SARDINHA, 2004).

De acordo com Thomas (2005), a variação do léxico de um *corpus* está dentro de um intervalo entre 0 e 1, sendo que quanto mais se aproxima de 0, o número mostra uma repetição contínua de palavras naquele determinado *corpus*, enquanto aproximando-se de 1, a variação de palavras é maior.

Segundo os números encontrados pela relação FI de cada *corpus* da presente pesquisa, em ordem decrescente, a maior variação lexical foi encontrada no *corpus Politics* com 0,86, seguida de 0,85 nos *corpora* dos temas *Technology* e *World* e de 0,84 para o *corpus Health*. Esses dados podem ser verificados na tabela abaixo.

Tabela 4 - Itens e formas dos *corpora* compilados

<i>Corpus</i>	<i>razão FI</i>
<i>Health</i>	0,84
<i>Politics</i>	0,86
<i>Tech</i>	0,85
<i>World</i>	0,85

Fonte: Dados dos *corpora* da presente pesquisa

Por meio dessa análise é possível afirmar que, apresentando uma variação entre 0,84 e 0,86 todos os *corpora* de estudo dessa pesquisa têm variação lexical consideravelmente alta, visto que os números se aproximam mais de 1 do que de 0.

A análise FI apresenta dados positivos para esta pesquisa, na medida em que foi possível utilizar diferentes agrupamentos lexicais nas atividades, possibilitando que os alunos entrassem em contato com um vocabulário mais diversificado. Caso a variação lexical dos *corpora* fosse muito baixa, as atividades desenvolvidas também teriam menos variação, já que os agrupamentos se repetiriam.

Assim sendo, foi possível escolher entre 5 a 10 agrupamentos para diversificar os exercícios propostos, justamente pela variação lexical encontrada em cada *corpus* de estudo. Mesmo que abrangendo um mesmo tema, cada atividade, de cada tópico, permitiu que diferentes ferramentas do *Sketch Engine*® e do *SkELL*® fossem utilizadas.

4.2 Elaboração das atividades por meio da análise lexical de cada *corpus*

Neste subcapítulo será explicado acerca de como se deu o processo de escolha das palavras-chave, dos agrupamentos lexicais e dos textos para os exercícios de CE de cada tema dos *corpora*. A ordem seguida será *Health*, *Politics*, *World* e *Tech*. Faz-se relevante mencionar que as duas primeiras atividades, dos temas *health* e *politics* foram mais longas e, assim, contaram com dois textos com exercícios de CE. Por outro lado, por motivo de menos tempo disponível em sala de aula, a professora-pesquisadora selecionou apenas um texto para os dois últimos temas – *world* e *tech*.

Assim, pretende-se mostrar como as escolhas lexicais de cada *corpus* de estudo resultaram em cada exercício, e, analisar esses exercícios dentro de cada etapa.

4.2.1 Atividade de *Warm-up*

Levando em consideração a importância do levantamento de conhecimento prévio dos alunos antes de realizar a leitura (FARRELL, 2003), dois passos foram considerados na etapa *warm-up* – o uso de palavras-chave para a inferência do assunto do texto e a discussão acerca do tema em busca do *background* dos alunos a respeito desse tema. É importante mencionar que o exercício de *warm-up* seguiu esse padrão em todas as atividades dos 4 temas, e isso pode ser verificado por meio da observação das atividades que se encontram na seção de apêndices.

Tratando-se, primeiramente, do tema *health*, foi analisado que o agrupamento lexical que apresentou frequência relativa mais alta foi *life expectancy*, com 647.402, como pode ser observado no Apêndice F. Posteriormente, foi analisado em qual texto do *corpus* esse agrupamento aparecia mais frequentemente a fim de selecioná-lo para a elaboração dos exercícios de CE. Por meio da imagem abaixo é possível verificar que *life expectancy* se encontra repetidas vezes no texto “*Life expectancy drops in the US and the UK*”, assim sendo, esse foi o texto abordado na atividade de CE.

Imagem 10 - Seleção de textos do *corpus Health*

File name	↓ Frequency	Relative [%]	
1 Life expectancy drops in the US and the UK.txt	21	3,073.9	...
2 UK among worst for life expectancy rises.txt	12	1,945.1	...
3 NHS prevention plan aims to boost healthy life expectancy.txt	4	917.6	...
4 Low and high carb diets increase risk of early death, study finds.txt	2	296.8	...
5 Childhood obesity America's 'true national crisis' measured state by state.txt	2	215.1	...

Fonte: Dados do *corpus Health*

A partir dessa seleção, o texto escolhido foi compilado isoladamente para extração dos agrupamentos lexicais que podem ser encontrados no primeiro exercício, em que é pedido ao aluno para inferir o assunto específico do texto a partir das palavras-chave apresentadas, que foram: *life expectancy*, *health-care provision*, *population mortality*, *rising concern* e *declining quality*.

Como explicado, esses agrupamentos foram utilizados para a elaboração do exercício de inferência do assunto do texto. Entretanto, outros agrupamentos do *corpus Health* foram selecionados para outros exercícios ao longo de toda a atividade.

No que diz respeito ao tema *politics*, o agrupamento em destaque foi *military parade*, devido sua frequência relativa de 154.295 no *corpus*, como é apresentado no Apêndice H. Após

a observação de qual texto esse agrupamento era mais frequente, obtivemos o seguinte resultado, apresentado abaixo:

Imagem 11 - Seleção de textos do *corpus Politics*

File name	↓ Frequency	Relative [%]
1 The Trump military parade was always a really bad idea.txt	10	6,376.8  ...

Fonte: Dados do *corpus Politics*

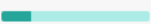
É possível verificar na imagem acima que o texto que possui maior concentração de *military parade* é “*The Trump military parade was always a really bad idea*”. Assim sendo, esse foi o texto selecionado para os exercícios de CE. É importante mencionar que outros agrupamentos com alta ou mesma frequência relativa levariam a outros textos, isso fica a critério do pesquisador-professor.

Desse modo, apesar de *administration oficial* ter apresentado frequência relativa mais alta no *corpus*, o agrupamento escolhido como principal para ser abordado nos exercícios de CE, foi *military parade*, pois, com segunda maior frequência, esse agrupamento possibilitava um *warm-up* interessante com os alunos acerca desse assunto, além de abrir espaço para atividades léxico-gramaticais que abordaram linhas de concordância.

A partir do texto selecionado, as palavras-chave (os agrupamentos-chave) extraídos foram: *military parade*, *bad local government*, *democratic-controlled government*, *French military roll* e *democratic city*.

Voltando-se para a atividade de tema *world* o agrupamento selecionado pela sua alta frequência relativa foi *warning system* com 171.073. Analisando cada um deles por meio da lista de agrupamentos, *warning system* foi o segundo agrupamento com frequência relativa mais alta, porém, foi o selecionado, pois, o primeiro era *bread roll*, não tendo nenhuma relação com os outros agrupamentos encontrados no *corpus*. Assim, o texto em que *warning system* possuía mais chavicidade, foi “*Indonesia accused of mishandling tsunami warnings*”, como visto na imagem abaixo:

Imagem 12 - Seleção de textos do *corpus World*

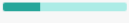
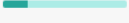
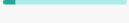
File name	↓ Frequency	Relative [%]
1 Indonesia earthquake and tsunami.txt	5	2,693.4  ...
2 Indonesia accused of mishandling tsunami warnings.txt	5	2,889.8  ...
3 Satellites warn African farmers of pest infestations.txt	1	1,102.3  ...

Fonte: Dados do *corpus World*

A partir da compilação desse texto, os agrupamentos extraídos foram *tide gauge*, *warning system*, *geophysics agency*, *national disaster agency* e *long narrow bay*.

Em relação ao *corpus Tech*, o agrupamento *artificial intelligence* teve frequência relativa mais alta em relação aos outros do *corpus*. Assim, o texto que tinha mais concentração desse agrupamento foi “*Facebook’s top AI scientist says it’s ‘dust’ without artificial intelligence*”, como pode-se notar na imagem abaixo:

Imagem 13 - Seleção de textos do *corpus Tech*

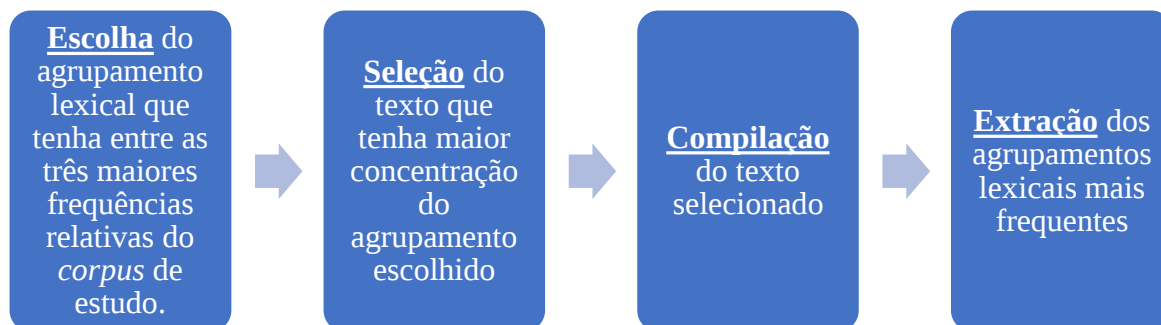
File name	↓ Frequency	Relative [%]
1 The A-Z of how artificial intelligence is changing the world.txt	10	1,075.4  ...
2 Facebook's top AI scientist says it's 'dust' without artificial intelligence.txt	3	1,443.8  ...
3 Rollable TVs and butler robots Gadgets that caught our eye at CES 2019.txt	2	368.4  ...
4 "The weird events" that make machines hallucinate".txt	1	109.9  ...
5 The world's most prolific writer is a Chinese algorithm .txt	1	107.4  ...
6 The designer changing the way aircraft are built.txt	1	114.8  ...
7 Saudi crisis means Masa Son.txt	1	314.8  ...
8 How Alexa knows the difference between a breaking window and a wine glass.txt	1	516.2  ...
9 How 2006 changed the internet.txt	1	228.7  ...
10 China's ByteDance is taking the social media world by storm.txt	1	242.6  ...

Fonte: Dados do *corpus Tech*

A partir dessa seleção, o texto escolhido foi compilado isoladamente para extração dos agrupamentos lexicais que podem ser encontradas no primeiro exercício, em que é pedido ao aluno para inferir o assunto específico do texto a partir das palavras-chave apresentadas, que foram: *deep learning*, *artificial intelligence*, *intelligence research*, *Facebook today*, e *face recognition*.

Tendo em vista que a seleção dos textos e dos agrupamentos lexicais passou por diferentes etapas até chegar à atividade, foi criada uma sequência para melhor visualização desse processo, como pode ser visto abaixo:

Imagem 14 - Sequência de seleção do texto e dos agrupamentos lexicais



Fonte: Sequência elaborada pela autora

Desse modo, a primeira etapa acontece por meio da escolha do agrupamento mais relevante para o *corpus*, através da análise de sua frequência relativa, e é seguido pelo segundo passo, que se dá por meio da seleção do texto que possui esse agrupamento com maior concentração em relação aos outros textos do *corpus*. A terceira etapa é a compilação separada do texto selecionado para extração dos agrupamentos lexicais que são considerados como palavras-chave no primeiro exercício de do texto – o de inferência do assunto.

4.2.2 Atividade de Vocabulário

Retomando os pensamentos de Coxhead (2013) que defende a ideia de que o *corpus* auxilia na identificação de léxico especializado, a seção de vocabulário segue dois critérios importantes: o uso de agrupamentos lexicais e o uso do contexto, todavia, não é necessariamente igual nas atividades dos diferentes temas. Por esse motivo, será abordado aqui como foi analisado cada agrupamento para se chegar no tipo de exercício final. Essa seção de questões relacionadas ao vocabulário teve como objetivo abordar o léxico não apenas do texto que os alunos leram, mas também do *corpus* em geral de determinado tema.

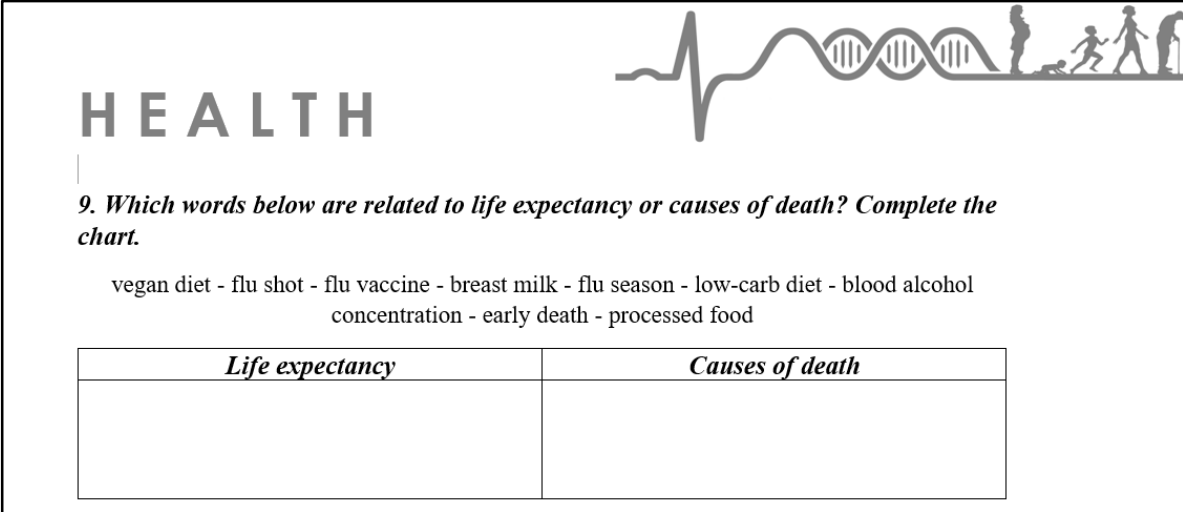
Como foi visto, alguns agrupamentos foram selecionados a fim de contribuírem para a inferência do texto na seção de *warm-up*, sendo considerados como palavras-chave, devido sua frequência no texto. Todavia, outros agrupamentos do *corpus Health* foram selecionados para outros exercícios ao longo de toda a atividade, como *vegan diet*, *flu shot*, *flu vaccine*, *breast*

milk, flu season, low-carb diet, blood alcohol concentration, early death e processed food. Para essa seleção, a frequência relativa também foi levada em consideração. Todavia, faz-se relevante mencionar que, dentre os 50 agrupamentos extraídos, alguns que não foram selecionados tinham a frequência um pouco mais alta do que determinadas unidades que foram escolhidas. Isso se deu pelo fato de que, ao desenvolver a atividade, alguns agrupamentos estavam relacionados uns aos outros, como por exemplo *life expectancy* e *population mortality*, facilitando, dessa forma, a elaboração dos exercícios.

Desse modo, por mais que a chavicidade seja um ponto de extrema importância na confecção de cada exercício, o olhar do professor-pesquisador no que diz respeito à conexão de um exercício com o outro, também é muito relevante. O que não deve ser feito, neste caso, é selecionar agrupamentos com frequência relativa muito baixa, pois, assim, os dados estatísticos oferecidos pelo *corpus* não terão tido a atenção necessária que a abordagem teórico-metodológica da LC defende.

Para ilustrar melhor esse fato, abaixo temos um exemplo de um exercício desenvolvido com os agrupamentos retirados do *corpus Health*:

Imagem 15 – Exemplo de atividades do *corpus Health* #1



HEALTH

9. Which words below are related to life expectancy or causes of death? Complete the chart.

vegan diet - flu shot - flu vaccine - breast milk - flu season - low-carb diet - blood alcohol concentration - early death - processed food

<i>Life expectancy</i>	<i>Causes of death</i>

Fonte: Atividade elaborada na presente pesquisa

Como pode-se notar, no exercício acima é pedido para que os alunos completem a tabela dividindo os agrupamentos oferecidos acima entre os que estão relacionados, de algum modo, à expectativa de vida (*life expectancy*) ou às causas de morte (*causes of death*). Em uma atividade como essa, não seria cabível inserir um agrupamento como *new study*, que aparece no *corpus* com frequência relativa alta. Por motivos como esse, que foram guiando o desenvolvimento de cada exercício, é que as escolhas finais se deram.

Já no exercício abaixo podemos ver que dois verbos foram abordados de modo acentuado: *decline* e *rise*. Isso se dá, pois, ao verificar os agrupamentos lexicais apresentados pelo *corpus* do texto, notou-se que, seis deles apresentavam verbos com sentido de “aumentar, crescer” ou “diminuir”: *simultaneous decline*, *declining quality of health-care provision*, *declining quality*, *rising concern*, *continued decline* e *continued rise*. Por esse motivo, ao perceber a relevância desse léxico para a compreensão do texto abordado, foram abordadas palavras sinônimas com o intuito dos alunos fazerem relação entre essas unidades.

Imagem 16 - Exemplo de atividades do *corpus Health #2*

2. Read the sentences below and complete the chart with the highlighted words.

<i>Decline</i>	<i>Rise</i>

he's now growing his business by buying even more apple seedlings to plant.

Australia has one of the fastest growing economies in the developed world.

The streets of our peer countries are becoming safer, faster. In Japan, the traffic death rate fell 47 percent between 2000 and 2011.

Europe dominated economic headlines on July 10, with worries about a large Portuguese bank and falling industrial production figures for France (down 1.7 percent), Germany (down 1.8 percent) and Italy (down 1.2 percent).

Facebook reached 1.35 billion monthly active users and counting. An overwhelming majority of consumers across the United States are dropping traditional landline phone service, according to updated US Telecom analyses

Needless to say, we are really excited to be part of this growing industry.

Fonte: Dados da presente pesquisa

A importância de se abordar sinônimos está no fato de que, é possível que os estudantes se deparem com leituras ao longo da vida e/ou em vestibulares, que não apresente exatamente as palavras *decline* e *rise*, mas sim sinônimos que carregam a mesma ideia. É relevante mostrar também que, esse vocabulário foi abordado de forma contextualizada, ou seja, os alunos precisavam observar as sentenças retiradas do *corpus* para realizar a atividade. A imagem acima apresenta apenas parte do exercício, sua totalidade pode ser vista no Apêndice B.

O segundo exercício de vocabulário foi elaborado por meio de nove agrupamentos lexicais do *corpus Health*, e com essa informação, considerada relevante para o tema, os alunos primeiramente tinham de relacionar os agrupamentos à expectativa de vida ou à causa de morte. Desse modo, os alunos já estabeleciam o contexto em que as palavras eram geralmente

utilizadas. Feito isso, os participantes deviam preencher as lacunas com os mesmos agrupamentos apresentados no exercício anterior, porém, nessa segunda atividade, o contexto já era oferecido ao estudante, como pode-se notar no Apêndice B (exercícios 9 e 10). Aqui, os alunos tinham de inferir determinados significados por meio do contexto que era oferecido a eles.

O exercício 11, também disponível no Apêndice B, trabalhou com o vocabulário abordado no exercício anterior em uma atividade de conexão entre o léxico e o seu contexto. Aqui, os alunos também tiveram a oportunidade de entrar em contato com palavras relacionadas ao léxico abordado. As sentenças apresentadas nesse exercício foram retiradas do *corpus Health*.

No que concerne o *corpus Politics*, outros agrupamentos também foram escolhidos para o desenvolvimento de outras atividades com foco no vocabulário: *trade war*, *nuclear deal*, *trade deal*, *new deal*, e *withdrawal deal* devido o envolvimento desses agrupamentos com o assunto já abordado e dentro do *corpus*. Assim, mesmo motivo ocorrido na atividade de tema *health*, a seleção de alguns dos agrupamentos para determinados exercícios se deu por afinidade com os outros já selecionados.

Imagem 17 - Exemplo de atividade do *corpus Politics* #1

2. Read the sentences below and then fill in the blanks with the expressions in the box:

To watch – was held – was staged – attended the – organized

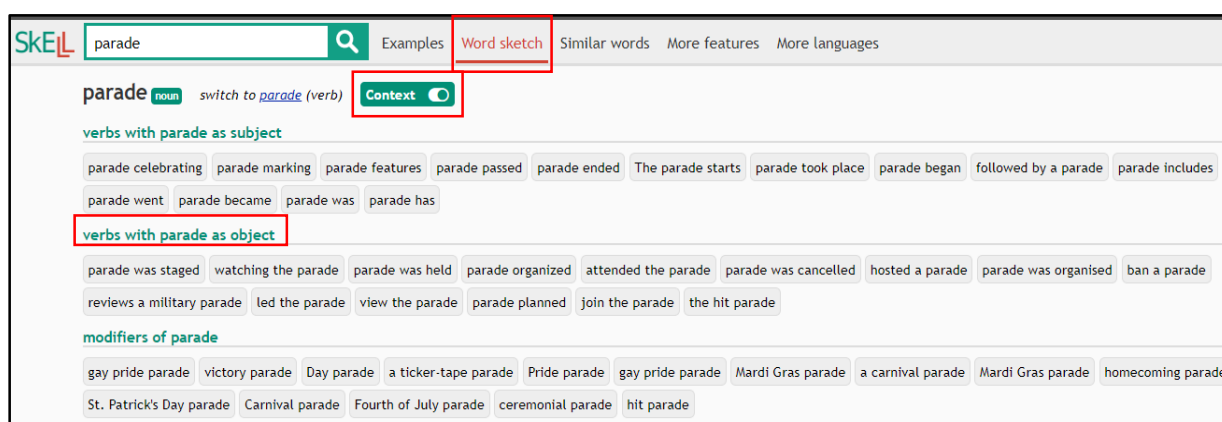
A military parade was staged on the former airfield, involving 270 armored vehicles and 800 troops.
 Kim Jong Un applauds as *he watches a military parade* in Pyongyang on April 15, 2012.
 More than 20,000 people *attended the military parade* organized with the occasion of the National Day.
A military parade was then held in a well-attended.
 More than 100,000 people *attend* notably *the main parade* with over 3,500 performers.

Another military parade _____ across downtown Seoul later in the afternoon.
 In 1987, an estimated 100,000 people came _____ the parade.
 An enormous parade _____ downtown the day of the game.
 Pro-administration officials _____ parades and demonstrations in favor of the tax.
 More than 20,000 people _____ military parade organized with the occasion of the National Day.

Fonte: Dados da presente pesquisa

Na imagem acima é possível ver que foi utilizado o agrupamento *military parade*. A partir da análise desse agrupamento, mais especificamente da palavra *parade* por meio da ferramenta *Word Sketch* do SkELL®, foi notado que os cinco verbos mais utilizados com *parade* (sendo este o objeto), foram *stage*, *watch*, *held*, *organize* e *attend*. Por meio dessa informação, o exercício acima foi desenvolvido. Abaixo é possível verificar a pesquisa feita com a palavra *parade* no SkELL®.

Imagem 18 - Ferramenta *Word Sketch*



Fonte: SkELL®

Um dos fatores interessantes no SkELL®, e que está marcado na imagem acima, é a possibilidade de ativar a função “contexto” e, assim, visualizar um conjunto de palavras, ou seja, ao pesquisar a palavra *parade*, na aba “verbo com *parade* como objeto” aparece a palavra *parade* já seguida dos verbos, como *parade was staged*. Visto que o contexto é um dos pontos essenciais para a elaboração das atividades, esse recurso foi ativado e utilizado no exercício.

O exercício 5 da atividade *Politics*, que pode ser visualizado no Apêndice D, enfatizou a importância dos agrupamentos ao separá-los no exercício para que os alunos pudessem ter a chance de juntá-los. O interessante dessa atividade é que os alunos, muitas vezes conhecem as palavras como unidades separadas, mas não sabem como combiná-las. A partir desses agrupamentos o exercício 6 e 7 foram elaborados, porém trazendo o contexto para o aluno fazer ligações entre os agrupamentos e as frases apresentadas, que também foram retiradas do *corpus Politics*.

Em relação ao *corpus World*, a partir da compilação do texto “*Indonesia accused of mishandling tsunami warnings*”, os agrupamentos extraídos foram *tide gauge*, *warning system*,

geophysics agency, *national disaster agency* e *long narrow bay*. A partir desses agrupamentos foi feita a atividade de *warm-up*, bem como a de vocabulário.

Foi possível desenvolver diferentes exercícios, como com nuvem de palavras, ou até mesmo com linhas de concordância, como aconteceu com a atividade de tema *politics*.

Por motivo de disponibilidade de tempo na aplicação das atividades, esses foram os únicos agrupamentos selecionados para a continuação da elaboração dos exercícios. Todavia, não deve-se pensar que os outros agrupamentos extraídos do *corpus Tech* foram ignorados, uma vez que alguns dos selecionados na compilação do texto “*Indonesia accused of mishandling tsunami warnings*”, também faziam parte do *corpus* de estudo *Tech*.

A atividade abaixo utilizou a nuvem de palavras, já apresentada no subcapítulo “SkELL”, para realizar esse exercício com a palavra *warning*. Todas as palavras encontradas na nuvem são cognatas ou estão, de alguma forma, relacionadas a *warning*, devido ao agrupamento *warning system*. A intenção foi discutir essas palavras a fim de voltar a atenção dos alunos aos sinônimos já que, muitas vezes, esses estão presentes em exames de vestibulares. Assim, os alunos deviam grifar as palavras conhecidas e as cognatas.

Imagem 19 – Exemplo de atividade do *corpus World* #1

2. A nuvem de palavras abaixo está relacionada com warning. Grife as palavras que você conhece e as que são cognatas.



Fonte: Dados retirados do SkELL

No exercício seguinte era pedido para que os estudantes completassem as lacunas com os adjetivos apresentados em uma caixa. Com esses adjetivos eles formavam agrupamentos importantes para o entendimento do assunto *warning system* – sistema de aviso – que também era o tema do texto abordado nas atividades de CE.

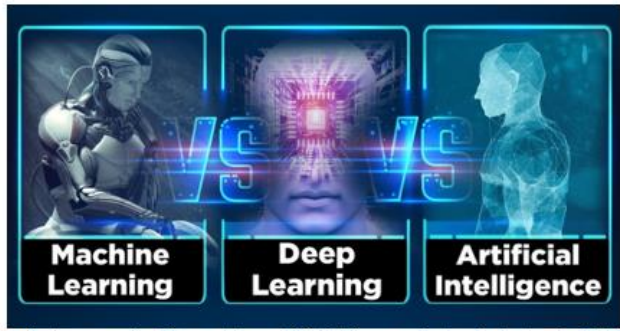
Nesse exercício de lacunas os alunos formavam agrupamentos como *repeated tsunami warning*, *advance warning*, *inadequate warning*, *extensive warning* e *early warning*. Para

chegar nas respostas certas, além do contexto das frases, os alunos também tinham palavras sinônimas para comparar com as suas opções e completar o exercício, como pode ser visto no Apêndice A. Ainda observando o mesmo apêndice, é possível verificar o exercício 4, que apresenta alguns verbos que “combinam”, ou seja, se agrupam com *warning*, e que foram extraídos do *corpus* de estudo *World*. Nesse exercício em específico, algumas frases são apresentadas aos estudantes, como “[...] *warnings were issued*” ou “*it lifted the warning*” e era necessário conectar cada agrupamento desse aos seus significados. Nesse exercício foi utilizado o dicionário online da Macmillan.

Sabendo que as palavras *machine learning*, *deep learning* e *artificial intelligence* foram consideradas chave no entendimento do texto “*Facebook’s top AI scientist says it’s ‘dust’ without artificial intelligence*”, foi feita uma pesquisa para saber a diferença entre essas três “vertentes” da área de Inteligência Artificial (IA). Essa diferença foi encontrada no *website* “*Hackernoon*” a partir das informações do texto “*Top Differences Between Artificial Intelligence, Machine Learning & Deep Learning*”. Essas informações foram utilizadas para a elaboração do exercício abaixo, em que é pedido para o aluno observar a imagem e tentar responder a diferença entre *Artificial Intelligence*, *Machine Learning* e *Deep Learning*.

Imagem 20 – Exemplo de atividade do *corpus Tech #1*

c. Observe a figura abaixo. Qual a diferença entre *machine learning*, *deep learning* e *artificial intelligence*?



<https://hackernoon.com/top-differences-between-artificial-intelligence-machine-learning-deep-learning-d39cb6f6feaa>

d. Ligue as expressões 1. *machine learning*, 2. *deep learning* e 3. *artificial intelligence* ao seus significados:

() The theory and development of computer systems that able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making and translation between languages.

() is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to “learn” (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

() is a subset of machine learning in Artificial Intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

Fonte: Dados da presente pesquisa (informações retiradas do *website* “*Hackernoon*”).

Após essa discussão, foi pedido para que os alunos ligassem os agrupamentos aos seus significados, também retirados do *website* citado acima.

No segundo exercício, também foi utilizado imagem, a fim de que os alunos ligassem os agrupamentos *artificial intelligence*, *generative design*, *machine learning* e *face recognition* com as figuras que representavam essas expressões. Esses agrupamentos foram retirados do *corpus Tech*, e não do *corpus* compilado apenas com o texto, e, seguindo o padrão das outras atividades, a seleção desse vocabulário se deu por meio da observação da maior chavicidade e frequência relativa dos agrupamentos, como pode ser visto no Apêndice G.

Tendo em vista a importância do contexto, enfatizado também nas outras atividades, o exercício 2 pedia aos alunos para que eles completassem os espaços dos trechos oferecidos na atividade com os agrupamentos *artificial intelligence*, *generative design*, *machine learning* e *face recognition*, como pode-se verificar no Apêndice C.

4.2.3 Exercícios de compreensão escrita

Neste subcapítulo não serão descritos todos os exercícios de CE elaborados nas atividades dos quatro diferentes temas, já que essas podem ser visualizadas nos apêndices, e, diferentemente dos exercícios de vocabulário que foram desenvolvidos de forma diversificada, os de CE foram baseados no texto escolhido (como descrito no subcapítulo “Análise lexical – extração das palavras-chave, agrupamentos lexicais e escolha textual”, e seguem o padrão do Quadro 6.

No Quadro 6 é possível ver que na seção de atividades de CE, há a obrigatoriedade de seguir dois passos: o desenvolvimento de questões mais gerais, que vão exigir do aluno o uso da técnica *skimming* e questões mais específicas do texto, que eles terão de utilizar o *scanning*, levando em consideração as técnicas descritas como relevantes no processo de CE (KLEIMAN, 1985; BROWN, 2001; FARELL, 2003; SANTOS, 2011),

Tomando como exemplo o tema *Health*, foram feitas questões do tipo “Segundo o estudo realizado” e “*Where did life expectancy decline?*”. Na primeira questão, o aluno tem 4 alternativas para responder a respeito do estudo realizado naquele texto. Aqui, o estudante não precisa analisar uma informação específica do texto para saber responder, mas sim analisar o texto de uma forma geral, como se estivesse perguntando acerca da ideia geral do texto, assim, a técnica utilizada deve ser o *skimming*. Já na segunda questão, o estudante também se depara com 4 alternativas, porém, dessa vez, ele tem de utilizar a técnica *scanning* em busca da resposta correta, já que ele vai ter que buscar por palavras-chave como *life expectancy* e *decline* para

encontrar o(s) local(is) específico(s) onde a expectativa de vida caiu. Do mesmo modo, nos outros textos também foram abordadas questões do mesmo tipo – de cunho geral e específico.

Dessa forma, analisando e adequando os quatro fatores apresentados no Quadro 1, pode-se afirmar que, em relação ao primeiro tópico *adequação do conteúdo*, os textos abordados nas atividades são relevantes para as necessidades dos vestibulandos, visto que textos jornalísticos são encontrados nos exames que os alunos prestam, bem como fora do contexto escolar, uma vez que se deparam com esses tipos de textos na leitura de jornais online, sendo, dessa forma, conteúdo autêntico. Em relação ao interesse do aluno acerca dos temas tratados, pode-se notar, por meio dos questionários (que podem ser encontrados na metodologia) respondidos pelos próprios estudantes, que o assunto foi de interesse dos mesmos.

O segundo tópico, *explorabilidade*, é abordado em todas as atividades, já que os textos são explorados para fins de ensino, desenvolvendo a habilidade de CE na LI, tendo seu conteúdo explorado para ensinar os agrupamentos lexicais encontrados por meio de palavras-chave que são de grande relevância para o entendimento do texto.

O tópico *legibilidade* foi levado em consideração nos questionários semanais entregues aos alunos, com a preocupação de que o texto não fosse estruturalmente muito complexo e que apresentasse vocabulário relevante para o ensino de determinado tema. Assim, foi questionado aos alunos acerca da dificuldade de leitura em relação a cada texto exposto, bem como se as atividades de vocabulário (que foi extraído dos próprios textos, ou dos *corpora*) estavam os auxiliando na CE. O vocabulário novo é abordado tanto em forma de exercícios quanto, posteriormente, na leitura do texto.

O quarto e último fator é a *apresentação*, que, segundo o autor, também se faz muito importante na escolha do material. Todas as atividades desenvolvidas para essa pesquisa tiveram o *layout* do texto copiado e colado no exercício, exatamente para parecer o mais autêntico possível, bem como atraente, já que os alunos olham para a atividade e têm certeza que o texto foi retirado de um contexto autêntico de leitura, um site jornalístico famoso, BBC® ou CNN®.

Um ponto muito relevante a ser mencionado é que, como visto, as atividades de tema *Health* e *Politics* tiveram duas seções de CE, ou seja, os alunos tiveram a oportunidade de fazer exercícios relacionados a dois diferentes textos. No que concerne o primeiro tema, a escolha do segundo texto “*Low and high carb diets increase risk of early death, study finds*” se deu pelo fato de que este também estava entre os 5 que continham o agrupamento principal, como pode ser visto na imagem 9 e também por conter outros agrupamentos que já haviam sido abordados nos exercícios de vocabulário, como *low-carb diet*.

Já no que diz respeito o tema *Politics*, a escolha do segundo texto “*Donalds Trump says new trade deal is most importante ever*” se deu pelo fato de que muitos agrupamentos já abordados nas atividades de vocabulário, ou léxico relacionado a eles estavam presentes nesse texto, como *new deal* e *trade deal*. Os dois temas restantes, *World* e *Tech* tiveram apenas um texto abordado ao longo de toda a atividade.

É importante enfatizar que os exercícios apresentados ao longo dessa seção e todos os outros contidos nas atividades aplicadas em sala de aula, são do tipo *hands-off*, como foi explicado anteriormente. Dessa maneira, como explanado na fundamentação teórica desta pesquisa, os exercícios foram desenvolvidos para estarem mais próximos possível da realidade do aluno, utilizando, até mesmo, palavras com que eles já estão familiarizados, sem o uso de um vocabulário mais técnico.

4.3 Análise dos questionários

Neste subcapítulo serão abordados os resultados da presente pesquisa alcançados até o momento. Para isso, os questionários semanais que foram entregues aos alunos a cada fim de atividade foram analisados, bem como o questionário final enviado aos alunos de forma online por meio da plataforma *SurveyMonkey*.

4.3.1 Análise dos questionários semanais

Os questionários semanais foram entregues ao final de cada atividade, ou seja, quando todos os alunos terminavam de realizar os exercícios de determinado tema, eles respondiam o questionário. Este contém três perguntas: 1. O que você achou mais interessante na aula de hoje? 2. A dificuldade das atividades (aqui, os participantes tinham de escolher qualquer número entre 1 e 5, sendo 1 “muito fácil” e 5 “muito difícil”) 3. As atividades de vocabulário te ajudaram a compreender o texto? (essa questão continha 3 alternativas: sim, em partes e não).

Por uma questão de praticidade e organização, as respostas dos alunos foram organizadas em tabelas, considerando as respostas de todas aulas juntas. Isso se deve pelo fato de que o interesse deste projeto não era saber as respostas dos alunos relacionadas às áreas separadamente, mas o impacto das atividades com *corpora* visto de maneira mais generalizada.

Temos, assim, três tabelas, cada uma referente às perguntas 1, 2 e 3 respectivamente. É relevante mencionar que alguns alunos, geralmente, têm mais de uma resposta para a primeira pergunta, por esse motivo o número de respostas excede o número de alunos. Abaixo, é possível

visualizar a tabela referente à pergunta número 1: O que você achou mais interessante na aula de hoje?

Tabela 5 – Pergunta #1 O que você achou mais interessante na aula de hoje?

<i>Total de respostas</i>	<i>Respostas</i>
18	Compreensão do texto / técnicas de leitura
34	Novo vocabulário / atividades de vocabulário
37	Tema / os textos

Fonte: Dados da presente pesquisa

Analisando a tabela acima, pode-se notar que há 18 respostas que mencionam a compreensão do texto ou as técnicas de leitura. Nessas últimas inclui-se as respostas acerca da utilização de cognatos, bem como *skimming* e *scanning*. Alguns alunos responderam a essa pergunta com “a busca de informações no texto”, e esse tipo de resposta foi considerada como técnica de leitura, já que se subentende que, de um modo mais simples, o aluno pretendia dizer que aprendeu e/ou utilizou as técnicas *skimming* e *scanning*.

Faz-se relevante mencionar que o fato de se utilizar agrupamentos lexicais e/ou palavras-chave na compreensão dos textos, foi considerado dentro das técnicas de leitura e não separadamente na análise. Isso se deve pelo fato de que, apesar de todos os alunos terem utilizado essas técnicas, já que as mesmas foram parte central das atividades de vocabulário, nem todos as colocaram nas respostas, exceto dois alunos que enfatizaram o fato de terem utilizados essas técnicas.

Por meio da análise desta tabela, é possível notar que, os alunos passaram a tirar o foco das técnicas de leitura em suas respostas, mesmo elas tendo sido abordadas em sala, como no primeiro dia. Respostas como “conhecimento na área de tecnologia” ou ainda “o diálogo e o conhecimento da tecnologia de IA” foram consideradas como resposta para “Tema/os textos”. A análise dessa pergunta em específico reflete no questionário final, em que a maioria dos participantes escolhem o tema *tech* como preferido.

No que diz respeito à segunda pergunta, referente à dificuldade das atividades, encontra-se a tabela abaixo:

Tabela 6 - Pergunta #2 – Qual a dificuldade das atividades?

<i>Total de respostas</i>	<i>Nível de dificuldade</i>
10	1 – muito fácil
41	2 – fácil
24	3 – intermediário
3	4 – difícil
0	5 – muito difícil

Fonte: Dados da presente pesquisa

Pode-se notar que as respostas dos alunos se concentraram entre “fácil” e “intermediário”, o que é considerado positivo para a pesquisa, visto que isso pode mostrar a facilidade que os alunos estão tendo em lidar com as questões de LE, ou que os exercícios de vocabulário os auxiliaram a responder a parte de CE. Se muitos alunos tivessem optado por “muito fácil” e/ou “muito difícil”, os dados mostrariam o contrário – ou que os exercícios não auxiliaram na CE, ou ainda que as atividades realmente não alcançavam o nível deles, sendo muito fácil ou muito difícil para a compreensão dos estudantes, o que nos retoma um ponto de extrema importância aqui: o nível dos alunos.

Os estudantes apresentaram níveis variados de inglês, assim, as questões não podiam ser fáceis demais e nem muito difíceis, pois estaríamos lidando com as extremidades, o que poderia prejudicar os alunos, já que eles não entenderiam (quase) nada ou iriam entender tudo, sem obter nenhum tipo de evolução.

Apesar de alguns alunos terem escolhido os níveis “muito fácil” e “muito difícil” no que diz respeito às atividades, voltamos ao ponto em que não há esse controle sobre uma sala heterogênea em relação ao nível da língua. Do mesmo modo que alguns alunos tinham contato com a LI fora de sala (em cursos, por meio de mídias, etc.) e, assim, apresentavam um nível mais elevado em inglês, outros estudantes não tinham a mesma realidade e acharam as atividades difíceis. Como dito, o objetivo era que as atividades tivessem, para os alunos, um nível pelo menos intermediário.

Em relação à terceira questão, “As atividades de vocabulário te ajudaram a compreender o texto?”, temos:

Tabela 7 – Pergunta #3 As atividades de vocabulário te ajudaram a compreender o texto?

<i>Total de respostas</i>	<i>Alternativas</i>
49	Sim
13	Em partes
0	Não

Fonte: Dados da presente pesquisa

Desse modo, fazendo uma análise geral dos questionários semanais, pode-se considerar que a maioria dos alunos acreditam terem adquirido e/ou entrado em contato com novas palavras, visto que, considerando as respostas dos alunos no primeiro questionário, pelo menos metade dos participantes responderam que o fator que acharam mais interessante na aula foi o novo vocabulário apresentado ou as atividades de vocabulário. Também é possível assegurar que os temas foram de interesse dos alunos, já que, pela análise, grande parte respondeu que gostaram do tema ou a discussão acerca do mesmo.

O grau de dificuldade das atividades ficou entre fácil e intermediário, o que pode ser considerado positivo pela perspectiva de que as atividades baseadas em *corpora* facilitam a compreensão textual, fato confirmado pela análise da terceira pergunta do questionário, em que nenhum aluno assinalou uma resposta negativa, ou seja, os exercícios de vocabulário ajudaram pelo menos em partes na compreensão do texto.

Claro que, deve-se levar em consideração os alunos que acharam as atividades fáceis, pois já conheciam aquele determinado vocabulário e/ou já tinha um nível mais avançado de leitura, porém, ainda que isso tenha acontecido em alguns casos, não se exclui a relevância das atividades com *corpora* para a compreensão do texto, visto que essas são facilitadoras nesse processo. Em outras palavras, mesmo que, por exemplo, um aluno já soubesse o significado de todas as palavras-chave que foram apresentadas no *warm-up*, o fato de estar consciente que essas palavras são chave no texto em que irá ler em seguida já facilita a sua leitura e compreensão.

Para um resultado mais concreto, será feita, no próximo subcapítulo, a análise do questionário final.

4.3.2 Análise do questionário final

O questionário final foi aplicado aos alunos de modo que eles pudessem refletir acerca das atividades aplicadas durante as 4 aulas em que a pesquisa estava sendo realizada. Desse modo, foi enviado o link da pesquisa para todos os participantes responderem a cinco perguntas: 1. Qual tema você mais gostou? 2. De alguma forma, as aulas te ajudaram a expandir seu vocabulário? 3. Você acredita que, de alguma forma, essas aulas irão te ajudar na resolução de próximos vestibular? 4. O que você achou de trabalhar com textos autênticos, ou seja, que são retirados do contexto real de uso da língua? No total, 24 alunos responderam o questionário.

Como resultado para primeira pergunta, a maioria dos alunos responderam que gostaram mais das atividades que abordaram o tema *Tech*, totalizando 41,67% dos participantes. Com 25%, o tema *Health* ficou em segundo lugar como preferido entre os alunos, seguido dos temas *Politics* e *World*, que totalizaram 16,67%. Abaixo é possível visualizar a tabela gerada pela própria plataforma *Survey monkey* que mostra a preferência dos alunos:

Tabela 8 - Pergunta #1 do questionário final

<i>Opções de respostas</i>	<i>Total de alunos</i>
<i>Health – saúde</i>	25%
<i>Politics – política</i>	16,67%
<i>Tech – tecnologia</i>	41,67%
<i>World - mundo</i>	16,67%

Fonte: Dados da presente pesquisa

Em relação à segunda questão do questionário, “De alguma forma, as aulas te ajudaram a expandir seu vocabulário”? 17 estudantes responderam que sim, as atividades ajudaram a expandir seu vocabulário de alguma forma. Apenas 5 dos respondentes afirmaram que teve uma expansão lexical parcial, enquanto 2 alunos responderam negativamente.

Acredita-se que esses dois alunos que responderam que não houve expansão em seu vocabulário seja o caso já explicado anteriormente, em que, talvez o aluno realmente já conhecia as palavras apresentadas nas atividades e/ou no texto, porém isso não diminui a relevância da ênfase nesse vocabulário para a compreensão textual. Vale lembrar que, mesmo nesses casos, a LC teve sua contribuição, já que as atividades só puderam ser desenvolvidas com mais precisão devido o *Sketch Engine*®. Abaixo temos a tabela relacionada à segunda pergunta:

Tabela 9 - Pergunta #2 do questionário final

<i>Opções de respostas</i>	<i>Total de alunos</i>
<i>Sim</i>	17
<i>Em partes</i>	5
<i>Não</i>	2
<i>Total</i>	24

Fonte: Dados dos *corpora* da presente pesquisa

Referente à terceira questão, “Você acredita que, de alguma forma, essas aulas irão te ajudar na resolução de próximos vestibular”?, não houve respostas negativas. 14 dos alunos responderam que as aulas ajudarão em vestibulares que irão prestar futuramente, enquanto 10 respondentes afirmaram que as aulas auxiliarão em partes nos testes futuros. Abaixo podemos visualizar a tabela referente à terceira pergunta do questionário:

Tabela 10 – Pergunta 3 do questionário final

<i>Opções de respostas</i>	<i>Total de alunos</i>
<i>Sim</i>	14
<i>Em partes</i>	10
<i>Não</i>	0
<i>Total</i>	24

Fonte: Dados dos *corpora* da presente pesquisa

No que diz respeito à autenticidade dos textos abordados nas atividades, 14 dos alunos achou muito bom, 8 consideraram bom, e 2 alunos afirmaram não ver diferenças nessa abordagem, como é possível verificar na tabela abaixo:

Tabela 11 - Pergunta 3 do questionário final

<i>Opções de respostas</i>	<i>Total de alunos</i>
<i>Muito bom</i>	14
<i>Bom</i>	8
<i>Não percebi diferenças</i>	2
<i>Não gostei</i>	0
<i>Total</i>	24

Fonte: Dados dos *corpora* da presente pesquisa

No início da primeira aula foi explicada a diferença entre um texto autêntico e um texto não autêntico, bem como os benefícios de se estar em contato com a língua real em uso. Como visto, a maioria dos estudantes tiveram uma recepção positiva em relação à questão da autenticidade, sendo que 22 alunos responderam entre “muito bom” e “bom”, reconhecendo a relevância dessa abordagem. Além disso, também foi dito em sala de aula que os textos

utilizados nos vestibulares também são autênticos, sejam eles notícias jornalísticas ou qualquer outro gênero textual, como música, poema entre outros.

Dessa forma, por meio dos resultados até aqui analisados, foi possível responder às duas perguntas de pesquisa e o alinhar a pesquisa ao seu objetivo inicial. As conclusões obtidas por meio das análises até aqui realizadas serão discutidas no próximo capítulo, bem como os encaminhamentos futuros da presente pesquisa.

5. CONCLUSÕES E ENCAMINHAMENTOS

Neste capítulo será discorrido acerca das conclusões desta pesquisa, levando em consideração as suas limitações e encaminhamentos de novas etapas para o futuro deste estudo. Assim, serão discutidos tanto os obstáculos que apareceram pelo percurso desta dissertação, apresentando alguns detalhes que são relevantes para o entendimento desse processo, bem como as conclusões que foram obtidas por meio das análises aqui feitas. Para finalizar, algumas ideias serão descritas como uma nova etapa para um trabalho de continuação desta pesquisa, que possivelmente será desenvolvido em um futuro Doutorado.

5.1 Limitações do estudo

Alguns pontos são necessários serem descritos aqui para apresentar as limitações encontradas durante todo o percurso pelo qual a presente pesquisa passou, que, de certo modo, justificam determinados dados obtidos na análise feita por este estudo.

Como já mencionado, esta pesquisa foi realizada em uma escola privada de ensino regular, durante as aulas de LI, e uma das primeiras considerações a serem feitas é em relação à falta dos alunos. No período de 8 aulas, sendo elas duplas, totalizando 4 encontros, houve uma variação significativa de presença dos alunos. Em outras palavras, não havia o mesmo número de alunos no primeiro, segundo, terceiro e quarto (e último) dia de aula.

Dessa maneira, não foi o mesmo total de alunos de assistiu às 8 aulas, podendo haver determinada influência na análise dos questionários, por exemplo. Assim, suponha-se que um aluno tenha faltado 2 dias (4 aulas), esse estudante não tem a mesma segurança para responder ao questionário do que um colega de classe que não teve falta alguma.

Por outro lado, outra decorrência no que concerne o processo desta pesquisa, se dá no total de aulas e atividades aplicadas. Já foi explicado que a quantidade de exercícios por atividade (que foram 4), variou. Enquanto as atividades de temas *Health* e *Politics* contaram com mais exercícios, abordando dois textos de cada tema, as atividades *Tech* e *World* abordaram apenas 1 texto e tiveram menos exercícios. Essa escolha foi feita, pois na primeira aula, os objetivos da pesquisa foram apresentados aos alunos, foi feita uma discussão acerca de alguns pontos importantes, como o material autêntico era utilizado em cada atividade, a estatística apresentada pelo computador na escolha das palavras-chave, bem como os alunos recolheram um documento de permissão de participação das aulas, o TCLE (Termo de

Consentimento Livre e Esclarecido aos Pais), por serem menores, e o TALE (Termo de Assentimento Livre e Esclarecido), que os estudantes mesmos assentiam sua participação. Assim, faz-se importante mencionar que o C.A.A.E. (Certificado de Apresentação para Apreciação Ética) da presente pesquisa é 01631718.8.0000.5466.

Por essa delonga necessária na primeira aula, a professora/pesquisadora teve menos tempo para desenvolver as atividades, sendo que a primeira aplicada com o tema *Health*, acabou durando mais tempo que o previsto. Assim, as atividades dos dois últimos temas tiveram de ser mais breves para caberem no prazo disponível de aplicação. Apesar de os alunos terem tido a oportunidade de entrar em contato com diversos agrupamentos lexicais de cada tema, é provável que os alunos teriam consolidado mais o conteúdo, caso tivessem tido mais contato com esses mesmos agrupamentos ao longo de mais atividades, por um tempo mais longo.

Outro ponto importante a ser considerado como uma das limitações desta pesquisa, está no fato de que o nível de dificuldade dos exercícios elaborados pela professora/pesquisadora não é exatamente igual ao dos vestibulares que os alunos prestam. Assim, um(a) aluno(a) deu a sugestão de, em próximas atividades, aumentar o grau de dificuldades dos exercícios de CE. Entretanto, voltamos a um fator também muito relevante: além de os alunos terem diferentes níveis de inglês na sala de aula, o objetivo principal desta pesquisa era abordar os agrupamentos de temas atuais de modo que os alunos tivessem contato com um vocabulário atual e de maneira contextualizada.

Desse modo, com as técnicas de leitura ensinadas e com os agrupamentos lexicais apresentados, os alunos tiveram a oportunidade de aprimorar sua CE em textos na LI. Ainda que o nível de dificuldade dos exercícios aplicados seja diferente dos exercícios encontrados no vestibular, os estudantes puderam entrar em contato com um novo vocabulário e desenvolver suas técnicas no momento da leitura.

5.2. Conclusões e Encaminhamentos

Por meio da análise da elaboração das atividades e também dos questionários entregues aos alunos, tanto os semanais quanto o questionário final, é possível afirmar que o objetivo inicial desta pesquisa que era “avaliar a aplicabilidade do levantamento de agrupamentos lexicais para o desenvolvimento de atividades de CE em LI a fim de contribuir para a identificação lexical de notícias da atualidade” foi obtido graças à união do IFE com o suporte teórico-metodológico da LC.

É possível notar que, pelo fato de os *corpora* estarem de acordo com o objetivo dos alunos e da pesquisa, os resultados foram positivos. Os agrupamentos lexicais extraídos por meio de cada *corpus* possibilitaram a elaboração de atividades de CE em LI mostrando sua relevância no desenvolvimento dos exercícios, visto que esses só eram feitos por meio da análise da frequência relativa de cada agrupamento.

Deve-se mencionar que a seleção de cada texto para os exercícios de CE também foi estatisticamente pensada, já que, os *corpora* são compostos por 66 textos, e, muitas vezes, mais de um texto possuía o agrupamento lexical escolhido para as atividades de CE. Dessa maneira, a análise estatística do *corpus* permitiu dados mais precisos para a elaboração de cada atividade desta pesquisa, desde a extração dos agrupamentos lexicais e das palavras-chave, quanto da seleção dos textos para os exercícios de CE.

Foi possível, portanto, contribuir, em certa medida, para a identificação lexical de notícias da atualidade. Afirma-se, aqui, que é em certa medida, pois, a compilação dos textos deu-se num determinado período de tempo e certa quantidade – 3 últimos meses de 2018 e 3 textos por semana, não sendo possível abranger todas as notícias dos temas abordados. Assim, o léxico extraído foi, sim, relevante para a identificação de notícias da atualidade, porém, para o contexto de compilação que foi realizado. Para uma pesquisa mais abrangente, faz-se necessário que os *corpora* sejam atualizados e, conseqüentemente aumentados, para se obter ainda mais agrupamentos que sejam relevantes para a identificação do tema em específico.

No que concerne às perguntas de pesquisa, pode-se afirmar que, em relação à pergunta 1 “Em que medida o uso de um *corpus* criado a partir de textos jornalísticos autênticos pode auxiliar no levantamento de agrupamentos lexicais para o vestibular?”, os *corpora* compilados podem auxiliar no levantamento de agrupamentos lexicais para o vestibular, na medida em que esse conteúdo é relevante para os alunos, visto que textos jornalísticos estão presentes nos vestibulares e, além disso, o conteúdo compilado era atual, ou seja, está em consonância com a proposta desta pesquisa: auxiliar na identificação lexical de notícias. Caso a pesquisa tivesse objetivo e público diferentes, os *corpora* compilados aqui não seriam úteis.

É desse modo que se obtém a resposta para a segunda pergunta da pesquisa “Em que medida as atividades baseadas em agrupamentos lexicais podem desenvolver o léxico dos alunos e serem relevantes para o contexto dos vestibulandos?”. Os agrupamentos lexicais possibilitam o desenvolvimento do léxico e são relevantes para o contexto dos alunos que prestam o vestibular na medida em que o *corpus* está em consonância com o que é exigido desses estudantes no momento do exame. Caso os *corpora* fossem de textos técnicos de uma

área específica, como medicina, por exemplo, os alunos até desenvolveriam o léxico deles, mas não necessariamente seria relevante para o contexto em que se encontram.

Ademais, como foi teoricamente explicado, os agrupamentos tendem a agir de forma mais efetiva no desenvolvimento lexical do estudante quando comparados às listas de palavras soltas, além do fato que os agrupamentos foram estatisticamente selecionados, ou seja, a frequência foi um ponto crucial no desenvolvimento dessas atividades. Não menos importante, deve-se levar em consideração que todo o processo de CE por meio dos agrupamentos levados aos alunos, foi fundado teoricamente em técnicas que auxiliam os estudantes a interpretar os textos com mais eficiência.

Como encaminhamento futuro, pretende-se, a partir desta pesquisa e dos resultados aqui encontrados, expandir o desenvolvimento de atividades com outros diversos temas, além dos aqui trabalhados, mantendo o foco na habilidade de CE da LI com base em *corpora*.

Busca-se, além da ampliação dos *corpora* e dos temas, também propor uma sequência pedagógica mais profundamente elaborada, com base nos estudos de Dolz, Noverraz e Schneuwly (2004) voltada especificamente para o ensino de CE em LI baseada em *corpora*. Será proposto, assim, a outros profissionais do ensino de inglês, como elaborar exercícios de CE voltados para um público específico (IFE) com materiais que sejam autênticos e atuais (*corpora* recicláveis) bem como com o uso de agrupamentos lexicais de determinada área (LC) por meio de uma sequência didática pronta a ser seguida e, se possível, adaptada a diversos contextos e línguas. Espera-se que esse ensino seja mediado por uma plataforma on-line como um MOOC, por exemplo, ou aulas via videoconferência, em que tenham a participação de professores de língua inglesa.

Resumidamente, objetiva-se ampliar a presente pesquisa tendo três bases teórico-metodológicas – o IFE, a LC e a sequência didática, que juntos serão o aparato para o desenvolvimento de atividades de CE que sejam aplicáveis em um contexto específico e não abra mão de textos autênticos – que serão compilados e analisados por meio de *corpora*. Portanto, acredita-se que, dessa maneira, tanto professor quanto aluno sejam beneficiados no processo de ensino-aprendizagem da língua, visto que farão uso de atividades embasadas teoricamente em sala de aula, com conteúdo autêntico, atual e relevante para o contexto do aluno.

REFERÊNCIAS

ALVES, A. L. L.; PINTO, P. T. *A utilização de that-clauses em abstracts escritos por alunos-pesquisadores brasileiros*. Revista Entrepalavras, v. 8, p. 288, 2018.

BARTHOLOMEU, M. A. A. N. *Prova de língua estrangeira (Inglês) dos vestibulares e sua influência nas percepções, atitudes e motivações de alunos do 3º ano do ensino médio*. 2002. 119 p. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. Disponível em: <<http://www.repositorio.unicamp.br/handle/REPOSIP/269565>>. Acesso em: 01 set. 2019.

BERARDO, S. A. *The use of authentic materials in the teaching of reading*. The reading Matrix, v.6 (2), p. 60-69, 2006.

BERBER SARDINHA, T. *Pesquisa em linguística de Corpus com WordSmith Tools*. Campinas, SP: Mercado de Letras, 2009.

BERBER SARDINHA, T. Como usar a Linguística de Corpus no ensino de língua estrangeira. Por uma Linguística de Corpus educacional brasileira. In: VIANA, Vander.; TAGNIN, Stella. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011.

BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Editora Manole, 2004.

BERBER SARDINHA, T. *Linguística de corpus: histórico e problemática*. Revista D.E.L.T.A., São Paulo, v. 16, n. 2, p. 323-367, 2000.

BERNARDINI, S. Corpora in the Classroom: an Overview and Some Reflections on Future Development. In: SINCLAIR, J. M. (Ed.) *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: Benjamins, p.15-36, 2004.

BIBER, D. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In: *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, p.159-192, 2012.

BIBER, D. *Representativeness in Corpus design*. Literary and Linguistic Computing, vol. 8, n. 4, 1993.

BOULTON, A. Hands-on / hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 152-168, 2012.

BOULTON, A. Data-driven learning: on paper, in practice. In: T. Harris M. Moreno Jaén. *Corpora in Language Teaching*, p. XX, 2009.

BOWKER, L. *Computer-aided Translation Technology: A Practical Introduction*. University of Ottawa Press: Ottawa, 2002.

BRASIL, *Parâmetros Curriculares Nacionais (PCN)*. Linguagens, Códigos e suas tecnologias. Brasília: MEC, 2000.

BRASIL, *Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais (PCN+)*. Linguagens, Códigos e suas tecnologias. Brasília: MEC, 2006.

BROWN, D. H. *Teaching by principles: and interactive approach to language pedagogy*. Second edition. New York: Longman, 2001.

CARTER, R.; MCCARTHY, M. J. 'Size Isn't Everything: Spoken English, Corpus and the Classroom', *TESOL Quarterly* v. 35(2), p. 337-40, 2001.

CARTER, R.; NUNAN, D. (Eds.) *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press, 2001.

CARVALHO, S. N. O enfoque instrumental na leitura. *SOLETRAS*, Ano V, n. 10. São Gonçalo: UERJ, jul./dez.2005.

CELANI, M. A. A. Considerações sobre a pesquisa: a necessidade e eficiência do ensino de inglês instrumental em Universidades brasileiras. *The ESspecialist*, v. 6, p. 2-9, 1983.

CELANI, M. A. A. Introduction. In: CELANI, M. A. A. et al. *ESP in Brazil: 25 years of evolution and reflection*. Campinas: Mercado de Letras, 2005.

CHAMBERS, A. What is data-driven learning? In: O'KEEFFE, A.; MCCARTHY, M. *The Routledge Handbook of Corpus Linguistics*. Routledge: New York, p. 345-358, 2010.

CHARLES, M. Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, p. 93–102, 2012.

COXHEAD, A. Vocabulary and ESP. In: PALTRIDGE, B.; STARFIELD, S. *The handbook of English for Specific Purpose*. p. 115-132, 2013.

CRYSTAL, D. *English as a Global Language* (Second edition). Cambridge: Cambridge University Press, 2003.

DELFINO, M. C. N. *O uso de música para o ensino de inglês como língua estrangeira em um ambiente baseado em corpus*. Orientador: Antônio Berber Sardinha. 2016. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo – PUC. São Paulo, 2016.

DOLZ, J.; NOVERRAZ, M.; SCHNEUWLY, B. Sequências didáticas para o oral e para o escrito: apresentação de um procedimento. In.: SCHNEUWLY, B.; DOLZ, J. *Gêneros orais e escritos na escola*. [Tradução e organização Roxane Rojo e Glais Sales Cordeiro] Campinas, SP: Mercado de Letras, p. 95-128, 2004.

DUDLEY-EVANS, T.; JOHN, M. J. S. T. *Developments in English for specific purposes: a multi-disciplinary approach*. Cambridge: Cambridge University Press, 1998.

DUTRA D.; SILERO, R. P. Descobertas linguísticas para pesquisadores e aprendizes: a linguística de corpus e o ensino de gramática. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 10, n. 4, p. 909-930, 2010.

DUTRA, D. P.; BERBER SARDINHA, T. Pacotes lexicais em corpora de aprendizes. In: IBANÔS, et al. *Pesquisas e perspectivas em linguística de corpus*. Campinas: Mercado de Letras, p. 57-80, 2014.

FARRELL, T. S. C. *Planejamento de atividades de leitura para aulas de idiomas*. São Paulo: SBS, 2003.

FRANKENBERG-GARCIA, A.; BOCORNY, A.E.P.; TAVARES-PINTO, P.; SARMENTO, S. (2019) *Supporting the Internationalization of Brazilian Research*. Workshops delivered at the Federal University of Rio Grande do Sul and at São Paulo State University, Porto Alegre and São José do Rio Preto, April-June 2019.

FREIRE, P. Desafios da educação de adultos ante a nova reestruturação tecnológica. In: *Pedagogia da Indignação: cartas pedagógicas e outros escritos*. São Paulo: UNESP, 2003.

Arquivo PDF. Disponível em:

http://comunidades.mda.gov.br/portal/saf/arquivos/view/ater/livros/Pedagogia_da_Indigna%C3%A7%C3%A3o.pdf. Acesso em: 10 de novembro de 2019.

GABRIELATOS, C. 2005. Corpora and language teaching: Just a fling or wedding bells? Teaching English as a Second Language. *Electronic Journal* v. 8(4). [on-line]
<http://teslej.org/ej32/a1.html>, acesso em: 08 jul 2019.

GAVIOLI, L. *Exploring Corpora for ESP Learning: Studies in Corpus Linguistics*. Amsterdam: John Benjamins, 2005.

GIRARD, M.; SIONIS, C. *The Functions of Formulaic Speech in the L2 Class*, 2004.

GREAVES, C.; WARREN, M. What can a corpus tell us about multi-word units? In: O'KEEFFE, A; MCCARTHY, M. (Eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, p. 255-269, 2010.

GRELLET, F. *Developing Reading Skills: A Practical Guide to Reading Comprehension Exercises*. Cambridge: Cambridge University Press, 1981.

HARMER, J. *The Practice of English Language Teaching*. London: Longman, 1994.

HIRVELA, A. ESP and Reading. In: PALTRIDGE, B; STARFIELD, S. (Ed.). *The handbook of English for specific purposes*. West-Sussex: Wiley-blackwell, 2013.

HOLMES, J. What do you mean by ESP? Working Papers. *CEPRIL*, n. 2. São Paulo: PUC-SP, 1981.

HUTCHINSON, T.; WATERS, A. *English for Specific Purposes: a learning-centred approach*. Cambridge University Press, 1987.

HYLAND, K. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, v. 18, p. 41-62, 2008.

JENKINS, J. English as a lingua franca: interpretations and attitudes. In: *World Englishes*, v. 28, n. 2, p. 200-207, 2009.

JOHNS, T. F. Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal*, v. 4, p.1–16, 1988.

JOHNS, T.; DAVIES, F. Text as a vehicle for information: the classroom use of written texts in teaching reading in a foreign language. *Reading in a Foreign Language*, v.1, p. 1-19, 1983.

JORDAN, R. R. *English for academic purposes: a guide and resource book for teachers*. Cambridge: Cambridge University Press, 1997.

KATO, M. *O aprendizado da leitura*. São Paulo: Martins Fontes, 1985.

KILGARRIFF, A., et al. *The Sketch Engine: ten years on*. Lexicography, p. 7-36, 2014.

KILICKAYA, F. (2004). *Authentic Materials and Cultural Content in EFL Classrooms*. Disponível em < <http://iteslj.org/Techniques/Kilickaya-AuthenticMaterial.html> > acesso em 20 abr 2019.

KLEIMAN, A. B. *Estratégias de inferência lexical na leitura de segunda língua*. Ilha do Desterro, p.67-82, 1985.

KOESTER, A. Building small specialised corpora. In: O'KEEFFE, A.; MCCARTHY, M. *The Routledge Handbook of Corpus Linguistics*. Routledge: New York, 2010.

MARSHALL, S.; GILMOUR, M. Lexical knowledge and reading comprehension in Papua New Guinea. *English for Specific Purposes*, p. 69-81, 1993.

MARTINEZ, A. *Authentic materials: An overview*, 2002. Disponível em <<http://www3.telus.net/linguisticsissues/authenticmaterials.html>> Acesso em 15 jan 2019.

MOORE, C. J. EST readers: Some principles for their design and use. *Reading in a Foreign Language*, p. 119-130, 1983.

NELSON, M. Building a written corpus: what are the basics? In: O'KEEFFE, A.; MCCARTHY, M. *The Routledge Handbook of Corpus Linguistics*. Routledge: New York, 2010.

NUTTALL, C. *Teaching Reading Skills in a foreign language* (New Edition) Oxford: Oxford University Press, 1996.

O'KEEFFE, A.; MCCARTHY, M. *The Routledge Handbook of Corpus Linguistics*. Routledge: New York, 2010.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. *From Corpus to Classroom*. Language Use and Language Teaching. Cambridge: Cambridge University Press, 2007.

PINTO, P.T. Um curso de Inglês com Fins Acadêmicos baseado em Corpus para alunos universitários das áreas de Humanas, Exatas e Biológicas. In: FERREIRA, M. M.; STELLA, V.C.R. (Org.). *Redação Acadêmica: múltiplos olhares para a produção textual e o seu ensino*. 01ed. São Paulo: Humanitas, v. 01, p. 122-136, 2018.

RAMOS, R. C. G. *Afinal, o que é LINFE?* Palestra apresentada no II Congresso Nacional de Línguas para Fins Específicos, FATEC-Tatuapé, São Paulo, 2012.

RAMOS, R. C. G. Instrumental no Brasil: a desconstrução de mitos e a construção do futuro. In: FREIRE, M. M.; ABRAHÃO, M. H. V.; BARCELOS, A. M. F. (Org.). *Linguística Aplicada e contemporaneidade*. São Paulo: ALAB; Campinas: Pontes, 2005.

RAMOS, R. C. G.; LIMA-LOPES, R. E.; GAZOTTI-VALLIM, M. A. Análise de necessidades: identificando gêneros acadêmicos em um curso de leitura instrumental. *The ESpecialist*, v. 25, n. 1, p.1-29, 2004.

REPPEN, R. *Using Corpora in the Classroom*. Cambridge: Cambridge University Press, 2010.

SANTOS, D. *Como ler melhor em inglês*. Barueri: São Paulo, 2011.

SANTOS, M. A. *Descrição do uso das conjunções but e however em redações acadêmicas em língua inglesa de nível b1 com base em corpus*. Orientadora: Paula Tavares Pinto. 2018. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”. São José do Rio Preto, 2018.

SCARAMUCCI, M. V. R. Vestibular e ensino de língua estrangeira (inglês) em uma escola pública. *Trab.Ling,Apl.*, Campinas, p. 7-20, 1999.

SCHMITT, N. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press, 2000.

SEIDLHOFER, B. Key concepts in ELT. English as a lingua franca. *ELT Journal*, v. 59/4 doi:10.1093/elt/cci06, 2005.

SILVA, L. F. *O uso de corpora na produção de atividades voltadas ao desenvolvimento da compreensão oral*. Orientadora: Paula Tavares Pinto. 2018. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”. São José do Rio Preto, 2018.

SINCLAIR, J. *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press, 1991.

SINCLAIR, J. *How to use corpora in language teaching*. Amsterdam: John Benjamins Publishing Company, 2004.

SINCLAIR, J. Preface. In: GHADESSY, M. HENRY, A; ROSEBERRY, R. L. (Eds.) *Small Corpus Studies and ELT: Theory and practice*. Amsterdam: John Benjamins, 2001.

SOLÉ, I. *Estratégias de leitura*. Porto alegre: Artes médicas, 1998.

TAGNIN, S. *A Identificação de equivalentes tradutórios em corpora comparáveis*. 2007. Disponível em:

<http://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/Stella_Abrapui%202007_artigo.pdf> Acesso em: 19 abr 2018.

TAHIR, M. M. English for specific purposes (ESP) and syllabus design. *Journal of the College of Languages (JCL)*, v. 20, p. 100-133, 2009. Disponível em: <http://jcolang.uobaghdad.edu.iq/index.php/JCL/article/view/411>. Acesso em 10 out 2019.

THOMAS, D. *Type-token Ratios in One Teacher's Classroom Talk: An Investigation of Lexical Complexity*, 2005. Disponível em <https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/languageteaching/DaxThomas2005a.pdf>, acesso em 21 out 2018.

TOMITCH, L. M. B. Aquisição de leitura em língua inglesa. In: LIMA, D. C. de. (Org.) *Ensino e aprendizagem de língua inglesa: conversas com especialistas*. São Paulo: Parábola Editorial, 2009.

TRIBBLE, C. Improving corpora for ELT: quick and dirty ways of developing corpora for language teaching. In: Lewandowska-Tomaszczyk, B. and Melia, P. J. (eds.) *Practical*

Applications in Language and Computers (PALC). Lodz: Lodz University Press, p. 106-117, 1997. Disponível em: <http://www.archive.tribble.co.uk/text/Tribble_C_Palc_97.pdf> Acesso em 20 jul 2019.

TSUI, A.B.M. What teachers have always wanted to know and how can corpora help. In: SINCLAIR, J. M. (Ed.), *How to Use Corpora in Language Teaching*. Amsterdam; Philadelphia: John Benjamins, p. 39-66, 2004.

UNESP. *Manual do Candidato 2019*. Disponível em <https://vestibular.unesp.br/Home/2019/manual_unesp_2019.pdf> Acesso em 15.jul.2019.

VIANA V.; TAGNIN S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011.

WIDDOWSON, H. *Aspects of Language Teaching*. Oxford: Oxford University Press, 1990.

WOOD, D. *Formulaic Language and Second Language Speech*. Fluency Background, Evidence and Classroom Applications. CPI Antony Rowe: Wiltshire, 2010.

ZANETTIN, F. DIY corpora: the WWW and the translator. In: MAIA, B.; HALLER, J.; ULRICH, M. (Eds.). *Training the Language Services Provider for the New Millennium*. Porto: Faculdade de Letras, Universidade do Porto, p. 239-248, 2002.

APÊNDICES



APÊNDICE A – Atividade do corpus World

1. As palavras abaixo são consideradas chave no texto que você irá ler a seguir. Você consegue imaginar do que se trata o texto?

tide gauge
warning system
geophysics agency
national disaster agency
long narrow bay

2. Das palavras-chave, alguma lhe é familiar ou um cognato?

1. Warm-up

a. Quais desastres naturais você já ouviu falar/conhece?

b. Quais fatores podem ocasionar um desastre natural?



2. A nuvem de palavras abaixo está relacionada com warning. Grife as palavras que você conhece e as que são cognatos.





3. Fill in the blanks with the most appropriate adjective. Pay attention to the synonyms in parentheses.

extensive – repeated – advance – early – inadequate

But there's a bigger problem - though the alert was sent out, and according to the communications ministry, _____ (continued) tsunami **warnings** were sent to residents via text message - they might not have been received.

This is something we must evaluate for the future. "Could more lives have been saved?" It's possible, as Indonesia does actually have a more advanced tsunami warning system, including a network of 21 buoys which would have dispatched _____ (previous) **warnings** based on data gathered by deep sea sensors.

Indonesia accused of mishandling tsunami warnings The Indonesian government has defended itself against accusations it issued _____ (ineffective) safety **warnings** as a powerful earthquake triggered a tsunami that killed at least 844 people.

The IPCC report, which was released earlier this month, has been called the most _____ (significant) **warning** yet on the risks of rising global temperatures. The report says that the impact of a 1.5C increase in global temperatures will "disproportionately affect disadvantaged and vulnerable populations through food insecurity, higher food prices, income losses, lost livelihood opportunities, adverse health impacts, and population displacements"

Present changes in heat waves and labor capacity "provide an _____ (advance) **warning** of the compounded and overwhelming impact on public health that is expected if temperatures continue to rise," said Hilary Graham, professor of health sciences at the University of York in the UK.

4. Read the sentences below. Match the highlighted expressions to their meanings.

() Numerous watches and **warnings were issued** during the hurricane's existence.
Numerous winter storm, heavy snowfall and freezing rain **warnings were issued**.

() Were people aware there was a tsunami? Many critics have accused BMKG of **lifting the warning** too early.
But it **lifted the warning** just over 30 minutes later.

() A tsunami **warning was sent out** - lasting just over 30 minutes - but it appears to have drastically underestimated the scale of the tsunami that would follow.
However, we do know that the system did in fact pick up the tsunami - because a **warning was sent out** - but what it failed to do was accurately gauge the scale of the tsunami.

1. to send a lot of copies of the same document to a large number of people.

2. to officially end a rule or law that stopped someone from doing something.

3. to announce something, or to give something to people officially.



5. Leia o título da notícia. O que você espera ler nesse texto?

 World » US | Africa | Americas | Asia | Australia | China | Europe | Middle East | India | UK

International Edition +  

1 (CNN) The Indonesian government has defended itself against accusations it issued inadequate safety warnings as a powerful earthquake triggered a tsunami that killed at least 844 people.

2 The 7.5 magnitude earthquake hit Sulawesi island on Friday causing three-meter-high (10 foot) waves to surge over parts of its northwestern shore, with groups of people caught on the coastline, apparently unaware of the danger.

3 Experts underscored the difficulties of predicting tsunamis and raising safety alerts across Indonesia's vast archipelago of over 17,000 islands and 261 million people, but others pointed to a lack of government coordination and funding. Online criticism from Indonesians centered on allegations that the country's meteorology and geophysics agency (BMKG) canceled the tsunami alert too early.

4 The alert was sounded shortly after 6 p.m. on Friday warning of potential waves of up to three meters. The warning was canceled at 6.36 p.m. However, the agency said the alert was lifted only after the tsunami hit. In a statement, the head of the organization, Dwikorita Karnawati, called the allegations "not correct."

5 "Our work is based on computer system/artificial intelligence. The warning system was lifted with the approval of the 28 other countries along the Indian ocean," said Karnawati.

6 Phil Cummins, professor of natural hazards at Australian National University told CNN that while all the facts have yet to emerge, he doubted the warning sign issue was a major concern. "During the time when a warning was potentially useful, it was apparently in effect, and that seems far more significant to me than having canceled it prematurely after the tsunami hit," he added.

Did the tidal buoys work?

7 Questions have focused on Indonesia's warning system, founded on a grid of 134 tide gauge stations and land seismographs, bolstered by education campaigns to train people in tsunami response. Its warnings are transmitted by sirens and text messages.

8 "From the reports I've heard in the media, they (sirens) weren't operable because the power was knocked out by the earthquake," said Cummins. "If true, that's something that should really be looked into." CNN has not been able to confirm these reports.

9 Experts also focused on the number of tidal gauge systems, which measure sea level at the coast, that are in use. "134 tide gauge stations are not adequate to confirm tsunami generation for the country which perhaps has the most islands in the world," said Professor Costas Synolakis, director of the University of South California Tsunami Warning Center. "To be effective, you need to have a tide gauge near every coastal community," he added.

10 Concerns over the government's readiness increased after Sutopo Purwo Nugroho, the spokesman for the country's national disaster agency (BNPB), admitted that none of the additional 22 deep water buoys connected to seafloor sensors to help monitor for tsunamis off Indonesia's coast had worked for the last six years.

11 **"We have around 60 buoys in Indonesia," Sutopo told a press conference in Jakarta on Monday. "Some of them are broken since 2012 because of vandalism. Local people steal them."**

12 The buoys are intended to work in addition to the tide gauge stations as part of Indonesia's Tsunami Early Warning System, designed after the devastating 2004 Aceh tsunami.



13 "Disaster funding continues to decline every year. The threat of a disaster increases, the incidence of disasters increases, the BNPB budget goes down," said Sutopo, according to a report by CNN Indonesia.

<https://edition.cnn.com/2018/10/01/asia/indonesia-earthquake-tsunami-warning-intl/index.html> (adapted)

4. Responda as questões seguintes, relacionadas ao texto:

a. According to the third paragraph,

- (A) There is a contrast between opinions about the warning system in Indonesia.
- (B) Experts defend the government and confirm the difficulties of predicting tsunamis and raising safety alerts in that area.
- (C) Indonesians criticized the BMKG on the internet for starting the alert too late.
- (D) The greatest problem was the need of warnings that did not work appropriately because of the bad government coordination.
- (E) The warning system did not work because of the area size, according to the government.

b. De acordo com Dwikorita Karnawati:

- (A) O sistema utilizado pela BMKG é baseado em inteligência artificial, não corroborando as falhas indicadas pela população.
- (B) Mesmo o sistema sendo computadorizado, não havia possibilidade de alertar toda a população indonésia acerca do tsunami.
- (C) 28 outros países aprovaram que o aviso de tsunami fosse cancelado às 6:36 da tarde.
- (D) As alegações acerca do trabalho da BMKG serão examinadas, pois não estão corretas.
- (E) A aprovação de outros 28 países confirma que a Indonésia, em toda sua extensão, foi corretamente alertada acerca do tsunami por volta de 6 horas antes do desastre.

c. O objetivo principal do texto é

- (A) Informar a respeito do tsunami na Indonésia que ocorreu devido à falta de alerta de segurança.
- (B) Defender o ponto de vista do governo, já que o aviso só não funcionou devidamente por conta do vandalismo praticado na área.
- (C) Mostrar como acontece um desastre natural e suas consequências para a população.
- (D) Informar os leitores acerca do tsunami que ocorreu na Indonésia e matou pelo menos 844 pessoas.
- (E) Informar a respeito das falhas do sistema de alerta durante o tsunami na Indonésia.

d. According to the text?

- (A) People only learn how to protect themselves at school, through education programs.
- (B) The tide gauge stations warn the population by sirens and text messages.
- (C) Tide gauges would be more effective if there were more of them along the islands.
- (D) 60 of the deep water buoys were not working correctly during the disaster.
- (E) 134 tide gauge stations are enough to confirm tsunami generation for the country.



APÊNDICE B – Atividade do corpus Health

1. As palavras abaixo são consideradas chave no texto que você irá ler a seguir. Você consegue imaginar do que se trata o texto?

life expectancy
health-care provision
population mortality
rising concern
declining quality

2. Das palavras-chave, alguma lhe é familiar ou um cognato?

1. Warm-up



Jeanne Daprano (born September 16, 1936) is a world record setting American masters trackand field athlete.

- Você acredita que a expectativa de vida tem mudado ao longo dos anos?
- Quais fatores podem mudar a expectativa de vida nos países?

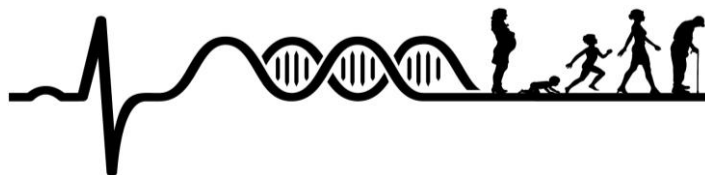
2. Read the sentences below and complete the chart with the highlighted words.

<i>Decline</i>	<i>Rise</i>

“he's now growing his business by buying even more apple seedlings to plant”.

“Australia has one of the fastest growing economies in the developed world”.

“The streets of our peer countries are becoming safer, faster. In Japan, the traffic death rate fell 47 percent between 2000 and 2011”.



“Europe dominated economic headlines on July 10, with worries about a large Portuguese bank and falling industrial production figures for France (down 1.7 percent), Germany (down 1.8 percent) and Italy (down 1.2 percent)”.

“Facebook reached 1.35 billion monthly active users and counting. An overwhelming majority of consumers across the United States are dropping traditional landline phone service, according to updated US Telecom analyses”.

“Needless to say, we are really excited to be part of this growing industry”.

“(…) be less able to negotiate the use of protection, putting them at increased risk for HIV and other sexually transmitted diseases”.

“We must increase our investment in education in a thoughtful manner. It is important to get the policy right – identify the problem, come up with a solution and ask if it is good for the kids and education”.

“These examples, and many more, emphasize the need to focus on prevention, on reducing the incidence of cancer, not just decreasing the mortality rate once cancer develops”.

“AMCOW is committed to working with partners such as the Bill & Melinda Gates Foundation to reduce this scourge and improve access to safe sanitation, thereby achieving our overall goal of decreasing poverty and disease in the continent”.

“The Association of American Railroads is touting new Federal Railroad Administration stats that show that train accidents declined by 7 percent in 2014, accidents caused by equipment dropped 6 percent, accidents caused by tracks declined 12 percent and accidents caused by people dropped 4 percent”.

3. Leia o texto (Skimming) e sublinhe os cognatos/palavras conhecidas que você encontrar.



Health » Food | Fitness | Wellness | Parenting | Vital Signs

International Edition + 🔍 ☰

LIVE LONGER

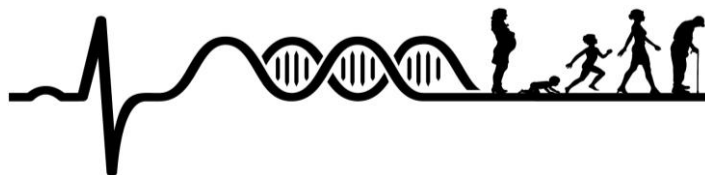
Life expectancy drops in the US and the UK, rises in Australia, a new study finds

By Rory Smith, CNN

🕒 Updated 0908 GMT (1708 HKT) August 17, 2018



1 Life expectancy fell across the majority of high-income countries, signaling a collective and simultaneous decline among affluent nations for the first time in decades, a new study finds. Among 18 high-income countries -- including Spain, Sweden, Japan, Australia, the UK and the United States -- most countries saw declines in life expectancy between 2014 and 2015, according to the study, published Wednesday in the British Medical Journal.



2 Australia, Japan, Denmark and Norway were the only countries in the study that showed an increase in life expectancy across all years for both men and women. The drop in the remaining 14 countries was "notable both for the number of countries and for the magnitude of the declines," the authors wrote.

3 Outside the United States, declines in overall life expectancy were focused among people 65 and older, with the rise in deaths among this demographic probably attributable to an unusually severe flu season, according to the study, which was co-authored by Jessica Ho of the University of Southern California and Arun Hendi of Princeton University.

4 It's a trend that highlights some potential issues around health-care provision within these countries, according to Domantas Jasilionis, a research scientist at the Max Planck Institute for Demographic Research in Germany who authored an editorial that accompanied the study.

5 "The fact that modern healthcare systems in the most advanced high-income countries were unable to cope with this unexpected challenge, resulting in the first reductions in longevity for decades, is striking and might signal more profound problems," Jasilionis wrote.

6 The study also suggested that respiratory disease, cardiovascular disease, Alzheimer's disease and other mental and nervous system disorders contributed in part to driving down life expectancy among those 65 and older.

7 In countries where environmental threats such as air pollution and poor air quality have become more problematic, deaths from these diseases may become more commonplace, suggests Holly Nelson-Becker, an expert in social gerontology at Brunel University in London, who was not involved in the study.

8 "The issue of susceptibility to influenza and respiratory problems speaks, in my view, to increasing issues of environmental problems such as smog and other forms of pollution," Nelson-Becker said.

(Rory Smith. <https://edition.cnn.com/2018/08/16/health/life-expectancy-uk-us-drop-study-intl/index.html>. 30.10.2018. Adaptado)

4. Qual o gênero desse texto?

() *resenha crítica jornalística* () *artigo jornalístico* () *crônica*

5. Qual a ideia geral do texto?

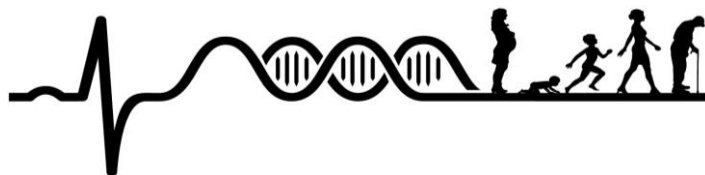
6. Responda as questões seguintes, relacionadas ao texto:

a. Where did life expectancy decline?

- (A) Spain, Sweden, Japan, Australia, the UK and the United States.
- (B) Australia, Japan, Denmark and Norway.
- (C) Spain, Sweden, UK and United States.
- (D) Australia and Japan
- (E) United States, UK and Australia.

b. According to the first and second paragraphs, a study shows that

- (A) Life expectancy is dropping in poor countries.
- (B) High income countries have always had problems with life expectancy.
- (C) 18 high income countries saw declines in life expectancy between 2014 and 2015.
- (D) The decline of life expectancy among affluent nations is something new.



(E) Low income countries are suffering because of drops in life expectancy.

c. De acordo com Domantas Jasilionis:

- (A) Apenas os sistemas mais modernos de saúde podem auxiliar no aumento de expectativa de vida.
- (B) Nem mesmo os sistemas mais modernos de saúde conseguiram lidar com o declínio na taxa de expectativa de vida.
- (C) A longevidade tem sido um desafio em países mais pobres.
- (D) Os sistemas de saúde pública enfrentarão problemas ainda mais sérios ao longo dos anos.
- (E) Já se era esperado que a taxa de longevidade fosse diminuir.

d. Segundo o estudo realizado

- (A) O meio ambiente pode ser um fator importante na expectativa de vida.
- (B) O Alzheimer é o fator que mais influencia na taxa de expectativa de vida entre pessoas de 65 anos ou mais.
- (C) Doenças respiratórias e cardiovasculares têm sido os únicos fatores que mudam a expectativa de vida em um país.
- (D) A qualidade do ar não influencia diretamente na taxa de expectativa de vida.
- (E) Apenas pessoas acima dos 65 anos podem adquirir doenças respiratórias e cardiovasculares.

7. Analise o texto abaixo e responda às questões:

Life expectancy in high income countries in 2016

Countries with highest life expectancy among women

1. Japan -- 87.17 years
2. Spain -- 85.84 years
3. France -- 85.50 years

Countries with highest life expectancy among men

1. Switzerland -- 81.63 years
2. Australia -- 81.49 years
3. Norway -- 80.61 years

Countries with lowest life expectancy among women

1. US -- 81.40 years
2. UK -- 82.72 years
3. Netherlands -- 83.16 years

Countries with lowest life expectancy among men

1. US -- 76.40 years
2. Portugal -- 78.11 years
3. Finland -- 78.43 years

(Rory Smith. <https://edition.cnn.com/2018/08/16/health/life-expectancy-uk-us-drop-study-intl/index.html>. 30.10.2018.)

a. Qual país configura maior taxa de expectativa de vida entre mulheres?



b. Qual o número que descreve o país que tem menor expectativa de vida entre homens?

8. Which words below are related to life expectancy or causes of death? Complete the chart.

vegan diet - flu shot - flu vaccine - breast milk - flu season - low-carb diet - blood alcohol concentration - early death - processed food

<i>Life expectancy</i>	<i>Causes of death</i>

9. Use the expressions above to fill in the blanks correctly:

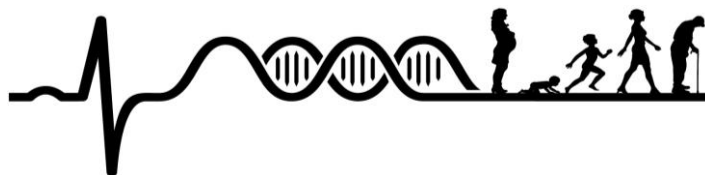
a. In addition to the finding an optimal range associated with a lower risk of _____, the team also calculated how many extra years a moderate-carb diet could provide at certain ages. From age 50, the average life expectancy was an extra 33 years for people who ate carbs in moderation.

b. The 2017-18 _____ ranked among the most severe on record, excluding pandemics. Dr. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, recalled that Australia had a "terrible season last year, and in fact, we had a very bad season last year," as well.

c. Within a few months, a well-balanced _____ which is low in salt and _____ may have impressive benefits for cardiovascular health, helping to prevent heart disease, stroke and reducing the risk of diabetes.

d. From age 50, the average life expectancy was an extra 33 years for people who ate carbs in moderation. This was four years longer than those on _____ (29 years) and one year longer than those on a high-carb diet (32 years), according to the study.

e. The researchers estimated that a hypothetical 3-month-old, weighing 6.1 kilograms or 13 pounds, would ingest about 0.040 nanograms of THC per milliliter of milk. "The question is,



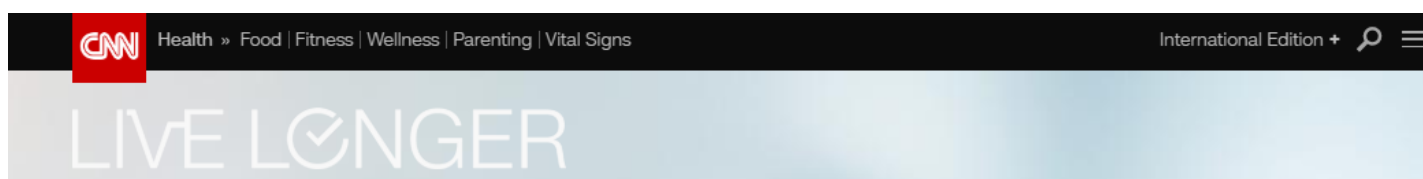
does it matter? ... Is it possible that even low levels in _____ may have an effect on a child's neurodevelopment? [...]"

10. Match the expressions to the correct context:

1. Marijuana use
2. Flu vaccine
3. Contraceptive pill
4. Low testosterone
5. Blood alcohol concentration

- () Men experience symptoms that can be directly correlated to it as they get older
- () Health implications of it while breastfeeding
- () It can take about four drinks for women or five for men consumed within about two hours for that to occur
- () It gives women less than a 1% chance of becoming pregnant
- () Children 9 and older require only one dose, no matter their health history

11. Leia o texto seguinte e tente aplicar as técnicas ensinadas até aqui:



Low and high carb diets increase risk of early death, study finds

By Meera Senthilingam, CNN

Updated 0907 GMT (1707 HKT) August 17, 2018



1 A low-carb or high-carb diet raises your risk of death, a new study suggests, with people eating the food staple in moderation seeing the greatest benefits to their health.

2 Less than 40% or more than 70% of your energy -- or calories -- coming from carbohydrates was associated with the greatest risk of mortality. Eating moderate levels between that range offered the best options for a healthy lifespan.

3 The lowest risk of an early death was seen where carbs made up 50-55% of a person's diet, according to the study published Thursday. However, the definition of a low-carb diet had some caveats as not all diets were equal.

4 People on low-carb diets who replaced their carbohydrates with protein and fats from animals, such as with beef, lamb, pork, chicken and cheese, had a greater risk of mortality than those whose protein and fats came from plant sources, such as vegetables, legumes, and nuts. "We need to look really carefully at what are the healthy compounds in diets that provide protection" said Dr. Sara Seidelmann -- a clinical and research fellow in cardiovascular medicine from Brigham and Women's Hospital in Boston -- who led the research.



5 Seidelmann warned about the widespread popularity of low-carb diets as a weight loss technique, with people giving up foods such as bread, pasta and potatoes.

6 Although previous studies have shown such diets can be beneficial for short-term weight loss and lower heart risk, the longer-term impact is proving to have more negative consequences, according to the study.

7 "Our data suggests that animal-based low carbohydrate diets, which are prevalent in North America and Europe, might be associated with shorter overall life span and should be discouraged," Seidelmann said.

8 "On an 'average' 2,000 kcal-a-day intake, a diet of 30% calories from carbs equates to only 150g a day, with sugars (natural or 'added') contributing around 50g of that total. With a mere 100g of complex carb a day to play with, a lower intake of cereals, grains, and starchy vegetables is inevitable," said Catherine Collins, a dietitian with the UK's National Health Service, who was not involved in the study.

9 She added that such diets compromise the essentials of a healthy diet -- dietary fiber to prevent constipation, support control of blood sugar and lower blood cholesterol levels.

10 Government guidelines in countries like the UK already recommend at least a third of the diet should consist of starchy foods.

11 The findings "will disappoint those who, from professional experience, will continue to defend their low carb cult, but contributes to the overwhelming body of evidence that supports a balanced approach to caloric intake recommended globally by public health bodies," Collins added.

12 The team studied more than 15,000 people aged between 45-64 years from diverse socioeconomic backgrounds across four regions of the United States. They then calculated average caloric intake and the proportion of calories coming from different food groups based on questionnaires the participants completed outlining the types of food and drink they consumed, portion sizes and how often they ate. They were then followed up for a median of 25 years, during which time, 6,283 people died.

Added years

13 In addition to the finding an optimal range associated with a lower risk of early death, the team also calculated how many extra years a moderate-carb diet could provide at certain ages.

14 From age 50, the average life expectancy was an extra 33 years for people who ate carbs in moderation. This was four years longer than those on a low-carb diet (29 years) and one year longer than those on a high-carb diet (32 years), according to the study.

15 The results were also combined with seven other studies on carbohydrate intake among people in North American, European and Asian countries, which revealed similar trends in life expectancy.

a. According to the first, second and third paragraphs

- (A) Only low-carb diets raise your risk of death
- (B) People should balance their diets between 40% and 70% of calories
- (C) Early death increases when carbs made up 50-55% of a person's diet
- (D) Eating moderate is not always healthy
- (E) High-carb diet is the main factor of early death, according to the new study.



b. What is the main idea of paragraph 4?

- (A) The compounds of our diet should be carefully selected.
- (B) Carbohydrates with fats from animals and plants offer the same risk of mortality.
- (C) Carbohydrates with proteins and fats from beef, lamb, pork, chicken and cheese drop the risk of mortality.
- (D) Low-carb diets are healthier when carbohydrates are replaced by animals fat.
- (E) Eating vegetables, legumes, and nuts is not the best option to drop the risk of early death.

c. Segundo Collins

- (A) É essencial que as dietas tenham pelo menos 30% de açúcar.
- (B) Os níveis de açúcar no sangue e de colesterol só podem ser controlados com dietas ricas em carboidratos.
- (C) Dietas ricas em fibras controlam doenças relacionadas à ingestão de açúcar.
- (D) Os estudos podem desapontar os profissionais que acreditam em uma dieta mais calórica.
- (E) A ingestão de carboidratos é necessária para uma saúde equilibrada

d. De acordo com os dois últimos parágrafos

- (A) A ingestão equilibrada de carboidratos pode aumentar a expectativa de vida em 50 anos.
- (B) Uma dieta moderada em carboidratos não é suficiente para aumentar a expectativa de vida.
- (C) A ingestão moderada de carboidratos pode aumentar a expectativa de vida 4 anos mais do que a baixa ingestão dos mesmos componentes.
- (D) A expectativa de vida para aqueles que ingerem carboidratos com moderação é de 33 anos.
- (E) Uma dieta com alto consumo em carboidratos pode aumentar a expectativa de vida em 29 anos, enquanto a de baixo consumo, 32 anos.



APÊNDICE C – Atividade do corpus Tech

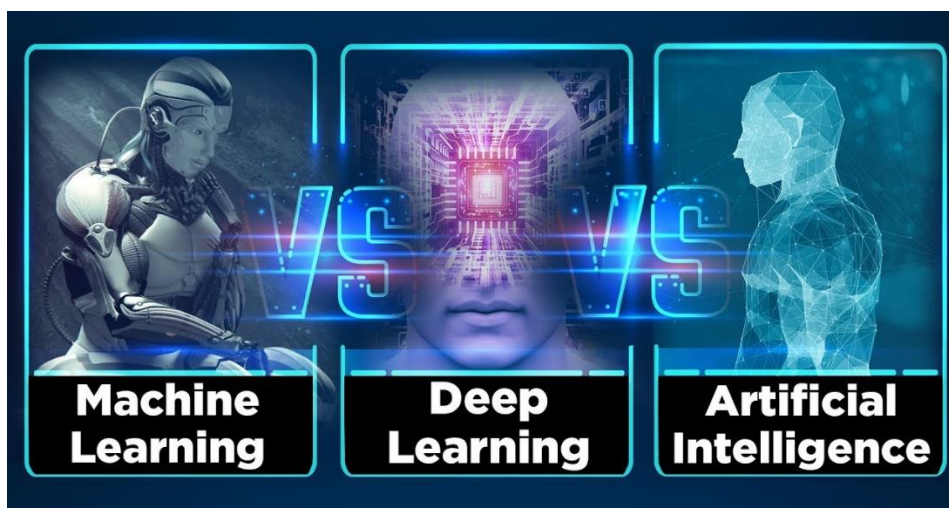
1. As palavras abaixo são consideradas chave no texto que você irá ler a seguir. Você consegue imaginar do que se trata o texto?

deep learning
artificial intelligence
intelligence research
Facebook today
face recognition

2. Das palavras-chave, alguma lhe é familiar ou um cognato?

1. Warm-up

- O que é inteligência artificial para você?
- Em que áreas a inteligência artificial pode ser aplicada/utilizada?
- Observe a figura abaixo. Qual a diferença entre *machine learning*, *deep learning* e *artificial intelligence*?



<https://hackernoon.com/top-differences-between-artificial-intelligence-machine-learning-deep-learning-d39cb6f6feaa>

d. Ligue as expressões 1. *machine learning*, 2. *deep learning* e 3. *artificial intelligence* ao seus significados:

() The theory and development of computer systems that able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making and translation between languages.

TECHNOLOGY



133

() is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to “learn” (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

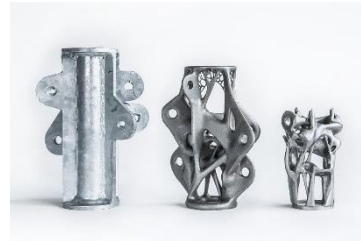
() is a subset of machine learning in Artificial Intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

2. Match the expressions to the pictures. Then, fill in the blanks with the same expressions.

Artificial intelligence – generative design – machine learning – face recognition



()



()



()



()

_____ is a design process that works in imitation of nature and is able to produce endless series of unique visual outputs. _____ is a fast method of exploring design possibilities, rapidly gaining ground, largely due to new programming environments such as Processing.

There also exist methods for extracting geometric features of facial images and local facial features for improving recognition performance and/or reducing processing time. This talk gives an introduction to _____ and its fundamental methods, implementations and applications. _____ can be cast as a pattern classification and machine learning problem: given a set of facial images labeled with subjects' identities [...]

Computer Science is concerned with the specification, design, construction and use of computer systems. It includes areas such as software engineering, hardware architecture and design, information systems, communications, graphics, simulation and modelling, _____ and human-computer interaction.

_____ is a data-driven approach for the development of technical solutions. Initially motivated by the adaptive capabilities of biological systems, _____ has increasing impact in many fields, such as vision, speech recognition, machine translation, and bioinformatics, and is a technological basis for the emerging field of Big Data.



3. Read the text and answer the questions:

a. O que você espera ler no texto levando em consideração o título?

 Markets Tech Media Success Perspectives Video

International Edition + 🔍 ☰

[INNOVATE]

Facebook's top AI scientist says it's 'dust' without artificial intelligence



By Rachel Metz, CNN Business
Updated 1659 GMT (0059 HKT) December 5, 2018



1 (CNN) Without artificial intelligence there wouldn't be much left of Facebook as we know it today.

2 That's according to Yann LeCun, who founded Facebook's artificial intelligence research lab five years ago. "If you take the deep learning out of Facebook today, Facebook's dust," LeCun, Facebook's chief AI scientist, recently told CNN Business. "It's entirely built around it now.

3 The technology is included in everything from the posts and translations you see in your news feed to advertisements. When LeCun established the lab, Facebook was already dabbling in deep learning — a type of machine learning he's worked on and championed since the 1980s. Deep-learning software, modeled after the way neurons work in the brain, ingests loads of data and learns to make its own predictions.

4 Back in 2013, the social network knew AI would be a key part of its future, and like a number of other tech companies, it looked at deep learning specifically for classifying photos and for face recognition.

5 While it appeared promising, it wasn't clear how useful it would be. But years later, aided by loads of data collected from users and increasingly powerful computers, the technology has improved rapidly. Facebook and other companies — such as Google, Microsoft and Amazon — are using it for many different things, such as tagging people in photos and enabling virtual assistants to tell you the weather.

6 LeCun said the social network in particular couldn't function today without deep learning. It's used "absolutely everywhere," he said. Yann Lecun, head of artificial intelligence research at Facebook, says deep learning plays a key role in how we use the social network.

7 This is true not just of what users can see but what they may not see. Deep learning aids Facebook's content filtering, too, and helps remove things like hate speech from the social network.

8 But Facebook's AI efforts have been met with criticism, too. For example, the company is turning to AI to help alert human moderators to hate speech shared on the platform, but plenty of these posts are able to slip through cracks of the system. While deep learning and other AI methods are evolving, it could take years for AI to excel at moderating content.

9 Despite the technology's increasing capabilities, however, LeCun stresses AI is nowhere close to what he likes to call a "Terminator scenario," during which robots would take over. Sure, it can beat humans at games like Go, but we're still far from creating what's known as artificial general intelligence. This type of AI can do human-like tasks and has enough common sense to help out in daily life rather than just performing fairly scripted tasks like Amazon's Alexa does today.

10 LeCun said even extraordinarily AI systems won't have the same drive to do the things humans do unless that's built into them. "The desire to dominate is not correlated with intelligence," he said. "In



fact, we have many examples of this ... in the world. It's not the smartest of us that necessarily wants to be the chief."

a. According to Yann LeCun:

- (A) Without AI Facebook would work in a good, but different way.
- (B) Facebook is partly dependent on AI.
- (C) AI is a very important part of Facebook.
- (D) Deep learning was first used by Facebook.
- (E) Deep learning is completely responsible for Facebook.

b. According to the third paragraph:

- (A) Everything deep-learning does is based on a human brain, including its predictions
- (B) Deep-learning is responsible for the data people see on social medias.
- (C) Facebook started using deep-learning when LeCun established the lab.
- (D) Facebook does not make predictions, but it works very similar to a human brain.
- (E) The brain neurons were modeled after the way deep-learning software works.

c. De acordo com texto:

- (A) A inteligência artificial começou a ser utilizada em 2013 pelo Facebook para postagem de fotos e reconhecimento facial.
- (B) A utilidade da inteligência artificial em redes sociais sempre foi evidente, especialmente no que diz respeito à coleta de dados.
- (C) O Facebook utiliza inteligência artificial para o reconhecimento facial, mas outras empresas utilizam essa tecnologia por outros motivos.
- (D) Com o passar dos anos, a quantidade de dados coletados por computador foi aumentando e a inteligência artificial passou a ser usada pelas empresas por diferentes motivos.
- (E) As redes sociais dificilmente funcionariam sem inteligência artificial, já que os sites utilizam os dados disponibilizados para agradar os seus usuários.

d. O texto defende a ideia de que:

- (A) Hoje em dia, o Facebook é totalmente dependente da inteligência artificial, ainda assim, há críticas sobre seu funcionamento.
- (B) O Facebook só tem reconhecimento facial devido a inteligência artificial, o que o diferencia de outros sites, sem isso seria apenas uma rede social comum.
- (C) Estamos chegando perto do que LeCun chama de "Terminator Scenario", com o avanço da tecnologia, em que os robôs vão assumir o controle de tudo.
- (D) O Facebook já utiliza a inteligência artificial geral, que faz com que a tecnologia reproduza tarefas iguais os humanos.
- (E) O Facebook já se aproxima da Alexa, da Amazon, por sua utilização de inteligência artificial.



APÊNDICE D – Atividade do corpus Politics

1. As palavras abaixo são consideradas chave no texto que você irá ler a seguir. Você consegue imaginar do que se trata o texto?

military parade
bad local government
democratic-controlled government
French military roll
democratic city

2. Das palavras-chave, alguma lhe é familiar ou um cognato?

1. Warm-up

a. O que são as paradas militares e por que elas acontecem?

b. Esse tipo de evento acontece(u) em seu país?



2. Read the sentences below and then fill in the blanks with the expressions in the box:

To watch – was held – was staged – attended the – organized

“A **military parade was staged** on the former airfield, involving 270 armored vehicles and 800 troops”.

“Kim Jong Un applauds as **he watches a military parade** in Pyongyang on April 15, 2012.”

“More than 20,000 people **attended the military parade** organized with the occasion of the National Day”.

“A **military parade was then held** in a well-attended”.

“More than 100,000 people **attend** notably **the main parade** with over 3,500 performers”.

Another military parade _____ across downtown Seoul later in the afternoon.

In 1987, an estimated 100,000 people came _____ the parade.

An enormous parade _____ downtown the day of the game.

Pro-administration officials _____ parades and demonstrations in favor of the tax.



More than 20,000 people _____ military parade organized with the occasion of the National Day.

a. O que você espera ler no texto levando em consideração o título?

Read the text and answer the questions:



45

CONGRESS

SUPREME COURT

2018 ELECTION RESULTS



THE POINT
WITH CHRIS CILLIZZA

The Trump military parade was always a really bad idea



Analysis by **Chris Cillizza**, CNN Editor-at-large
Updated 1657 GMT (0057 HKT) August 17, 2018



1 On Thursday night, the Pentagon⁶³ said that it was shelving the much-ballyhooed military parade in Washington -- until at least next year. On Friday morning, President Donald Trump tried to explain why.

2 "The local politicians who run Washington, D.C. (poorly) know a windfall when they see it. When asked to give us a price for holding a great celebratory military parade, they wanted a number so ridiculously high that I cancelled it," he tweeted. "Never let someone hold you up! I will instead attend the big parade already scheduled at Andrews Air Force Base on a different dates & go to the Paris parade, celebrating the end of the War, on November 11th. Maybe we will do something next year in D.C. when the cost comes WAY DOWN. Now we can buy some more jet fighters!"

3 Which, um, probably not.

4 DC Mayor Muriel Bowser tweeted a response to that effect. "Yup, I'm Muriel Bowser, mayor of Washington DC, the local politician who finally got thru to the reality star in the White House with the realities (\$21.6M) of parades/events/demonstrations in Trump America (sad)," she said. *(The total cost of the parade was estimated at \$92 million, roughly \$80 million more than original estimates, according to several news outlets.)*

5 Here's the thing: This military parade has been a very bad idea since Trump first conceived of it. What he's doing now is using a scapegoat that he knows will work with his base -- bad local government in an overwhelmingly Democratic city -- to get out from under something that was fundamentally unworkable -- and unpopular!

6 The genesis of the military parade, which was originally scheduled for Veterans Day this year, speaks to how this grew from a Trump whim.

7 He was by French President Emmanuel Macron's side last summer in Paris to celebrate Bastille Day -- the annual celebration commemorating the storming of the Bastille, a seminal

⁶³ Pentagon, large five-sided building in Arlington county, Virginia, near Washington, D.C., that serves as the headquarters of the U.S. Department of Defense, including all three military services—Army, Navy, and Air Force. (<https://www.britannica.com/topic/Pentagon>)



moment in the French Revolution. Watching the French military roll through the streets of Paris, Trump could barely contain his excitement; he was like a kid on Christmas morning. "It was one of the greatest parades I've ever seen," Trump said at a United Nations gathering a few months later. "It was military might."

8 In February, The Washington Post first reported that Trump had asked Defense Secretary James Mattis and other top military brass to organize a parade "like the one in France," according to one military official. White House spokeswoman Sarah Sanders sought to frame the entire thing as a demonstrating of appreciation and patriotism; "President Trump is incredibly supportive of America's great service members who risk their lives every day to keep our country safe," she said. "He has asked the Department of Defense to explore a celebration at which all Americans can show their appreciation."

9 The idea was almost immediately panned by, well, almost everyone not named Donald Trump.

10 "I think confidence is silent and insecurity is loud," Louisiana Republican Sen. John Kennedy said. "America is the most powerful country in all of human history; you don't need to show it off."

11 Mark Hertling, a retired Army general and a CNN contributor, added: "We're the only nation in the world that defends a piece of paper. An ideology. And to say we're going to strut our stuff with tanks and rocket launchers and things like that, is just not a good representation of what the military does in a democratic nation."

12 Democrats were even more critical, suggesting that Trump's desire to have tanks rolling through the streets of Washington had a distinct authoritarian air.

13 "A military parade in DC would shut down the nation's capital and waste taxpayer dollars just to feed Trump's ego," said DC delegate Eleanor Holmes Norton.

14 The public also disliked the idea. A lot.

15 A Quinnipiac University poll released in February showed that just 26% of the public approved of the military parade while 61% disapproved. Told that the cost of the parade was estimated to be between \$10 and \$30 million, the public was even less interested in the parade. Just 18% said that expenditure was a good use of government dollars; 75% said it was a bad way to spend federal dollars.

16 To put a fine point on it: The military parade idea was an albatross from the start. The "rah rah" patriotism Trump thought he could gin up with it wasn't building, and the costs -- as they always do when you are talking about massive movements of our military -- were going through the roof.

17 There are only so many unpopular ideas created on a whim that even a president like Trump is willing to keep pushing on. With the Space Force already in motion -- and unpopular -- Trump needed a way to cut bait on the military parade. The cost estimates that began leaking out this week were the perfect exit strategy for Trump. Lay the blame on a local, Democratic-controlled government, banking on the built-in belief among many Republicans that Democrats love to spend money unwisely.



18 The reality is this: The military parade was a wildly bad idea from the start that was unworkable for lots of reasons -- from logistical ones to symbolic ones. It just took the President of the United States longer than most people to realize that fact.

(Chris Cillizza. <https://edition.cnn.com/2018/08/17/politics/parade-trump/index.html>. 17.08.2018. Adaptado)

a. Why did Trump cancel the parade?

Because of...

- (A) The price.
- (B) Problems with dates.
- (C) The politicians did not agree.
- (D) Washington DC is not prepared for this kind of events.
- (E) There are not weapons enough.

b. According to the seventh paragraph:

- (A) France was the inspiration for a new political change in the US.
- (B) Trump wants to hold the Bastille Parade in the US.
- (C) For Trump, the best moment to have a military parade is on Christmas.
- (D) Trump's idea of a military parade was inspired by the one in France.
- (E) The Bastille Parade on Christmas made Trump feel very excited.

c. De acordo com Mark Hertling:

- (A) Com as paradas militares, a nação americana não está defendendo apenas um pedaço de papel.
- (B) A ideologia americana deve ser defendida por meio da parada militar.
- (C) É necessário adquirir mais tanques e lançadores de foguetes para a realização da parada militar.
- (D) A parada militar pode arruinar a democracia americana.
- (E) Uma parada militar não é a melhor maneira de representar o serviço dos militares.

e. No trecho “When asked to give us a price for *holding* a great celebratory military parade [...]” **o verbo em destaque poderia ser substituído, sem mudança de sentido, por:**

- (A) celebrating
- (B) watching
- (C) attending
- (D) organizing
- (E) going to



d. O texto defende a ideia de que

- (A) As paradas militares sempre farão parte da cultura americana.
- (B) Realizar uma parada militar nos EUA é desnecessário.
- (C) As paradas militares são realizadas apenas para exibição de uma nação.
- (D) Não há dinheiro suficiente para realizar uma parada militar no atual momento.
- (E) Apesar de não ser popular, a maioria da população americana é a favor da parada militar.

5. Which words below are related to “trade”? Complete the chart.

pact – record – future – peace – police – war – deal – parties

<i>Trade</i>

6. Use the expressions above to fill in the blanks correctly. More than one correct answer is possible.

- a. German Chancellor Angela Merkel said "substantial progress" was needed on the UK's withdrawal agreement by the next European Council meeting in October, with "still a large piece of work" on the separate issue of _____ relations with the UK.
- b. Moreover, with the end of the Obama-negotiated Trans-Pacific Partnership, there are really no other reliable _____ involving the US across Asia. Still, as Trump told a news conference the morning after his NAFTA victory, Japan and India have indicated an interest in negotiating tariff agreements.
- c. During this time, the two sides hope to negotiate a permanent _____. The UK and the EU want to avoid a hard Northern Ireland border whatever happens, so they agreed to a "backstop" - described as an insurance policy by Mrs May.
- d. Finally, of course, there is the mother of all trade disputes, already developing into a _____, between the world's two leading trading nations, China and the US. Neither side here seems to be anywhere near an early round of productive negotiations.



7. Match the expressions to the correct context:

- () Nuclear deal
- () Trade deal
- () New deal
- () Withdrawal deal

1. "Democrats and some within Mr Trump's own party insist they will not approve it. Mr Trump argued on Friday that the US had lost out to Mexico by \$75bn a year under the North American Free Trade Agreement (Nafta), which he replaced with a _____ with Mexico and Canada last month".

2. "(...) ought to be sending chills up the spines of diplomats and trade negotiators around the world. Trump largely got his way. And now, no one can tell him his bull-in-a China-shop way won't work. Should Sunday night's sudden 11th-hour approval of the U.S.-Mexico-Canada Agreement give hope or dismay to Europeans hoping to do a _____ with America"

3. "Winners of the 2014 international digital design competition were announced by the Academy [...] A new poll has shown that Israelis are overwhelmingly of the view that President Barack Obama's impending _____ with Iran will hurt Israel".

4. "Speaking at a press conference in Brussels after meeting the 27 other EU member states, he said the _____ was "fair and balanced" but made clear a transition period extension could not be indefinite, "it has to be a fixed period of time. During Monday's meeting, the Spanish government raised concerns about two articles in the withdrawal agreement".

8. Read the text and answer the questions:

[Sign in](#)

[News](#)
[Sport](#)
[Weather](#)
[Shop](#)
[Reel](#)
[Travel](#)
[More](#)

NEWS

[Home](#)
[Video](#)
[World](#)
[UK](#)
[Business](#)
[Tech](#)
[Science](#)
[Stories](#)
[Entertainment & Arts](#)
[Health](#)
[World News TV](#)
[More](#)

[Business](#)
[Market Data](#)
[Global Trade](#)
[Companies](#)
[Entrepreneurship](#)
[Technology of Business](#)
[More](#)

Donald Trump says new trade deal is 'most important ever'

🕒 1 October 2018

Top Stories

China's Huawei rejects US criminal charges

🕒 1 hour ago

1 Donald Trump said the new trade deal struck with Canada and Mexico was "the most important" ever agreed by the US.

2 The president said the United States-Mexico-Canada Agreement (USMCA), which replaces Nafta, would bring thousands of jobs back to North America.

3 Speaking at the White House, Mr Trump said the new pact vindicated his threats over trade tariffs.



4 The new deal covered trade between the three countries worth \$1.2trn and was "truly historic", he said.

5 It was also "the biggest trade deal in the United States history", he told a press conference.

6 Mr Trump spoke after earlier posting tweets claiming the new trade pact solved the "deficiencies and mistakes" in Nafta, which has governed trade between the three countries since 1994.

7 He said the new deal was "much more reciprocal" than Nafta, which he described as "perhaps the worst trade deal ever made". The measures in the new deal - which Mr Trump dubbed "USMCA" - "will support many - hundreds of thousands - American jobs".

8 The president has adopted an "America First" policy and launched a trade war against China, as well as imposed tariffs on steel and aluminium imports from Mexico and Canada. "Without tariffs we wouldn't be talking about a deal," Mr Trump said.

9 Canadian Prime Minister Justin Trudeau said the trade deal was "profoundly beneficial" to Canadians. "We had to make compromises, and some were more difficult than others," he said. "We never believed that it would be easy, and it wasn't, but today is a good day for Canada."

10 However, Mr Trump said it was too soon to talk to China, on which the US has imposed three rounds of tariffs this year, about a trade deal. "China wants to talk, very badly ... [we] can't talk now because they're not ready," Mr Trump said.

11 Other trading partners had been tough on the US, he said, including the European Union, which introduced retaliatory tariffs on US goods in June.

12 The president said he was having a "successful negotiation" with the EU. He said it was "privilege" for other countries to trade with the US: "So we have negotiated this new agreement [with Mexico and Canada] based on the principle of fairness and reciprocity - to me it's the most important world in trade, because we've been treated so unfairly by so many nations all over the world."

13 The new deal was "terrific" for all three countries, he added. Hundreds of pages of the agreement released early on Monday contain updated arrangements for Canada's dairy industry and measures aimed at shifting lower-paid car manufacturing jobs from Mexico.

14 US farmers will gain access to 3.5% of Canada's dairy market worth \$16bn a year, while another requirement stipulates that 40% of components for vehicles produced in the USMCA area must be made in areas paying wages of \$16 an hour.

'More jobs'

15 Meredith Crowley, international trade economist at the University of Cambridge, said the agreement on dairy appeared to be a "cosmetic concession", suggesting Canada had done well out of the pact.

16 She said the insertion of a minimum wage level for car parts workers could lead to similar clauses in other trade deals: "If it turns out to be very politically popular in the US it could [be used again]."

17 Mr Trump said the deal was a victory for farmers, car workers and the US manufacturing industry: "It means far more American jobs, and these are high-quality jobs."



a. According to the first and second paragraphs

- (A) The new deal is not good for Mexico and Canada.
- (B) The trade deal will be important only for the US.
- (C) The trade deal will bring a lot of jobs to these countries.
- (D) USMCA will not be as important as the Nafta deal.
- (E) USMCA will bring some problems to North America.

b. What is the main idea of the paragraphs 5 and 6?

- (A) Nafta was the best deal for the US.
- (B) Nafta was much more reciprocal than USMCA.
- (C) "deficiencies and mistakes" in Nafta were solved by USMCA deal.
- (D) Low-carb diets are better when carbohydrates are replaced by animals fat.
- (E) Eating vegetables, legumes, and nuts increases the risk of early death.

c. De acordo com o Primeiro Ministro canadense Justin Trudeau

- (A) O novo acordo foi muito benéfico para o Canadá.
- (B) O novo acordo foi o mais difícil para os canadenses.
- (C) Os EUA serão os únicos beneficiados no novo acordo.
- (D) É um dia marcante para os canadenses, apesar de ruim.
- (E) O Canadá não fará mais parte das decisões do novo acordo.

d. Para Meredith Crowley

- (A) A inserção de salário mínimo só cabe aos trabalhadores de autopeças.
- (B) Não há possibilidade de o salário mínimo não se tornará popular nos EUA.
- (C) O salário mínimo deve aumentar para se tornar popular nos EUA.
- (D) Popular ou não, o salário mínimo será adotado em todas as esferas trabalhistas nos EUA.
- (E) Se o salário mínimo for muito popular na política dos EUA, poderá ser usado novamente.

e. No trecho “The new deal covered trade between the three countries worth \$1.2trn and was “truly historic”, he said.” a palavra destacada pode ser substituída sem mudança de sentido, por:

- (A) agree
- (B) pact
- (C) divergence
- (D) promise
- (E) leadership

APÊNDICE E – Chavicidade relativa do corpus World

	<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>		<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>
1	<i>bread roll</i>	217.729	0.037	26	<i>tidal gauge</i>	46.656	0.005
2	<i>crown prince</i>	155.521	0.198	27	<i>initial quake</i>	46.656	0.009
3	<i>warning system</i>	171.073	0.683	28	<i>Thai immigration</i>	46.656	0.011
4	<i>Chinese state</i>	93.313	0.2	29	<i>early warning system</i>	62.208	0.344
5	<i>aircraft carrier</i>	139.969	0.819	30	<i>warning level</i>	46.656	0.02
6	<i>tide gauge</i>	77.76	0.058	31	<i>monsoon season</i>	62.208	0.373
7	<i>migrant caravan</i>	62.208	0	32	<i>satellite photo</i>	46.656	0.036
8	<i>blue flu</i>	62.208	0.002	33	<i>airport hotel</i>	46.656	0.059
9	<i>state tv</i>	62.208	0.003	34	<i>local time</i>	108.865	1.465
10	<i>change skepticism</i>	62.208	0.006	35	<i>Saudi-led coalition</i>	46.656	0.092
11	<i>climate change skepticism</i>	62.208	0.006	36	<i>climate change agreement</i>	46.656	0.096
12	<i>Chinese scientist</i>	62.208	0.008	37	<i>change agreement</i>	46.656	0.098
13	<i>immigration police</i>	62.208	0.014	38	<i>sandwich shop</i>	46.656	0.098
14	<i>gay sex</i>	93.313	0.545	39	<i>extreme heat</i>	62.208	0.515
15	<i>river dolphin</i>	62.208	0.042	40	<i>source close</i>	46.656	0.245
16	<i>powerful earthquake</i>	62.208	0.082	41	<i>southern city</i>	46.656	0.269
17	<i>tsunami warning</i>	62.208	0.091	42	<i>fighter jet</i>	46.656	0.283
18	<i>refugee agency</i>	62.208	0.242	43	<i>heavy snow</i>	46.656	0.343
19	<i>satellite image</i>	62.208	0.294	44	<i>news agency</i>	108.865	2.228
20	<i>live ammunition</i>	62.208	0.318	45	<i>climate policy</i>	62.208	0.968
21	<i>sea clash</i>	46.656	0	46	<i>geophysics agency</i>	31.104	0
22	<i>re-education school</i>	46.656	0	47	<i>issued non-compulsory evacuation</i>	31.104	0
23	<i>Saudi teenager</i>	46.656	0	48	<i>non-compulsory evacuation</i>	31.104	0
24	<i>Chinese state tv</i>	46.656	0	49	<i>inoffensive foodstuff</i>	31.104	0
25	<i>Saudi consulate</i>	46.656	0.002	50	<i>first gene-edited</i>	31.104	0

APÊNDICE F – Chavicidade relativa do corpus Health

	<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>		<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>
1	<i>swine fever</i>	315.806	0.102	26	<i>opioid crisis</i>	63.161	0.008
2	<i>African swine fever</i>	252.645	0.061	27	<i>blood alcohol</i>	78.952	0.269
3	<i>African swine</i>	252.645	0.062	28	<i>premature menopause</i>	63.161	0.026
4	<i>flu vaccine</i>	221.064	0.295	29	<i>low-carb diet</i>	63.161	0.066
5	<i>vegan diet</i>	205.274	0.254	30	<i>research brief</i>	63.161	0.099
6	<i>life expectancy</i>	647.402	3.087	31	<i>new study</i>	315.806	4.548
7	<i>feeding style</i>	157.903	0.005	32	<i>parenting style</i>	63.161	0.131
8	<i>raw turkey</i>	126.322	0.006	33	<i>wild boar</i>	78.952	0.436
9	<i>peanut allergy</i>	126.322	0.071	34	<i>health board</i>	63.161	0.182
10	<i>flu season</i>	142.113	0.347	35	<i>plant-based diet</i>	63.161	0.184
11	<i>flu shot</i>	126.322	0.217	36	<i>marijuana use</i>	78.952	0.555
12	<i>scheduling sex</i>	94.742	0.001	37	<i>true national crisis</i>	47.371	0
13	<i>childhood obesity</i>	221.064	1.336	38	<i>pediatrics group</i>	47.371	0
14	<i>low testosterone</i>	94.742	0.06	39	<i>weekly newsletter sign</i>	47.371	0
15	<i>abortion pill</i>	94.742	0.061	40	<i>smoke-free society</i>	47.371	0.001
16	<i>breast milk</i>	189.484	1.193	41	<i>raw ground turkey</i>	47.371	0
17	<i>gut microbiota</i>	94.742	0.176	42	<i>huge epidemic</i>	47.371	0.003
18	<i>acute flaccid myelitis</i>	78.952	0.002	43	<i>sexual frequency</i>	47.371	0.005
19	<i>flaccid myelitis</i>	78.952	0.002	44	<i>medical uncertainty</i>	47.371	0.004
20	<i>last flu season</i>	78.952	0.003	45	<i>allergy risk</i>	47.371	0.007
21	<i>last flu</i>	78.952	0.004	46	<i>amyloid hypothesis</i>	47.371	0.008
22	<i>beta amyloid</i>	78.952	0.023	47	<i>raw ground</i>	47.371	0.01
23	<i>ground turkey</i>	78.952	0.111	48	<i>low-carbohydrate diet</i>	47.371	0.022
24	<i>fever crisis</i>	63.161	0	49	<i>youth obesity</i>	47.371	0.022
25	<i>swine fever crisis</i>	63.161	0	50	<i>vitamin d</i>	47.371	0.025

APÊNDICE G – Chavacidade relativa do corpus Tech

	<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>		<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>
1	<i>generative design</i>	160.245	0.035	26	<i>solar energy capacity</i>	43.703	0.015
2	<i>fake news</i>	160.245	0.079	27	<i>electric aircraft</i>	43.703	0.016
3	<i>artificial intelligence</i>	364.193	2.066	28	<i>5g network</i>	43.703	0.017
4	<i>5g service</i>	101.974	0.001	29	<i>Chinese tech</i>	43.703	0.019
5	<i>facial recognition</i>	131.109	0.365	30	<i>urban mining</i>	43.703	0.021
6	<i>coal plant</i>	116.542	0.357	31	<i>blue origin</i>	43.703	0.024
7	<i>facial recognition technology</i>	87.406	0.052	32	<i>legitimate news</i>	43.703	0.025
8	<i>recognition technology</i>	87.406	0.204	33	<i>using artificial intelligence</i>	43.703	0.029
9	<i>machine vision</i>	87.406	0.296	34	<i>American drone</i>	43.703	0.042
10	<i>Saudi government</i>	72.839	0.192	35	<i>global tech</i>	43.703	0.056
11	<i>solar orbiter</i>	58.271	0.007	36	<i>visual recognition</i>	43.703	0.058
12	<i>budding relationship</i>	58.271	0.03	37	<i>investment conference</i>	43.703	0.059
13	<i>machine learning</i>	131.109	1.299	38	<i>big tech</i>	43.703	0.06
14	<i>deep learning</i>	72.839	0.308	39	<i>lunar orbit</i>	43.703	0.062
15	<i>false news</i>	58.271	0.093	40	<i>old coal</i>	43.703	0.062
16	<i>composite image</i>	58.271	0.135	41	<i>solar park</i>	43.703	0.063
17	<i>energy capacity</i>	58.271	0.229	42	<i>space tourism</i>	43.703	0.072
18	<i>saudi investment conference</i>	43.703	0	43	<i>emotional bond</i>	43.703	0.091
19	<i>chinese internet</i>	43.703	0	44	<i>autonomous driving</i>	43.703	0.093
20	<i>indian internet</i>	43.703	0	45	<i>capacity factor</i>	43.703	0.118
21	<i>rocket-powered plane</i>	43.703	0	46	<i>space travel</i>	58.271	0.488
22	<i>dummy satellite</i>	43.703	0.001	47	<i>sovereign wealth fund</i>	43.703	0.141
23	<i>four-seat aircraft</i>	43.703	0.001	48	<i>wealth fund</i>	43.703	0.155
24	<i>saudi investment</i>	43.703	0.005	49	<i>tech industry</i>	58.271	0.564
25	<i>compromise measure</i>	43.703	0.012	50	<i>hate speech</i>	72.839	0.996

APÊNDICE H – Chavicidade relativa do corpus Politics

	<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>		<i>Palavra</i>	<i>Corpus de estudo</i>	<i>Corpus de referência</i>
1	<i>special counsel</i>	154.295	0.096	26	<i>government shutdown</i>	92.577	0.733
2	<i>trade war</i>	154.295	0.099	27	<i>southern border</i>	77.147	0.491
3	<i>administration official</i>	169.724	0.211	28	<i>simple majority</i>	77.147	0.543
4	<i>military parade</i>	154.295	0.157	29	<i>new deal</i>	77.147	0.549
5	<i>no-deal brexit</i>	108.006	0	30	<i>former campaign chairman</i>	46.288	0.001
6	<i>senior administration</i>	123.436	0.172	31	<i>withdrawal deal</i>	46.288	0
7	<i>9th circuit</i>	108.006	0.092	32	<i>nomination record</i>	46.288	0
8	<i>trade deal</i>	154.295	0.576	33	<i>election interference</i>	46.288	0.001
9	<i>withdrawal agreement</i>	92.577	0.005	34	<i>voting eligible population</i>	46.288	0.002
10	<i>senior administration official</i>	92.577	0.063	35	<i>political litmus test</i>	46.288	0.005
11	<i>crown prince</i>	92.577	0.198	36	<i>caucus vote</i>	46.288	0.005
12	<i>campaign chairman</i>	77.147	0.026	37	<i>steadfast partner</i>	46.288	0.006
13	<i>campaign trail</i>	108.006	0.473	38	<i>political litmus</i>	46.288	0.008
14	<i>defense official</i>	77.147	0.078	39	<i>former campaign</i>	46.288	0.038
15	<i>transition period</i>	123.436	0.82	40	<i>9th district</i>	46.288	0.039
16	<i>security clearance</i>	92.577	0.394	41	<i>border wall</i>	46.288	0.052
17	<i>new election</i>	77.147	0.2	42	<i>press secretary</i>	61.718	0.4
18	<i>no-deal scenario</i>	61.718	0	43	<i>future trade</i>	46.288	0.071
19	<i>saudi consulate</i>	61.718	0.002	44	<i>fake news</i>	46.288	0.079
20	<i>eligible population</i>	61.718	0.052	45	<i>lead attorney</i>	46.288	0.091
21	<i>political settlement</i>	77.147	0.327	46	<i>six-year term</i>	46.288	0.118
22	<i>future relationship</i>	61.718	0.078	47	<i>resignation letter</i>	46.288	0.142
23	<i>today programme</i>	61.718	0.126	48	<i>trade pact</i>	46.288	0.146
24	<i>defense secretary</i>	61.718	0.141	49	<i>unpaid leave</i>	46.288	0.213
25	<i>new leadership</i>	108.006	0.983	50	<i>watchdog group</i>	46.288	0.249