UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Luana Viana dos Santos

Atividades de leitura em língua inglesa direcionadas por dados:
o uso de um *corpus* com a temática dos ODS

São José do Rio Preto
2024

Luana Viana dos Santos

Atividades de leitura em língua inglesa direcionadas por dados:

o uso de um *corpus* com a temática dos ODS

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 - 88887.671444/2022-00

Orientador: Profª. Drª. Paula Tavares Pinto

São José do Rio Preto
2024

S237a

Santos, Luana Viana dos
Atividades de leitura em língua inglesa direcionadas por dados : o uso de um corpus com a temática dos ODS / Luana Viana dos Santos. -- São José do Rio Preto, 2024
153 p.

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. Data driven learning. 3. Termos multipalavras. 4. Estratégias de leitura. 5. ODS. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

Luana Viana dos Santos

Atividades de leitura em língua inglesa direcionadas por dados:

o uso de um *corpus* com a temática dos ODS

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 - 88887.671444/2022-00

Comissão Examinadora
Titulares

Prof.ª Dr.ª Paula Tavares Pinto
UNESP – Câmpus de São José do Rio Preto
Orientador

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

Prof.ª Dr.ª Sandra Mari Kaneko Marques
UNESP – Câmpus de Araraquara

São José do Rio Preto
30 de agosto de 2024

Dedico este trabalho à vó Marlene, ao vô Isaías (*in memorian*), à vó Deral (*in memorian*) e ao vô Chiquinho (*in memorian*).

AGRADECIMENTOS

Agradeço a Deus e a N. Sra. Aparecida pela minha vida e por todas as oportunidades que a mim foram dadas.

Aos meus pais, por terem me incentivado a trilhar o caminho dos estudos, por todo o apoio e por terem feito de tudo para que eu chegasse à universidade.

À minha orientadora, Paula Tavares Pinto, por ter me convidado para fazer iniciação científica na graduação e por ter me conduzido até o Mestrado, pela paciência e pelos conselhos nos momentos de insegurança e pela orientação humana e atenciosa durante os anos de pesquisa.

Ao Daniel pela parceria, pelo apoio carinhoso e pela paciência nos momentos em que a vida acadêmica fez com que eu me ausentasse.

Aos meus familiares, em especial a Laura, que sempre acreditaram em mim e que torciam para que eu concluísse a jornada do Mestrado.

Ao Davi e à Márcia, pelo incentivo e pelas palavras sempre doces e encorajadoras.

Aos amigos da graduação que continuaram me apoiando na pós, Yngrid, Raquel, Pedro, Johnathan, Rafael e Ana Cristina.

À Melanie e à Naiara, pela amizade que me deu fôlego nos momentos em que os estudos pesavam.

Aos colegas do grupo de pesquisa En-Corpora pela acolhida e pelas reuniões em que compartilhávamos experiências e conhecimento, em especial à Carol, ao Luciano e ao José Victor pelo tempo dedicado a tirar minhas dúvidas por meio de encontros via *Google Meet* e por meio de mensagens, e à Silmara e à Jeane, pela parceria nos eventos acadêmicos.

Aos alunos da disciplina Estratégias de Leitura em Língua Inglesa – Nível Básico de 2023, pela experiência e por toda a troca de conhecimentos, assim como às professoras-tutoras de 2024 pela contribuição dada por meio da participação de suas turmas na pesquisa.

Aos meus alunos e aos colegas de jornada da escola pública, pela força para continuar atuando na área da educação como professora-pesquisadora.

Aos professores do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL), Edson Rosa, Marize Hattnher, Fabiana Komesu e demais professores da disciplina de Letramentos Acadêmicos-Científicos, Lília Santos Abreu-Tradelli, Solange Aranha, Talita Serpa, Adriane Orenha, Maurizio Babini e Celso Rocha pelas aulas e por todos os valiosos ensinamentos.

À professora Sandra Mari Kaneko Marques pelas contribuições na apresentação de painel no Selin, em 2022, e ao professor Tony Berber Sardinha pelos apontamentos e orientações dadas no debate do Selin, em 2023.

Ao Ibilce, por ter me proporcionado o ensino de excelência da universidade pública na graduação e na pós-graduação, por ter me levado a lugares que não imaginava alcançar e por ter me apresentado pessoas queridas e inesquecíveis.

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

“Nos olhos da ciência que desvenda a alegoria
Soam os tambores em anúncio à boa nova
Impávido diante do que não se conhecia
Avança no sendeiro o homem forte a cada prova
E quando sofre a alma nessa casa de argila
Há num mergulho de olho aberto a mão amiga que conduz
Do cosmonauta que avisou: Cada um carrega a sua cruz” (Black Alien; Forfun, 2011)

RESUMO

Diante do atual cenário de internacionalização das universidades (Pinto *et al*, 2021), que tem evidenciado a importância da compreensão escrita em Língua Inglesa (LI) (Heemann, 2009; Tomitch, 2009), este estudo teve como objetivo desenvolver atividades de leitura dirigidas por dados, com base nos trabalhos da Linguística de *Corpus* (LC) e da abordagem Aprendizagem Direcionada por Dados, *Data Driven Learning* (DDL), para alunos de níveis A1 e A2 em contexto de ensino de Inglês Para Fins Acadêmicos (IFA) (Pinto *et al*, 2023; 2021; Mark; O'Keeffe, 2022; O'Boyle; Viana, 2022; Anthony, 2018; Boulton; Cobb, 2017; Tagnin; Viana, 2011; Berber Sardinha, 2011; 2005; 2004; Biber, 2009; Sinclair, 2005; 1991; Johns, 1988). Os participantes da pesquisa são 68 alunos dos Programas de Pós-Graduação da Universidade Estadual Paulista (Unesp), que se matricularam na disciplina on-line "Estratégias de Leitura em Língua Inglesa – Nível Básico". A elaboração do material didático se deu a partir do levantamento das palavras-chave de um *corpus* de artigos científicos produzidos por autores que abordam os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), estabelecidos e firmados na Agenda 2030. Por estarem em trabalhos de pesquisadores internacionais de diversas áreas, as palavras-chave dos ODS podem ser entendidas como termos multipalavras recorrentes no vocabulário acadêmico, o que as torna relevantes para aprendizes de inglês com fins específicos. O *corpus* de estudo, o *SDG-Plos* (Pinto *et al*, no prelo), contém 400 artigos publicados na revista *PLOS-ONE* entre os anos de 2011 e 2021 nas áreas de ciência e medicina. As palavras-chave foram extraídas da lista de termos multipalavras gerada pela ferramenta *Keywords* do *software Sketch Engine* (Kilgarriff *et al*, 2014), que guiou a extração dos textos a serem usados no desenvolvimento de atividades *hands on* e *hands off* com auxílio das ferramentas *CorpusMate* (Baisa; Crosthwaite, 2023), *Skell* (Baisa; Suchomel, 2014) e *VersaText* (Thomas, 2020). As atividades foram aplicadas para as turmas do segundo semestre de 2023 e do primeiro semestre de 2024, e avaliadas, com relação ao uso das ferramentas e da temática dos textos lidos, na percepção dos alunos, por meio de um questionário. Os resultados apontam como a DDL pode ser utilizada para beneficiar aprendizes de nível básico, desde que sejam feitas as adaptações necessárias, como os ODS podem atuar como recurso didático multidisciplinar em contexto de IFA e como materiais baseados em *corpora* podem ser replicados por professores de línguas não especialistas em LC.

Palavras–chave: Linguística de *corpus*. *Data driven learning*. Termos multipalavras. Estratégias de leitura. Objetivos de desenvolvimento sustentável.

ABSTRACT

Given the current scenario of university internationalization (Pinto et al, 2021), which has highlighted the importance of reading comprehension in English (Heemann, 2009; Tomitch, 2009), this study aimed to develop corpus-driven reading activities for A1 and A2 level students in the context of English for Academic Purposes (EAP) based on the works of Corpus Linguistics (CL) and the Data Driven Learning (DDL) approach (Pinto *et al*, 2023; 2021; Mark; O'Keeffe, 2022; O'Boyle; Viana, 2022; Anthony, 2018; Boulton; Cobb, 2017; Tagnin; Viana, 2011; Berber Sardinha, 2011; 2005; 2004; Biber, 2009; Sinclair, 2005; 1991; Johns, 1988). The research participants are 68 students from the Graduate Programs of São Paulo State University (Unesp), who enrolled in the online elective course "Reading Strategies in English – Basic Level". The activities were developed based on the keywords extracted from a corpus of scientific articles authored by researchers addressing the United Nations' Sustainable Development Goals (SDGs), as established in the 2030 Agenda. Since these keywords appear in the works of international researchers from various fields, they can be understood as recurrent multi-word terms in academic vocabulary, making them relevant for learners of English for Specific Purposes. The study corpus, SDG-Plos (Pinto et al, in press), contains 400 articles published in the PLOS ONE journal between 2011 and 2021 in the fields of science and medicine. The keywords were extracted from the list of multi-word terms generated by the Keywords tool of the Sketch Engine software (Kilgarriff et al, 2014), which guided the extraction of texts to be used in developing hands-on and hands-off activities with the help of the tools CorpusMate (Baisa; Crosthwaite, 2023), Skell (Baisa; Suchomel, 2014), and VersaText (Thomas, 2020). The activities were implemented in the classes of the second semester of 2023 and the first semester of 2024 and evaluated, regarding the use of the tools and the themes of the texts read, from the students' perspectives through a questionnaire. The results indicate how DDL can be beneficial for basic-level learners provided the necessary adaptations, how the SDGs can act as a multidisciplinary teaching resource in the EAP context, and how corpora-based materials can be replicated by language teachers who are not specialists in LC.

Keywords: Corpus linguistics. Data driven learning. Multiword terms. Reading strategies. Sustainable development goals.

LISTA DE ILUSTRAÇÕES

Figura 1 – Resultados do Programa de Língua Inglesa da Unesp (PLIU) promovido pela parceria Unesp-Santander.....25
Figura 2 – Número de Exames TOEFL aplicados na Unesp entre 2017 e 2019.....26
Figura 3 – Os 17 objetivos da Agenda 2030.....29
Figura 4 – Procedimento de palavras-chave.....37
Figura 5 – Estágios metodológicos para aulas de CE.....46
Figura 6 – Quadro de autoavaliação do Quadro Comum Europeu de Referência para Línguas – Habilidade de Compreensão.....48
Figura 7 – Modelo de aquisição, complexidade e formulaicidade para *design* de DDL.....50
Figura 8 – O *AntCorGen*.....55
Figura 9 – O *Sketch Engine*.....56
Figura 10 – Ferramenta *Keywords*.....60
Figura 11 – Palavras-chave da lista *Multi-Word Terms*.....61
Figura 12 – Cronograma do curso.....65
Figura 13 – Como usar o *VersaText* para leitura.....68
Figura 14 – O *CorpusMate*.....69
Figura 15 – O *Skell*.....70
Figura 16 – O *VersaText*.....71
Figura 17 – Estágios de elaboração das atividades.....72
Figura 18 – Exercício 1 da seção *pre-reading* da atividade 2.....74
Figura 19 – Exercícios 2 e 3 da seção *pre-reading* da atividade 2.....75
Figura 20 – Exercício 4 da seção *pre-reading* da atividade 2.....75
Figura 21 – Recurso *Distribution by topic* da ferramenta *CorpusMate*.....76
Figura 22 – Exercício 5 da seção *pre-reading* da atividade 2.....77
Figura 23 – Momento *Reading* da atividade 2.....77
Figura 24 – Exercício 1 da seção *post-reading* da atividade 2.....78
Figura 25 – Exercício 2 da seção *post-reading* da atividade 2.....79
Figura 26 – Exercícios 3 e 4 da seção *pre-reading* da atividade 2.....79
Figura 27 – Exercício 3 da seção *pre-reading* da atividade 3.....80
Figura 28 – Exercício 5 da seção *post-reading* da atividade 4.....81

Figura 29 – *Final Survey*.....82
Figura 30 – *Final Survey* - 2024.....84
Figura 31 – Linhas de concordância com *ecosystem* service geradas pelo *Concordance* no *Sketch Engine*.....98
Figura 32- Exercício 1 da atividade elaborada com o termo multipalavra *ecosystem service*.....99

Gráfico 1 – Respostas dadas na questão 1 do *Final Survey* (Turma 3).....102
Gráfico 2 – Respostas dadas na questão 2 do *Final Survey* (Turma 3).....102
Gráfico 3 – Respostas dadas na questão 3 do *Final Survey* (Turma 3).....103
Gráfico 4 – Respostas dadas na questão 4 do *Final Survey* (Turma 3).....103
Gráfico 5 – Respostas dadas na questão 1 do *Final Survey* (Turma 4).....106
Gráfico 6 – Respostas dadas na questão 2 do *Final Survey* (Turma 4).....106
Gráfico 7 – Respostas dadas na questão 3 do *Final Survey* (Turma 4).....107
Gráfico 8 – Respostas dadas na questão 4 do *Final Survey* (Turma 4).....107
Gráfico 9 – Respostas à pergunta 1 do *Final Survey* – 2024.....112
Gráfico 10 – Respostas à pergunta 2 do *Final Survey* – 2024.....112
Gráfico 11 – Respostas à pergunta 4 do *Final Survey* – 2024.....114
Gráfico 12 – Respostas à pergunta 5 do *Final Survey* – 2024.....115
Gráfico 13 – Respostas à pergunta 6 do *Final Survey* – 2024.....116
Gráfico 14 – Respostas à pergunta 7 do *Final Survey* – 2024.....117
Gráfico 15 – Respostas à pergunta 9 do *Final Survey* – 2024.....118
Gráfico 16 – Respostas à pergunta 10 do *Final Survey* – 2024.....119

Quadro 1 – Níveis de Proficiência do Quadro Comum Europeu de Referência para Línguas.....27
Quadro 2 – Critério obrigatórios para atividades com base em *corpora*.....52
Quadro 3 – Características do *corpus* utilizado na pesquisa.....58
Quadro 4 – Palavras-chave selecionadas.....62
Quadro 5 – Relação das palavras-chave e dos conteúdos referentes a cada atividade.....73
Quadro 6 – Dados da pergunta 2 do *Needs Analysis* – Turmas 3 e 4 de 2023.....87
Quadro 7 – Dados acadêmico-profissionais dos alunos da Turma 3.....88
Quadro 8 – Dados acadêmico-profissionais dos alunos da Turma 4.....90

Quadro 9 – Dados da pergunta 4 do *Needs Analysis* – Turmas 3 e 4 de 2023.....92
Quadro 10 – Dados da pergunta 7 do *Needs Analysis* – Turmas 3 e 4 de 2023.....93
Quadro 11 – Dados da pergunta 8 do *Needs Analysis* – Turmas 3 e 4 de 2023.....94
Quadro 12 – Frequências relativas dos termos multipalavras selecionados.....96
Quadro 13 – Relação entre as ferramentas de *corpora* voltadas ao ensino e seu emprego nas atividades.....120

LISTA DE TABELAS

Tabela 1 – Infomações do *corpus* de referência *enTenTen21*.....57
Tabela 2 – Itens e Formas do *corpus SDG-Plos*.....58

LISTA DE ABREVIATURAS E SIGLAS

ASL Aquisição de Segunda Língua

BraVE *Brazilian Virtual Exchange*

CsF Ciências sem Fronteiras

CLDP Centro de Línguas e Desenvolvimento de Professores

DDL *Data Driven Learning*

EAP *English for Academic Purposes*

EGAP *English for General Academic Purposes*

ELF *English as a Lingua Franca*

ESAP *English for Specific Academic Purposes*

ESP *English for Specific Purposes*

EST *English for Science and Technology*

IELTS *International English Language Testing System*

IES Instituição de Ensino Superior

IFA Inglês para Fins Acadêmicos

IFAG Inglês para Fins Acadêmicos Gerais

IFE Inglês para Fins Específicos

IFAE Inglês para Fins Acadêmicos Específicos

IFG Inglês para Fins Gerais

ILF Inglês como Língua Franca

ILI Inglês como Língua Internacional

IsF Idioma sem Fronteiras

LA Linguística Aplicada

LE Língua Estrangeira

LI Língua Inglesa

LinFE Línguas para Fins Específicos

LC Linguística de *Corpus*

MEC Ministério da Educação e Cultura

ODS Objetivos de Desenvolvimento Sustentável

ONU Organização das Nações Unidas

PLEU Programa de Línguas Estrangeiras da Unesp

PLIU Programa de Língua Inglesa da Unesp

PPG Programa de Pós-Graduação

PROPG Pró-Reitoria de Pós-Graduação

QCER Quadro Comum Europeu de Referência para Línguas

SLA *Second Language Aquisition*

TOEFL *Test of English as a Foreign Language*

UB *Usage-Based*

Unesp Universidade Estadual Paulista

SUMÁRIO

1 INTRODUÇÃO 17

2 FUNDAMENTAÇÃO TEÓRICA 21

2.1 Globalização e Internacionalização: impactos no ensino de Língua Inglesa 21

2.2 O ensino de Língua Inglesa e os Objetivos de Desenvolvimento Sustentável 26

2.3 O Inglês para Fins Específicos e o Inglês para Fins Acadêmicos 31

2.4 A Linguística de *Corpus* e suas contribuições para o ensino de Língua Estrangeira . 34
2.4.1 *Data Driven Learning* 40

2.5 O processo de leitura e a habilidade de compreensão escrita 43
2.5.1 Materiais didáticos baseados em *corpora*: foco na compreensão escrita no nível básico 49

3 MATERIAIS E MÉTODO 54

3.1 Extração dos dados 54
3.1.1 O *AntCorGen* 54
3.1.2 O *Sketch Engine* 55
3.1.3 O *corpus SDG-Plos* 57
3.1.4 O levantamento das palavras-chave do *corpus* 59

3.2 Aplicação pedagógica 63
3.2.1 O contexto de aplicação das atividades 63
3.2.2 As ferramentas de *corpora* voltadas ao ensino: *CorpusMate*, *Skell* e *VersaText* 68
3.2.3 A elaboração das atividades de leitura direcionadas por dados 71
3.2.4 Os questionários aplicados ao final do curso 82

4 ANÁLISE DOS DADOS 86

4.1 Dados do questionário *Needs Analysis* aplicado em 2023 86

4.2 As palavras-chave do *corpus SDG-Plos* entendidas como termos multipalavras dos ODS 95

4.3 Os princípios e os critérios que embasam a elaboração das atividades 100

4.4 A percepção dos alunos acerca das atividades 101
4.4.1 Dados do *Final Survey*: Turma 3 102
4.4.2 Dados do *Final Survey*: Turma 4 105
4.4.3 Dados do *Final Survey* – 2024 111

4.5 A aplicabilidade das atividades 121

5 CONSIDERAÇÕES FINAIS E ENCAMINHAMENTOS 123

REFERÊNCIAS 127

APÊNDICES 133

APÊNDICE A – Atividade elaborada a partir do termo multipalavra *protected area* .. 134

APÊNDICE B – Atividade elaborada a partir do termo multipalavra *waterbird habitat* 138

APÊNDICE C – Atividade elaborada a partir do termo multipalavra *ocean health* 142

APÊNDICE D – Atividade elaborada a partir do termo multipalavra *human well-being* 147

APÊNDICE E – Atividade elaborada a partir do termo multipalavra *ecosystem service* 151

1 INTRODUÇÃO

A Língua Inglesa (doravante LI) passou a ser considerada indispensável para o sucesso acadêmico de graduandos, pós-graduandos e demais pesquisadores (Hyland, 2018). O *status* de língua global que o inglês possui influencia diversos setores da sociedade, já que se impõe como a língua de intercâmbio econômico, político e cultural entre os países, fator que motiva o oferecimento de cursos de curta duração com propósitos específicos e acadêmicos nas universidades brasileiras.

Compreendido sob a perspectiva dos pressupostos teóricos e metodológicos das abordagens de Línguas Para Fins Específicos (LinFE), o ensino de línguas caracteriza-se por colocar as necessidades específicas dos alunos no centro da aprendizagem. A abordagem *English for Specific Purposes* (ESP), ou Inglês Para Fins Específicos (IFE), baseia-se no ensino de habilidades, gêneros e aspectos da língua dos quais os alunos mais necessitam, com o intuito de desenvolver capacidades linguísticas específicas para a comunicação em diversas atividades, sejam elas acadêmicas ou profissionais (Heemann, 2009).

No contexto acadêmico, com base no aporte teórico do *English for Academic Purposes* (EAP), ou Inglês Para Fins Acadêmicos (IFA), uma ramificação do ESP (O'Boyle; Viana, 2022), dá-se ênfase para as habilidades comunicativas próprias das atividades realizadas no âmbito universitário, nos cursos de graduação e pós-graduação.

Na Unesp, foram oferecidos cursos de LinFE no âmbito do programa Ciências sem Fronteiras (CsF), que acabaram tendo continuidade por meio do Programa Idiomas sem Fronteiras (IsF) e, posteriormente, com a extinção de ambos os programas, por meio do Programa de Língua Inglesa da Unesp (PLIU), parceria público-privada que se deu com o Convênio UNESP-Santander (Pinto *et al*, 2021; Baffi-Bonvino; Kaneko-Marques, 2020).

Diante do cenário de crescente incentivo à internacionalização das instituições de ensino superior, em 2018 a Unesp "[...] elaborou seu Plano Estratégico de Internacionalização para todos os setores da universidade com objetivo de desenvolver talentos globais, de pesquisa e de parcerias com instituições internacionais" (Pinto *et al*, 2021, p. 5621). A internacionalização das universidades depende do nível de proficiência linguística de sua comunidade, uma vez que as instituições que oferecem oportunidades de intercâmbio e mobilidade têm como língua franca o inglês (Baffi-Bonvino; Kaneko-Marques, 2020).

Dada a necessidade de desenvolver a proficiência de graduandos, pós-graduandos e do quadro de funcionários, a Unesp passou a investir no mapeamento da proficiência em LI e a

oferecer diversos cursos para a aprendizagem do idioma com fins acadêmicos. Conforme Pinto *et al* (2021, p. 5624):

> a universidade passou a investir em ações que pudessem aumentar o nível de proficiência em língua inglesa em todos os seus setores, seja de alunos de graduação e pós, docentes e pesquisadores quanto de técnicos-administrativos. Para tanto, contou com a especialidade de docentes-pesquisadores dos três cursos de graduação e pós-graduação em Letras e Linguística para observarem mais atentamente as necessidades de cada setor.

O curso Estratégias de Leitura em Língua Inglesa – Nível Básico, contexto em que se insere os participantes da presente pesquisa, ocorreu no âmbito do PLIU e, em 2020, teve como público-alvo 200 graduandos dos diversos campi da Unesp (Santos; Pinto, 2022). Em 2023 e 2024, o curso aconteceu no formato de disciplina optativa on-line (com aulas síncronas e assíncronas) para os alunos da pós-graduação de todos os campi, compondo o Portfólio de Integração de Disciplinas da Pós-Graduação da Unesp (PROPG), conjunto de disciplinas instrumentais, transversais e específicas voltadas para o corpo discente da instituição, no formato de ensino híbrido, com o objetivo de gerar a integração entre os Programas de Pós-Graduação[1] (PPGs) (Unesp, 2023).

A leitura de textos em inglês é considerada imprescindível para alunos dos cursos de nível superior, ainda que as outras habilidades linguísticas como compreensão oral, escrita e fala também sejam importantes. Por essa razão, esta pesquisa visa desenvolver atividades didáticas movidas por *corpus*, com base na abordagem *Data Driven Learning* (DDL), voltadas para a compreensão escrita em LI de alunos do curso Estratégias de Leitura em Língua Inglesa - Nível Básico, da Unesp.

Conforme Nesi (2016), a união entre a Linguística de *Corpus* (LC) e o IFA tornou-se comum uma vez que os *corpora* podem fornecer informações quantitativas sobre o discurso acadêmico, embasando informações qualitativas, além de possibilitarem oportunidades de aprendizagem baseadas em dados provindos do uso real da língua. A associação entre a DDL e o IFA é reconhecida em razão do fato de o caráter exploratório de atividades baseadas em *corpora* render bons frutos com aprendizes de níveis intermediários e avançados (Boulton; Cobb, 2017), mas o desempenho de alunos de níveis básicos por meio da DDL ainda é algo a se investigar (Mark; O'Keffe, 2022).

[1] Disponível em: https://www2.unesp.br/portal#!/propg/disciplinas-portfolio.

Da relação entre a LC e o ensino de línguas, surgem possibilidades de pesquisas como esta, que se configura como um dos desdobramentos do projeto "*The United Nations Sustainable Development Goals (SDG) in the Brazilian research: a corpus-based approach to support research-paper writing and translation*" coordenado por Pinto (2021; 2024), que tem como objetivo desenvolver materiais direcionados e baseados em *corpora* para o alunado de contexto acadêmico por meio da exploração de *corpora* especializados com foco nos objetivos da Agenda 2030. O projeto visa apoiar o processo de internacionalização das universidades por meio do desenvolvimento da proficiência linguística de graduandos e pós-graduandos com o apoio do ensino baseado e dirigido por *corpora* (Pinto, 2024).

O *corpus* de estudo utilizado neste trabalho é constituído por artigos científicos de pesquisadores internacionais que abordaram os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU) em suas pesquisas. Tendo servido como guia internacional para ações da sociedade civil, a Agenda 2030, documento em que estão inscritos os 17 objetivos, vem sendo seguida também por muitos pesquisadores, que têm abordado em seus trabalhos assuntos que remetem aos ODS, o que faz com que utilizem expressões recorrentes dentro de determinadas áreas de especialidade.

O Plano Estratégico de internacionalização da Unesp está alinhado às demandas do mundo globalizado, que toma estudantes e pesquisadores como "cidadãos globais", conceito que está "baseado no princípio de que todos os cidadãos do planeta têm os mesmos direitos e responsabilidades, um dos pilares para a construção dos Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da ONU" (Unesp, 2023). A associação do ensino aos objetivos da ONU pode ser entendida como uma das ações para internacionalização tomadas pela universidade, o que aproxima as pesquisas desenvolvidas nos PPGs aos temas abordados pela Agenda 2030.

Isso posto, é possível afirmar que os ODS são conhecidos pelos graduandos e pós-graduandos da instituição e que fazem parte do trabalho por eles realizado. Esses objetivos seriam, então, uma possível porta de entrada para o ensino-aprendizagem da LI para alunos/pesquisadores de diferentes áreas, que necessitam do inglês com propósitos específicos.

Esta pesquisa tem como objetivo geral elaborar atividades de leitura movidas por *corpus* com a temática dos ODS para estudantes de nível superior que apresentem nível básico de proficiência em LI, como mencionado, assim como avaliar sua aplicabilidade em contexto de um curso de IFA de uma universidade multicampi, a Unesp. São objetivos específicos:

i. levantar as palavras-chave do *corpus*;

ii. desenvolver atividades de leitura em LI por meio das ferramentas *CorpusMate* (Baisa; Crosthwaite, 2023), *Skell* (Baisa; Suchomel, 2014) e *VersaText* (Thomas, 2020), e;

iii. investigar a aplicabilidade das atividades desenvolvidas, com relação ao uso das ferramentas educacionais de *corpus* empregadas e aos temas dos ODS, por meio da análise de um questionário a ser respondido pelos alunos.

Diante do contexto de pesquisa apresentado e de nossos objetivos, foram feitas as seguintes perguntas para nortear este estudo:

1. Quais itens lexicais podem servir para o desenvolvimento de atividades didáticas voltadas para a compreensão escrita?
2. De que modo os ODS podem ser associados ao ensino de inglês em contexto acadêmico?
3. Em que medida atividades didáticas direcionadas por dados podem influenciar a aprendizagem de leitura em LI de estudantes de nível básico?

Esta pesquisa se caracteriza como quali-quantitativa, uma vez que os procedimentos de coleta e análise de dados se deram a partir da investigação de base empirista dos itens lexicais de um *corpus* e da avaliação de atividades didáticas por meio da interpretação de questionários voltados a estudantes.

Nas seções a seguir, serão apresentadas as bases teóricas das quais a pesquisa se vale, as ferramentas de *corpus* utilizadas para o trabalho, o contexto de aplicação em que os participantes estavam inseridos e em que as atividades foram aplicadas e os resultados encontrados.

2 FUNDAMENTAÇÃO TEÓRICA

A seguir, são apresentados os pressupostos teóricos que embasam a presente pesquisa. São discutidos o fenômeno da globalização e da internacionalização das universidades, a associação entre o ensino de LI e os ODS, a importância da abordagem Inglês Para Fins Específicos (IFE) e Inglês Para Fins Acadêmicos (IFA), as contribuições teórico-metodológicas da Linguística de *Corpus* (LC) para ensino de Língua Estrangeira (LE) e da abordagem Aprendizagem Direcionada por Dados, *Data Driven Learning* (DDL), as definições de leitura conforme os modelos ascendente, descendente e interativo e a adotada no estudo, bem como a importância de um *corpus* sobre os ODS.

2.1 Globalização e Internacionalização: impactos no ensino de Língua Inglesa

Entendida como o processo que "implica o livre fluxo de bens e serviços, e aliada à tecnologia e ao rápido fluxo de informação entre fronteiras que vão gradativamente desaparecendo, pode resultar numa sociedade mundial integrada" (Finardi; Porcino, 2014, p. 241) a globalização estabelece uma língua de uso comum (Siqueira, 2011), o inglês, idioma historicamente relacionado ao imperialismo cultural, que antes se caracterizava pelo colonialismo britânico, mas que agora se apresenta sob a perspectiva da hegemonia norte-americana.

Conforme Siqueira (2011), o inglês, quando passa a carregar o *status* de língua global, se sujeita às mudanças provindas das marcas de primeira língua dos falantes não-nativos, que se apropriam do idioma para integrar o mundo globalizado, para participar da vida acadêmico-científica, por exemplo. De acordo com Crystal (2003, p. 2, tradução nossa), "[...] se há uma consequência previsível de uma língua se tornar uma língua global, é que ninguém mais a possui. Ou melhor, todos que a aprendem passam a possuí-la"[2].

Crystal (2003, p. 3, tradução nossa) aborda o conceito de língua global e o define como uma língua que "desenvolve um papel especial que é reconhecido em todos os países"[3], papel esse que é influenciado pelo poderio político, militar, econômico e cultural de determinado país cuja língua tenha alcançado o caráter mencionado (Crystal, 2003). Já Jenkins (2014, p. 2), ao analisar o ensino-aprendizagem de LI como uma das demandas da globalização e do contexto

[2] No original: *[...] if there is one predictable consequence of a language becoming a global language, it is that nobody owns it any more. Or rather, everyone who has learned it now owns it [...]* (Crystal, 2003, p. 2)
[3] No original: *[...] it develops a special role that is recognized in every country.* (Crystal, 2003, p. 3)

de internacionalização, aborda o Inglês como Língua Franca (ILF), ou *English as a Lingua Franca* (ELF), entendido como "o inglês quando usado como língua de contato entre pessoas que possuem idiomas diferentes como primeira língua (incluindo falantes nativos de inglês)"[4].

À medida que a globalização se intensifica e as relações econômicas, políticas e socias se estreitam, a corrida pela aprendizagem do inglês também ganha força. As Instituições de Ensino Superior (doravante IES) demonstram forte interesse em garantir que a LI seja amplamente difundida em suas unidades para que as parcerias com grandes centros de pesquisa ao redor do mundo sejam possíveis. Ao processo de inserção das IES ao contexto de compartilhamento e construção do saber científico a nível internacional dá-se o nome de internacionalização. Conforme Santos, *et al* (2021, p. 61),

> a internacionalização das IES é um passo na direção da busca pela construção de uma sociedade mais igualitária, em que os alunos possam ter melhores condições para compartilhamento de conhecimento científico com seus pares, independente da língua em que as trocas possam ser estabelecidas. A partir do momento em que a língua não se constitui em uma barreira de acesso ao conhecimento, todos podem interagir igualmente, defender suas pesquisas e conquistar novos espaços.

É sabido que a língua já se manifesta como uma barreira ao acesso à informação, uma vez que a proficiência em LI representou e ainda representa um desafio para a internacionalização das IES brasileiras (Baffi-Bonvino; Kaneko-Marques, 2020; Santos *et al*, 2021; Pinto *et al*, 2021).

Ao compreender o inglês como língua franca, Jenkins (2014) visa problematizar a forma como ele tem sido ensinado nas universidades, de modo a refletir sobre "o tipo de inglês que é requisitado e promovido no ES (Ensino Superior) global, e até que ponto reflete ou não reflete a composição da instituição, e, por sua vez os tipos de inglês usados quando o idioma atua como uma língua franca internacional [...]"[5] (Jenkins, 2014, p. 2, tradução nossa). A autora revela uma preocupação com o que de fato as instituições de ensino superior internacionais têm tomado como língua franca: o inglês de falantes nativos ou de falantes não-nativos, visto que a aprendizagem de inglês como língua franca é reflexo de uma necessidade imposta aos falantes

[4] No original: *[...] English when it is used as a contact language between people from different first languages (including native English speakers).* (Jenkins, 2014, p. 2)
[5] No original: *[...] the kind of English that is required and promoted in global HE, and the extent to which it does/does not reflect the diversity of the institution's composition and, in turn, the kinds of English used when the language performs an international lingua franca [...]* (Jenkins, 2014, p. 2)

do segundo grupo, fator que influencia sua integração a determinados espaços considerados privilegiados.

Ao se pensar o ensino de ILF ou mesmo como língua internacional (ILI), segundo Siqueira (2011, p. 347), "o que importa [...] são as mútuas acomodações para promover inteligibilidade, condição que garante ao falante de ILF certa vantagem que, necessariamente, não é a mais internacionalmente inteligível". Nessa perspectiva, o ensino de LI se dá a partir de políticas pedagógicas que valorizem o caráter cultural da língua de quem a fala, ou seja, de um currículo internacionalizado, que, para Baffi-Bonvino e Kaneko-Marques (2020, p. 514) "[...] prevê o atendimento de alunos locais e estrangeiros de forma a prepará-los para atuarem em contextos multiculturais e internacionais".

Como afirmam Santos *et al* (2021, p. 59),

> Ao trabalharmos com a internacionalização e o ensino de línguas, devemos estar atentos a esses processos transculturais, de modo que nosso(a) aluno(a) seja capaz não apenas de interagir na língua alvo, como também de se posicionar criticamente e de perceber o papel do outro e dos espaços conquistados.

Fica clara a importância do ensino de forma crítica e atenta às relações de poder implícitas ao processo de aprendizagem de uma língua internacional. A internacionalização se dá por meio de políticas asseguradoras do desenvolvimento de atividades diversificadas a partir de determinadas estratégias tomadas em âmbito institucional voltadas para a mobilidade de estudantes e professores, para a parceria em pesquisas e para publicação de resultados, por exemplo (Santos *et al*, 2021). Para as autoras, o programa IsF "pode ser um importante auxiliar no processo de internacionalização desenhado por cada IES" (Santos *et al*, 2021, p. 62) pois, por meio dele, as universidades

> podem incluir em seus currículos temas que contribuam para o desenvolvimento das habilidades esperadas entre os membros da comunidade acadêmica que estão envoltos em programas de internacionalização, e esses temas podem estar disseminados nas ações propostas pelo programa. (Santos *et al*, 2021, p. 63)

O IsF surgiu a partir do Programa Ciências sem Fronteiras (CsF), criado em 2011 como uma iniciativa do Ministério da Educação (MEC), do Ministério da Ciência, Tecnologia e Inovação (MCTI) e de instituições de fomento a pesquisa como o CNPq e a CAPES para "[...] acompanhar os avanços na produção de conhecimento com a globalização e

internacionalização, tendo em vista o investimento em áreas de ciências exatas, biológicas e tecnológicas (*Hard Sciences*)" (Garcia; Kaneko-Marques, 2019, p. 243).

Segundo Garcia e Kaneko-Marques (2019, p. 244), em relação às IES globais, "muitas instituições e centros em países de língua inglesa não recebiam o número esperado de candidaturas de alunos brasileiros" e muitos estudantes não tinham acesso a testes de proficiência. O CsF representou a consolidação da internacionalização das IES brasileiras e tinha como um de seus objetivos ampliar as oportunidades de acesso a cursos de ensino de LE, pricipalmente o inglês, além de divulgar e oferecer, por meio dos centros de línguas das universidades, os testes niveladores de proficiência mundialmente reconhecidos (Gracia; Kaneko-Marques, 2019).

O IsF surgiu em 2012 a partir da criação de centros de aplicação do teste TOFL-ITP (*Test of English as a Foreign Language - Institutional Testing Program*) e do oferecimento de cursos de outras LEs além do inglês, como espanhol, francês, italiano, alemão, japonês e português para estrangeiros, de forma presencial e on-line, por meio do *My English Online* (Garcia; Kaneko-Marques, 2019).

Conforme as autoras,

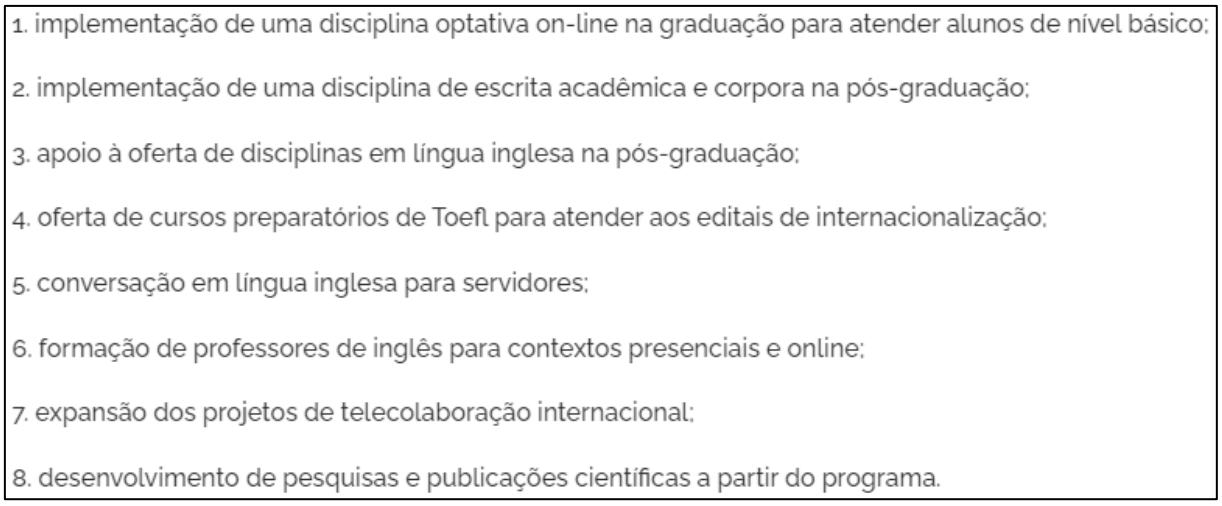
> O IsF tornou-se uma importante iniciativa para auxiliar o processo de internacionalização e contribuir para o desenvolvimento de uma política linguística nas universidades brasileiras, uma vez que visa promover a residência docente para os profissionais do ensino de línguas estrangeiras, a capacitação da comunidade universitária em línguas estrangeiras, bem como o ensino de língua portuguesa para estrangeiros. (Garcia; Kaneko-Marques, 2019, p. 244-245).

Tanto o CsF como o IsF foram interrompidos pelo Ministério da Educação e Cultura (MEC), o primeiro tendo perdurado entre o período de 2011 a 2016 e o segundo de 2012 a 2019. O IsF oportunizou a aprendizagem de um segundo idioma para muitos brasileiros e representa um marco das ações em prol da internacionalização das IES no Brasil. Conforme Pinto *et al* (2021, p. 5620), "Diante da grandiosidade do Programa, é importante destacar que esta rede contemplou parcerias nacionais e internacionais que vislumbraram uma visão eficiente e especializada acerca do processo de ensinar e aprender línguas estrangeiras".

Além desses programas, uma parceria público-privada surgiu em 2017 entre a Unesp e o banco Santander para promover cursos de inglês com fins específicos e acadêmicos, voltados para a melhoria da proficiência em inglês da comunidade acadêmica, o Programa de Língua Inglesa da Unesp: internacionalização, formação e transversalidade (PLIU) (Unesp, 2021). De

acordo com a Unesp (2021), “O programa partiu do Plano de Internacionalização da Unesp e do mapeamento da proficiência em língua inglesa da comunidade unespiana para a implementação de ações transversais apoiadas em sua Política Linguística de Internacionalização” e garantiu oito ações desenvolvidas na universidade:

Figura 1 - Resultados do Programa de Língua Inglesa da Unesp (PLIU) promovido pela parceria Unesp-Santander



Fonte: Captura de tela do site da Unesp (Unesp, 2021)[6].

O Estratégias de Leitura em Língua Inglesa – Nível Básico é resultado dessas ações implementadas na Unesp. No ano de 2020, o curso teve 200 alunos matriculados e 179 concluintes (Santos; Pinto, 2022). Atualmente, o PLIU se denomina Programa de Línguas Estrangeiras da Unesp (PLEU), pois contempla outras línguas além do inglês.

No âmbito do PLEU, outras iniciativas voltadas para o aprimoramento da proficiência linguística com vistas à promoção da internacionalização acontecem na instituição, tais como o oferecimento de curso de escrita acadêmica, de Inglês para Fins Ocupacionais por meio do Centro de Línguas e Desenvolvimento de Professores (CLDP) e de intercâmbio virtual por meio do Teletandem e do *Brazilian Virtual Exchange* (BraVE) (Unesp, 2020), programas de aprendizagem on-line colaborativa entre pelo menos dois aprendizes cujas línguas maternas (LM) sejam as línguas-alvo um do outro.

Esta pesquisa tem como contexto de realização o atual cenário de ensino-aprendizagem de LI como língua internacional. Assim, se vale das teorias de ensino de línguas baseadas no

[6] Disponível em: https://www.ibilce.unesp.br/#!/centro-de-linguas-1/idiomas-sem-fronteiras.

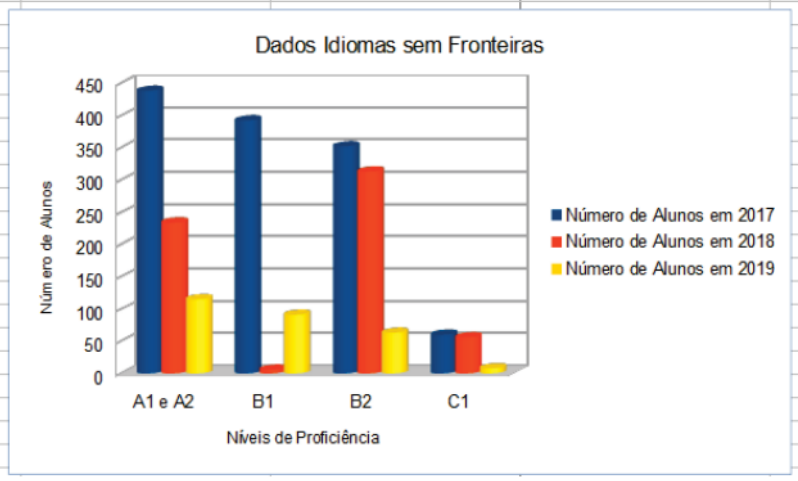
IFE e no IFA, no uso de TICs (Tecnologias de Informação e Comunicação) e TDICs (Tecnologias Digitais da Informação e Comunicação), e na Linguística de *Corpus*, área da linguística que contribui muito com o ensino de línguas, especialmente no campo acadêmico.

2.2 O ensino de Língua Inglesa e os Objetivos de Desenvolvimento Sustentável

Programas de mobilidade e de parcerias internacionais requerem o conhecimento intermediário ou avançado da LI, conforme Pinto *et al* (2021, p. 5623), que afirmam que "[...] os níveis C1 e C2 são mais apropriados" para as avaliações de proficiência em Língua Estrangeira (LE), tais como as provas do *International English Language Testing System* (IELTS) e do *Test of English as a Foreign Language* (TOEFL). No entanto, os resultados de uma pesquisa feita pelo IsF revelou que entre os anos de 2017 e 2019 cerca de 40% dos graduandos dos campi da Unesp demonstraram os níveis de proficiência A1 e A2, referentes ao nível básico conforme o Quadro Comum Europeu de Referência para Línguas (QCER)[7] (Pinto; *et al*, 2021), como é possível observar na imagem a seguir:

Figura 2 - Número de exames TOEFL aplicados na Unesp entre 2017 e 2019

Dados Idiomas Sem Fronteiras				
Níveis de Proficiência	Número de Alunos em 2017	Número de Alunos em 2018	Número de Alunos em 2019	Total de Alunos no Triênio
A1 e A2	438	234	115	
B1	392	6	91	
B2	352	313	63	
C1	60	56	8	
Total de Alunos	1242	609	277	2128



Fonte: Dados da plataforma IsF Gestão do Ministério *apud* Pinto *et al*, 2021, p. 5623.

[7] Quadro Comum Europeu de Referência para Línguas (Common European Framework of References for Languages – CEFR): https://www.britishcouncil.org.br/quadro-comum-europeu-de-referencia-para-linguas-cefr. O QCER será apresentado e retomado no decorrer do trabalho.

Conforme Pinto *et al* (2021), o gráfico apresentado na imagem indica a quantidade de exames do Toefl que foram aplicados na comunidade da Unesp entre 2017 e 2019 e revela o seu nível de proficiência em inglês:

> nos três anos em que houve aplicação da prova de TOEFL na UNESP, os resultados mostraram que a porcentagem de alunos com níveis básicos de proficiência em língua inglesa variou entre, aproximadamente, 35% a 41%, já a porcentagem de alunos nos níveis intermediários variou entre 22% e 55% e para o nível avançado, a porcentagem variou entre, aproximadamente, 3% e 9,20%. (Pinto *et al*, 2021, p. 5624)

Ainda que os resultados da pesquisa indiquem os níveis de proficiência em LI de forma geral, ou seja, não foquem na proficiência das quatro habilidades (*listening*, *speaking*, *writing*, *reading*) isoladamente, direcionar a atenção para alunos iniciantes se mostra imprescindível dado que esse perfil de alunado pode não contar com um repertório de vocabulário amplo, fator que influencia em sua compreensão leitora (Baldo, 2009).

A respeito das denominações para níveis básico, intermediário e avançado usadas pelo QCER, cabe informar o que indicam e qual sua importância no que se refere aos conhecimentos ligados às habilidades linguísticas.

O QCER é definido como "[...] um padrão internacionalmente reconhecido para descrever a proficiência em um idioma" (*British Council*, 2023). Ele apresenta três níveis de proficiência, cada um com duas subdivisões:

Quadro 1 - Níveis de Proficiência do Quadro Comum Europeu de Referência para Línguas

Níveis de Proficiência conforme QCER	**Subdivisões**
Básico	A1
	A2
Intermediário	B1
	B2
Avançado	C1
	C2

Fonte: Elaborado com base em *British Council* (2023).

Essa organização dos níveis de proficiência de línguas recebe o nome de escala global (*British Council*, 2023) e indica o que é esperado do falante/aprendiz que se encontra em cada uma das subdivisões. A escala global não avalia cada uma das habilidades linguísticas separadamente, mas traz algumas tarefas que são capazes de ser desempenhadas com maior ou

menor ênfase pelo estudante de cada nível. Como se sabe, os níveis intermediário e avançado são os requeridos pelos testes de proficiência aplicados nas universidades, pois indicam maiores capacidades e habilidades referentes ao uso das línguas em diferentes contextos de comunicação.

As habilidades de fala e de escrita são muito importantes para aqueles que almejam participar de um programa de mobilidade, pois atividades como escrita de artigos científicos e de relatórios e apresentações orais em congressos internacionais são atividades comuns para os pesquisadores. Contudo, dado o nível de proficiência apresentado pela pesquisa do IsF, faz-se necessário voltar o olhar para os graduandos e pós-graduandos de nível básico que já vivenciam o contexto acadêmico e têm que lidar com as demandas necessárias à pesquisa no mundo atual. O curso Estratégias de Leitura em Língua Inglesa é o resultado dessa necessidade real do alunado da Unesp e o presente estudo visa contribuir com a aprendizagem desses estudantes.

A pesquisa assume as abordagens IFA e DDL como abordagens de ensino ao direcionar o foco de análise para uma necessidade de uso real de uma habilidade específica, a leitura em LI, para um público específico determinado, estudantes de nível básico. Com relação ao uso de um *corpus* que tematize os ODS, pode-se afirmar que o fato de os objetivos da ONU abrangerem áreas diversas pode conferir aos textos em que são abordados vocabulário relevante para os estudantes do curso em que as atividades serão aplicadas, que também são de cursos de áreas diferentes entre si.

Guia para a comunidade internacional e diretriz de ações para uma sociedade mais sustentável, a Agenda 2030 foi criada em 2015, em Nova Iorque, por representantes de 193 Estados-Membros da ONU, que juntos firmaram o compromisso de atingir 17 objetivos, até o ano de 2030, que promovam o desenvolvimento, a parceria entre os países e um futuro melhor para o planeta (*United Nations*, 2023). Esses objetivos são os chamados Objetivos de Desenvolvimento Sustentável (ODS), ou em inglês, *Sustainable Development Goals* (SDG).

Conforme consta no documento, os ODS englobam ações para acabar com a pobreza, proteger o meio ambiente e o clima e garantir condições dignas de vida para todos os povos (*United Nations*, 2023). Os objetivos estabelecidos entre a ONU, o Brasil e outros países do mundo para o enfrentamento de vários problemas sociais são: 1. Erradicação da pobreza; 2. Fome zero e agricultura sustentável; 3. Saúde e bem-estar; 4. Educação de qualidade; 5. Igualdade de gênero; 6. Água potável e saneamento; 7. Energia limpa e acessível; 8. Trabalho decente e crescimento econômico; 9. Indústria, inovação e infraestrutura; 10. Redução das desigualdades; 11. Cidades e comunidades sustentáveis; 12. Consumo e produção sustentáveis;

13. Ação contra a mudança global do clima; 14. Vida na água; 15. Vida terrestre; 16. Paz, justiça e instituições eficazes; e 17. Parcerias e meios de implementação. A imagem a seguir ilustra os objetivos citados:

Figura 3 - Os 17 objetivos da Agenda 2030

Fonte: *United Nations* (2023).

De acordo com Pinto (*et al*, 2023, p. 2), "os Objetivos de Desenvolvimento Sustentável tornaram-se um aspecto central discutido por investigadores de diferentes países para encontrar soluções para as principais questões sociais, a nível local e global"[8]. Conforme os autores, as publicações científicas do período pós-pandemia de Covid-19 têm abordado os objetivos da ONU nas áreas de Saúde, Sociologia, Economia, Química, Matemática, Linguística e algumas outras (Pinto *et al*, 2023). Tratar, nas aulas, dos problemas e questões que tocam as sociedades atuais mostra-se de extrema importância, já que o uso da língua deve refletir a realidade dos falantes.

A associação entre os ODS e o inglês com propósitos específicos e acadêmicos pode ser observada em materiais de ensino internacionalmente, como é o caso do *British Council* (2017), que publicou um livro organizado por Alan Maley e Nik Peachey com planos de aulas voltados para as quatro habilidades linguísticas e para alunos do nível iniciante ao avançado e que

[8] No original: *[...] the UN Sustainable Development Goals have become a central aspect discussed by researchers of different countries to find solutions for major social issues, in local and global levels [...].* (Pinto *et al*, 2023, p. 2)

tematizam os objetivos da Agenda 2030[9]. Conforme Bolster e Livrai (2017, p. 195, tradução nossa),

> [...] encontrar tópicos que são relevantes e atraentes para estudantes de uma variedade de disciplinas pode ser uma tarefa difícil para professores e desenvolvedores de cursos. Os ODS prestam-se ao estudo multidisciplinar e pode ser abordado através das lentes de diferentes campos de estudo acadêmico, o que os torna uma excelente área de conteúdo para um curso de EAP, envolvendo os alunos intelectualmente, academicamente e emocionalmente.[10]

Para os autores, os objetivos da ONU são particularmente relevantes para alunos em contexto acadêmico devido a sua capacidade de integrar propostas de aulas para alunos de cursos variados. Além disso, as ações do Plano Estratégico de Internacionalização da Unesp buscam um alinhamento das atividades acadêmicas com os ODS (Pinto *et al*, 2021). Conforme consta no site oficial da universidade (Unesp, 2018),

> Os ODS irão contribuir para a internacionalização da UNESP em dois sentidos: o primeiro é a adesão da Universidade às preocupações que pautam uma agenda mundial, posicionando-a em consonância com a pesquisa e o ensino atualmente desenvolvidos em muitas das principais universidades de classe mundial.
> O segundo aspecto positivo da inserção dos ODS à internacionalização da UNESP é a reorganização dos atuais programas de pós-graduação da Universidade, que somando esforços entre si, cada um em sua expertise, facilitam o alinhamento da instituição a iniciativas análogas de instituições internacionais parceiras, para compartilhar ciência e tecnologia, gerar conhecimento e buscar soluções para os principais problemas globais. (Unesp, 2018)

Os estudantes de pós-graduação já estão vivenciando o cenário de implementação do plano supracitado, logo, já leem textos que tratam dos ODS referentes aos seus trabalhos e às suas pesquisas. Sendo assim, é proveitoso utilizar artigos científicos, um dos gêneros mais lidos em contexto acadêmico, que associam os objetivos da ONU às áreas de especialidade dos alunos para o ensino e aprendizagem de compreensão escrita. Desta forma, estariam em contato com o vocabulário recorrente em suas respectivas áreas, desenvolveriam as estratégias de

[9] Disponivel em: www.teachingenglish.org.uk/publications/resource-books/integrating-global-issues-creative-english-language-classroom.

[10] No original: *[...] finding topics that are relevant and appealing to students from a variety of disciplines can be a difficult task for course designers and teachers. The SDGs lend themselves to multidisciplinary study and can be approached through the lens of different fields of academic study, which makes them an excellent content area for an EAP course, engaging students intellectually, academically and emotionally.* (Bolster; Livrai, 2017, p. 195)

leitura necessárias para o IFE e mobilizariam reflexões pertinentes a respeito de temas que estão no palco dos debates sociais.

Tendo em vista a relevância dos ODS e da Agenda 2030 enquanto meios de viabilização da internacionalização na universidade além do desenvolvimento da proficiência em LI, apresentamos, na seção seguinte, a abordagem Inglês para Fins Acadêmicos, que embasa o ensino e o desenvolvimento de material didático em inglês para aprendizes em contexto acadêmico, perfil dos estudantes que participam da presente pesquisa.

2.3 O Inglês para Fins Específicos e o Inglês para Fins Acadêmicos

Após o fim da Segunda Guerra Mundial, em 1945, o mundo vivenciou uma expansão sem precedentes na ciência, tecnologia e economia, o que gerou uma demanda por uma língua internacional que, devido ao poderio econômico dos Estados Unidos no pós-guerra, entre outros fatores, veio a ser o inglês (Hutchinson; Waters, 1987). De acordo com Hutchinson e Waters (1987, p. 6, tradução nossa), "o inglês se tornou a língua internacional da tecnologia e do comércio aceita, e isso criou uma nova geração de aprendizes que sabiam especificamente por que estavam aprendendo uma língua"[11].

Uma mudança no modo de pensar a linguagem e, consequentemente, no modo de ensinar línguas, também motivou o surgimento de uma nova uma abordagem, segundo Hutchinson e Waters (1987). A tradicional preocupação em descrever regras gramaticais de uso do inglês passa a dar lugar para uma nova perspectiva de análise: a investigação do modo como a língua era de fato usada na comunicação (Hutchinson; Waters, 1987), pois como afirmam os autores, "se a língua varia de uma situação de uso para outra, seria possível determinar as características de situações específicas e então torná-las a base do curso dos aprendizes"[12] (Hutchinson; Waters 1987, p. 7, tradução nossa)

Na década de 1960, professores de LE "viram a necessidade de ensinar a língua alvo para áreas especializadas; ou seja, inglês para Negócios, inglês para Medicina, inglês para Hotelaria, etc." (Heemann, 2009, p. 139). Assim, na iminência de se pensar o ensino de inglês focado nas necessidades específicas dos aprendizes, são dadas as condições para a consolidação

[11] No original: *[...] English became the accepted international language of technology and commerce, it created a new generation of of learners who knew specifically why they were learning a language [...]* (Hutchinson; Waters, 1987, p. 6)

[12] No original: *[...] if language varies from one situation of use to another, it should be possible to determine the features of specific situations and make these features the basis of the learners' course.* (Hutchinson; Waters, 1987, p. 7)

e o avanço da abordagem *English for Specific Purposes (ESP)*, ou Inglês para Fins Específicos (IFE).

O IFE tem sido reconhecido pela predominância das pesquisas em gêneros textuais, mas também pelo expressivo aumento das pesquisas baseadas em *corpus*, especialmente para a análise do discurso acadêmico em sua modalidade escrita (Johns, 2013). A LC proporcionou amostras reais do uso corrente da língua aos linguistas, que tiveram acesso aos padrões linguísticos comuns dos gêneros e puderam embasar com dados estatísticos as pesquisas a respeito das suas características prototípicas (Nesi, 2016), logo, "os estudos sobre gêneros textuais e a Linguística de Corpus [...] contribuem, sobremaneira, a essa modalidade de ensino" (Nadin, 2022, p. 103).

No século XXI, a LC tem se aproximado cada vez mais das questões relacionadas ao ensino e aprendizagem de línguas, seja por meio das teorias/abordagens de ensino seja por meio da produção de material didático (Tagnin; Finatto; From, 2021). De acordo com Bocorny e Welp (2021, p. 1594-1595),

> Os estudos a partir de grande volume de dados (*big data*) com a utilização de ferramentas potentes de análise de *corpora*, conduzidos pelos princípios e pelas metodologias propostos pela Linguística de *Corpus*, possibilitam a extração de dados linguísticos que podem contribuir com o desenvolvimento de aplicações pedagógicas usadas no ensino de diferentes gêneros textuais.

Isso posto, entende-se o papel fundamental da associação entre a LC e a Linguística Aplicada (LA) no que diz respeito ao ensino de línguas, especialmente no contexto acadêmico. A partir das ferramentas de *corpora* é possível investigar o discurso acadêmico de modo a promover uma conscientização dos aprendizes a respeito do modo como a língua é empregada na academia e, em especial, a LI. Assim, as tarefas baseadas e direcionadas por dados ganham destaque por serem pautadas no empirismo e não na memorização de regras.

Como se vê, o IFE é uma das abordagens mais bem estabelecidas no ensino de LI (Anthony, 2018), que coloca as necessidades dos alunos no centro da aprendizagem (Hutchinson; Waters, 1987). Conforme Anthony (2018, p. 10-11, tradução nossa),

> Inglês para Fins Específicos (IFE) é uma abordagem de ensino de línguas que visa às necessidades acadêmicas ou ocupacionais atuais e/ou futuras dos alunos, concentra-se no idioma, nos gêneros e nas habilidades necessários para atender a essas necessidades e auxilia os alunos a atenderem a essas demandas

> através do uso de materiais e métodos de ensino gerais e/ou específicos de disciplinas.[13]

Trata-se de uma abordagem que se vale da análise prévia de necessidades de aprendizagem, na análise de gêneros textuais, nos métodos comunicativos de ensino e no construtivismo (Hyland, 2022), diferenciando-se do Inglês para Fins Gerais (IFG), que está pautado no ensino das quatro habilidades linguísticas e na formação de falantes proficientes sem propósitos acadêmicos ou profissionais previamente definidos.

Por dar ênfase às necessidades de aprendizes de contextos acadêmicos e profissionais variados, o IFE desdobrou-se em várias outras abordagens consideradas ramificações. O EAP, Inglês para Fins Acadêmicos (IFA), em português, é talvez a ramificação mais influente do IFE (Anthony, 2018). Segundo Viana e O'Boyle (2022, p. 2, tradução nossa), o IFA

> [...] refere-se ao aprendizado, ensino e pesquisa de práticas de linguagem e discurso em inglês que as pessoas usam para estudar e/ou trabalhar no ensino superior. Ele se baseia em uma variedade de áreas temáticas, como Linguística Aplicada, Educação e Linguística.[14]

De acordo com Hyland (2006), o IFA pode ainda ser dividido em mais duas ramificações, o *English for General Academic Purposes* (EGAP), ou Inglês para Fins Acadêmicos Gerais (IFAG), e o *English for Specific Academic Purposes* (ESAP), Inglês para Fins Acadêmicos Específicos (IFAE). O primeiro compreende o ensino focado em habilidades, conteúdos linguísticos e atividades comuns a todas as disciplinas de forma geral, enquanto o segundo compreende o ensino da língua e das habilidades relacionadas às demandas particulares de uma disciplina (Hyland, 2006).

No Brasil, o IFE foi introduzido por meio do Projeto "Ensino de Inglês Instrumental em Universidades Brasileiras" a partir de 1970 na Pontifícia Universidade Católica (PUC-SP) sob a orientação e supervisão da Prof.ª Dr.ª Maria Antonieta Alba Celani (Celani *et al*, 2005). A denominação "Inglês Instrumental" se justificava pelo caráter funcional da língua: o de usá-la como instrumento para a execução de tarefas dentro de uma determinada área de especialidade.

---

[13] No original: *English for Specifi c Purposes (ESP) is an approach to language teaching that targets the current and/or future academic or occupational needs of learners, focuses on the necessary language, genres, and skills to address these needs, and assists learners in meeting these needs through the use of general and/or discipline-specifi c teaching materials and methods.* (Anthony, 2018, p. 10-11)

[14] No original: *[...] refers to the learning, teaching and research of language and discourse practices in English that people use to study and/or work in higher education. It draws on a range of subject areas such as Applied Linguistics, Education and Linguistics.* (Viana; O'Boyle, 2022, p. 2)

Conforme Celani (1983, p. 4), o inglês já era uma necessidade naquele momento para os estudantes brasileiros e a habilidade mais requerida era a leitura: "a necessidade de inglês, particularmente para a finalidade de leitura, é indiscutível, quer em relação a um curso determinado, quer para complementação dos estudos [...]". Por ter chegado ao Brasil como uma disciplina com foco para a CE, o IFE acabou sendo associado ao ensino de leitura instrumental, no entanto, já é de amplo conhecimento que a abordagem visa ao ensino de quaisquer habilidades, desde que sejam elencadas como prioritárias a determinado grupo de aprendizes.

Este estudo se vale da abordagem IFE, pois objetiva a elaboração de material didático baseado em *corpus* para o estudo de compreensão escrita, assim como se apoia no IFA e, mais especificamente no IFAG, por ter como participantes da pesquisa pós-graduandos de cursos diversificados que precisam da leitura em inglês com finalidades acadêmicas, sejam elas ter acesso a bibliografia publicada em LI ou realizar exames de proficiência em formato de prova escrita, entre outras.

A seguir, são apresentadas as aplicações pedagógicas dos dados provindos de *corpora* no ensino de IFA, que fundamentam e norteiam as atividades propostas na pesquisa.

2.4 A Linguística de *Corpus* e suas contribuições para o ensino de Língua Estrangeira

Conforme Berber Sardinha (2004, p. 3), a Linguística de *Corpus* (LC) é uma área da Linguística que se volta para a "exploração da linguagem por meio de evidências empíricas, extraídas por computador". Trata-se de uma corrente linguística empirista que se opõe às teorias racionalistas da linguagem por priorizar a investigação de extensas amostras da língua em uso a partir da qual inúmeros trabalhos renomados na Linguística e no ensino de línguas surgiram, como o dicionário *Cobuild*, de Sinclair (1987), primeiro a ser compilado com base em *corpus* computadorizado (Berber Sardinha, 2004), e a Gramática da Língua Inglesa baseada em *corpus*, de Biber *et al* (1999).

De acordo com Berber Sardinha (2004, p. 3), *corpora* são "conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística". Por não se pautar em especulações a respeito da linguagem, mas sim na análise dos padrões efetivamente frequentes nas línguas por meio de ferramentas computacionais, a LC associada ao ensino possibilita o uso de *corpora* em sala de aula, que permite ao aluno entrar em contato com a língua em uso e criar maior consciência léxico-gramatical e autonomia na aprendizagem (Zanettin, 2009).

Assim, a LC contribui com o desenvolvimento deste trabalho por meio do fornecimento de materiais autênticos e representativos, organizados sob a forma de um *corpus*. Ainda conforme Berber Sardinha (2000, p. 338), "[...] o *corpus* deve ser composto de textos autênticos, em linguagem natural. Dessa forma, os textos não podem ter sido produzidos com o propósito de serem alvo de pesquisa linguística. E não podem ter sido criados em linguagem artificial [...]".

Corpora são construídos para o estudo da linguagem, então, seu conteúdo deve ser selecionado de modo a contribuir para o alcance dos propósitos de uma pesquisa (Sinclair, 2005). Em outras palavras, os textos selecionados para a compilação de um *corpus* devem ser representativos de determinada amostra da língua que se pretende investigar.

Com relação à noção de representatividade de um *corpus*, Berber Sardinha (2004, p. 22) afirma que "Na sua essência, um *corpus*, seja de que tipo for, é tido como representativo da linguagem, de um idioma, ou de uma variedade dele" (Berber Sardinha, 2004, p. 22) e que a função representativa que carrega está associada à sua extensão, pois para a LC a linguagem é um sistema probabilístico em que "certos traços são mais frequentes que outros" (Berber Sardinha, 2004, p. 23). Desse modo, quanto maior o *corpus*, maior a probabilidade de ocorrência de palavras com as quais se pretende trabalhar.

Para alcançar a representatividade, olha-se para a amostragem: importa conhecer a probabilidade de ocorrência de traços lexicais, estruturais, pragmáticos e discursivos (Berber Sardinha, 2004). Para Berber Sardinha (2004), a representatividade pode ser tomada como um ato de fé, ou seja, é uma responsabilidade designada ao linguista, que deve estabelecer critérios coerentes com seus objetivos na hora de construir um *corpus*.

A depender dos objetivos de determinado estudo, alguns tipos de *corpora* podem ser compilados. Na presente pesquisa, fazemos o uso de um *corpus* acadêmico. *Corpora* desse tipo "são usados para informar nossa compreensão do discurso acadêmico, a partir de perspectivas lexicais, gramaticais, fraseológicas e de gênero"[15] (Nesi, 2016, p. 211, tradução nossa).

O *corpus Bank of English*, criado na Universidade de Birmingham para fornecer informações para o *Collins Cobuild English Language Dictionary* possui extensão suficiente para conter milhares de exemplos de itens lexicais mais frequentes, apesar disso, *corpora* como o mencionado não foram compilados para a exploração de características específicas de textos acadêmicos, por isso, tornou-se comum o uso de *corpora* pequenos "para revelar padrões de

[15] No original: *[...] are used to inform our understanding of academic discourse, from lexical, grammatical, phraseological, and genre perspectives*. (Nesi, 2016, p. 211)

uso da linguagem acadêmica em uma ampla gama de gêneros”[16] (Nesi, 2016, p. 207, tradução nossa).

O *corpus SDG-Plos* (Pinto *et al*, no prelo), utilizado para este trabalho, é representativo da amostra da língua inglesa que objetivamos investigar na medida em que atende aos nossos objetivos de pesquisa: é uma compilação de textos acadêmicos (artigos científicos), escritos por pesquisadores internacionais que abordam os ODS em suas publicações. Conforme Nesi (2016, p. 207, tradução nossa), o uso de *corpora* constituídos por materiais de periódicos acadêmicos como artigos é recorrente “porque são fáceis de coletar em formato eletrônico e podem ser selecionados para representar domínios de pesquisa altamente específicos, e podem produzir resultados de grande relevância para determinados contextos de EAP”[17]. Descreveremos o *corpus* de estudo com mais detalhes na seção Materiais e Método, em que apresentamos a metodologia empregada.

Os linguistas de *corpus* utilizam ferramentas computacionais “essencialmente quantitativas, realizando cálculos com base na frequência de itens lexicogramaticais especificados”[18] (Nesi, 2016, p. 211, tradução nossa). Logo, no que tange ao léxico, pode-se falar em palavras de maior e de menor frequência, assim como em sentidos mais frequentes e menos frequentes dos itens lexicais.

A partir de análises quantitativas, de acordo com Nesi (2016, p. 211, tradução nossa), “A contagem de frequência pode ser usada para gerar vários tipos de listas de palavras acadêmicas, desde simples discriminações de cada forma de palavra até comparações mais complexas dentro e entre *corpora* e listas de termos restritos a domínios específicos.” As ferramentas computacionais de LC ajudam os pesquisadores a acessarem listas de palavras, para observar sua frequência, e listas de palavras-chaves, para observar sua chavicidade, por exemplo.

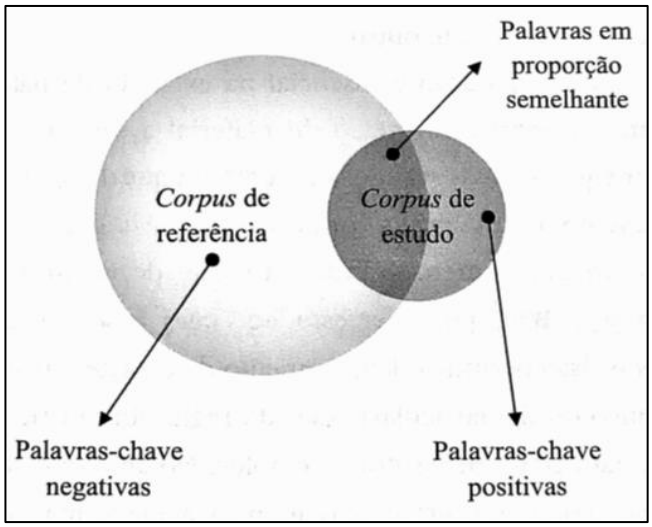
No presente estudo, damos destaque para o levantamento de palavras-chave devido ao seu caráter representativo do conteúdo de um *corpus*. De acordo com Berber Sardinha (2005, p. 238), as palavras-chave são “[...] aquelas cujas frequências são estatisticamente diferentes no corpus de estudo e no corpus de referência”. Segundo o linguista, a lista de palavras-chave é

[16] No original: *[...] have been used to reveal patterns of academic language use in a wide range of genres [...]* (Nesi, 2016, p. 207)
[17] No original: *[...] because they are easy to collect in electronic form, can be selected to represent highly specific research domains, and can yield findings of great relevance to certain EAP contexts.* (Nesi, 2016, p. 207)
[18] No original: *[...] the software tools they use are essentially quantitative, performing calculations based on the frequency of specified lexicogrammatical items.* (Nesi, 2016, p. 211)

gerada a partir de uma comparação entre a lista de palavras do *corpus* de estudo e a lista de palavras do *corpus* de referência (Berber Sardinha, 2005).

As palavras-chave podem ser separadas em três grupos: 1. o das "empregadas em proporção semelhante em ambos os conjuntos de texto", 2. o das "frequentemente associadas ao *corpus* de estudo" e 3. o das "mais particulares ao *corpus* de referência", segundo Viana (2011, p. 61). Pode-se dizer que as palavras-chave do terceiro grupo (negativas) são as menos relevantes para o *corpus* de estudo, já as do segundo grupo (positivas) são aquelas consideradas importantes porque "servem para identificar linguisticamente o corpus que se deseja descrever" (Viana, 2011, p. 62). A imagem a seguir ilustra a forma como o grupo de palavras-chave do *corpus* de referência e o do de estudos se interseccionam:

Figura 4 – Procedimento de palavras-chave



Fonte: Viana (2011, p. 61)

As palavras-chave positivas apresentam alta chavicidade, conceito entendido como, segundo Bondi e Scott (2010, p. 43, tradução nossa), "[...] uma qualidade possuída por palavras, grupos de palavras, frases etc."[19] Para os autores, a teoria da chavicidade indica haver itens lexicais proeminentes nos textos (Bondi; Scott, 2010).

A ferramenta computacional utilizada para este trabalho, *Sketch Engine* (Kilgarriff *et al*, 2014), calcula a chavicidade conforme a seguinte fórmula:

[19] No original: *Keyness is a quality possessed by words, word-clusters, phrases etc. [...].* (Bondi; Scott, 2010, p. 43)

$$\frac{fpm_{focus} + n}{fpm_{ref} + n}$$

em que fpm_{focus} diz respeito à frequência das palavras do *corpus* de estudo, fpm_{ref} à frequência das palavras do *corpus* de referência, e *n* ao número padrão de aproximação de valores (n=1).

Palavras-chave podem ser consideradas palavras que tornam determinada amostra da língua, de certa maneira, única. Elas podem ser analisadas conforme diferentes abordagens teóricas. A depender do olhar que recebem, são concebidas como expressões multipalavras e podem ser analisadas e entendidas enquanto termos, colocações, expressões idiomáticas, agrupamentos lexicais, *clusters* etc. As combinatórias de palavras-chave também podem ser entendidas como *multiword terms*, ou em português, termos multipalavras (Ramisch; Villavicencio, 2010; *Sketch Engine*, 2014).

Nazzi-Laranja (2020) propôs em seu estudo um levantamento dos agrupamentos lexicais de quatro *corpora* cujas temáticas eram "saúde", "tecnologia", "política" e "mundo" para o desenvolvimento de atividades de leitura para o vestibular, por exemplo. Já Milizia (2010) analisou as combinações feitas com a palavra *climate* (clima) nos discursos de Tony Blair e George W. Bush e verificou a emergência do padrão *climate change* (mudança climática), o que demonstrou "como as palavras preferem significativamente a companhia umas das outras"[20] (Milizia, 2010, p. 142, tradução nossa) e "convencionalmente se atraem umas às outras por fatores que vão além das normas gramaticais"[21] (Milizia, 2010, p. 142, tradução nossa).

Segundo Muller (*et al*, 2011), as abordagens de pesquisa da LC levam em consideração, principalmente, a frequência de dados, além da chavicidade para aferir a força de combinação entre palavras, o que permite que os estudos conduzidos por *corpora* abranjam uma ampla quantidade de expressões multipalavras de naturezas variadas. Conforme Biber (2009, p. 277) há dois tipos de expressões multipalavras: *multiword lexical collocations* (colocações lexicais multipalavras) e *multiword formulaic sequences* (sequências formulaicas multipalavras). Essas expressões diferenciam-se pelo fato de a primeira referir-se a termos técnicos formados por sequências de palavras lexicais e a segunda referir-se a sequências de alta frequência que incluem palavras funcionais e palavras de conteúdo (Biber, 2009).

[20] No original: *[...] how words significantly prefer each other's company [...]* (Milizia, 2010, p. 142)
[21] No original: *[...] and are conventionally found to attract each other for factors that go beyond grammatical norms.* (Milizia, 2010, p. 142)

Nas pesquisas orientadas por evidências, com base na LC, adota-se o princípio da idiomaticidade, definido por Sinclair (1991, p. 110, tradução nossa), como o princípio em que "um usuário de uma língua tem à sua disposição um grande número de frases pré-construídas que constituem escolhas únicas, mesmo que possam parecer analisáveis em segmentos"[22], o que implica em dizer que os falantes dispõem de blocos pré-fabricados recorrentes na língua ao invés de ter de fazer escolhas aleatórias diante das muitas combinações de palavras possíveis em seu idioma. Por essa razão, as expressões multipalavras são tão caras ao ensino de línguas.

Esses padrões podem ser analisados sob a persectiva da Fraseologia, disciplina da Lexicologia que tem como objeto de estudo unidades fraseológicas (UFs) (Corpas Pastor; Mitkov, 2019), expressões formadas por mais de duas palavras caracterizadas por possuírem certa fixidez em seu modo de ocorrência nas línguas. Elas são tomadas como blocos pré-fabricados, que configuram os fraseologismos, termo que usualmente se refere a "Um conjunto de frases feitas, locuções figurativas, metáforas e comparações fixas, expressões idiomáticas e ditados populares, existentes em uma língua, seja em uso individual ou em grupo[23]" (Corpas Pastor, 1996, p. 20, tradução nossa).

A presente pesquisa não visa aprofundar-se no campo da Fraseologia, nem fazer um levantamento de termos de determinada área de especialidade, como se faz nos trabalhos de Terminologia, vasta área que se volta para o "estudo científico dos termos usados nas línguas de especialidade, ou melhor, empregados em discursos e textos de áreas técnicas, científicas e especializadas" (Barros, 2007, p. 9). Aqui, nos apoiamos na Pedagogia do Léxico (Orenha-Ottaiano; Pinto, 2018) para fazer um levantamento dos itens lexicais propícios para o desenvolvimento de atividades de leitura, as expressões multipalavras, tendo como base teórico-metodológica a *Linguística de Corpus*.

Para Villavicencio e Ramisch (2010, p. 33) as expressões multipalavras e os termos, unidades lexicais "que designa(m) um conceito de um domínio de especialidade" (Barros, 2007, p. 9), diferem em vários aspectos, dentre essas diferenças está o fato de as expressões multipalavras estarem presentes na linguagem especializada e na linguagem cotidiana, diferentemente dos termos, que se restringem à comunicação especializada. No entanto, os autores afirmam que há um ponto em comum: "O objeto que pertence à intersecção entre termo

[22] No original: *[...] a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments*. (Sinclair, 1991, p. 110)

[23] No original: *[...] Conjunto de frases hechas, locuciones figuradas, metáforas y comparaciones fijadas, modismos y refranes, existentes en una lengua, en el uso individual o en el de algún grupo.* (Corpas Pastor, 1996, p. 20)

e EM é usualmente denominado Termo Multipalavra (TM)" (Villavicencio; Ramisch, 2010, p. 41), caracterizado como "uma locução ou expressão multipalavra que possui estatuto terminológico" (Villavicencio; Ramisch, 2010, p. 41).

Os termos multipalavras, para os autores, deveriam ser considerados nos estudos das fraseologias de domínio (Villavicencio; Ramisch, 2010, p. 42). Neste estudo, analisamos as palavras-chave do *corpus SDG-Plos* e utilizamos a expressão "termos multipalavras" com base na concepção do *Sketch Engine* (Kilgarriff *et al*, 2014, tradução nossa), segundo a qual "Termos são expressões multipalavras que aparecem com mais frequência no *corpus* em foco do que no *corpus* de referência [...]"[24]. A plataforma nomeia de "palavras-chave" unidades formadas por palavras únicas e de "termos" unidades multipalavras: "1. Palavra-chave = item de palavra única - 2. Termo = expressão multipalavra (consiste em 2 ou mais palavras)"[25] (*Sketch Engine*, 2024, tradução nossa).

O reconhecimento dessas estruturas pré-fabricadas pode facilitar a compreensão de textos em LE, já que torna possível que os aprendizes lancem mão de estratégias de identificação de padrões presentes nas línguas ao invés de ter que lidar com a possibilidade de combinações aleatórias. Para apresentar esses itens lexicais de forma guiada aos estudantes, nos apoiamos na abordagem cuja máxima é a primazia da exploração dos dados linguísticos apresentados sob a forma de um *corpus*, *Data Driven Learning* (DDL), Aprendizagem Direcionada por Dados, descrita na seção seguinte.

2.4.1 *Data Driven Learning*

Desde as décadas de 1980 e 1990, a LC tem sido usada no ensino de LE, especialmente nos contextos de ensino de línguas para fins específicos, primeiramente no ensino de inglês (Tagnin; Viana, 2011), por meio do *classroom concordancing*, ensino baseado no uso de concordâncias, e do *Data Driven Learning* (DDL), aprendizagem direcionada por dados, introduzida pelo linguista Tim Johns (Berber Sardinha, 2011). Conforme Scott (2011, p. 8),

> Na DDL, o aluno observa alguns ou muitos exemplos das formas linguísticas (dados) que precisa aprender, e chega às conclusões que consegue (aprendizado). São os dados que, junto com o esforço e o raciocínio do aluno, determinam a aprendizagem, não as regras dadas pelo professor.

[24] No original: *Terms are multi-word expressions which appear more frequently in the focus corpus than in the reference corpus [...]* (*Sketch Engine*, 2014)
[25] No original: *1. Keyword = single-word item - 2. term = multi-word expression (it consists of 2 or more words).* (*Sketch Engine*, 2024)

A DDL fundamenta-se na exposição dos aprendizes à língua (por meio do uso de *corpora*) e na sua exploração de forma organizada e guiada, de modo a fazer com que tenham acesso a coletâneas de textos de seus interesses e a frequências e distribuição das ocorrências de itens léxico-gramaticais de forma contextualizada (Boulton, 2017). Para Boulton (2017), trata-se de uma abordagem construtivista e indutiva na qual os alunos chegam às suas próprias conclusões a respeito dos padrões da língua ao invés de esperarem por regras prontas e acabadas.

Na visão de Berber Sardinha, no entanto, a DDL pode seguir uma orientação não somente indutiva, "[...] em que o aluno precisa encontrar evidências de uso e delas formar generalizações sobre formas e sentidos da língua" (Berber Sardinha, 2011, p. 314), mas também dedutiva, em que "[...] antes de iniciar a análise da concordância, o aluno formula hipóteses sobre a questão/ponto em foco na atividade. Em seguida, procura confirmar ou rejeitar essa hipótese com base nos dados" (Berber Sardinha, 2011, p. 314).

O aluno, dentro do escopo da DDL, é visto como um pesquisador ou mesmo um detetive da língua (Berber Sardinha, 2011). Nas atividades baseadas em *corpora*, os estudantes exploram os dados fornecidos pelo *corpus* a fim de entrar em contato com o léxico e com a gramática de determinados gêneros textuais, amostras da linguagem que contêm as informações morfológicas, gramaticais, fraseológicas e discursivas tomadas como objeto de ensino pelos professores.

De acordo com Berber Sardinha (2011), há três tipos de atividades de ensino com *corpora*: as centradas na concordância, as centradas no texto e as multimídias/multigêneros. O primeiro tipo foca na observação das linhas de concordância de um *corpus* para a identificação de padrões de uso e, consequentemente, na descoberta do significado de palavras e expressões e na análise das classes gramaticais (Berber Sardinha, 2011). Essa vertente pode ser considerada a mais comum dentro da DDL, pois prioriza a aprendizagem de lexicogramática, o pilar da abordagem (Berber Sardinha, 2011).

O segundo tipo surge para suprir algumas desvantagens que o anterior pode apresentar, como falta de familiaridade dos estudantes com as linhas de concordância de um *corpus* e a "[...] falta de ligação entre as concordâncias e uma situação concreta de uso da língua, onde os padrões estejam representados" (Berber Sardinha, 2011, p. 316), ou seja, a falta de um contexto situacional de uso da língua que o texto proporciona. Essa vertente apresenta o texto para os

aprendizes e os conduz para a observação e análise dos padrões nele contidos, sem que se perca de vista a essência de uma atividade baseada em *corpus* (Berber Sardinha, 2011).

As atividades do terceiro tipo, multimídias/multigêneros visam "[...] capacitar os alunos a lidar com (produzir e/ ou receber) a linguagem envolvida em determinada atividade social" (Berber Sardinha, 2011, p. 329), tendo em vista a diversificação do registro da língua veiculado nas diferentes mídias sociais, a saber, a profusão e a complexificação dos gêneros (Berber Sardinha, 2011). Ainda que as atividades baseadas em *corpus* se beneficiem de textos escritos (e não da compilação de sons e imagens), as atividades multidimídias/multigêneros se valem do "texto escrito para ser lido (ex. texto de jornal), o escrito para ser ouvido (ex. notícia de TV), o preparado para ser visto como se fosse espontâneo (ex. cinema), o escrito para ser lido como se tivesse sido ouvido (ex. muitos *emails*, *tweets* etc.), entre muitos outros" (Berber Sardinha, 2011, p. 324).

As três vertentes de atividades explicitadas tanto podem seguir uma metodologia *hands on*, por meio da qual o aluno tem acesso ao computador, às ferramentas de *corpora* e ao próprio *corpus*, como *hands off*, por meio da qual eles realizam os exercícios impressos no papel (Boulton, 2012).

Berber Sardinha (2000, p. 61) considera que a DDL gera impacto no aluno, no professor e no ensino como um todo:

> (1) Aluno: Assume um papel mais central e independente, tornando-se um pesquisador dotado de habilidades que lhe permitem obter respostas para suas dúvidas e verificação de suas hipóteses. (2) Professor: Deixa de ser a fonte única de informação na sala de aula. Por conseguinte diminui a pressão sobre ele para saber todas as respostas. O computador, enquanto informante, assume um papel de destaque como provedor de informação. (3) Ensino: Centra-se mais no aluno e permite que se fundamente pelo princípio de descoberta. Deixa, portanto, de estar baseado na transmissão de regras prontas.

A DDL possui um caráter transformador, pois redimensiona os papeis tradicionalmente estabelecidos para professores e alunos. Essa característica também pode ser muito desafiadora, uma vez que diante das informações contidas em um *corpus*, os alunos podem se deparar com dados que podem ser inusitados tanto para eles quanto para os professores. A riqueza linguística contida em amostras reais de uso da língua não cabe nos livros didáticos ou nas gramáticas, por

isso, tanto o aluno quanto o professor descobrirão novas informações sobre a língua, segundo Johns (1988, p. 3, tradução nossa[26]):

> O que há de distintivo na abordagem DDL para o ensino indutivo de línguas é o princípio de que os dados são primários e o professor não consegue antecipar exatamente quais regras ou padrões os alunos descobrirão: na verdade, eles muitas vezes perceberão coisas que são desconhecidas não apenas pelo professor, mas também pelas obras padrão de referência sobre o idioma. É este elemento de desafio e de descoberta que dá a DDL o seu sabor especial e estimulante.

A DDL tem sido difundida gradativamente no Brasil e a elaboração de materiais didáticos baseados em *corpora* tem crescido (Berber Sardinha, 2000; 2011; Delfino, 2016). Apesar dos desafios, a abordagem vem ganhando espaço aos poucos no ensino de línguas brasileiro, um exemplo são as publicações de materiais produzidos pelo grupo de pesquisa Ensino Baseado e Dirigido por Corpora (En-Corpora)[27] da Unesp de São José do Rio Preto. A elaboração de materiais baseados em *corpora* será aprofundada no decorrer deste trabalho.

Na pesquisa proposta, nos voltamos para a elaboração de atividades centradas no texto por priorizarmos a habilidade de leitura, e do tipo *hands off* e *hands on*, pois o primeiro tipo de atividade pode ser aplicado tanto em contexto de aula presencial quanto on-line, além de ser de acesso facilitado para os alunos de forma geral, e o segundo proporciona um contato direto dos alunos com ferramentas da LC, apesar dos desafios que traz consigo, como a necessidade de acesso à internet e ao computador. Conforme Nazzi-Laranja (2020, p. 49), "cabe ao professor pensar em formas de ensinar vocabulário aos seus alunos e fazer desse vocabulário parte importante para a compreensão textual", seja por meio de atividades *hands on* ou *hands off*.

Aprofundamos o conceito de compreensão leitora na seção a seguir.

2.5 O processo de leitura e a habilidade de compreensão escrita

Para a elaboração de atividades didáticas voltadas para o aprimoramento de CE, cabe entender quais são os processos de leitura. Leffa (1996) considera a existência de duas acepções

[26] No original: *What is distinctive about the DDL approach to inductive language teaching is the principle that the data is primary, and the teacher does not know in advance exactly what rules or patterns the learners will discover: indeed, they will often notice things that are unknown not only to the teacher, but also to the standard works of reference on the language. It is this element of challenge and of discovery that gives DDL its special flavour and stimulus.* (Johns, 1988, p. 3)

[27] Disponível em: https://sites.google.com/unesp.br/en-corpora-ensino-baseado-e-di/in%C3%ADcio

do que seja a leitura, consideradas definições restritas e antagônicas, a de que "(a) ler é extrair significado do texto" (p. 11) e a de que "(b) ler é atribuir significado ao texto" (p. 11). Na primeira, a leitura é tomada como um processo ascendente, uma vez que o significado parte do texto para o leitor (Leffa, 1996), já na segunda, a leitura é tomada como um processo descendente, em que os significados relacionados às hipóteses do leitor são confirmados ou não pela leitura do texto (Leffa, 1996). Ambos os processos possuem suas limitações, já que "A complexidade do processo da leitura não permite que se fixe em apenas um de seus pólos" (Leffa, 1996, p.17) e que "Para compreender o ato da leitura temos que considerar então (a) o papel do leitor, (b) o papel do texto e (c) o processo de interação entre o leitor e o texto" (Leffa, 1996, p. 17).

No contexto de IFE, o ensino de compreensão escrita volta-se para as estratégias de leitura, cuja função é

> "[...] desenvolver habilidades tais como deduzir o significado e uso de itens lexicais não familiares; entender a informação explícita e a implícita dentro da sentença; entender o valor comunicativo das sentenças; extrair seletivamente pontos relevantes do texto; distinguir ideias principais das secundárias, etc. [...] (Heemann, 2009, p. 144).

Algumas das estratégias de leitura são o reconhecimento do *layout* característico a determinados gêneros textuais; a identificação de cognatos, marcadores do discurso, elementos de referência e seus efeitos de sentido; os tipos de coesão (lexical e gramatical); coerência e as noções implícitas a esse fenômeno (cronologia, importância, comparação); *skimming*, técnica de leitura rápida cujo objetivo é depreender a ideia principal de um texto por meio do uso do título, de imagens e de palavras-chave; e *scanning*, técnica na qual o leitor lê de forma rápida com a intenção de encontrar informações específicas no texto. Além dessas, a inferência também é uma estratégia que se presta como recurso facilitador na leitura em LE, pois por meio dela o leitor deduz "[...] o significado e uso de itens lexicais não familiares por meio de dicas contextuais [...]" (Heemann, 2009, p. 144).

Conforme Baldo (2009), a leitura bem sucedida em LE está relacionada ao processo de aquisição de vocabulário, principalmente contextual, de modo que o conhecimento de estruturas sintáticas teria menor influência na proficiência leitora, pois essa é influenciada pelo repertório de informações garantido pelo acervo lexical dos aprendizes. Leitores mais proficientes na LE conseguem mobilizar as estratégias de leitura com que dispõem na LM, nesse sentido, os recursos facilitadores da compreensão são eficientes quando se é capaz de reconhecê-los, pois

não importa a quantidade de vezes que se lance mão deles sem que se consiga enxergar sua presença nos textos (Baldo, 2009).

Assim, é possível afirmar que no processo de leitura "o leitor passa a ser visto como um sujeito ativo, um bom usuário de estratégias e um aprendiz cognitivo" (Heemann, 2009, p. 138), que é capaz de relacionar seu conhecimento de mundo, ou conhecimento prévio, com as informações do texto. A leitura concebida como processo de produção de sentidos, resultado de uma atividade interativa entre o texto, o autor, o leitor e o contexto, mobiliza tanto o processo ascendente, do inglês *bottom up*, quanto o descendente, *top down* (Leffa, 1996; Heemann, 2009).

Conforme Brown (2001, p. 306-311, tradução nossa[28]) algumas estratégias relacionam-se com o processo ascendente e outras com processo descendente, tais como:

> 1. identificar o propósito da leitura (identificar claramente o propósito da leitura de modo que se saiba o que está procurando); 2. identificar padrões e regras grafêmicas na decodificação *bottom up*, especialmente para aprendizes iniciantes; 3. usar técnicas eficientes de leitura silenciosa para a compreensão rápida (de intermediário para níveis avançados); 4. olhar o texto em busca de ideias principais (correr os olhos rapidamente por todo o texto em busca de sua essência) para fazer previsões; 5. escanear o texto e busca de informações específicas (procurar rapidamente por informações no texto); 6. usar mapeamento ou *clustering* (agrupar ideias em grupos significativos ajuda os leitores a trazerem ordem ao caos); 7. adivinhar quando não se tem certeza (adivinhar o significado de uma palavra, adivinhar uma combinação gramatical, adivinhar uma relação discursiva, inferir significado subentendido, inferir referências culturais e conteúdos de mensagens; 8. analisar vocabulário (procurar por prefixos e sufixos, procurar por radicais que são familiares, procurar por contextos gramaticais que sinalizem informações, observar o contexto semântico em busca de pistas); 9. distinguir entre sentido literal e sentido figurado; 10. reconhecer marcadores do discurso para interpretar relações semânticas. Brown (2001, p. 306-311)

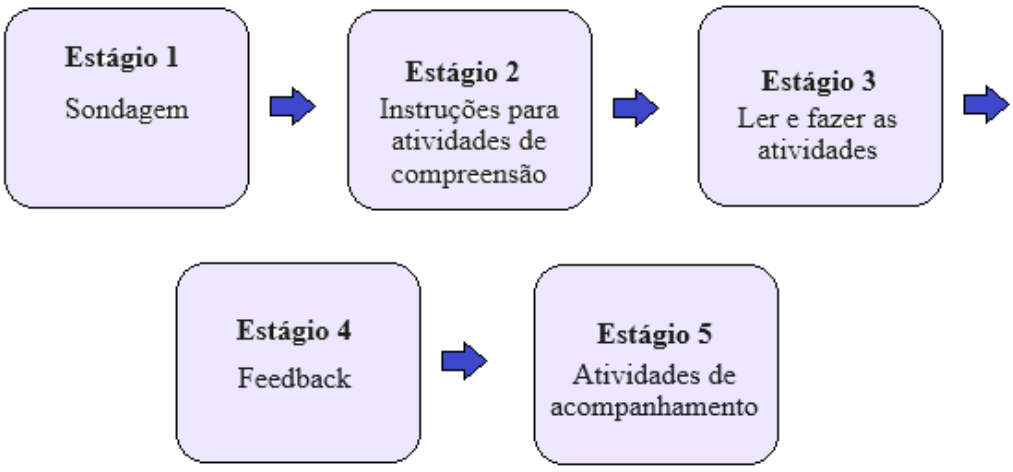
[28] No original: *1. identify the purpose of reading (clearly identify the purpose in reading something so you will know what you are looking for); 2. use graphemic rules and patterns to aid bottom-up decoding, especially for beginning level learners; 3. use efficient silent reading techniques for rapid comprehension (for intermediate to advanced levels); 4. skim the text for main ideas (quickly running one's eyes across whole text for its gist) for prediction; 5. scan the text for specific information (quickly searching for some particular piece or pieces of information in a text); 6. use semantic mapping or clustering (grouping ideas into meaningful clusters, helps readers to provide some order to the chaos); 7. guess when you aren't certain (guess a meaning of a word, guess a grammatical relationship, guess a discourse relationship, infer implied meaning, guess about a cultural reference and guess content messages); 8. analyze vocabulary (look for prefixes, look for suffixes, look for roots that are familiar, look for grammatical contexts that may signal information, look at semantic context for clues); 9. distinguish between literal and implied meanings; 10. capitalize on discourse markers to process relationships*. (Brown, 2001, p. 306-311)

A partir das estratégias elencadas por Brown (2001) são elaboradas atividades didáticas com a finalidade de desenvolver a compreensão escrita e a autonomia dos estudantes. Conforme Khand (2004), o ensino de leitura em LE pode se dar por meio da abordagem comunicativa que, entre outros fatores, caracteriza-se por assumir que "o ensino de línguas deve promover a competência comunicativa do aprendiz" (Abrahão, 2015, p. 26) e que a aprendizagem parta do "desenvolvimento das 4 habilidades, de forma integrada, desde o início do processo de ensino e aprendizagem" (Abrahão, 2015, p. 26).

Ao assumir as proposições feitas por Khand (2004) entende-se que o ensino de leitura pode acontecer de forma integrada às outras habilidades, ou seja, por meio de atividades que envolvam os aprendizes de tal forma que sejam capazes de mobilizar estratégias para além da leitura, tais como discussões em grupo, leitura compartilhada, sumarização de conteúdos, apresentações orais etc. Seguindo esse princípio, o autor elabora um processo de produção de materiais para as aulas de leitura orientadas conforme o Ensino Baseado em Tarefas (EBT), que concebe que a aprendizagem se dá de forma mais natural à medida em que o aluno realiza atividades relacionadas ao contexto real de uso da língua (Finardi; Porcino, 2014).

A imagem a seguir ilustra os cinco estágios metodológicos para aulas de leitura em LE propostos por Khand (2004):

Figura 5 - Estágios metodológicos para aulas de CE



Fonte: Elaborado com base em Khand (2004, p. 48).

Conforme o esquema apresentado na imagem, o Estágio 1, de sondagem, é o momento utilizado para apresentar o tema que será abordado no texto e fazer perguntas oralmente a fim de detectar o conhecimento prévio dos alunos; o Estágio 2 serve para que sejam passadas as

instruções sobre as atividades de compreensão; o Estágio 3 é relativo ao momento em que os aprendizes leem o texto e realizam as atividades; o Estágio 4 serve para que os alunos exponham suas respostas e suas impressões do texto com base na leitura que fizeram e para o que o professor dê a eles um *feedback*, um direcionamento; já o Estágio 5 é dedicado às atividades de acompanhamento, ou *follow-up tasks*, que são relacionadas às atividades e os textos já passados no Estágio 3 (Khand, 2004).

O QCER também leva em consideração o desenvolvimento da competência comunicativa, por isso avalia a proficiência linguística levando em consideração as quatro habilidades. Contudo, o documento traz uma análise pormenorizada sobre as atividades comunicativas relacionadas à compreensão e à produção (oral e escrita), por isso, além de trazer a escala global com os Níveis Comuns de Referência, apresentada anteriormente neste trabalho, também propõe um quadro com as "as principais categorias de uso da língua em cada um dos seis níveis" (Quadro Comum Europeu de Referência para Línguas, 2001, p. 50), que abarca as quatro habilidades linguísticas e as atividades comunicativas relativas à competência comunicativa do aluno nos níveis do A1 ao C2 (Quadro Comum Europeu de Referência para Línguas, 2001). Conforme o documento, o aprendiz é capaz de realizar determinadas atividades comunicativas a depender do nível em que se encontra. A imagem a seguir apresenta o quadro para autoavaliação do estudante com relação à compreensão, foco desta pesquisa:

Figura 6 – Quadro para autoavaliação do Quadro Comum Europeu de Referência para Línguas – Habilidade de Compreensão

Compreender		
	Compreensão do oral	Leitura
A1	Sou capaz de reconhecer palavras e expressões simples de uso corrente relativas a mim próprio, à minha família e aos contextos em que estou inserido, quando me falam de forma clara e pausada.	Sou capaz de compreender nomes conhecidos, palavras e frases muito simples, por exemplo, em avisos, cartazes ou folhetos.
A2	Sou capaz de compreender expressões e vocabulário de uso mais frequente relacionado com aspectos de interesse pessoal como, por exemplo, família, compras, trabalho e meio em que vivo. Sou capaz de compreender o essencial de um anúncio e de mensagens simples, curtas e claras.	Sou capaz de ler textos curtos e simples. Sou capaz de encontrar uma informação previsível e concreta em textos simples de uso corrente, por exemplo, anúncios, folhetos, ementas, horários. Sou capaz de compreender cartas pessoais curtas e simples.
B1	Sou capaz de compreender os pontos essenciais de uma sequência falada que incida sobre assuntos correntes do trabalho, da escola, dos tempos livres, etc. Sou capaz de compreender os pontos principais de muitos programas de rádio e televisão sobre temas actuais ou assuntos de interesse pessoal ou profissional, quando o débito da fala é relativamente lento e claro.	Sou capaz de compreender textos em que predomine uma linguagem corrente do dia-a-dia ou relacionada com o trabalho. Sou capaz de compreender descrições de acontecimentos, sentimentos e desejos, em cartas pessoais.
B2	Sou capaz de compreender exposições longas e palestras e até seguir partes mais complexas da argumentação, desde que o tema me seja relativamente familiar. Consigo compreender a maior parte dos noticiários e outros programas informativos na televisão. Sou capaz de compreender a maior parte dos filmes, desde que seja utilizada a língua-padrão.	Sou capaz de ler artigos e reportagens sobre assuntos contemporâneos em relação aos quais os autores adoptam determinadas atitudes ou pontos de vista particulares. Sou capaz de compreender textos literários contemporâneos em prosa.
C1	Sou capaz de compreender uma exposição longa, mesmo que não esteja claramente estruturada ou quando a articulação entre as ideias esteja apenas implícita. Consigo compreender programas de televisão e filmes sem grande dificuldade.	Sou capaz de compreender textos longos e complexos, literários e não literários, e distinguir estilos. Sou capaz de compreender artigos especializados e instruções técnicas longas, mesmo quando não se relacionam com a minha área de conhecimento.
C2	Não tenho nenhuma dificuldade em compreender qualquer tipo de enunciado oral, tanto face a face como através dos meios de comunicação, mesmo quando se fala depressa, à velocidade dos falantes nativos, sendo apenas necessário algum tempo para me familiarizar com o sotaque.	Sou capaz de ler com facilidade praticamente todas as formas de texto escrito, incluindo textos mais abstractos, linguística ou estruturalmente complexos, tais como manuais, artigos especializados e obras literárias.

Fonte: Captura de tela do Quadro Europeu Comum de Referência para Línguas (2001, p. 53).

Como se pode notar, foi destacado no quadro as atividades comunicativas que alunos dos níveis A1 e A2 são capazes de realizar em relação à habilidade de leitura. Com base nessas

informações foi possível delinear os critérios de elaboração dos exercícios que compõem as atividades, como a adequação do material ao nível de proficiência do alunado.

Além do QCER, nos baseamos no modelo de aquisição, complexidade e formulaicidade para o *design* de DDL proposto por Mark e O'Keeffe (2022) e nos critérios obrigatórios para elaboração de materiais baseados em *corpora*, desenvolvidos por Delfino (2016), os quais explicitamos na próxima seção.

2.5.1 Materiais didáticos baseados em *corpora*: foco na compreensão escrita no nível básico

De acordo com com Boulton e Cobb (2017), a DDL concentra seus melhores resultados no contexto de ensino superior, em que se encontram estudantes entre os níveis intermediário e avançado de proficiência, do B1 ao C2, embora ressaltem que já existem pesquisas voltadas para experiência com DDL no ensino básico e nos cursos livres de idiomas. Mark e O'Keeffe (2022) reforçam que os princípios da abordagem, como a frequência de exposição à linguagem natural no processo de aprendizagem de línguas, podem ser amplamente aproveitados por todos os aprendizes de LI, desde que haja uma curadoria do material retirado dos *corpora* de acordo com o que se espera dos alunos em cada nível proposto pelo QCER.

Conforme Mark e O'Keeffe (2022, p. 2, tradução nossa), "[...] o ethos tradicional da DDL é promover a descoberta indutiva independente por meio da exploração livre de *corpus*, no entanto, isso requer pelo menos um nível intermediário de competência"[29]. As autoras, então, argumentam a favor de orientações para o ensino dirigido por *corpora,* conforme as quais, os professores desempenham um papel mais influente na seleção e mediação das tarefas, diferenciadas por níveis, para que os estudantes possam de fato se beneficiar dos recursos fornecidos pela LC.

Mark e O'Keeffe (2022) associam descobertas provindas das pesquisas em Aquisição de Segunda Língua (ASL), *Second Language Acquisition* (SLA), especialmente entre os modelos de aquisição *usage-based* (UB), que concebem a aquisição como um resultado do uso e da exposição do aprendiz à língua (Mark; O'Keeffe, 2022). Os modelos UB mostram que "os aprendizes de segunda língua normalmente passam de um repertório de sequências holofrásticas fixas em níveis baixos para um sistema expandido de esquemas e *frames* para

[29] No original: *[...] the traditional ethos of DDL is to promote independent inductive discovery through free corpus foraging, however, this requires at least an intermediate level of competency.* (Mark; O'Keeffe, 2022, p. 2)

padrões totalmente abstratos (geralmente figurativos)"[30] (Mark; O'Keeffe, 2022, p. 2, tradução nossa), ou seja, inicialmente, descobrem padrões da língua, como combinações de palavras, para então se apropriarem de sequências mais complexas que requerem do falante um conhecimento maior do idioma.

As autoras propõem um modelo para aprimorar a curadoria de DDL e o *design* de atividades baseado nos princípios da aquisição, da complexidade e da formulaicidade (Mark; O'Keeffe, 2022), isso porque: i. a aquisição nos níveis mais baixos se difere da aquisição nos níveis mais avançados (Mark; O'Keeffe, 2022); ii. consequentemente, nos níveis mais baixos, a complexidade varia, e iii. a formulaicidade é um marcador de um aluno avançado, pois o uso de linguagem formulaica, como colocações e expressões idiomáticas, requer maior conhecimento e consciência da língua (Mark; O'Keeffe, 2022).

O esquema contido na imagem a seguir ilustra a organização do processo de aquisição de LE com base na variabilidade da complexidade e da formulaicidade inerente a cada nível do QCER conforme o modelo para *design* de materiais orientados pela DDL proposto por Mark e O'Keeffe (2022):

Figura 7 - Modelo de aquisição, complexidade e formulaicidade para o *design* de DDL

Princípio	A1	A2	B1	B2	C1	C2
Aquisição	combinação de palavras		identificação de slots e frames		saber formulaico: refinamento de padrões abstratos para a percepção do registro de associação de palavras	
Complexidade	foco em padrões verbais		mudança de foco em padrões verbais para padrões nominais		foco na complexidade nominal	
Formulaicidade	foco no sentido literal		ampliação do foco no sentido literal para foco no sentido figurado		foco no sentido figurado	

Fonte: Elaborado com base em Mark e O'Keeffe (2022, p. 10).

[30] No original: *[...] second language learners typically move from a repertoire of fixed holophrasal sequences at low levels to an expanded slot and frame system to fully abstracted (often figurative) patterns [...]* (Mark; O'Keeffe, 2022, p. 2)

De acordo com as autoras,

> Esta estrutura se afasta do ethos original do DDL, no qual o aluno se envolve em um processo de descoberta de aprendizagem indutiva. Embora este ainda seja um ideal alcançável, argumentamos que, para que o DDL funcione em todos os níveis, há necessidade de uma mediação baseada em princípios para diferenciar dados e tarefas para diferentes níveis. (Martin; O'Keeffe, 2022, p. 9, tradução nossa[31])

O modelo proposto por Mark e O'Keeffe (2022) é relevante para esta pesquisa porque, embora esteja situada no arcabouço indutivo da abordagem DDL, que prima pela livre exploração dos *corpora* e das linhas de concordância em sua forma "original", denota a importância de uma curadoria para a elaboração de atividades para o público de níveis A1 e A2. Isso não afasta as atividades produzidas dos princípios de autonomia e descoberta da língua, apenas as adapta para estudantes que não estão familiarizados com uso de *corpus* nem com vocabulário e estruturas gramaticais complexas.

As atividades de leitura devem facilitar o reconhecimento dos recursos facilitadores da compreensão e, para o caso do público-alvo desta pesquisa, devem trabalhar com os itens lexicais relacionados às estratégias de leitura contidas no quadro de conteúdos previstos no curso. Como nos propomos a desenvolver material didático com auxílio da LC, nos pautamos no modelo proposto por Mark e O'Keeffe (2022) e nos critérios obrigatórios para atividades baseadas em *corpus* propostos por Delfino (2016):

[31] No original: *This framework moves away from the original ethos of DDL in which the student engaged in a discovery process of inductive learning. While this is still an attainable ideal, we argue that if DDL is to work across levels, there is need for principled mediation to differentiate data and tasks for different levels.* (Mark; O'Keeffe, 2022, p. 9)

Quadro 2 - Critérios obrigatórios para elaboração de atividades com base em *corpora*

O1 – O exercício faz uso de corpus;
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).

Valendo-nos dos dados linguísticos obtidos pelo *corpus* e da concepção de leitura como processo interativo de produção de sentidos, elaboramos as atividades de compreensão escrita em LI, conforme os critérios acima apresentados. Buscamos também investigar os recursos fornecidos pelas ferramentas on-line *Corpus Mate* (Baisa; Crosthwaite, 2023), *Skell* (Baisa; Suchomel, 2014) e *VersaText* (Thomas, 2020) com o propósito de desenvolver os exercícios do tipo *hands on* e *hands off* para que os alunos leiam trechos selecionados dos artigos e entrem em contato com ferramentas simples e de fácil manuseio.

Os recursos oferecidos pelas ferramentas citadas favorecem a produção de atividades para três momentos importantes de uma aula de leitura em LE, segundo Tomitch (2009): a pré-leitura, o momento anterior à leitura de textos na língua alvo em que o professor faz uma sondagem a respeito do conhecimento prévio dos alunos acerca do tema da aula/do texto e em que se faz o acionamento de esquemas relevantes para a construção do significado; a leitura, em que o aluno recebe o texto e o lê efetivamente; e a pós-leitura, fase em que faz-se "a ligação entre o texto e a realidade do aluno" (Tomitch, 2009, p. 197). A pós-leitura é um momento muito importante, pois é a fase de consolidação do que foi apreendido (Tomitch, 2009).

Conforme sugere Tomitch (2009), visamos elaborar as atividades levando em consideração o objetivo do curso em que serão aplicadas, o interesse e nível de proficiência dos alunos, a adequação ao seu nível linguístico em LI, o gênero escolhido para leitura e o tamanho dos textos levados para as aulas.

É importante mencionar que não é nosso objetivo testar o nível de proficiência leitora dos alunos, uma vez que entendemos que as atividades propostas devem apresentar as estratégias de leitura como ferramentas para alcançar a compreensão leitora, ou seja, não as tomar como o objetivo final do curso.

De acordo com Tomitch (2009, p. 201),

> A proficiência na compreensão leitora em LE não é um fim, mas um meio que vai então possibilitar o acesso livre do leitor aos textos escritos que ele escolher, seja para adquirir conhecimento sobre determinado assunto, para entretenimento, enfim, para seu crescimento como parte integrante e atuante da sociedade em que vive.

Assim, entendendo que as atividades serão voltadas para aprendizes de nível básico, nos voltamos para o ensino das estratégias de leitura como ferramentas "que levem o aluno a otimizar o conhecimento que já possui na língua materna e na estrangeira, seja esse conhecimento linguístico, metalinguístico ou de mundo [...]" (Tomitch, 2009, p. 195), ou seja, recursos identificáveis nos textos e possíveis de serem investigados com o apoio de ferramentas de *corpora*.

Na seção a seguir, são explicitadas as ferramentas utilizadas no levantamento de itens lexicais do *corpus* e no desenvolvimento das atividades de leitura, assim como os procedimentos empregados em cada uma das etapas.

3 MATERIAIS E MÉTODO

Nesta seção, serão apresentados os materiais e os procedimentos metodológicos utilizados para desenvolvimento da pesquisa. Em um primeiro momento, na subseção 3.1, nos voltamos para a etapa de levantamento dos dados do *corpus*, descrevendo os *softwares* de compilação e análise de *corpora*, o *design* do *corpus* de estudo e a extração de termos multipalavras. Em seguida, na subseção 3.2, nos debruçamos sobre a etapa de aplicação pedagógica dos dados linguísticos, apresentando o contexto em que os estudantes estavam inseridos, as ferramentas voltadas ao ensino e o processo de elaboração das atividades didáticas, bem como os instrumentos de coleta de dados dos participantes.

3.1 Extração dos dados

A seguir, apresentamos as ferramentas computacionais de criação e compilação de *corpora* utilizadas, as características do *corpus* de estudo, o levantamento de suas palavras-chave e a seleção de seus termos multipalavras.

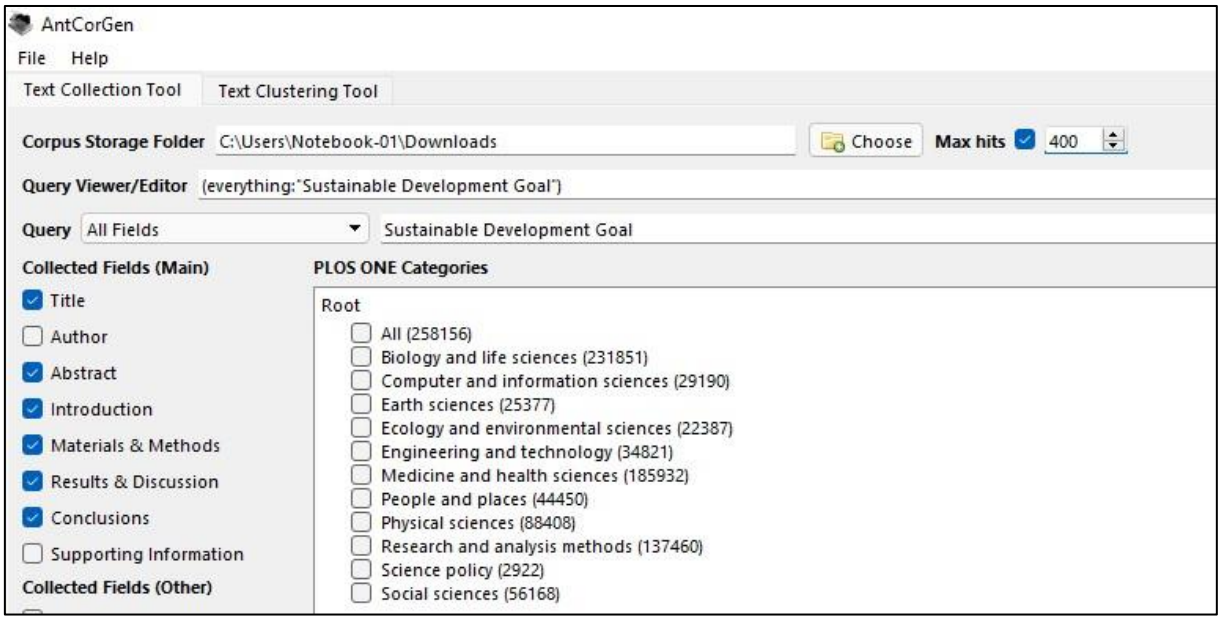
3.1.1 O *AntCorGen*

O *AntCorGen* (Anthony, 2022) é um programa de criação de c*orpora* específicos para diferentes disciplinas. Ele pode "pesquisar, coletar, limpar, marcar e anotar automaticamente *corpora* de disciplinas específicas"[32] (Anthony; Thompson, 2017, tradução nossa). O *corpus* utilizado na pesquisa foi criado por meio desse programa para a pesquisa de Pinto *et al* (no prelo) para o desenvolvimento de uma análise multidimensional lexical. Como o estudo em questão faz parte de uma pesquisa mais ampla sobre léxico dos ODS em artigos científicos, utilizamos o mesmo *corpus* para o levantamento de itens lexicais para serem usados em atividades didáticas.

A imagem a seguir ilustra o programa quando as buscas pelos textos que iriam compor o *corpus* eram feitas:

[32] No original: *[...] search, collect, clean, tag, and annotate very large discipline-specific corpora.* (Anthony; Thompson, 2017, p. 1)

Figura 8 - O *AntCorGen*



Fonte: Captura de tela do programa *AntCorGen* (Anthony, 2022).

Em "*Query Viewer/Editor*" coloca-se uma palavra de busca que remeta ao assunto sobre o qual os textos constituintes do *corpus* irão tratar. Os artigos são extraídos da Plataforma *PLOS ONE*[33], revista acadêmica multidisciplinar que publica trabalhos internacionalmente. Por meio desse programa é possível selecionar as partes dos artigos que interessam ao pesquisador, como o resumo, a introdução, a metodologia, os resultados, a conclusão, ou ainda o texto completo.

Também é possível escolher as áreas dos artigos, que são áreas de atuação e publicação da *PLOS*. Para a criação do *corpus* empregado na pesquisa, foram selecionadas todas as seções dos artigos e todas as áreas em que a palavra de busca encontrasse publicações. Aprofundaremos as características do *corpus* de estudo no item 3.3.

3.1.2 O *Sketch Engine*

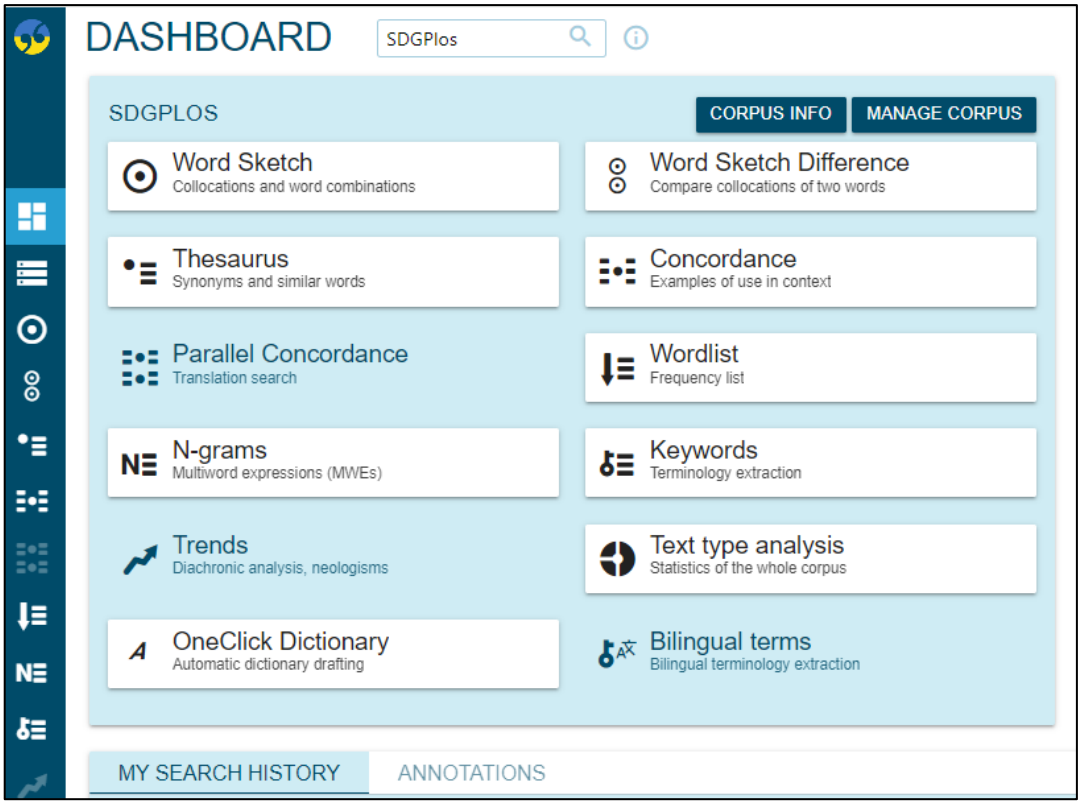
A plataforma *Sketch Engine* oferece ferramentas para a exploração do funcionamento da linguagem. Por meio de análises possibilitadas por seus algoritmos, textos de bilhões de palavras podem ser estudados com a finalidade de "[...] identificar instantaneamente o que é

[33] PLOS-ONE: https://journals.plos.org/plosone/.

típico na linguagem e o que é de uso raro, incomum ou emergente" (*Sketch Engine*, 2024, tradução nossa[34]).

Esse *software* foi utilizado para a análise dos dados do *corpus* sobre os ODS e para a elaboração de atividades didáticas, pois por meio dele fizemos o levantamento das palavras-chave do *corpus* usando a ferramenta *Keywords*. A imagem a seguir traz as informações a respeito do *layout* da plataforma:

Figura 9 – O *Sketch Engine*



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

O *Sketch Engine* é um concordanciador que oferece diversas ferramentas além do *Keywords*, como o *Word List*, que contém as palavras do *corpus* ordenadas pela frequência, o *N-Grams*, que traz listas com os agrupamentos lexicais identificados nos textos compilados, o *Concordance*, que traz as expressões em linhas de concordância para serem analisadas dentro de seu cotexto, entre outras. O *software* abriga vários *corpora* de referência, como o *English Web* 2021 (*enTenTen21*), o utilizado na pesquisa.

[34] No original: […] *to identify instantly what is typical in language and what is rare, unusual or emerging usage.* (*Sketch Engine*, 2024)

O *enTenTen21* é um *corpus* compilado com textos da *web* publicados/divulgados entre o período de novembro de 2021 a janeiro de 2022, considerado tanto de uso geral como de linguagem especializada por abranger fontes da internet e gêneros variados (*Sketch Engine*, 2024). Na tabela a seguir, são listadas as informações desse *corpus*:

Tabela 1 – Informações do *corpus* de referência *enTenTen21*

Itens (*Tokens*)	**Formas (*Types*)**	**Frases**	**Parágrafos**	**Documentos**
61,585,997,113	52,268,286,493	2,852,972,274	1,034,640,844	120,252,162

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

Os dados da tabela evidenciam a extensão do *enTenTen21* (mais de 52 bilhões de palavras), que confere a esse *corpus* confiabilidade para servir como um *corpus* de referência de determinadas amostras da língua.

Como foi mencionado na Fundamentação Teórica deste trabalho, na seção 2.4, o *Sketch Engine* compara a lista de palavras do *corpus* de referência com a do *corpus* de estudo para gerar a lista de palavras-chave. O *software* em questão cria duas listas porque entende como palavras-chave (*Keywords*) palavras únicas e como termos (*Terms*) expressões multipalavras que só aparecem no *corpus* selecionado, o que evidencia o que pode ser considerado típico no *corpus* em foco em relação ao *corpus* de referência (*Sketch Engine*, 2024). Aprofundaremos essa discussão mais adiante, na apresentação das listas geradas pela ferramenta, a *Keywords* e a *Multi-Word Terms*.

3.1.3 O *corpus SDG-Plos*

Conforme os critérios de compilação de *corpora* da taxonomia proposta por Viana (2011), o *corpus* utilizado no estudo, o *SDG-Plos* (Pinto *et al*, no prelo), caracteriza-se por ser de abrangência especializada (ou seja, não pretende documentar o uso de uma língua de forma geral, mas o seu uso em contexto específico por meio de amostras reduzidas da linguagem), de meio escrito, sincrônico, contemporâneo, dinâmico, monolíngue (inglês), e de estudo.

Trata-se de um *corpus* acadêmico, de artigos científicos de pesquisadores internacionais que abordaram os ODS em suas pesquisas. Foi compilado por meio da busca pela expressão

Sustainable Development Goals na ferramenta *AntCorGen*, que cria *corpora* de disciplinas específicas (Anthony, 2022), e depois foi compilado no *Skecth Engine*. Os artigos são provenientes da plataforma on-line da editora *PLOS*.

No quadro a seguir é possível observar as características do *corpus* conforme a tipologia de Viana (2011):

Quadro 3 - Características do *corpus* utilizado na pesquisa

Abrangência	Especializado - divulgação científica de várias áreas de pesquisa (conteúdo acadêmico: artigos científicos provenientes do periódico eletrônico *PLOS*).
Meio	Escrito.
Tempo	Sincrônico; contemporâneo.
Renovação	Dinâmico.
Línguas	Monolíngue; de falantes nativos e não-nativos.
Emprego	Estudo - desenvolvimento de atividades didáticas direcionadas por dados.

Fonte: Elaborado com os dados da pesquisa com base na tipologia de Viana (2011).

Com relação ao tamanho do *corpus*, pode-se dizer que se trata de um *corpus* pequeno, característica comum a *corpora* de estudo. A tabela seguinte traz a relação de itens e formas encontradas no *SDG-Plos*:

Tabela 2 - Itens e Formas do *corpus SDG-Plos*

Formas (*Types*)	**Itens (*Tokens*)**	**Textos**
2,044,674	2,640,307	400

Fonte: Dados do *corpus* da pesquisa.

As formas, ou *types*, são as palavras diferentes contidas no *corpus*, sua quantidade sem repetição. Já os itens, ou *tokens*, são todas as palavras, contando as que se repetem. De acordo com Berber Sardinha (2004), a razão forma/item (FI) indica a variação lexical do *corpus*. Os valores são baseados em um intervalo entre 0 e 1, assim, quanto mais próximo de 1 for o valor de FI, maior a variação lexical, ou seja, maior a riqueza de vocabulário apresentada pelo *corpus*. A razão FI apresentada pelo *SDG-Plos* é de 0,77, indicando uma variação lexical relativamente alta.

Os textos contidos no *corpus* abrangem as áreas demarcadas no *software AntCorGen*: Biologia, Ciências da Computação, Ciências da Informação, Ecologia e Ciências Ambientais, Engenharia e Tecnologia, Medicina, Física, Ciências Políticas e Ciências Sociais. Todos os textos do *corpus* relacionam-se com a temática dos ODS, evidenciando a interconexão entre as áreas mencionadas por meio do elo central que os objetivos da Agenda 2030 representam.

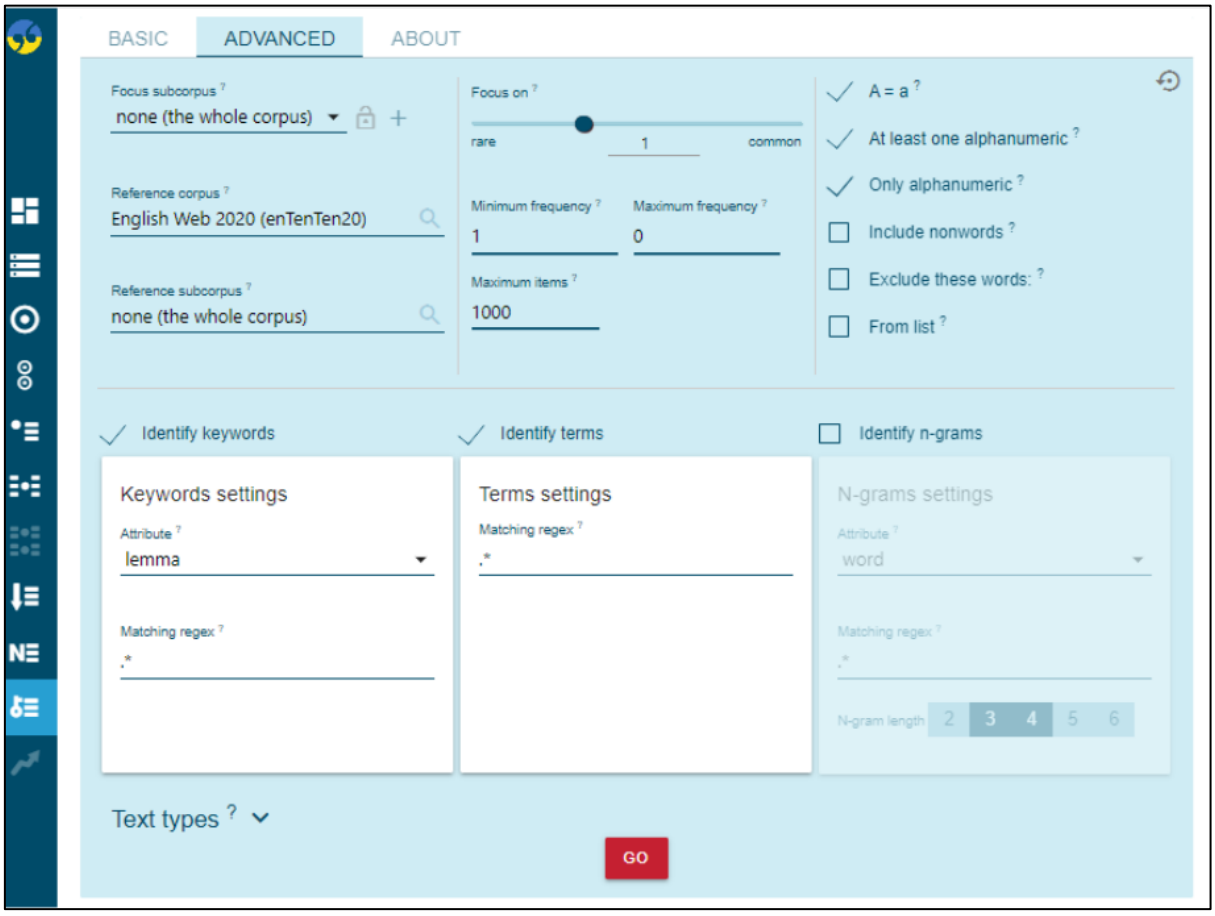
Dessa forma, não foram selecionadas quantidades exatas de textos para cada uma das áreas, visto que não era o objetivo, no momento de criação do *corpus*, selecionar disciplinas específicas, mas obter textos de cientistas que estavam publicando trabalhos voltados aos ODS para identificar os temas mais abordados por pesquisadores (saúde, educação, combate às desigualdades, por exemplo) e as áreas em que essas produções eram mais recorrentes (Pinto *et al*, no prelo; Pinto, 2021).

3.1.4 O levantamento das palavras-chave do *corpus*

Para selecionarmos as palavras-chave que nos interessam para a presente investigação foi preciso escolher aquelas com chavicidade mais alta e que, ao mesmo tempo, apresentassem relação semântica com os ODS. Isso porque a lista gerada pode trazer algumas palavras com alta ocorrência, mas que são apenas itens lexicais comuns ao gênero dos textos contidos no *corpus* ou expressões de outra natureza que acabaram sendo muito utilizadas pelos autores por algum motivo.

Para realizar o levantamento das palavras-chave, utilizamos a ferramenta *Keywords* do *Sketch Engine*. A imagem a seguir traz as informações a respeito dos critérios utilizados para criação da lista de palavras-chave e o *layout* da ferramenta:

Figura 10 – Ferramenta *Keywords*



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

A ferramenta gera duas listas de palavras, a lista *Single Words* e a lista *Multi-Word Terms*. A primeira traz palavras únicas organizadas por ordem decrescente de chavicidade. A segunda lista traz palavras-chave, também organizadas por ordem decrescente de chavicidade, formadas por mais de uma palavra, no caso deste trabalho, formadas por duas.

A lista *Multi-Word Terms* traz palavras-chave mais específicas do que a lista de *Single Words*, pois para que sejam identificadas combinatórias consideradas importantes dentro do campo dos ODS, temática dos artigos contidos no *corpus*, a busca feita pela ferramenta é mais refinada. Por essa razão optamos por selecionar as palavras dessa lista e não da primeira.

A ferramenta *Keywords*, então, gerou a seguinte lista de *Multi-Word Terms*:

Figura 11 – Palavras-chave da lista *Multi-Word Terms*

	Term	Frequency	Frequency per million	Score ?	
1	ecosystem service	465	176.12	85.0	•••
2	s1 file	185	70.07	70.9	•••
3	s1 table	163	61.74	62.6	•••
4	chw program	151	57.19	57.5	•••
5	sensitivity analysis	213	80.67	54.5	•••
6	health facility	290	109.84	52.4	•••
7	maternal health	173	65.52	50.0	•••
8	key informant	157	59.46	49.4	•••
9	effect size	178	67.42	44.8	•••
10	sustained attention	121	45.83	43.1	•••
11	knowledge product	115	43.56	42.5	•••
12	protected area	339	128.39	42.5	•••
13	preventive chemotherapy	104	39.39	40.1	•••
14	s1 appendix	103	39.01	40.0	•••
15	intervention group	134	50.75	39.6	•••

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

Dada a extensão da lista, para ilustrar a forma como as palavras-chave são dispostas na lista *Multi-Word Terms*, das com chavicidade (*score*) mais alta para as com chavicidade mais baixa, apresentamos apenas a primeira página gerada pelo *Keywords*.

Os 17 objetivos são inter-relacionados e interdependentes, isso é, para que se atinjam determinados objetivos, outros precisam ser alcançados. Por exemplo, para que o objetivo 1, Erradicação da Pobreza, seja atingido, é necessário que o 2, Fome Zero, tenha sido alcançado antes. Da mesma forma, para que o objetivo 3, Boa Saúde e Bem-Estar, seja atingido, o 6, Água Limpa e Saneamento também precisa.

Dada a aproximação dos ODS, tanto em relação às ações que implicam na prática quanto ao campo semântico dos nomes que recebem, torna-se difícil elencar palavras-chave do *corpus SDG-Plos* exclusivas para cada um deles. Um exemplo disso é a palavra-chave *ecosystem service*, que pode ser considerada um termo multipalavra do objetivo 7, Energia Acessível e

Limpa, do 11, Cidades e Comunidades Sustentáveis, e do 13, Combate às Alterações Climáticas, e até mesmo outros.

Sabendo disso, selecionamos as palavras-chave consideradas pela ferramenta como termos multipalavras sem a intenção de designar uma para cada um dos ODS exclusivamente, mas com a consciência de que apenas uma delas pode ser representativa de mais de um objetivo.

A seleção das palavras-chave do *corpus*, com base nos critérios esclarecidos, nos levou à seguinte lista de expressões:

Quadro 4 – Termos multipalavras selecionados

(continua)

Termo Multipalavra	**Chavicidade (*score*) no *corpus***
1. *ecosystem service*	89.6
2. *maternal health*	49.4
3. *protected area*	39.7
4. *ocean health*	38.7
5. *waterbird habitat*	35.7
6. *maternal mortality*	35.1
7. *human well-being*	35.1
8. *meat consumption*	34.1
9. *conservation action*	33.2
10. *food insecurity*	33.1
11. *ghg emission*	31.4
12. *sustainable employability*	31.2
13. *water service*	30.9

(conclusão)

14. *health system*	30.5
15. *biodiversity conservation*	30.0

Fonte: Dados da pesquisa.

O Quadro 3 apresenta quinze palavras-chave selecionadas do *corpus* que podem ser entendidas como termos multipalavras, uma vez que são formadas pela combinação de duas palavras lexicais (composto nominal) e podem ser consideradas específicas do *corpus* de estudo (*Sketch Engine*, 2024). Como é possível observar, o valor da chavicidade das expressões escolhidas vai diminuindo, na ordem do número de chavicidade mais alta para o de chavicidade mais baixa. Isso se deve ao fato de termos seguido com rigor o critério de seleção número 2, que aborda a necessidade de relação semântica das expressões com os ODS.

A partir do levantamento dos dados do *corpus SDG-Plos*, os termos multipalavras representativos dos ODS, materiais didáticos podem ser desenvolvidos. As seções seguintes dão ênfase para a elaboração das atividades de leitura, seu contexto de aplicação e os instrumentos de análise de sua aplicabilidade.

3.2 Aplicação pedagógica

Tendo feito o levantamento dos termos multipalavras do *corpus SDG-Plos*, partimos para a criação das atividades de leitura dirigidas por dados com a temática dos ODS. Nesta seção, apresentamos o contexto de aplicação do material desenvolvido, o formulário de Análise de Necessidades respondido pelos participantes, as ferramentas *CorpusMate*, *Skell* e *VersaText* e sua utilidade nos procedimentos de elaboração das atividades, assim como o formulário de avaliação e *feedback* dos alunos.

3.2.1 O contexto de aplicação das atividades

O curso Estratégias de Leitura em Língua Inglesa – Nível Básico foi oferecido no segundo semestre de 2023 e no primeiro semestre de 2024 como disciplina instrumental do Portfólio de Integração de Disciplinas da PROPG da Unesp, como mencionado na introdução deste trabalho. Foram oferecidas quatro turmas com 30 vagas cada uma, intituladas turmas 1, 2, 3 e 4, com horários nos períodos da manhã e da tarde.

As atividades desenvolvidas nesta pesquisa foram aplicadas primeiramente paras as turmas 3 e 4 do segundo semestre de 2023, em que a pesquisadora atuou como tutora. As aulas dessas turmas aconteciam às segundas e às quartas das 16h30 às 18h30, por meio da ferramenta de videoconferências *Google Meet*, durante o período de 7 de agosto a 29 de setembro. A turma 3 era composta por 26 alunos e a turma 4 por 29 alunos, totalizando 55 participantes da pesquisa. O curso utilizava a plataforma *Google Classroom* como ambiente virtual de aprendizagem, onde eram postados os materiais e os demais recursos utilizados.

Na intenção de colher novos dados para analisar as impressões dos alunos acerca das atividades quando aplicadas por professores diferentes, que não fosse a pesquisadora, incluímos como participantes os estudantes das quatro turmas oferecidas no primeiro semestre de 2024, cujas tutoras não eram especialistas em Linguística de *Corpus* nem em DDL.

As aulas oferecidas no primeiro semestre de 2024 aconteceram durante o período de 4 de março a 26 de abril, também com turmas nos períodos da manhã e da tarde. Nesta segunda aplicação das atividades, a pesquisadora apresentou o material desenvolvido e as ferramentas utilizadas nos exercícios (*CorpusMate*, *Skell* e *VersaText*) para as novas tutoras por meio de um vídeo explicativo. Em seguida, elas incluíram tais recursos no currículo do curso e os apresentaram para seus alunos.

Da turma 1 – 2024, 15 alunos aceitaram participar da pesquisa por meio de Termo de Consentimento Livre e Esclarecido (TCLE), e da turma 2 – 2024, 17 alunos aceitaram participar. Na turma 3 – 2024, obtivemos 20 participantes, e da turma 4 – 2024, 18 participantes. No total, 70 alunos das turmas do primeiro semestre de 2024 aceitaram participar da presente pesquisa.

Os mesmos conteúdos estavam previstos para os quatro grupos. Para ilustrar os temas abordados nas aulas, apresentamos o cronograma do curso de 2023 com o conteúdo programático previsto na imagem a seguir:

Figura 12 - Cronograma do curso

Estratégias de Leitura em Língua Inglesa - Nível Básico

Turmas do 2° Semestre de 2023

Cronograma

Período	Programação
De 07/08 a 11/08	Introdução às estratégias de leitura: layout, inferência e cognatos (+1 atividade assíncrona)
De 14/08 a 18/08	Palavras-chave e falsos cognatos (+1 atividade assíncrona)
De 21/08 a 25/08	Skimming e Scanning (+1 atividade assíncrona)
De 28/08 a 01/09	Formação de palavras em inglês: Processos de afixação; e Análise de textos acadêmicos com foco para os abstracts (+1 atividade assíncrona)
De 04/09 a 08/09	Coesão e coerência (+1 atividade assíncrona)
De 11/09 a 15/09	Marcadores do discurso (+1 atividade assíncrona)
De 18/09 a 22/09	Elementos de referência (+1 atividade assíncrona)
De 25/09 a 29/09	Avaliação final

Fonte: Captura de tela do Cronograma do curso Estratégias de Leitura em Língua Inglesa – Nível Básico.

As atividades desenvolvidas foram aplicadas juntamente com outras já previstas no material didático do curso, então os alunos fizeram exercícios baseados em *corpus* e exercícios de outras naturezas, o que diversificou as metodologias empregadas nas aulas. Antes de aplicar o material preparado com base em *corpus*, a tutora juntamente com a professora responsável pela disciplina, em 2023, deu uma aula com tarefas simples de investigação das ferramentas *CorpusMate*, *Skell* e *VersaText* para que os alunos pudessem se familiarizar com elas. Em 2024, cada uma das tutoras responsáveis orientou suas turmas com total autonomia, com base no vídeo explicativo oferecido pela pesquisadora.

No início do curso, os estudantes responderam a um questionário de Análise de Necessidades, atividade comum tanto para as turmas de 2023 quanto de 2024. Como não foi possível ter acesso ao formulário das turmas em que a pesquisadora não atuou como tutora, apresentamos as questões e uma discussão sobre as respostas dos alunos apenas das Turmas 3 e 4 de 2023.

O formulário *Needs Analysis* não é de autoria da pesquisadora, ele já era um instrumento usado por tutores da Unesp que participavam dos cursos de IFE na universidade. Ele foi aplicado no primeiro dia de aula síncrona das turmas 3 e 4 como atividade inicial. As questões contidas nesse questionário eram as seguintes:

1- Nome:
2- Qual é o seu nível de proficiência em inglês? (Se você não tem certeza, indique o que você acredita que possa ser seu nível de proficiência)
3- Fale um pouco sobre sua atuação na pós-graduação (curso, programa, atividades desenvolvidas, leituras específicas, contato com estrangeiros etc.).
4- Já teve que se comunicar em inglês com um estrangeiro? 4.1- Em caso de resposta afirmativa na questão anterior, qual foi sua maior dificuldade?
5- Você já estudou inglês anteriormente? Por quanto tempo? Em qual contexto (aula particular, escola de idiomas, escola regular...)?
6- Comente um pouco sobre a sua experiência de aprendizagem de inglês. Você gosta de estudar inglês? Sente que tem facilidade ou dificuldade com o idioma?
7- Em qual das habilidades abaixo você tem mais **facilidade** ao se comunicar em inglês?
8- Em qual das habilidades abaixo você tem mais **dificuldade** ao se comunicar em inglês?
9- Quais dos itens abaixo você acredita serem os maiores obstáculos para sua comunicação em inglês?
10- O que você espera deste curso? Que objetivos você quer alcançar com o inglês?

Apesar de o curso ser focado na habilidade de leitura, o questionário *Needs Analysis* foi aplicado dada a necessidade de mapear características do alunado, como seus conhecimentos prévios em LI, seu contato com a língua-alvo, suas expectativas em relação ao curso etc. O item 1 pede o nome dos estudantes apenas para que os professores os conheçam melhor, vale lembrar que o formulário em questão não continha perguntas em que os alunos avaliassem o curso, por isso, descarta-se a possibilidade de possível constrangimento por parte dos respondentes.

A pergunta 2 serve para o professor ter uma noção do nível de variação em proficiência linguística dos alunos. Ainda que o curso tenha sido ofertado para estudantes de nível básico com base nos resultados das pesquisas em proficiência na universidade (Pinto *et al*, 2021), muitos alunos se inscrevem nos cursos de IFE pela oportunidade de estar em contato com o idioma e para aprofundar seus conhecimentos.

Nas perguntas 3 e 4 os estudantes dão informações a respeito de suas áreas de formação/atuação, como Programa de Pós-Graduação, áreas de pesquisa, câmpus e cidade, assim como informações a respeito do uso que fazem da LI no contexto de pesquisa em que estão inseridos. As perguntas 5, 6, 7 e 8, de forma geral, abordam o repertório de estudo que os estudantes possam ter tido, suas crenças a respeito da LI e de seu próprio processo de aprendizagem.

Já na questão 9, os aprendizes podem elencar fatores que consideram obstáculos para sua comunicação em inglês. Os itens apresentados para os alunos são:

- Falta de vocabulário;
- Pouco conhecimento gramatical;
- Pouco incentivo ou poucas oportunidades para usar o idioma;
- Dificuldade para compreender o que leio;
- Dificuldade para compreender o que escuto;
- Dificuldade para falar; Insegurança ao falar;
- Dificuldade para escrever em outro idioma;
- Medo de cometer erros (na modalidade escrita e/ou falada);
- Medo de ser criticado;
- Pouca experiência com o idioma;
- Pouco tempo para me dedicar aos estudos.

Os itens listados ajudam a traçar o perfil dos alunos e a pensar nas estratégias pedagógicas a serem colocadas em prática nas aulas.

A questão 10 é a mais cara para esta pesquisa, uma vez que revela o que os alunos almejam aprender mesmo sabendo que estão matriculados em um curso com objetivos previamente estabelecidos – o de apresentar estratégias de leitura para aprendizes iniciantes de inglês em contexto acadêmico. Os resultados obtidos a partir da coleta de dados iniciais dos estudantes por meio do formulário supracitado são apresentados na seção 4, Análise dos Dados.

As mesmas atividades elaboradas nesta pesquisa, aplicadas para as Turmas 3 e 4 de 2023, foram aplicadas para as quatro turmas de 2024. Todas pautavam-se nos conteúdos previstos no cronograma. A imagem seguinte apresenta um *slide* de uma aula em que a professora e a tutora das Turmas 3 e 4, de 2023, ensinaram como buscar por palavras-chave no *VersaText*:

Figura 13 - Como usar o *VersaText* para leitura

Fonte: Captura de tela de um dos *slides* usados nas aulas do curso.

Na aula anterior a de apresentação do *VersaText*, os alunos realizaram a leitura em grupo de textos selecionados e tiveram que escolher palavras-chave para o texto lido. Na aula seguinte, depois de terem assistido à explicação da professora e da tutora, puderam gerar uma nuvem de palavras-chave no *VersaText* e comparar as palavras dadas pela ferramenta com as que eles haviam selecionado, assim, puderam discutir sobre as noções de frequência de palavras e de palavras-chave para a LC.

Nas seções a seguir, serão aprofundadas as características de cada ferramenta utilizada e o processo de elaboração das atividades.

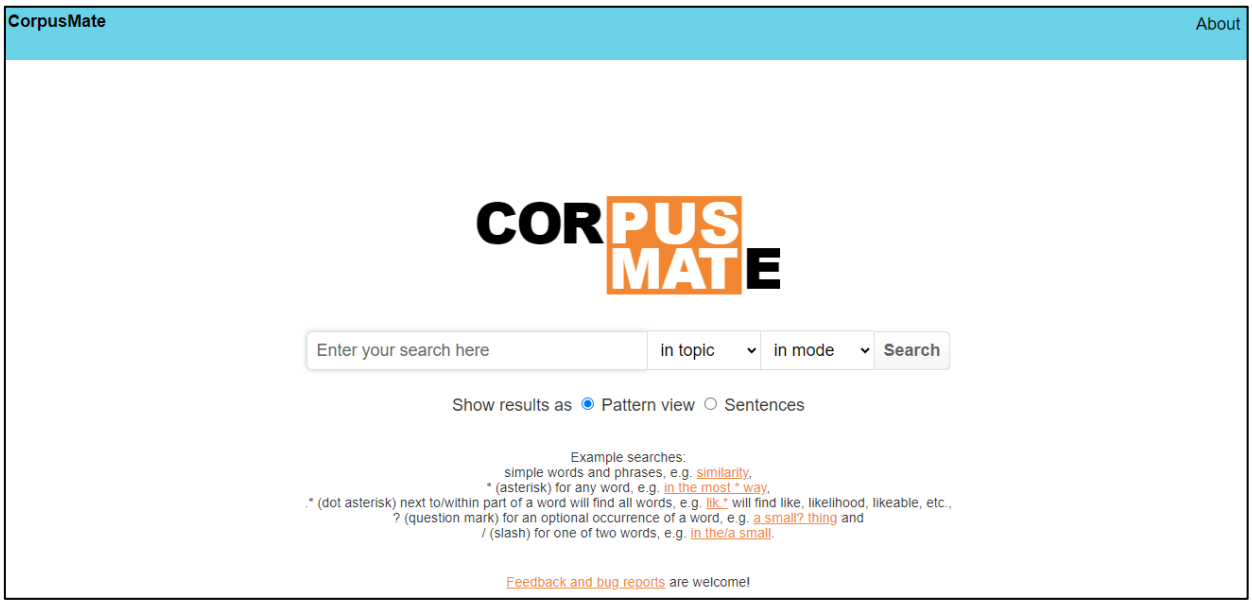
3.2.2 As ferramentas de *corpora* voltadas ao ensino: *CorpusMate*, *Skell* e *VersaText*

O *CorpusMate* é uma plataforma desenvolvida pelos pesquisadores Peter Crosthwaite e Vít Baisa (2022) pensada para o ensino de línguas baseado e dirigido por *corpora* a alunos de ensino básico (Baisa; Crosthwaite, 2022). Seu *corpus* foi compilado a partir de cinco recursos,

de acordo com os autores, o *corpus* BAWE, o *corpus* de *TED Talks*, o *Simple English Wikipedia*, o *corpus BBC Tech*, o *Elsevier Corpus* e o *BNC 2014 Spoken*, o que fornece uma gama de dados da modalidade escrita e falada da língua.

A imagem a seguir apresenta o *layout* da plataforma:

Figura 14 – O *CorpusMate*



Fonte: Captura de tela da ferramenta CorpusMate (Baisa; Crosthwaite, 2022).

O *CorpusMate* abriga textos de cerca de 21 áreas diferentes, entre elas História, Geografia, Política, Química, Matemática, Psicologia e Enhenharia (Baisa; Crosthwaite, 2022). Entre os recursos oferecidos pela plataforma, podemos citar as estatísticas referentes à expressão de busca, os padrões encontrados com a palavra pesquisada e a sua distribuição pelas áreas em que ocorre. Um dos exercícios elaborados explora esse último recurso e será apresentado mais adiante.

O *CorpusMate* é uma plataforma muito proveitosa para este estudo por ser pensado para alunos de ensino básico, ou seja, fora do contexto acadêmico. Por essa característica, o *software* é mais simples que os concordanciadores de análise de *corpus* tradicionais, o que facilita o uso para estudantes entre os níveis A1 e A2.

Outra plataforma de *corpora* voltada ao ensino utilizada para o desenvolvimento das atividades é o *Skell* (*Sketch Engine for Language Learners*) (Baisa; Suchomel, 2014), uma vertente simplificada do *Sketch Engine* com *layout* amigável e com recursos que são aproveitados dos *corpora* de seu *software* de origem. A imagem que segue apresenta seu *layout*:

Figura 15 – O *Skell*

Fonte: Captura de tela da ferramenta *Skell* (Baisa; Suchomel, 2014).

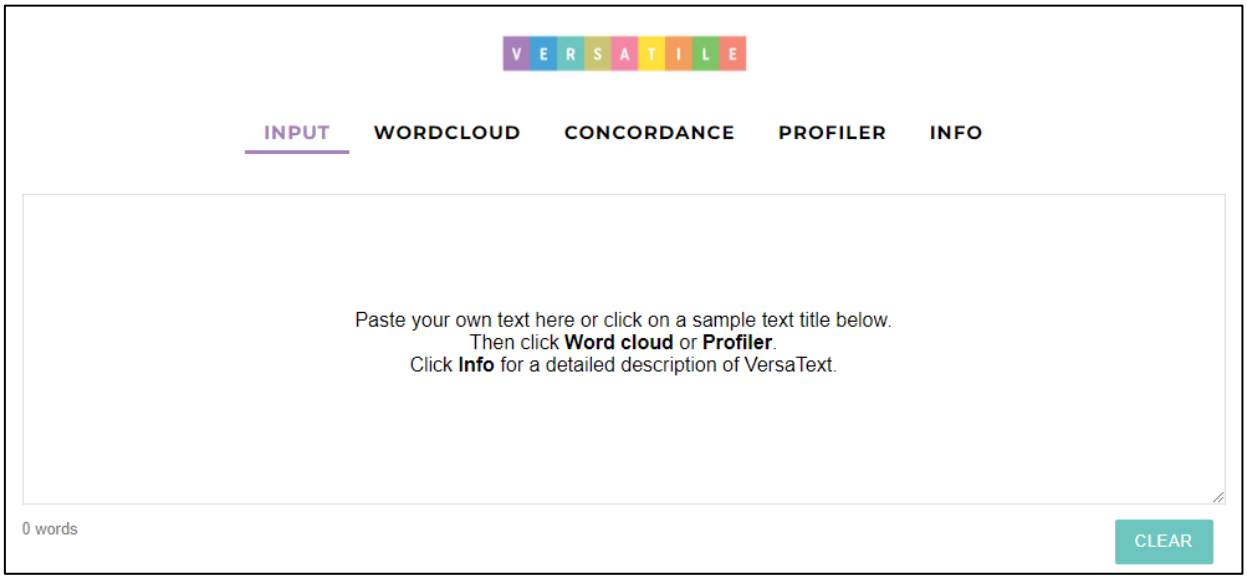
Os recursos do *Skell* são o *Examples*, que apresenta frases retiradas dos *corpora* do *Sketch Engine* com as palavras de busca para que o aluno as veja em contexto; o *Word sketch*, que oferece ao aluno as combinações da palavra de busca com outras que podem atuar como seu sujeito, seu objeto, seu modificador etc; e o *Similar Words*, que traz sinônimos da palavra de busca.

A terceira ferramenta é o *VersaText*, que "oferece ferramentas para a exploração linguística de um único texto"[35] (Thomas, 2020, p. 9, tradução nossa). Nele, o aluno pode colar amostras de textos e investigá-las quanto ao seu vocabulário por meio da criação de nuvem de palavras através do *Wordcloud*, da análise estatística e morfológica do léxico por meio do *Profiler* ou pela leitura detalhada das linhas de concordância, com os termos coloridos a partir da etiquetagem automática já disponibilizada na ferramenta, por meio do *Concordance*.

O *VersaText* é totalmente pensado para facilitar a exploração da linguagem feita por alunos ou do ensino básico ou de níveis de proficiência mais baixos, isso porque apresenta vocabulário menos específico do que o *Skell*, que traz exemplos provindos dos vários *corpora* do *Sketch Engine*, e porque explora recursos visuais, como o uso de cores diferentes para as palavras da nuvem de palavras e das concordâncias geradas a partir das amostras de textos coladas na ferramenta. A seguir, apresentamos o *layout* da plataforma:

[35] No original: *[...] offers tools for exploring the language of a single text.* (Thomas, 2020, p. 9)

Figura 16 – O *VersaText*



Fonte: Captura de tela da ferramenta *VersaText* (Thomas, 2020).

Na imagem, é possível ver os ícones dos recursos *Wordcloud*, *Concordance* e *Profiler*. O espaço em branco, do item *Input*, é destinado para a colagem da amostra do texto que se deseja investigar.

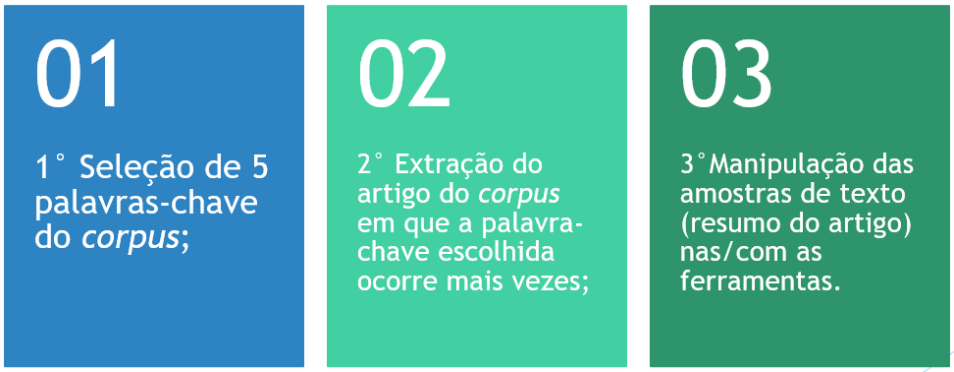
As atividades, exemplificadas na próxima seção, demostram como os recursos mencionados podem ser utilizados.

3.2.3 A elaboração das atividades de leitura direcionadas por dados

Apresentamos, a seguir, as etapas de seleção dos textos do *corpus* que seriam lidos nas aulas. Inicialmente, foram selecionadas os termos multipalavras considerados representativos dos ODS. Como a lista gerada pelo *Keywords* é muito longa, foi necessário selecionar uma quantidade de textos viável para o trabalho em sala de aula. Então, selecionamos cinco termos multipalavras que gerariam as buscas pelos textos a serem levados para as aulas. Em seguida, foram extraídos os artigos do *corpus* nos quais os termos multipalavras elencados eram mais frequentes. Estes artigos foram usados para a elaboração das atividades.

A imagem a seguir ilustra as etapas de elaboração das atividades:

Figura 17 - Etapas da elaboração das atividades



Fonte: Elaboração própria.

As cinco palavras-chave selecionadas foram retiradas da lista *Multi-Word Terms*, logo, consideradas termos multipalavras conforme os critérios do *Sketch Engine*. Os termos multipalavras dos ODS escolhidos para a extração dos textos foram: *ecosystem service* (com chavicidade igual a 89.6), *protected area* (com chavicidade igual a 39.7), *ocean health* (com chavicidade igual a 38.7) e *waterbird habitat* (com chavicidade igual a 35.7). Essas são as palavras-chave do topo da lista apresentada no Quadro 3. Não selecionamos a palavra-chave *maternal health* porque, o artigo em que ela ocorre mais vezes traz vocabulário relacionado às áreas de obstetrícia e ginecologia, que se distancia do campo semântico relacionado às outras palavras-chave do início da lista.

Selecionados os textos, optamos por trabalhar com a seção de *abstracts*, ou seja, os resumos dos artigos, que são de extensão curta e contêm de forma sintetizada os objetivos, o referencial teórico, a metodologia e os resultados dos trabalhos. Essa organização do material que seria lido e utilizado nas atividades foi fundamental dado o tempo de duração do curso, que não permitia a leitura detalhada de todas as seções dos artigos.

Os resumos foram manipulados no *VersaText*, e a partir dos recursos contidos na ferramenta, bem como dos recursos do *Corpus Mate* e do *Skell*, elaboramos os exercícios de cada atividade. Foram desenvolvidas cinco atividades, mas foi possível aplicar somente quatro em razão da carga horária do curso. O quadro a seguir mostra as atividades feitas a partir de cada termo multipalavra:

Quadro 5 - Relação dos termos multipalavras e dos conteúdos referentes a cada atividade

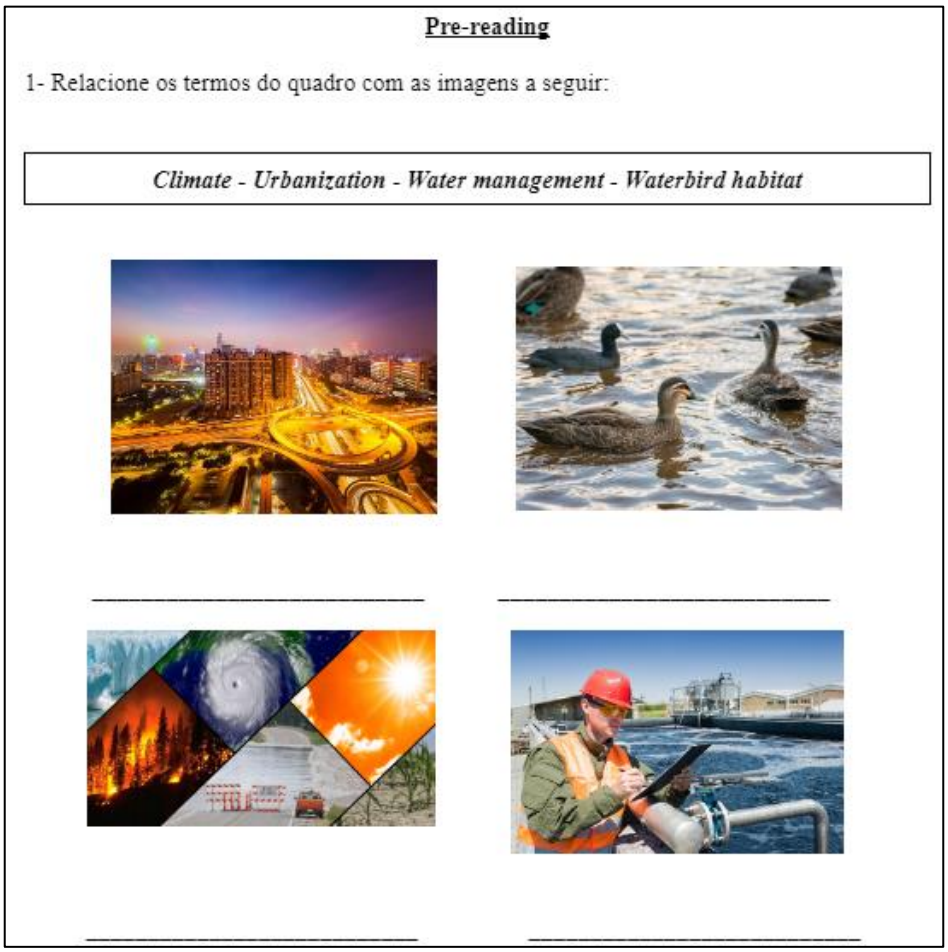
Atividade	**Termo multipalavra a partir do qual foi elaborada**	**Conteúdos do cronograma do curso abordados na aula em que foram aplicadas**
1	*protected area*	Características dos *abstracts*
2	*waterbird habitat*	Coesão e coerência
3	*ocean health*	Marcadores do discurso
4	*human well-being*	Elementos de referência

Fonte: Elaboração própria.

Essa relação entre a atividade desenvolvida a partir dos termos multipalavras dos ODS e os conteúdos previstos nas semanas em que foram empregadas nas aulas ilustra o cenário de aplicação encontrado pela pesquisadora em 2023. As tutoras de 2024 escolheram a ordem de aplicação das atividades conforme seus próprios critérios.

Cada uma das quatro atividades está estruturada em exercícios *pre-reading*, *reading* e *post-reading*. Todas as atividades se encontram nos apêndices ao final deste trabalho, mas para ilustrar sua organização, iremos descrever a atividade número 2. A imagem a seguir apresenta o primeiro exercício da seção *pre-reading*, em que o aluno deve relacionar as palavras do quadro com as imagens dadas:

Figura 18 - Exercício 1 da seção *pre-reading* da atividade 2

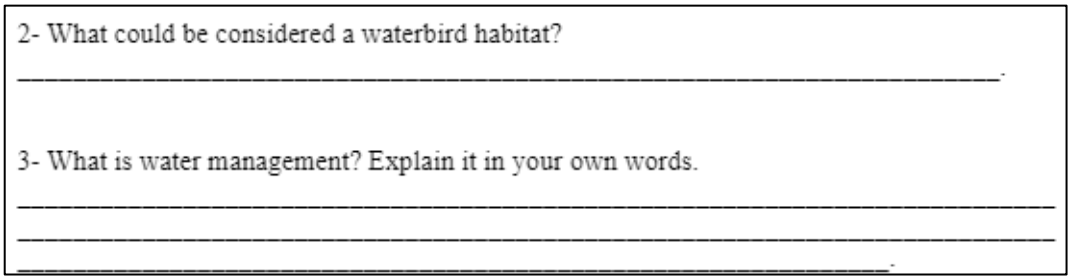


Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Para criar esta atividade, manipulamos no *VersaText* o resumo do artigo cujo termo multipalavra *waterbird habitat* era mais recorrente e geramos uma *wordcloud*, nuvem de palavras. As palavras que mais se destacaram na nuvem de palavras foram as escolhidas para estarem no quadro do exercício apresentado na imagem. Este exercício fez com que os alunos entrassem em contato com o vocabulário do texto que iriam ler e fizessem o acionamento de esquemas para ativarem seu conhecimento prévio acerca do assunto tratado (Leffa, 1996) por meio da observação e análise das imagens.

Os exercícios seguintes, 2 e 3, exploram as palavras indicadas no exercício 1 e o conhecimento prévio dos alunos. Eles deveriam dar possíveis definições para *waterbird habtat* e *water management* com base nas informações que já possuíam sobre o tema abordado no texto:

Figura 19 – Exercícios 2 e 3 da seção *pre-reading* da atividade 2

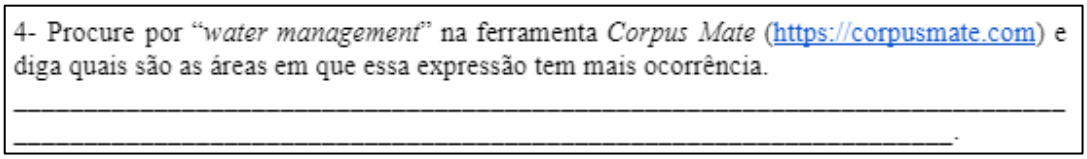


Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Os enunciados dessas questões estão em inglês porque essa já era a segunda atividade a ser feita pelos alunos e porque são perguntas consideradas simples. À medida em que o curso se aprofundava nas estratégias de leitura, os enunciados em inglês foram sendo incorporados nas atividades como uma forma de incentivar a leitura e a escrita dos estudantes na língua alvo. Todavia, eles não eram obrigados a responderem esses exercícios em inglês.

As perguntas 3 e 4 funcionaram como uma pesquisa para o professor e para os alunos, que puderam conferir o que se sabia sobre o tema. Depois disso, os alunos tiveram a oportunidade de conferir suas respostas pesquisando o vocabulário ao qual foram expostos com ferramentas de aprendizagem de *corpora*. O Exercício 4, ainda da seção *pre-reading*, solicita ao aluno que procure uma combinação de palavras, padrão lexical importante para a compreensão do texto:

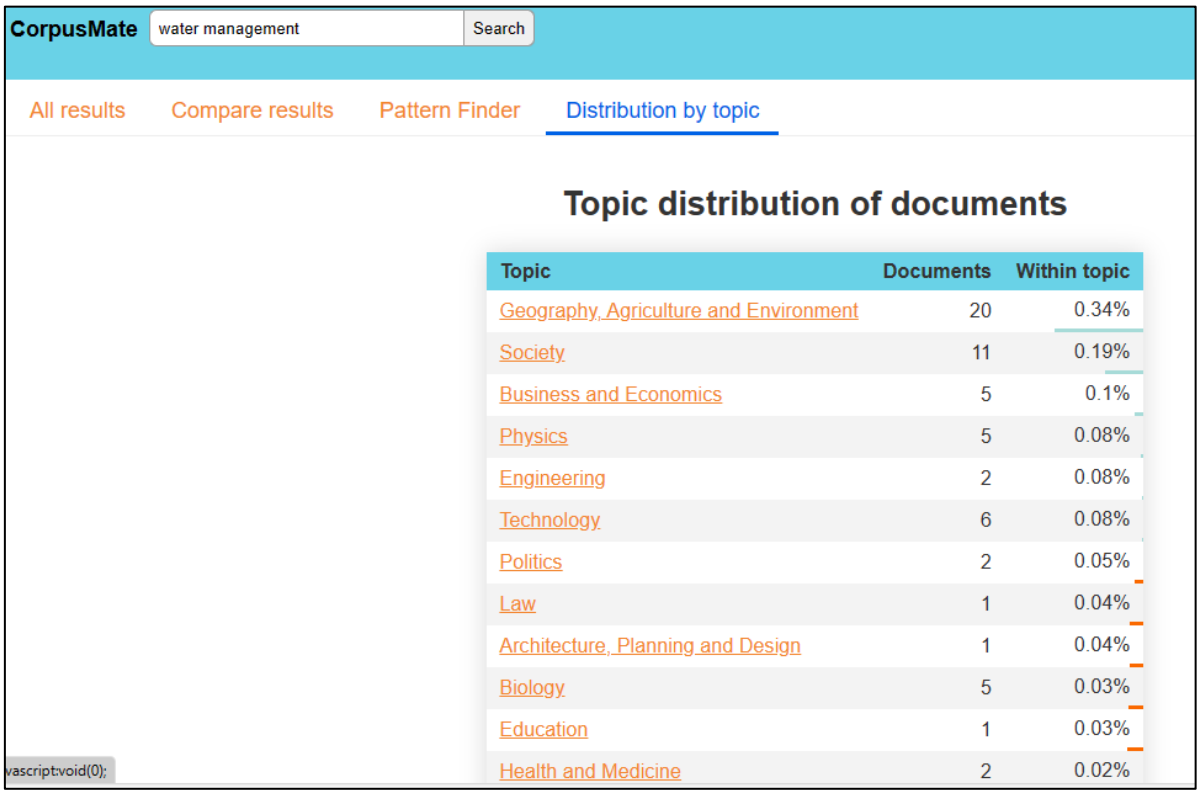
Figura 20 - Exercício 4 da seção *pre-reading* da atividade 2



Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Neste exercício, o aluno explora a ferramenta *CorpusMate* ao procurar pelas áreas em que a combinação mais aparece, o que permite que antecipe informações sobre o conteúdo do texto. Ao fazer a busca na ferramenta, o aluno pode encontrar informações como as que seguem na imagem:

Figura 21 - Recurso *Distribution by topic* da ferramenta *CorpusMate*

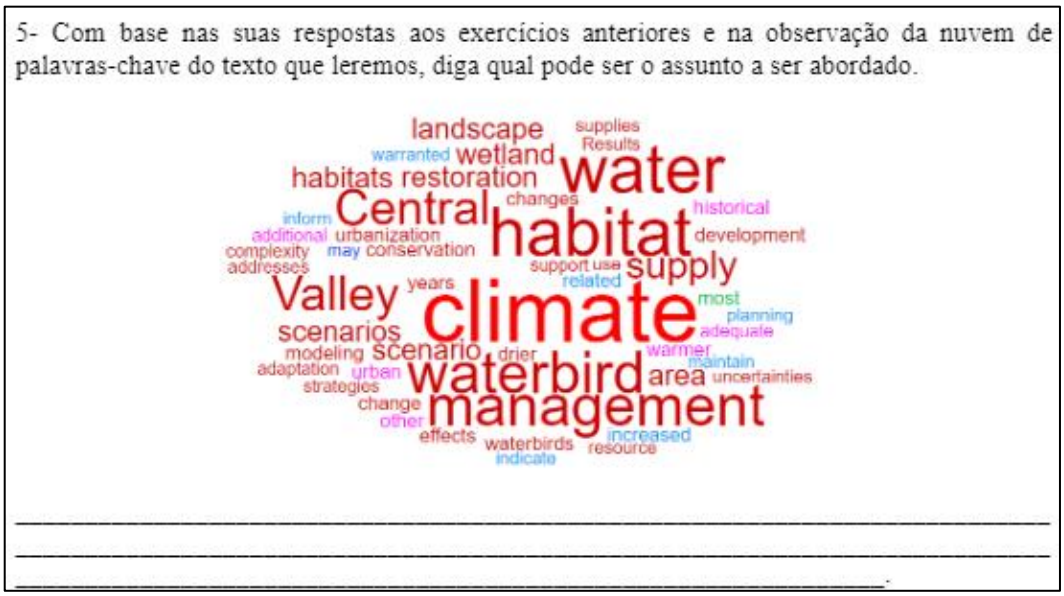


Fonte: Captura de tela da ferramenta *CorpusMate*.

O aluno pode clicar nos nomes das áreas, em laranja, e ver as linhas de concordância para comparar como diferentes pesquisadores estão utilizando a combinação *water management* (gestão da água). Eles podem aprender diferentes maneiras de usar uma expressão e ver se estão corretos quanto ao seu significado. Este exercício ofereceu aos alunos a oportunidade de partilhar e discutir ideias, uma vez que podem ser dadas múltiplas respostas.

O exercício 5 do *pre-reading* pede a observação da nuvem de palavras do resumo do artigo gerada pelo *VersaText* para que o aluno, com base nos conhecimentos acionados nos exercícios anteriores, diga qual pode ser o assunto tratado no texto.

Figura 22 - Exercício 5 da seção *pre-reading* da atividade 2



Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

A questão 5 finaliza a seção *pre-reading* com um exercício de escrita para a sistematização de todo o conhecimento prévio levantado e discutido ao longo da sondagem inicial feita por meio de cada um dos exercícios. Ao analisar a nuvem de palavras, os estudantes também devem ser capazes de identificar palavras cognatas, bem como se atentar para as diferentes cores em que as palavras se apresentam.

A seção seguinte é reservada para a leitura do resumo do artigo, que já compõe um dos exercícios da atividade. A próxima imagem traz as instruções dadas para o momento de leitura:

Figura 23 – Momento *Reading* da atividade 2

Reading

- Click here to read the text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169780;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

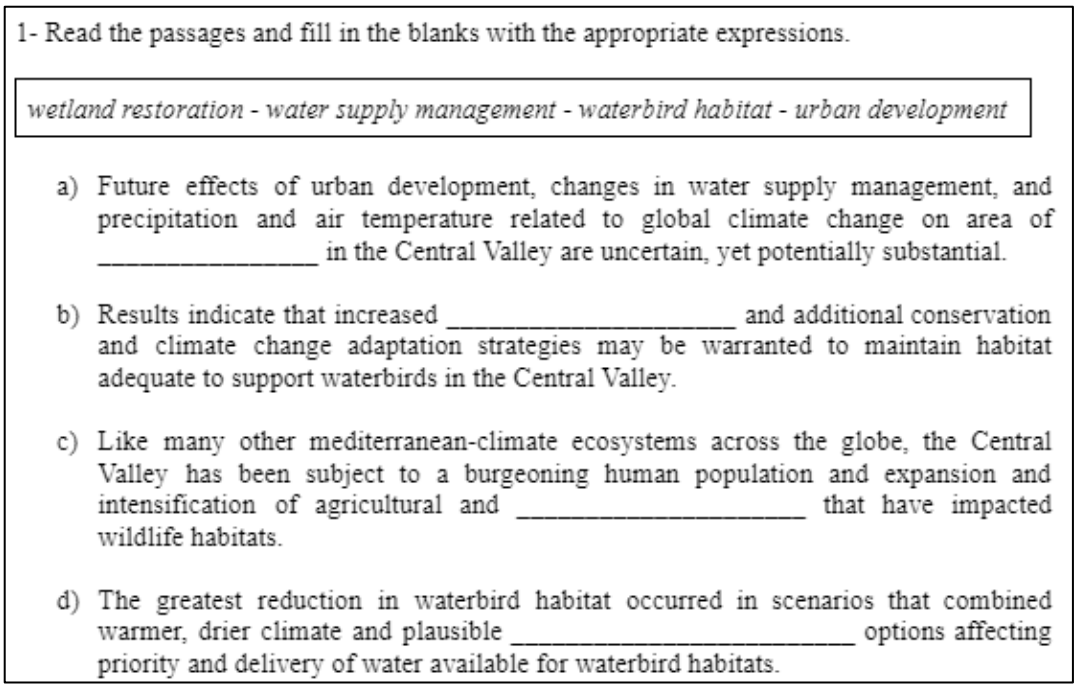
Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Os alunos puderam ler o resumo no próprio site em que foi publicado, seu suporte de veiculação original, já que compõe uma atividade movida por dados centrada no texto. Nas

instruções, pede-se que os estudantes observem o *layout* do texto e façam uma leitura rápida, *skimming*, para mobilizar as estratégias já vistas no início do curso.

A última seção da atividade é a *post-reading*, em que constam exercícios de consolidação da compreensão a partir da leitura feita. No exercício 1, os alunos deveriam completar lacunas com padrões de combinações importantes que aparecem ao longo do resumo do artigo. Para identificar esses padrões, no momento de elaboração do exercício, clicamos nas palavras que se destacavam na *wordcloud* e observamos o modo como eram usadas nas linhas de concordância geradas pelo *VersaText*. Assim, ao clicarmos na palavra *development*, por exemplo, verificamos que esta ocorria com o adjetivo *urban*.

Figura 24 - Exercício 1 da seção *post-reading* da atividade 2



1- Read the passages and fill in the blanks with the appropriate expressions.

wetland restoration - water supply management - waterbird habitat - urban development

a) Future effects of urban development, changes in water supply management, and precipitation and air temperature related to global climate change on area of _____ in the Central Valley are uncertain, yet potentially substantial.

b) Results indicate that increased _____ and additional conservation and climate change adaptation strategies may be warranted to maintain habitat adequate to support waterbirds in the Central Valley.

c) Like many other mediterranean-climate ecosystems across the globe, the Central Valley has been subject to a burgeoning human population and expansion and intensification of agricultural and _____ that have impacted wildlife habitats.

d) The greatest reduction in waterbird habitat occurred in scenarios that combined warmer, drier climate and plausible _____ options affecting priority and delivery of water available for waterbird habitats.

Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Os trechos de cada alternativa são partes retiradas do resumo em que as combinações dadas no quadro aparecem originalmente. Nesse exercício, o aprendiz pode focar sua atenção para as expressões dadas e refletir sobre seus sentidos no contexto em que foram empregadas de forma mais atenciosa do que em uma primeira leitura rápida.

O exercício 2 da seção *post-reading* focaliza o conteúdo de marcadores do discurso, ou *linking words*:

Figura 25 – Exercício 2 da seção *post-reading* da atividade 2

2- Assinale a alternativa em que a *linking word* destacada exerça a mesma função que "despite" no trecho a seguir: *The Central Valley of California is one of the most important regions for wintering waterbirds in North America* ***despite*** *extensive anthropogenic landscape modification and decline of historical wetlands there.*

a) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **because of the** extensive anthropogenic landscape modification and decline of historical wetlands there.
b) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **as well as** extensive anthropogenic landscape modification and decline of historical wetlands there.
c) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **such as** extensive anthropogenic landscape modification and decline of historical wetlands there.
d) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **even with** extensive anthropogenic landscape modification and decline of historical wetlands there.

Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Já os exercícios 3 e 4, do *post-reading*, compõem exercícios de interpretação em que os alunos deveriam mobilizar tanto o *skimming* quanto o *scanning* para a identificação de informações contidas no texto:

Figura 26 – Exercícios 3 e 4 da seção *post-reading* da atividade 2

3- O *Central Valley*, ou Vale Central, em português, é um ecossistema

a) de clima equatorial, onde se encontra boa parte da vegetação nativa dos Estados Unidos.
b) de clima mediterrâneo, em que houve intensificação do desenvolvimento agrícola e urbano dos Estados Unidos.
c) de clima semi-árido, que concentra boa parte do desenvolvimento urbano e agrícola dos Estados Unidos.
d) de clima mediterrâneo, onde se encontra boa parte da vegetação nativa dos Estados Unidos.

4- De acordo com o texto, são considerados efeitos futuros do desenvolvimento urbano no habitat de aves aquáticas no Vale Central:

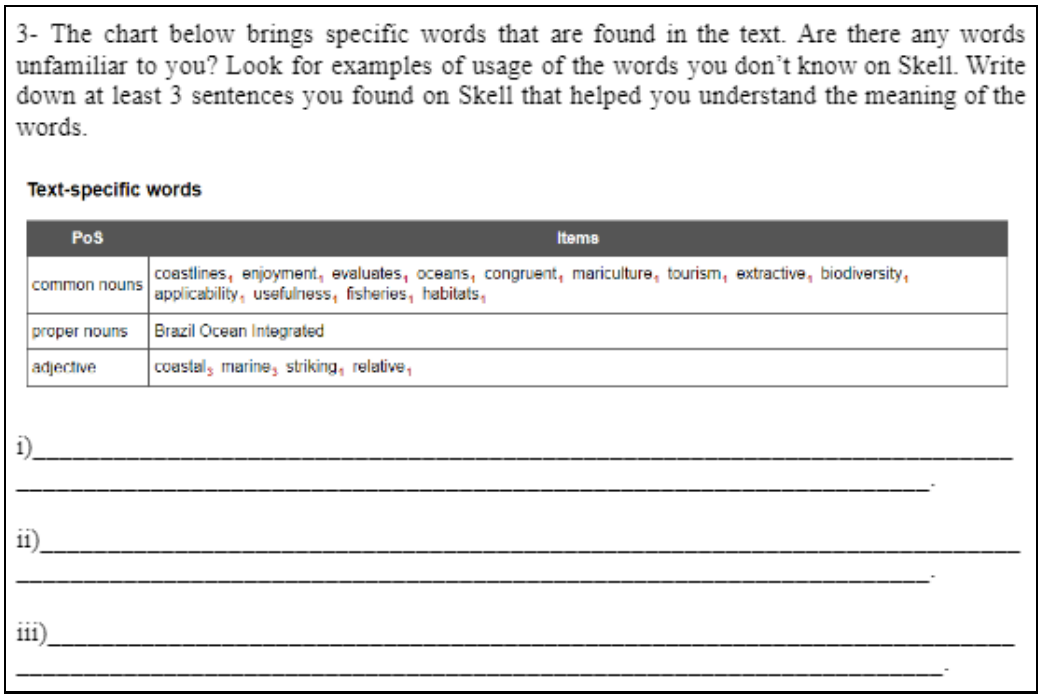
a) manutenção dos modelos de gestão do abastecimento de água e alterações de temperatura relacionadas à mudança climática global.
b) manutenção dos modelos de gestão do abastecimento de água e alterações de temperatura e precipitação relacionadas à mudança climática global.
c) alterações na gestão do abastecimento de água e alterações de temperatura e precipitação relacionadas à mudança climática global.
d) alterações na gestão do abastecimento de água e manutenção da temperatura e dos índices de precipitação.

Fonte: Captura de tela da atividade 2 desenvolvida na pesquisa.

Para ilustrar um exercício em que o aluno deveria utilizar a ferramenta *Skell*, apresentamos a seguir os exercícios 3 da *seção pre-reading* da atividade 3, elaborada a partir do termo-multipalavra *ocean health*, e o 5 da seção *pre-reading* da atividade 4, elaborada a partir do termo multipalavra *human well-being*.

Ao fornecer uma amostra de texto ao *VersaText*, ele apresenta um quadro com as palavras que podem ser consideradas pertencentes ao grupo do vocabulário acadêmico. Isso é feito a partir de uma análise estatística da ferramenta. A imagem a seguir ilustra o quadro de palavras acadêmicas gerado pelo *VersaText*:

Figura 27 – Exercício 3 da seção *pre-reading* da atividade 3



3- The chart below brings specific words that are found in the text. Are there any words unfamiliar to you? Look for examples of usage of the words you don't know on Skell. Write down at least 3 sentences you found on Skell that helped you understand the meaning of the words.

Text-specific words

PoS	Items
common nouns	coastlines$_1$, enjoyment$_1$, evaluates$_1$, oceans$_1$, congruent$_1$, mariculture$_1$, tourism$_1$, extractive$_1$, biodiversity$_1$, applicability$_1$, usefulness$_1$, fisheries$_1$, habitats$_1$
proper nouns	Brazil Ocean Integrated
adjective	coastal$_3$, marine$_3$, striking$_1$, relative$_1$

i)_____
_____.

ii)_____
_____.

iii)_____
_____.

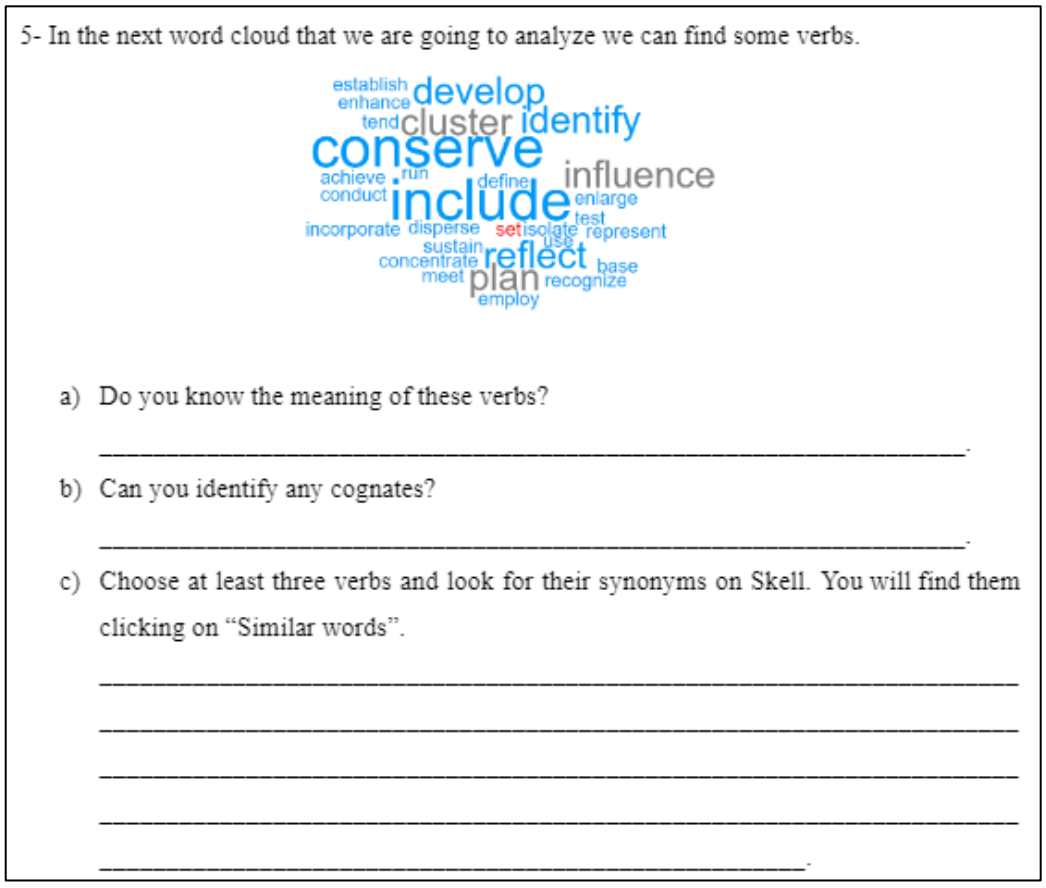
Fonte: Captura de tela da atividade 3 desenvolvida na pesquisa.

No exercício exemplificado acima, os estudantes analisam as palavras do texto lido consideradas do campo acadêmico. Então, sua tarefa é identificar aquelas cujos significados não conheçam para buscar por exemplos de usos no *Skell*. O aluno é induzido a pesquisar informações a respeito do vocabulário acadêmico em uma ferramenta que traz amostras correntes da língua, assim, aprende a investigar o idioma ao invés de procurar pelos seus

significados isolados. Esta também pode ser uma forma de se preparar para a leitura do texto, que apresentará vocabulário específico e, muitas vezes, desconhecido pelo aluno.

O exercício 5 de *pre-reading* da atividade 4 foi criado a partir da nuvem de palavras do *VersaText* com apenas os verbos do resumo lido. A ferramenta em questão oferece a opção de criação de nuvens de palavras com palavras de conteúdo (substantivos, verbos, adjetivos, advérbios) e com palavras funcionais (preposições, conjunções, pronomes, artigos), de modo que o usuário possa selecionar as palavras da *wordcloud*.

Figura 28 – Exercício 5 da seção *pre-reading* da atividade 4



Fonte: Captura de tela da atividade 4 desenvolvida na pesquisa.

Nesse exercício, o aluno deve tentar reconhecer os cognatos dos verbos e tentar encontrar sinônimos para eles por meio do recurso *Similar words* do *Skell*. O enunciado das questões da atividade 4 aparece em inglês, mas os alunos tinham a liberdade de responder em português se quisessem, como pontuamos anteriormente.

Na seção a seguir, apresentamos os questionários aplicados aos alunos para colher informações a respeito de suas impressões sobre as atividades.

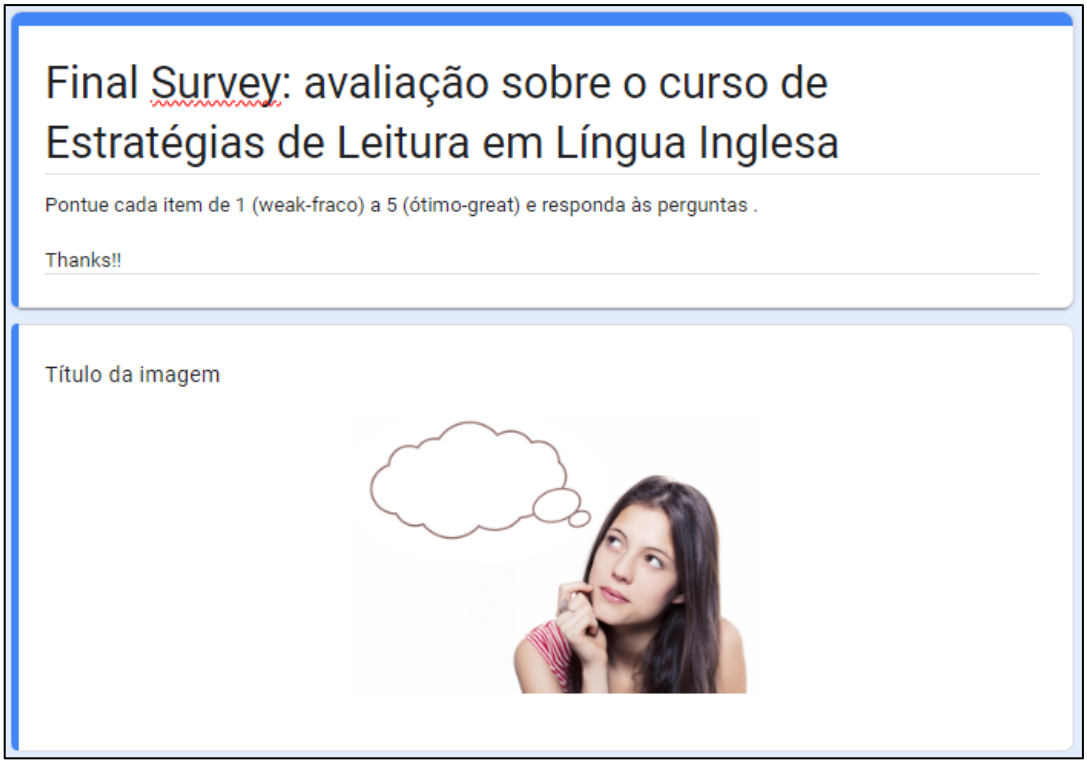
3.2.4 Os questionários aplicados ao final do curso

Como instrumento de análise da percepção dos alunos a respeito das atividades, utilizamos dois questionários em formato de formulário do *Google Forms* com os participantes da pesquisa. Apresentaremos primeiro o questionário utilizado com as turmas de 2023 e depois o utilizado com as turmas de 2024.

O formulário aplicado ao final do curso para as Turmas 3 e 4 do segundo semestre de 2023, intitulado *Final Survey*, não era uma atividade avaliativa, por isso, nem todos os alunos o responderam. Obtivemos 22 respostas da Turma 3 e 19 da Turma 4, totalizando 41 respostas.

A imagem a seguir ilustra o questionário utilizado:

Figura 29 – *Final Survey*



Fonte: Captura de tela do formulário *Final Survey*.

Esse questionário já foi usado em 2020 na primeira oferta do curso, por isso já se constituía como material de avaliação do trabalho realizado para professores e para o próprio PLIU, que oferecia o curso. O formulário contém 6 itens, sendo 4 itens de avaliação objetiva e 2 questões dissertativas. São eles:

1- Qualidade das atividades propostas durante o curso;
2- Adequação das atividades para o nível básico;
3- Qualidade das aulas síncronas;
4- Qual é o seu nível de satisfação com o curso? (1- pouco satisfeito/ 5- muito satisfeito);
5- Justifique sua resposta dada na questão anterior;
6- Quais sugestões você daria para aprimorarmos este curso?

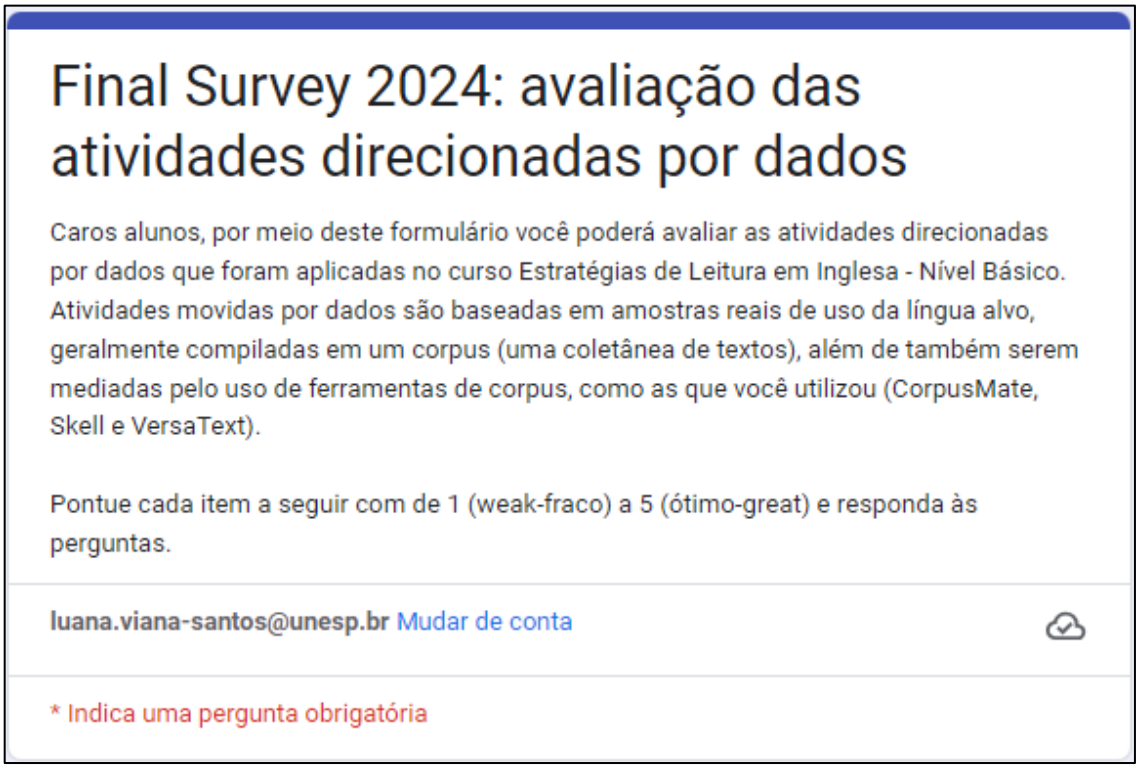
Para as questões de 1 a 4 os alunos deveriam pontuar o item analisado de 1, considerado *weak* ou fraco, a 5, considerado *great* ou ótimo. Essas avaliações dadas pelos alunos assim como as sugestões oferecidas serão analisadas na seção seguinte, no item 4.2.

Como o *Final Survey* não foi desenvolvido pela pesquisadora, julgamos necessário criar um novo questionário para as turmas da oferta seguinte de cursos no qual os estudantes pudessem avaliar as ferramentas de *corpora* apresentadas e os temas dos textos lidos nas atividades. Assim, foram formuladas novas questões, um pouco mais específicas do que as do primeiro questionário para que pudéssemos colher dados mais alinhados com os objetivos deste trabalho.

O segundo formulário foi nomeado *Final Survey - 2024* e foi aplicado também ao final do curso, no primeiro semestre de 2024, pela pesquisadora, que entrou em contato por e-mail com os estudantes que haviam confirmado participação via TCLE. Infelizmente, apesar da alta adesão de participação dos alunos na pesquisa (70 alunos confirmaram o aceite por meio do TCLE), o número de respondentes para o segundo questionário foi baixo, com apenas 27 respostas.

Na imagem a seguir consta uma captura de tela do instrumento de coleta de dados:

Figura 30 – *Final Survey* - 2024



Fonte: Captura de tela do formulário *Final Survey* - 2024.

Como é possível notar na imagem, há um breve comentário introdutório explicando a função do questionário para retomar informações importantes que os alunos podem ter esquecido com o término do curso. A seguir, apresentamos as questões contidas neste segundo questionário:

1- Como você avalia o curso? (1 a 5);
2- Como você avalia as atividades direcionadas por dados? (1 a 5);
3- Justifique sua resposta dada na questão anterior;
4- Como você avalia as ferramentas apresentadas? (*CorpusMate*, *Skell* e *VersaText*). (1 a 5);
5- Qual das ferramentas apresentadas você mais gostou de utilizar? (Pense na eficácia e na facilidade de uso da ferramenta para o estudo da língua) (Opções dadas: *CorpusMate* – *Skell* – *VesaText* – Nenhuma);
6- Para você, o grau de complexidade de uso das ferramentas foi: (Opções dadas: Fácil – Moderado – Difícil);

7- Como você avalia os temas trazidos nos textos das atividades? (Considere a relevância dos assuntos abordados, seu nível de interesse pelos temas dos textos, a complexidade do vocabulário etc.) (1 a 5);
8- Você pretende usar as ferramentas apresentadas nas atividades nos seus estudos futuros? (Opções dadas: Sim – Não);
9- Na sua opinião, as atividades direcionadas por dados ajudaram na sua prática de leitura em inglês?;
10- Nas atividades, vocês utilizaram o *CorpusMate*, o *Skell* e o *VersaText* para colocar em prática diferentes estratégias de leitura, lendo resumos de artigos científicos que tratavam de assuntos como saúde dos oceanos, mudança climática e saúde e bem-estar humano. Indique quais são as ferramentas que, na sua opinião, auxiliaram melhor a prática das estratégias estudadas ao longo do curso.

Por meio das perguntas desse segundo formulário, mais alinhadas aos objetivos da pesquisa, esperávamos colher dados mais robustos com os participantes de 2024. Acreditamos que o contato via e-mail depois do fim das aulas possa ser uma das razões para o baixo número de respostas colhidas, mesmo com o aceite de 70 alunos.

Na seção seguinte, serão apresentados e discutidos os dados encontrados.

4 ANÁLISE DOS DADOS

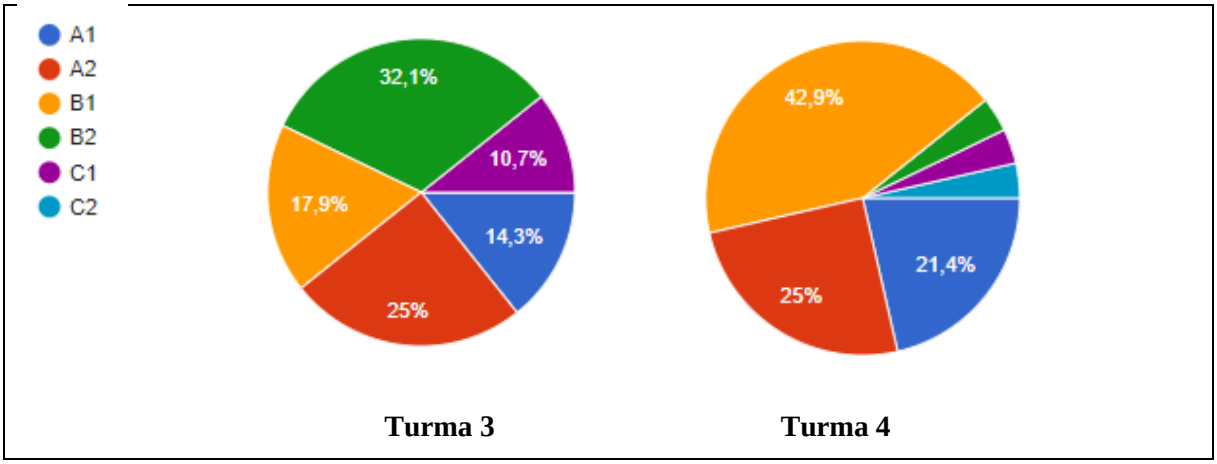
Nesta seção, são apresentados os dados do questionário de Análise de Necessidades de 2023, as análises dos itens lexicais do *corpus* utilizados na elaboração do material didático, os princípios e os critérios que os embasam, a percepção dos alunos a respeito da realização dos exercícios direcionados por dados e, por fim, as considerações que podem ser feitas sobre sua aplicabilidade.

4.1 Dados do questionário *Needs Analysis* aplicado em 2023

A partir do questionário de análise de necessidades, o formulário *Needs Analysis*, aplicado no início das aulas, foi possível colher informações importantes sobre o perfil de alunos que se matricularam no curso. Lembramos que esse tipo de mapeamento, fase necessária para qualquer curso de IFE, foi realizado tanto com as turmas de 2023 quanto com as de 2024, no entanto, apresentamos e exploramos os dados colhidos apenas das turmas 3 e 4 de 2023 pelo fato de termos tido acesso apenas ao questionário aplicado para esses grupos mencionados. Cada uma das turmas continha 28 alunos e todos responderam ao questionário, o que indica um total de 56 respostas.

Como a primeira pergunta do questionário solicita o nome dos estudantes, começaremos apresentando os dados obtidos pelas respostas dadas na pergunta 2, “Qual é o seu nível de proficiência em inglês? (Se você não tem certeza, indique o que você acredita que possa ser seu nível de proficiência)”. O quadro a seguir traz o percentual de nível de proficiência em que os alunos acreditavam estar:

Quadro 6 – Dados da pergunta 2 do *Needs Analysis* - Turmas 3 e 4 de 2023



Fonte: Questionário *Needs Analysis*.

Segundo as respostas dadas na questão 2, a maioria dos alunos da turma 3 acredita apresentar o nível B2 de proficiência, enquanto a maioria da turma 4 acredita apresentar o nível B1. Detalhando os dados, cerca de 18% da turma 3 indicou o nível B1 (5 alunos) e cerca de 32% (9 alunos) indicou no nível B2. Na turma 4, 43% (12 alunos) indicou o nível B1 e cerca de 4% (1 aluno) indicou o nível B2, ambos níveis intermediários de proficiência conforme QCER. Não é possível afirmar que esses sejam de fato o nível de proficiência dos estudantes, uma vez que não foi aplicado nenhum teste de nivelamento para as turmas. A própria questão pede para que seja dado o nível que o respondente acredita ter.

O segundo nível de proficiência mais indicado pelos alunos nas respostas à pergunta 2 é o A2, 25% na turma 3 (7 alunos) e, coincidentemente, também 25% (7 alunos) na turma 4. O nível A1 foi o terceiro nível mais indicado, com cerca de 14% dos estudantes da turma 3 (4 alunos) e cerca de 21% na turma 4 (6 alunos). Os níveis avançados foram os menos indicados: 3 alunos da turma 3 marcaram o C1, mas nenhum marcou o C2; na turma 4, 1 aluno marcou o C1 e 1 aluno marcou o C2.

De modo geral, somando-se os dados das turmas, entre os níveis básicos (A1 e A2), foram identificadas 24 respostas, entre os intermediários (B1 e B2) observou-se 27 indicações, e entre os níveis avançados (C1 e C2), apenas 5 respostas. Os dados demostram que cerca de metade dos estudantes das duas turmas podem se enquadrar como falantes intermediários e metade como falantes iniciantes. Isso pode ser justificado pelo fato de alguns alunos fora do nível alvo do curso terem se inscrito pela oportunidade de contato com a língua e com um curso para fins específicos, como mencionado anteriormente neste trabalho.

O Estratégias de Leitura em Língua Inglesa foi primeiramente oferecido para os cursos de graduação da Unesp em razão do percentual de proficiência em LI apresentado pela pesquisa do IsF, conforme discutido na introdução e na fundamentação teórica deste estudo. Os resultados positivos demonstrados pela oferta de 2020 e a demanda por um curso voltado especificamente para a leitura em inglês foram alguns dos fatores que levaram a PROPG a incluir o curso como uma das disciplinas instrumentais dos cursos de pós-graduação.

Apesar de termos colhido informações de apenas duas turmas de 2023, os dados possibilitaram uma generalização dos níveis de proficiência dos pós-graduandos matriculados naquele momento: ainda que se perceba uma maior quantidade de alunos intermediários, principalmente do nível B1, pouco menos da metade dos alunos das turmas 3 e 4 juntas

considerou-se de nível básico, mais especificamente o A2. Não se pode atestar os níveis informados pela não necessidade de testes de proficiência como requisito para ingresso no curso, porém, é possível afirmar que aprendizes de nível básico integram consideravelmente os cursos de pós-graduação da universidade e sentem a necessidade de aprender o inglês com propósitos acadêmicos.

Na pergunta 3, "Fale um pouco sobre sua atuação na pós-graduação (curso, programa, atividades desenvolvidas, leituras específicas, contato com estrangeiros, etc)" os alunos puderam falar sobre sua área de pesquisa de forma mais detalhada, pois essa era uma questão dissertativa. Alguns alunos mencionaram os campi onde estudam, os cursos de pós-graduação (mestrado ou doutorado), os programas de pós-graduação do qual fazem parte e até mesmo o tempo de experiência que possuem com seus temas de pesquisa. Por ser uma pergunta aberta que deu margem tanto para respostas curtas como para respostas extensas, organizamos um quadro[36] com a relação de cursos, campi e cidade dos estudantes:

Quadro 7 – Dados acadêmico-profissionais dos alunos da Turma 3

(continua)

Programa de Pós-Graduação e Campus	**Curso**		**Quantidade de Alunos**
	Mestrado	**Doutorado**	
Agricultura – Unesp (Botucatu)	-	-	1
Animais Selvagens – Unesp (Botucatu)	3		3
Biociências – Unesp (S. J. Rio Preto)	1		1
Biotecnologia Animal – Unesp (Botucatu)	1		1
Biotecnologia – Unesp (Botucatu)		1	1
Ciência Florestal – Unesp (Botucatu)	2		2
Ciências Sociais - Unesp (Marília)		1	1

[36] Onde se lê "-" na coluna "Curso", entende-se que o aluno informou o PPG do qual faz parte, mas não indicou se cursa Mestrado ou Doutorado. Já onde se lê "(?)", na coluna "Programa de Pós-Graduação e Campus", entende-se que o estudante não indicou em sua resposta o campus da Unesp em que faz o curso de Pós-Graduação.

(conclusão)

Ciências Veterinárias – Unesp (Jaboticabal)		1	1
Clínica Veterinária – Unesp (Bauru)	1		1
Comunicação – Unesp (Bauru)		1	1
Engenharia Civil – Unesp (?)	1		1
Engenharia de Bioprocessos e Biotecnologia – Unesp (?)	1		1
Fonoaudiologia – Unesp (Marília)		2	2
Matemática – Unesp (Bauru)		1	1
Odontologia – Unesp (Araraquara)		1	1
Odontologia – Unesp (Araçatuba)	1		1
Oncologia Veterinária – Unesp (Jaboticabal)		1	1
Reprodução Animal – Unesp (?)		1	1
Total de respostas colhidas		22	

Fonte: Questionário *Needs Analysis*.

Na questão 3, 6 alunos falaram sobre suas pesquisas e sobre suas experiências no campo acadêmico sem mencionar o PPG ou o curso do qual fazem parte, por isso, foi preciso desconsiderar suas respostas para a elaboração do quadro, o que totalizou 22 respostas. Como se pode observar, a turma 3 é bastante heterogênea quando se pensa no campo de atuação dos estudantes, com pós-graduandos de cursos e campi variados. Isso demonstra o alcance da disciplina Estratégias de Leitura em Língua Inglesa – Nível Básico no estado e a demanda por um curso como esse.

A turma 4 também era constituída por alunos vindos dos mais diversos PPGs e campi, como se pode observar no quadro a seguir:

Quadro 8 - Dados acadêmico-profissionais dos alunos da Turma 4

(continua)

Programa de Pós-Graduação e Campus	**Curso**		**Quantidade de Alunos**
	Mestrado	**Doutorado**	
Agricultura – Unesp (Botucatu)	1		1
Anestesiologia – Unesp (Botucatu)		3	3
Biociências e Biotecnologia Aplicadas à Farmácia – Unesp (?)	1		1
Biodiversidade de Ambientes Costeiros – Unesp (São Vicente)	2		2
Biometria – Unesp (Botucatu)	-	-	1
Ciências Biológicas – Unesp (Botucatu)	1		1
Ciências Sociais – Unesp (Araraquara)		1	1
Ciência e Tecnologia de Materiais – Unesp (Presidente Prudente)		1	1
Desenvolvimento Territorial na América Latina e Caribe – Unesp (São Paulo)	1		1
Ecologia, Evolução e Biodiversidade – Unesp (Rio Claro)	-	-	1
Educação para a Ciência – Unesp (Bauru)		1	1
Educação para a Ciência – Unesp (?)	-	-	1
Engenharia de Bioprocessos e Biomateriais – Unesp (?)	1		1
Engenharia de Produção – Unesp (Bauru)	-	-	1

(conclusão)

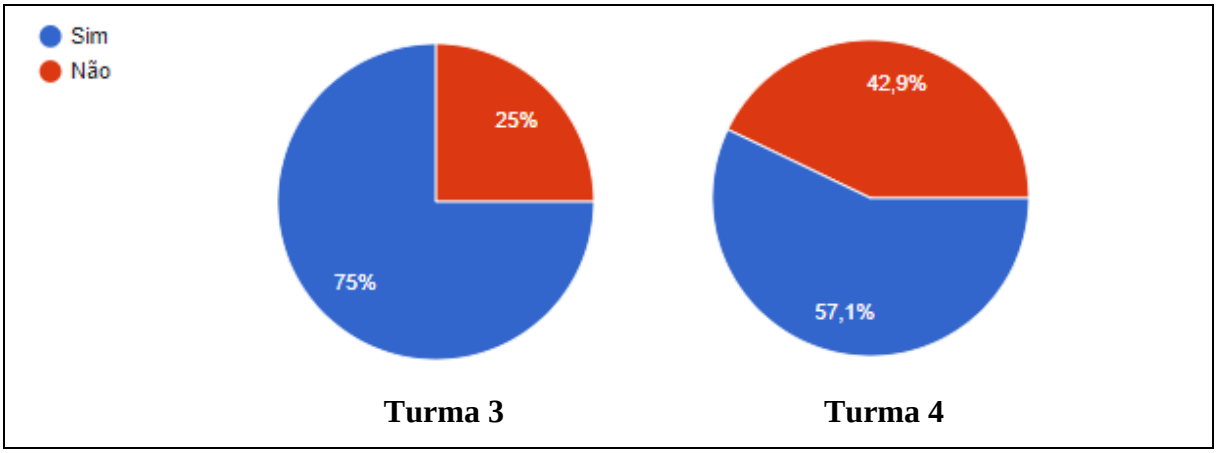
Fonoaudiologia – Unesp (Marília)		2	2
Gestão e Regulação de Recursos Hídricos – Unesp (São Paulo)	1		1
História – Unesp (Franca)		1	1
Planejamento e Análise de Políticas Públicas – Unesp (Franca)	-	-	1
Saúde Animal – Unesp (?)	-	-	1
Saúde Coletiva em Odontologia – Unesp (Araraquara)	1		1
Total de respostas colhidas		24	

Fonte: Questionário *Needs Analysis*.

Na turma 4, apenas 4 estudantes deixaram de mencionar seus campi e PPGs, então, desconsideramos suas respostas. As 24 respostas consideradas mostram a diversidade das áreas e dos campi dos quais os alunos fazem parte. Muitos discorreram sobre sua atuação acadêmica e em sua maioria relataram a necessidade de comunicação em LI em suas áreas de pesquisa, principalmente da habilidade de compreensão escrita no idioma para a leitura de artigos científicos.

A questão que aborda a necessidade do inglês é a 4, “Já teve que se comunicar em inglês com um estrangeiro?”, para a qual, nas duas turmas, a maioria dos alunos respondeu que sim:

Quadro 9 - Dados da pergunta 4 do *Needs Analysis* - Turmas 3 e 4 de 2023



Fonte: Questionário *Needs Analysis*.

Como é possível observar, 75% dos estudantes da turma 3 e cerca de 57% dos da turma 4 já tiveram que se comunicar com um estrangeiro em inglês. A pergunta 4.1, "4.1- Em caso de resposta afirmativa na questão anterior, qual foi sua maior dificuldade?", aprofunda as informações colhidas por meio da questão 4 ao pedir para que o aluno especifique aquilo que ele mais sente falta no momento da comunicação na língua alvo. Entre os obstáculos citados pelos estudantes na questão 4.1, o que apareceu mais vezes tanto nas respostas da turma 3 quanto nas da turma 4 foi a habilidade de fala, justificada na maioria das respostas, em primeiro lugar, pela falta de vocabulário específico e, em segundo lugar, pela dificuldade de formar frases.

A falta de vocabulário, citada pelos alunos, pode, sim, ser considerada um empecilho para a comunicação oral e pode ser estendida para a habilidade de leitura, já que o repertório vocabular é um dos elementos decisivos para o estabelecimento da compreensão de um texto (Baldo, 2009). Tanto a lacuna de vocabulário quanto a dificuldade de construção de frases são objetos de ensino alvo da abordagem DDL, que prima pelo desenvolvimento da léxico-gramática dos aprendizes.

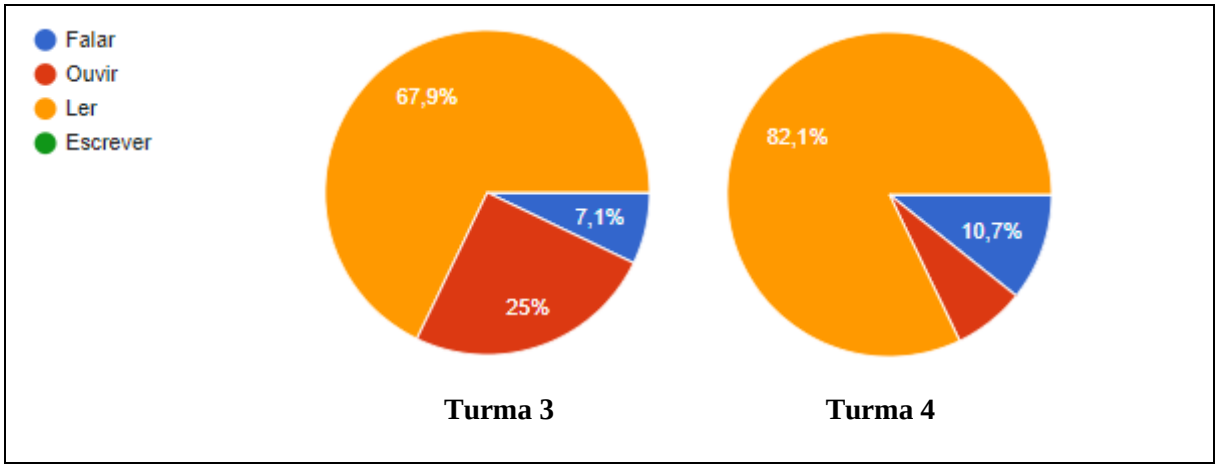
A pergunta 5, "Você já estudou inglês anteriormente? Por quanto tempo? Em qual contexto (aula particular, escola de idiomas, escola regular...)?", também era uma questão aberta. De modo geral, a maioria dos estudantes respondeu afirmativamente e citou a aprendizagem do inglês tanto em escolas de idiomas quanto em cursos oferecidos em projetos da universidade. Os poucos alunos que afirmaram nunca ter feito curso de inglês informaram que contavam com o que aprenderam somente na escola (Ensino Básico) e com estudos de forma autônoma/autodidata.

A pergunta 6, “Comente um pouco sobre a sua experiência de aprendizagem de inglês. Você gosta de estudar inglês? Sente que tem facilidade ou dificuldade com o idioma?”, tinha o objetivo de identificar a relação que os estudantes já tinham com a LI, um fator muito importante para o aprendizado, já que a falta de afinidade com o idioma e/ou possíveis bloqueios em vista da procura por um curso somente pela necessidade de aprender podem influenciar o rendimento da turma.

Na turma 3, de 28 respostas, identificamos 4 em que os respondentes afirmaram não gostar de estudar em inglês em razão de suas dificuldades relacionadas a falta de tempo para se dedicar ao estudo da língua e à inibição causada por inseguranças, como falta de vocabulário e vergonha ao falar. Na turma 4, foi possível identificar 5 respostas, entre 28, em que os alunos diziam não gostar de estudar inglês. Entre esses, a justificativa que se repetiu foi a insegurança relacionada à habilidade de fala.

Essa é uma característica comum entre aprendizes de inglês, que tendem a sentir mais dificuldade com o *speaking* do que com outras habilidades comunicativas. A maioria dos alunos demonstrou gostar de aprender inglês e compreender a importância do idioma para sua vida profissional e acadêmica. A tendência de sentir mais facilidade com a leitura e mais dificuldade com a fala pôde ser confirmada com os grupos analisados por meio de suas respostas às questões 7, “Em qual das habilidades abaixo você tem mais **facilidade** ao se comunicar em inglês?”, e 8, “Em qual das habilidades abaixo você tem mais **dificuldade** ao se comunicar em inglês?”, como se pode observar nos quadros a seguir:

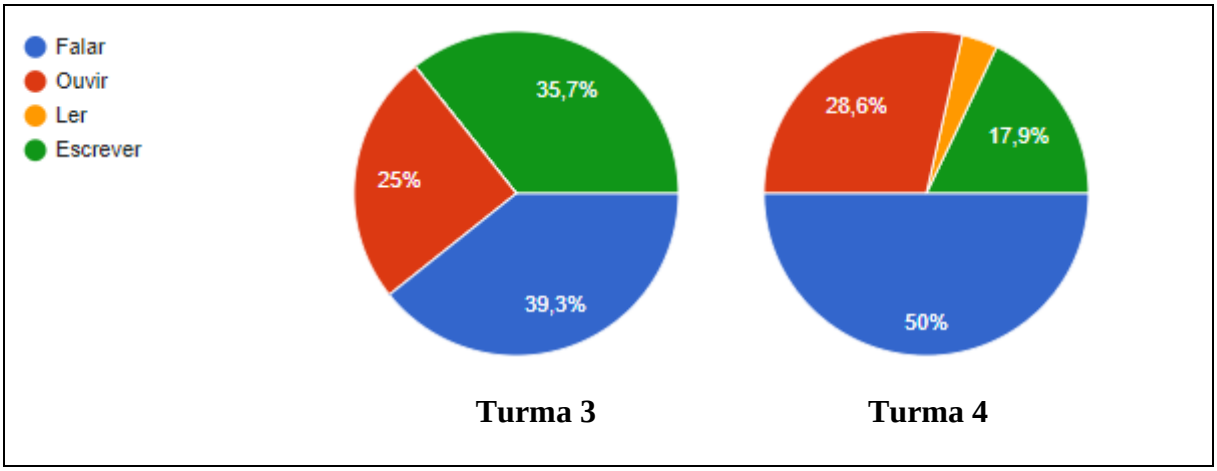
Quadro 10 - Dados da pergunta 7 do *Needs Analysis* - Turmas 3 e 4 de 2023



Fonte: Questionário *Needs Analysis*.

Cerca de 68% dos alunos da turma 3 e cerca de 82% dos alunos da turma 4 acreditam que a habilidade de leitura é a que têm mais facilidade. A segunda habilidade considerada mais fácil para os estudantes da turma 3 foi a de compreensão oral, com 25% de votos, e para os da turma 4 foi a de fala, com cerca de 10% de votos. Nenhum aluno de nenhuma das turmas indicou a escrita como habilidade de mais fácil aprendizagem. As habilidades consideradas por eles como mais difíceis são apresentadas a seguir:

Quadro 11 - Dados da pergunta 8 do *Needs Analysis* - Turmas 3 e 4 de 2023



Fonte: Questionário *Needs Analysis*.

Na questão 8, a habilidade considerada mais difícil nos estudos de inglês pela turma 3 foi a de fala, com aproximadamente 39% de votos, mas outra habilidade também foi indicada como a mais difícil de forma considerável, a escrita, com cerca de 35% de votos. Metade da turma 4 considerou a fala a habilidade com a qual tem mais dificuldade, seguida pela compreensão oral, com aproximadamente 28% de votos.

A facilidade com a leitura entre as demais habilidades linguísticas é consenso entre os estudantes. Apesar disso, não se pode dizer que a maioria dos alunos conhecem as estratégias que fazem do aprendiz de LE um leitor proficiente, por isso, importa saber os aspectos da língua e da comunicação que são vistos por eles como obstáculos e quais são suas expectativas com o curso.

Na questão 9, “Quais dos itens abaixo você acredita serem os maiores obstáculos para sua comunicação em inglês?”, entre os itens listados, os três mais indicados pelos estudantes da turma 3 foram: “Insegurança ao falar”, com 75% de votos, “Medo de cometer erros (na modalidade escrita e/ou falada)”, com cerca de 53% de votos, e “Falta de vocabulário”, com 50% de votos. Já na turma 4, os três fatores mais recorrentemente indicados pelos alunos foram:

“Dificuldade para falar”, com cerca de 68% de votos, “Insegurança ao falar”, também com aproximadamente 68% de votos, e “Medo de cometer erros (na modalidade escrita e/ou falada)”, com cerca de 60% de votos. Os dados demonstram o quanto a insegurança relacionada principalmente à produção oral influencia as crenças dos estudantes a respeito da aprendizagem de LI.

Entre as expectativas e os objetivos em relação ao curso, tanto os alunos da turma 3 quanto os da turma 4 apontaram, na questão 10, “O que você espera deste curso? Que objetivos você quer alcançar com o inglês?”, i. melhorar a leitura de textos acadêmicos em língua inglesa, ii. aprender estratégias de leitura voltadas para a compreensão de textos com vocabulário complexo e iii. praticar a comunicação em inglês durante as aulas. Foi recorrente, nas respostas das duas turmas, a menção à necessidade de estratégias de leitura para artigos científicos especificamente.

O questionário *Needs Analysis* permitiu a identificação do repertório com a LI e dos objetivos dos estudantes com o curso, de modo que foi possível alinhar as atividades elaboradas na pesquisa com suas necessidades.

Na seção seguinte, apresentamos as análises dos termos multipalavras do *corpus SDG-Plos* usados para o desenvolvimento das atividades.

4.2 As palavras-chave do *corpus SDG-Plos* entendidas como termos multipalavras dos ODS

Pela observação dos termos multipalavras contidos no Quadro 3, apresentado na seção anterior, nota-se que se tratam de expressões formadas pela combinação Adjetivo + Substantivo (como em *protected area*) e Substantivo + Substantivo (como em *ecosystem service*), ambas atuando como sintagmas adjetivais na língua inglesa, caracterizando-se, portanto, como *multiword lexical collocations*, mais especificamente, *multiword technical terms* (Biber, 2009), por serem constituídas apenas por palavras lexicais e por poderem ser compreendidas também como termos multipalavras.

Com base nos parâmetros de análise do *Sketch Engine* (2024), os termos multipalavras do *corpus SDG-Plos* podem assim ser definidos porque atendem a duas condições estabelecidas pelo *software* como parâmetro para identificação de termos: 1. Aparecem no *corpus* de estudo com mais frequência do que na linguagem geral ou no *corpus* de referência; 2. Possuem

estrutura permitida para termos na língua (definida em uma gramática de termos) (*Sketch Engine*, 2024).

A primeira condição elencada pelo *Sketch Engine*, a alta frequência da expressão no *corpus* de estudo quando comparada com sua frequência no *corpus* de referência pode ser atestada pela comparação da frequência relativa das palavras selecionadas no *SDG-Plos* com sua frequência relativa no *enTenTen21*, *corpus* de referência utilizado neste estudo.

O quadro a seguir traz a frequência relativa dos termos multipalavras selecionados no *SDG-Plos* e no *corpus* de referência:

Quadro 12 – Frequência relativa dos termos multipalavras selecionados

(continua)

Termo Multipalavra	**Frequência Relativa no *corpus* de estudo**	**Frequência Relativa no *corpus* de referência**
ecosystem service	176.12	1.08
maternal health	65.52	0.33
protected area	128.39	2.05
ocean health	39.77	0.05
waterbird habitat	34.84	<0.01
maternal mortality	51.89	0.46
human well-being	42.04	0.23
meat consumption	40.90	0.22
conservation action	37.50	0.16
food insecurity	68.93	1.3

(conclusão)

ghg emission	55.68	0.83
sustainable employability	30.30	<0.01
water service	48.10	0.58
health system	143.92	3.67

Fonte: *Sketch Engine* (Kilgarriff *et* al, 2014).

É possível dizer que as palavras-chave do *corpus* de estudo podem ser entendidas como termos multipalavras, conforme os critérios do *software Sketch Engine*, se considerarmos sua frequência relativa, ou seja, a proporção de ocorrência em relação ao total de *tokens* do *corpus*. Assumindo-se esse princípio, pode-se afirmar que *ecosystem service* é mais frequente no *corpus* de estudo do que no *corpus* de referência, pois sua ocorrência no *SDG-Plos* é de aproximadamente 0.018%, enquanto no *enTenTen21* é de 0.00011%. Ainda que a frequência absoluta (*hits*) seja maior no *corpus* de referência, deve-se considerar sua distribuição em relação à extensão do *corpus*.

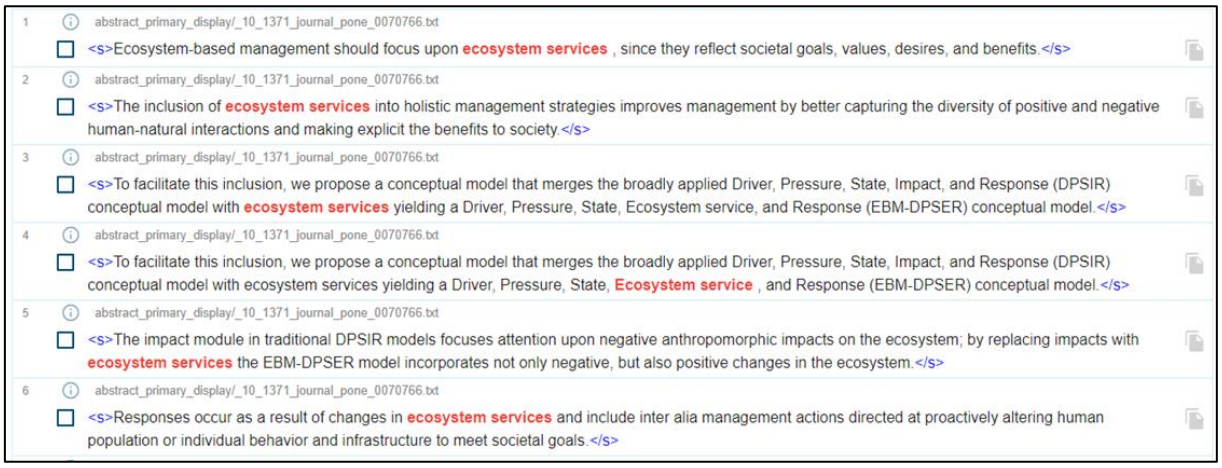
A segunda condição estabelecida pelo *Sketch Engine* versa sobre a estrutura de termo na língua, determinada por uma gramática de termo, que define uma estrutura permitida, ou seja, que de fato ocorre na língua analisada. Essa combinação aceita como termo na LI, como a estrutura adjetivo + adjetivo (opcional) + substantivo (*Sketch Engine*, 2024) já é pré-programa e aplicada automaticamente pela ferramenta *Keywords* quando gera as listas *Single Words* e *Multi-Word Terms*.

Ressaltamos que a presente pesquisa não ambiciona ser um trabalho terminográfico, portanto, utilizamos a nomenclatura e a fundamentação de base teórica do próprio *software* de análise de *corpus* empregado, com a intenção de selecionar os itens lexicais que se mostram relevantes para o propósito do estudo, a elaboração de atividades direcionadas por dados. Assim, ao selecionarmos os termos multipalavras do *SDG-Plos*, consideramos seu potencial representativo do *corpus* com o intuito de servirem como guia para a extração dos textos a serem levados para as aulas e usados no desenvolvimento das atividades.

A observação e a análise das linhas de concordância em que os termos multipalavras aparecem podem informar de que modo os pesquisadores-autores dos artigos estão

empregando-as. Na imagem seguinte, é possível ver as linhas de concordância com *ecosystem service*, por exemplo:

Figura 31 - Linhas de concordância com *ecosystem service* geradas pelo *Concordance*



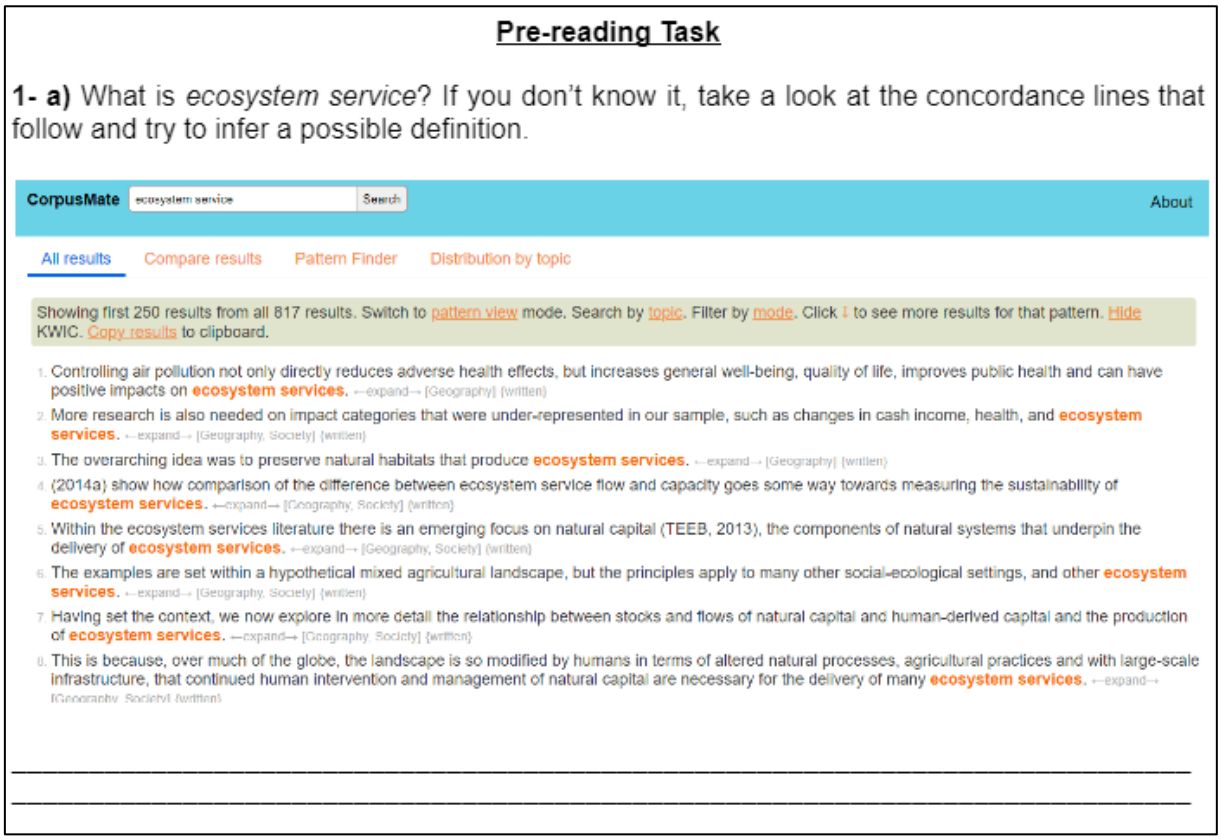
Fonte: *Sketch Engine* (Kilgarriff *et al*, 2014)

A leitura das linhas de concordância nos permite observar que o termo multipalavra *ecosystem service* tem sido empregado para tratar de serviços que auxiliam a preservação ambiental, já que "refletem os objetivos, os valores, os desejos e os benefícios sociais" (tradução nossa), como aponta a primeira linha de concordância. Essa é uma interpretação que se pode fazer com base nos dados fornecidos pelo *corpus*, o que compõe uma atividade de pesquisa possível de ser passada a alunos de áreas de especialidade focalizadas em cursos de IFA, por exemplo.

Muitos são os exercícios para ensino de vocabulário possíveis de serem aplicados a aprendizes de inglês em contexto específico e acadêmico. É possível, por exemplo, pedir para que os alunos observem os termos multipalavras selecionados do *corpus* e busquem inferir um possível significado com base em seus conhecimentos enquanto membros de determinada comunidade científica e com base nas informações encontradas nas linhas de concordância do *corpus* utilizado.

Na imagem a seguir, ilustramos como o termo multipalavra *ecosystem service* foi empregado em uma das atividades elaboradas com auxílio do *CorpusMate* (Baisa; Crosthwaite, 2023) quando estávamos investigando as possibilidades de uso da ferramenta. Essa atividade não chegou a ser aplicada no curso devido à carga-horária, mas ela está presente na seção Apêndices.

Figura 32 – Exercício 1 da atividade elaborada com o termo multipalavra *ecosystem service*



Fonte: Captura de tela da atividade.

O exercício ilustrado na imagem compõe uma sequência de atividades elaboradas com base no artigo do *corpus* escolhido a partir do levantamento do texto em que mais ocorria o termo multipalavra *ecosystem service*. Na pergunta 1, o aluno deve observar as linhas de concordância geradas pelo *CorpusMate* para tentar inferir um possível significado para a expressão dada. Esse é um exercício de *pre-reading*, em que se faz o levantamento de conhecimento prévio dos leitores.

Com base na observação dos dados fornecidos pelo *corpus* de estudo, foi possível descrever as características dos termos multipalavra dos ODS, de acordo com a definição do *Sketch Engine* (2024). A análise desenvolvida permitiu a identificação dos termos multipalavras do *corpus* como *multiword technical terms*, segundo a concepção proposta por Biber (2009), expressões fixas formadas pela combinação de duas palavras lexicais (ou de conteúdo).

Na seção a seguir, analisamos os fundamentos teóricos da LC e da DDL que fundamentam as aplicações pedagógicas dos dados do *corpus* na elaboração de materiais didáticos.

4.3 Os princípios e os critérios que embasam a elaboração das atividades

As atividades baseiam-se nos princípios de Mark e O'Keeffe (2022), pois foram elaboradas a partir de combinações de palavras, entendidas e analisadas enquanto termos multiplavras, padrões relevantes para este estudo por permitirem a criação de atividades movidas por dados centradas no texto. Por meio dos termos multipalavras do *corpus* de estudo, os textos com temas dos ODS foram selecionados.

As ferramentas *CorpusMate*, *Skell* e *VersaText* foram associadas ao ensino das estratégias de leitura quando empregadas na seção de *pre-reading* para mobilizar o reconhecimento de cognatos e falsos cognatos, a classe morfológica das palavras do texto lido, assim como para direcionar o processo de inferência dos leitores. O nível de proficiência do público-alvo foi respeitado por meio do uso de ferramentas de *corpora* próprias para o ensino e por meio da estruturação das tarefas a serem realizadas em etapas com instruções e direcionamentos da tutora.

O material desenvolvido também está fortemente ancorado nos critérios obrigatórios para elaboração de exercícios baseados em *corpora* de Delfino (2016) porque:

- o critério 1, que exige o uso de *corpus* foi seguido, uma vez que a análise da lista de *Multi-Word Terms* do *corpus* de estudo guiou a extração dos textos;
- o critério 2, que diz respeito à necessidade de enunciados claros foi respeitado através da curadoria dos materiais e ferramentas utilizados nos exercícios, de sua divisão em três etapas (*pre-reading*, *reading* e *post-reading*) e da inclusão de questões em inglês aos poucos, no desenrolar das atividades;
- baseia-se no padrão léxico-gramatical, pois foi criado a partir dos termos multipalavras do *SDG-Plos*, portanto respeita o critério 3;
- respeita a integridade do aluno e foi pensado para o seu nível de dificuldade, portanto é ético, conforme orienta o critério 4;

- segue o critério 5, pois é replicável e adaptável por outros professores por utilizar ferramentas de *corpora* para o ensino consideradas abertas, ou seja, gratuitas e de fácil acesso;
- é motivador ao se considerar que as ferramentas de *corpora* apresentadas aos alunos são utilizadas como instrumentos para a construção de seu aprendizado por meio da prática das estratégias de leitura, portanto, respeita o critério 6;
- apresenta material autêntico, logo, respeita o critério 7, que aborda a importância de não simplificar os textos ou linhas de concordância utilizados;
- foi criado com o uso de ferramentas de *corpora* próprias para o ensino de línguas e com foco nos conteúdos de estratégias de leitura para os níveis A1 e A2 do currículo do curso em que foi aplicado, então segue o que diz o critério 8, o respeito do nível de proficiência dos aprendizes;
- foi elaborado a partir das necessidades e especificidades do alunado em foco, conforme exige o critério 9, pensado para aprendizes iniciantes em contexto acadêmico;
- coloca o professor como mediador do uso das ferramentas e dos conteúdos apresentados, pois são dadas as instruções para que os estudantes investiguem o texto de forma autônoma e saiam da posição de receptáculo de conhecimento, como se pede no critério 10;
- e, portanto, coloca o aluno na posição de "descobridor" da língua ao se oferecer ferramentas de *corpora* nos exercícios, conforme consta no critério 11.

Tendo explicitado como as atividades se valem do modelo de aquisição, complexidade e formulaicidade de Mark e O'Keeffe (2022) e dos critérios de Delfino (2016), partimos para a análise da avaliação dada pelos alunos ao material desenvolvido.

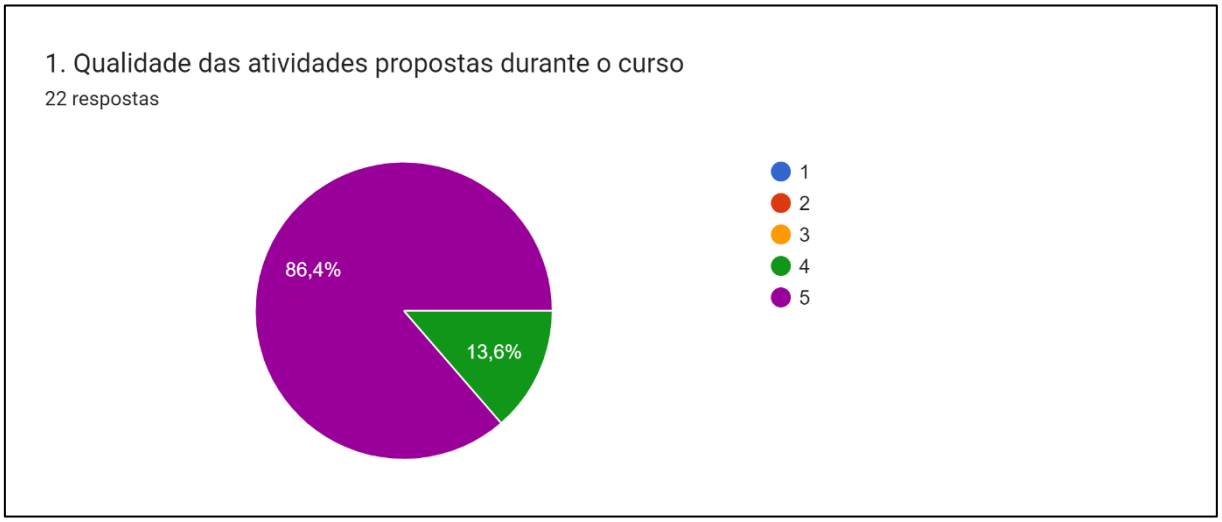
4.4 A percepção dos alunos acerca das atividades

Iremos analisar a percepção dos alunos com base em suas respostas dadas no *Final Survey* e no *Final Survey - 2024*, questionários aplicados ao final do curso no segundo semestre de 2023 e no primeiro semestre de 2024, respectivamente. Serão analisadas primeiramente as respostas dadas pelos alunos participantes da pesquisa em 2023, turmas 3 e 4, e em seguida as respostas dadas pelos participantes das quatro turmas em 2024.

4.4.1 Dados do *Final Survey*: Turma 3

Para a primeira questão, “Qualidade das atividades propostas durante o curso”, foram marcadas 19 respostas “5” (*great*/ótimo) e 2 respostas “4”, um nível abaixo de “*great*” ou “ótimo”.

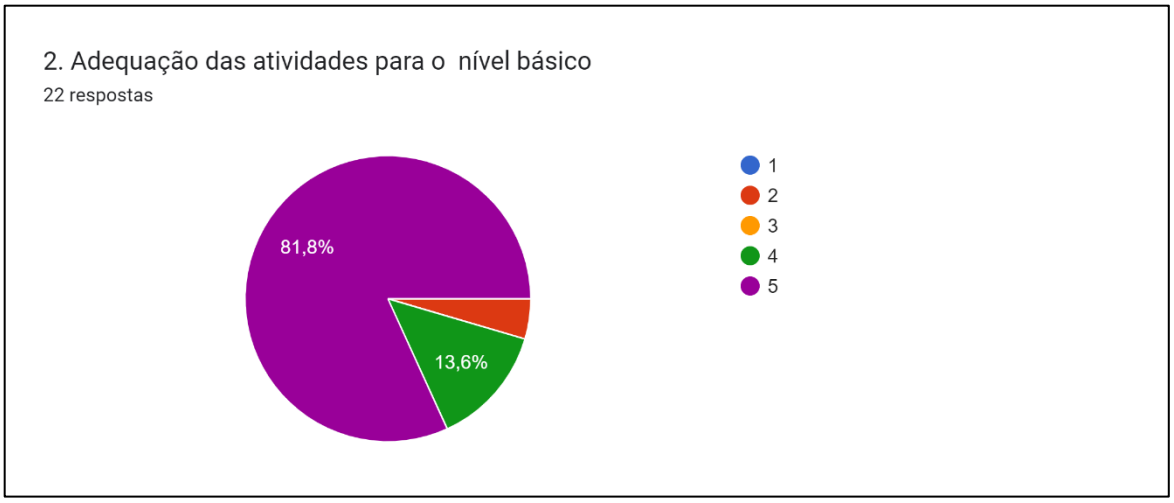
Gráfico 1 – Respostas dadas na questão 1 do Final Survey (Turma 3)



Fonte: *Final Survey* aplicado no final do curso.

Para a questão 2, “Adequação das atividades para o nível básico”, 18 alunos marcaram “5”, 3 marcaram “4” e 1 marcou “3”:

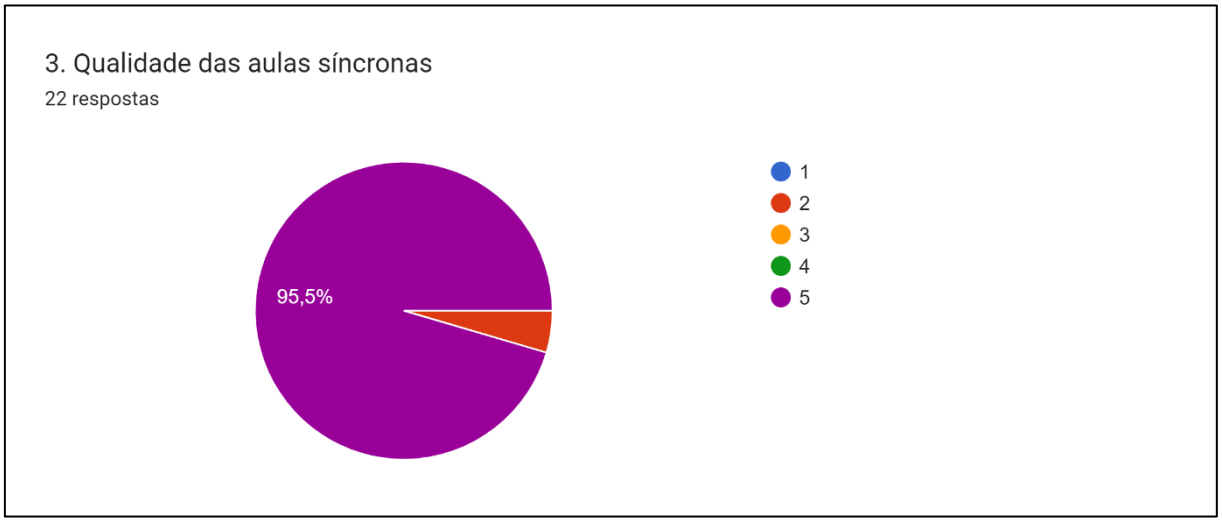
Gráfico 2 – Respostas dadas na questão 2 do *Final Survey* (Turma 3)



Fonte: *Final Survey* aplicado no final do curso.

Na questão 3, “Qualidade das aulas síncronas”, 21 alunos avaliaram com “5” e 1 aluno avaliou com “2”:

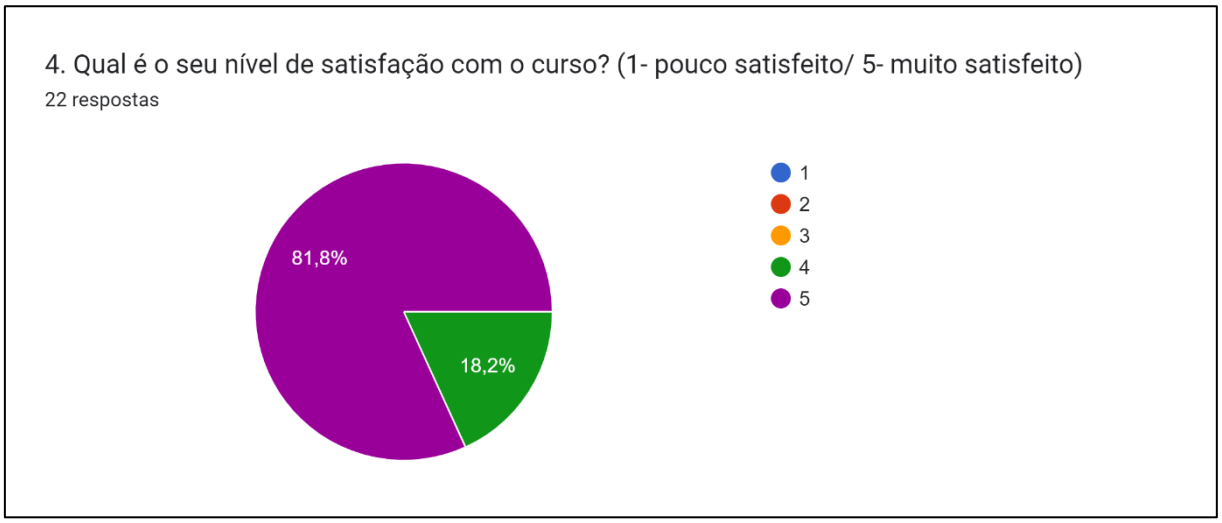
Gráfico 3 – Respostas dadas na questão 3 do *Final Survey* (Turma 3)



Fonte: *Final Survey* aplicado no final do curso.

Para a questão 4, “Qual é o seu nível de satisfação com o curso? (1- pouco satisfeito/ 5- muito satisfeito)”, 18 alunos marcaram “5” e 4 alunos marcaram “4”:

Gráfico 4 – Respostas dadas na questão 4 do *Final Survey* (Turma 3)



Fonte: *Final Survey* aplicado no final do curso.

Na questão 5, “Justifique a resposta dada na questão anterior.”, das 22 respostas, podemos considerar quase todas positivas. Para resguardar o sigilo da identidade dos participantes, apresentamos alguns exemplos de respostas dadas identificando os respondentes com a utilização de letras maiúsculas:

Aluno A: “O curso superou todas as expectativas, professoras atenciosas e solícitas, material de apoio e atividades complexas, mas acessível a todos os níveis.”

Aluno B: “Desde o planejamento do curso, a elaboração de um livro, a disponibilidade de sites e diversas ferramentas, a organização das aulas com tempo, interação e disponibilização dos materiais e slides, as aulas em si, com dinâmicas bem adequadas para os conteúdos de cada uma. Eu entendo que foi muito adequada toda organização de ensino para a proposta do curso. Cumpriu toda ementa satisfatoriamente. As professoras são excelentes.”

Aluno C: “Gostei da maneira como os temas foram trazidos e principalmente o uso de diferentes ferramentas.”

Aluno D: “Muito satisfeita com o curso, me ajudou a praticar a leitura de textos de forma mais simples utilizando os recursos ensinados.”

Uma das respostas diz mais sobre o aluno do que sobre o curso em si e outra informa que a pessoa esperava mais leituras científicas no curso:

Aluno E: “Gostei bastante do curso mas esperava maior leitura científica.”

Aluno F: “Infelizmente, não consegui acompanhar 100% as atividades presenciais.”

As interpretações que podem ser feitas a respeito da expressão “leitura científica” são amplas, visto que o aluno pode estar se referindo à leitura de artigos científicos, mas também de outros gêneros acadêmicos, ou ainda de gêneros acadêmicos de forma geral, mas com temas relacionados à sua área de pesquisa e atuação. A inclusão de gêneros acadêmicos nas leituras previstas no conteúdo programático do curso é coerente, já que dessa vez foi oferecido a alunos de pós-graduação. Contudo, ressaltamos que é necessária uma curadoria detalhada para que

sejam respeitados os limites que um aluno de nível básico pode apresentar e para que o vocabulário e a léxico-gramática do texto não sejam empecilhos para a compreensão e para a aprendizagem, como afirmam Mark e O'Keeffe (2022).

A justificativa dada pelo aluno F parece conter ou um erro de escrita ou um problema de interpretação da pergunta, pois não houve aula presencial, todo o curso foi ministrado de modo 100% on-line. Como apenas um aluno avaliou com "1" a questão 3, "Qualidade das aulas síncronas", podemos interpretar que esta possivelmente seria a justificativa dada pela avaliação mais baixa.

Com relação à pergunta 6, "Quais sugestões você daria para aprimorarmos este curso?", os alunos mencionaram o oferecimento de mais horários de aulas, o aumento da carga horária do curso, a inclusão de mais atividades durante as aulas síncronas e mais leituras, inclusive de base teórica a respeito de CE, o uso de ferramentas interativas para as aulas síncronas, e o oferecimento do curso para os níveis intermediário e avançado, o que pode ser entendido como uma demanda dos alunos que se consideravam de outros níveis além do A1 e do A2.

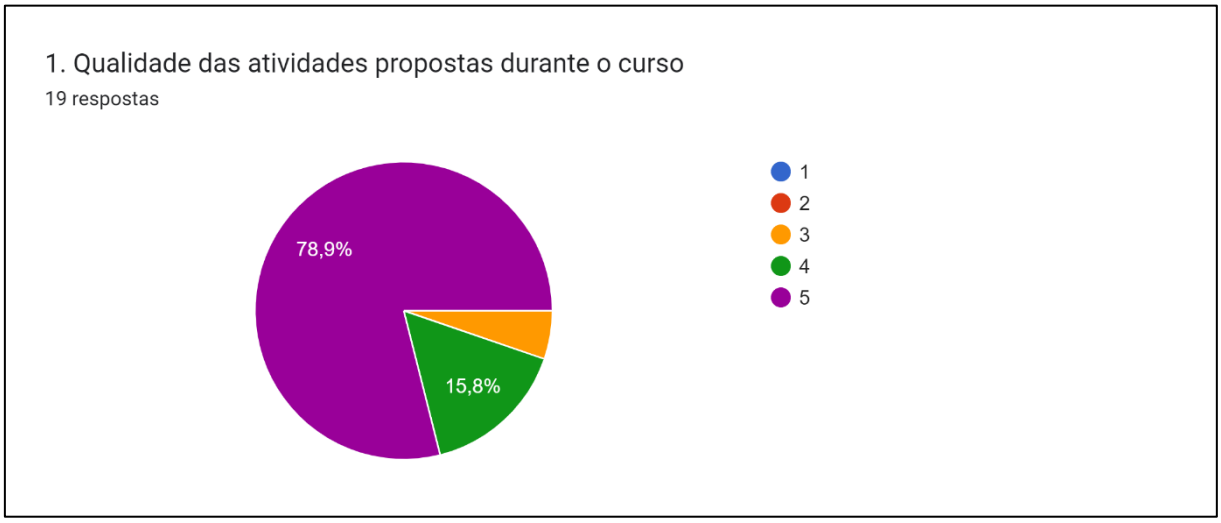
As sugestões dadas podem servir como uma análise de necessidade para o planejamento de uma terceira edição do curso. Elas também evidenciam o quanto os alunos se mantiveram engajados, pois foram capazes de apontar melhorias em relação ao tempo de duração, aos materiais e à dinâmica das aulas.

A seguir, apresentamos os dados da turma 4 e as análises que podem ser feitas com base em suas respostas.

4.4.2 Dados do *Final Survey*: Turma 4

Para a pergunta 1, "Qualidade das atividades propostas durante o curso", na turma 4 tivemos 15 respostas com "5", 3 respostas com "4" e 1 resposta com "3":

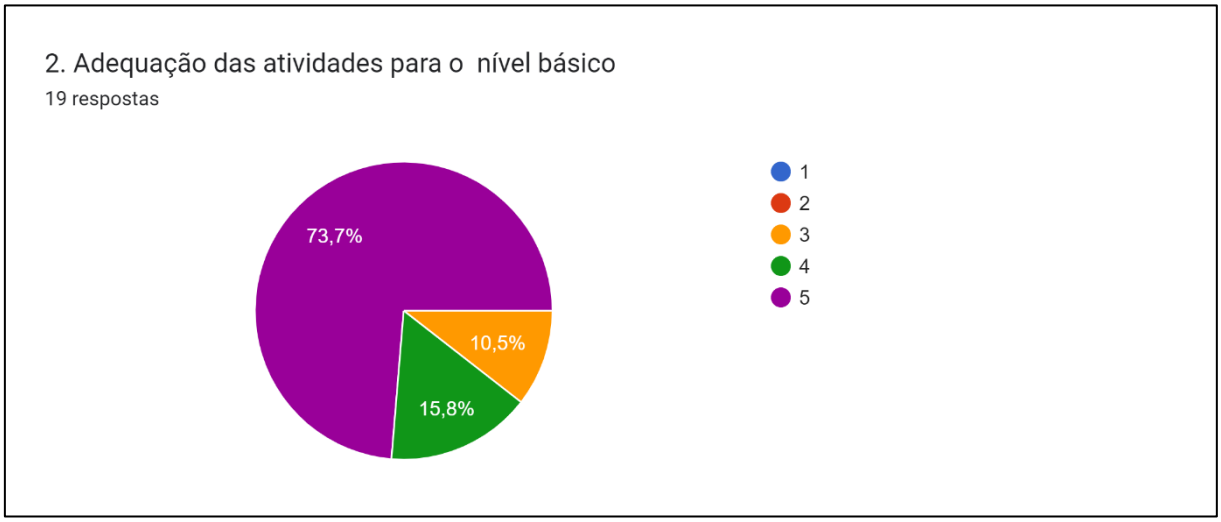
Gráfico 5 – Respostas dadas na questão 1 do *Final Survey* (Turma 4)



Fonte: *Final Survey* aplicado no final do curso.

Na pergunta 2, “Adequação das atividades para o nível básico”, 14 alunos responderam “5”, 3 alunos responder “4” e 2 responderam “3”:

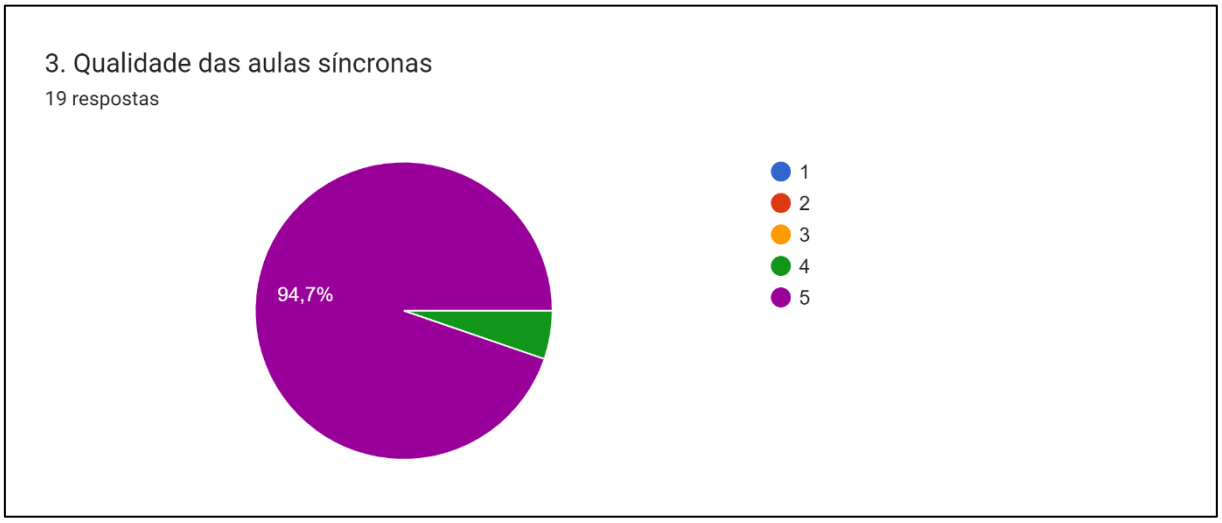
Gráfico 6 – Respostas dadas na questão 2 do *Final Survey* (Turma 4)



Fonte: *Final Survey* aplicado no final do curso.

A pergunta 3, “Qualidade das aulas síncronas”, obteve 18 respostas marcadas com “5” e 1 marcada com “4”:

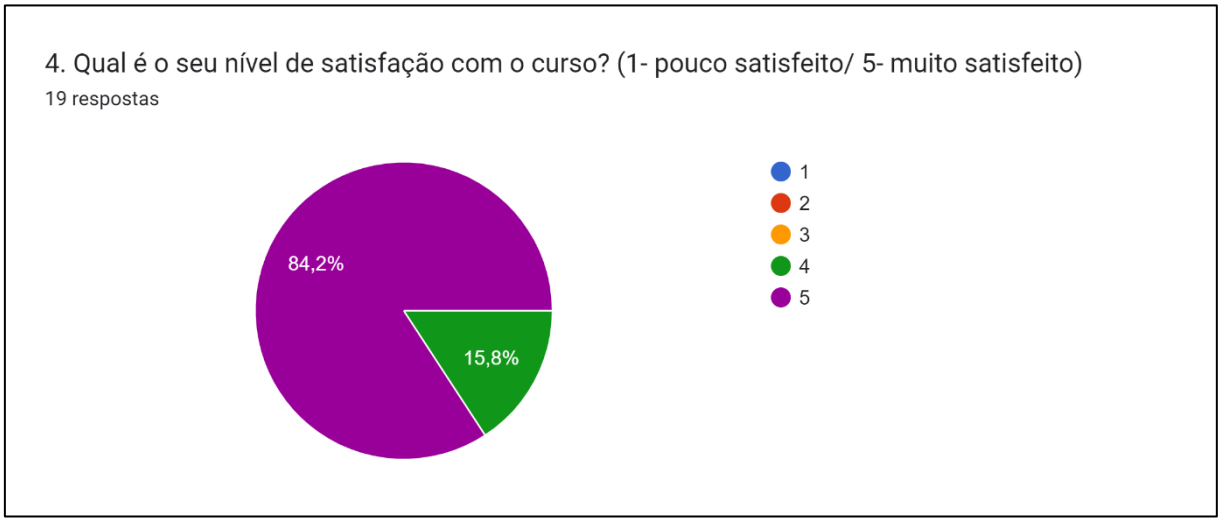
Gráfico 7 – Respostas dadas na questão 3 do *Final Survey* (Turma 4)



Fonte: *Final Survey* aplicado no final do curso.

Na pergunta 4, "Qual é o seu nível de satisfação com o curso? (1- pouco satisfeito/ 5- muito satisfeito)", obtivemos 16 respostas "5" e 3 respostas "4":

Gráfico 8 – Respostas dadas na questão 4 do *Final Survey* (Turma 4)



Fonte: *Final Survey* aplicado no final do curso.

A pergunta 5, "Justifique a resposta dada na questão anterior.", obteve respostas que remetem tanto aos materiais e a dinâmica do curso quanto a avaliação passada ao final da disciplina. De modo geral, as respostas foram majoritariamente positivas, assim como as da Turma 3. Apresentaremos alguns exemplos de respostas dadas pelos estudantes e, novamente,

para resguardar o sigilo de sua identidade, utilizaremos letras, agora duplicadas, para denominá-los:

Aluno AA: “As primeiras aulas foram maravilhosas e consegui acompanhar muito bem, no entanto, no decorrer do curso, acredito que pelo meu nível na língua inglesa, comecei a ter muita dificuldade para compreender as aulas e realizar os trabalhos propostos. Mas no geral, tenho somente agradecimentos às professoras [...]”

Aluno BB: “A partir das aulas e até mesmo das próprias atividades, vejo a prova em nível muito acima do exigido durante todo o curso. Avaliação desproporcional, me fazendo pensar que se gostariam que os alunos fosse capazes de resolver a prova apenas com o que foi ensinado nas aulas, faltou conteúdo e treino dos mesmos.”

Aluno CC: “Estou muito satisfeito com o curso em diversos aspectos. Primeiramente, a qualidade das atividades propostas ao longo do programa de estudos tem sido excepcional. Elas são desafiadoras, envolventes e proporcionam uma aprendizagem significativa. Além disso, a adequação das atividades para o nível básico de conhecimento tem sido uma vantagem significativa. Isso me permitiu acompanhar o curso de forma eficaz e construir uma base sólida de compreensão. As aulas síncronas também merecem destaque, a interação direta com os colegas, enriqueceu minha experiência de aprendizado, tornando o processo mais dinâmico e participativo.”

Aluno DD: “Estou muito satisfeita com o curso, para mim agregou demais. Estou iniciando um curso de inglês e ainda sou nível básico, assim o curso me ajudou muito a aprimorar as técnicas de leitura, principalmente através de ferramentas diferentes que auxiliam no entendimento do significado de palavras sem ter que usar tradutor.”

Aluno EE: “Curso é didático e mantém uma qualidade extraordinária como base de estratégias de leitura em língua inglesa tão importante para os alunos de pós-graduação.”

As justificativas dadas pelos alunos da turma 4 variaram bastante, pois encontramos desde elogios ao uso das ferramentas de *corpora*, assim como observamos na turma 3, a reclamação sobre uma possível disparidade entre o nível de dificuldade das atividades feitas

nas aulas síncronas e assíncronas e o da avaliação. Não analisamos a avaliação nesta pesquisa porque seu escopo restringe-se às atividades direcionadas por dados feitas a partir dos termos mutipalavras dos ODS, no entanto, validamos a resposta dada pelo aluno BB uma vez que entendemos que os exercícios feitos em aula devem ser compatíveis com aqueles contidos em um teste final.

Mesmo obtendo apenas um comentário que relaciona o nível de dificuldade da avaliação ao das atividades feitas, compreendemos que ele revela em algum grau a percepção dos alunos da turma 4 a respeito das atividades. Aliás, obtivemos justificativas que mencionaram a adequação dos materiais ao nível básico como um diferencial do curso, como relatou o aluno DD, e ao mesmo tempo comentários que revelaram certa dificuldade no acompanhamento do curso relacionada ao nível de proficiência, como demonstrou em sua resposta o aluno AA. Esta característica da turma 4 revela o quanto seus integrantes são mais heterogêneos em comparação com a turma 3.

As justificativas dadas pelos alunos na pergunta 5 foram positivas, com apenas um comentário que pode ser considerado uma crítica negativa, sem deixar de ser construtiva na medida em que aponta melhorias a serem feitas no curso. A questão 6, “Quais sugestões você daria para aprimorarmos este curso?”, também recebeu mensagens com elogios e sugestões que demonstram o nível de envolvimento dos alunos com o curso, entre elas o uso de artigos relacionados às áreas de pesquisa dos alunos, como apontou um estudante: “Como é um curso que abrange vários programas da pós, seria legal uma atividade onde pudéssemos trazer artigos de nossa área e apresentar sobre o que se trata (pelo menos o abstract) ou ao substituir alguns textos dados em atividades por artigos.”; tradução de alguns textos devido a dificuldades de compreensão; gravação das aulas síncronas; e aumento do tempo de duração do curso.

As sugestões dadas reforçam como a percepção dos alunos da turma 4 a respeito do curso e, embora possam ser consideradas amplamente positivas, variaram bastante. Percebe-se que enquanto boa parte dos alunos acharam o nível das atividades adequado, alguns ainda sentirem dificuldade para acompanhar o curso, reforçando mais uma vez a necessidade de uma mediação cuidadosa por parte dos professores.

Como o questionário *Final Survey* já era um instrumento do curso, não havia em sua estrutura perguntas que direcionassem o olhar dos alunos para a avaliação das atividades direcionadas por dados especificamente, por isso, entendemos que o formulário fornece informações sobre a percepção das atividades de forma geral e não com especificidade sobre o uso de cada uma das ferramentas, por exemplo. A avaliação dos alunos sobre o curso já pode

ser considerada um indicativo de que as atividades foram proveitosas por constituírem boa parte das aulas, mas as perguntas 1 e 2, voltadas à qualidade do material e à adequação ao nível básico podem ajudar a responder as perguntas de pesquisa norteadoras do estudo com mais detalhes.

Ao todo, obtivemos 41 respostas, o que significa que, somando os alunos das turmas 3 e 4 (55), apenas 14 alunos deixaram de responder ao questionário. Não se pode afirmar categoricamente que todos os alunos aprovaram as atividades, contudo, consideramos que em sua maioria, os estudantes demonstraram estar satisfeitos com os materiais empregados, pois mesmo sem contar com as respostas dos que não responderam ao formulário, do total de respostas dadas, nenhuma pode ser considerada negativa. O apontamento em relação ao nível de dificuldade da avaliação não fez com que o aluno BB considerasse que não se sentia satisfeito com o curso, visto que para esta questão (4) tivemos apenas respostas com avaliações altas, "4" e "5".

Poucos alunos revelaram dificuldade em acompanhar o curso em razão do nível de proficiência. Como os gráficos apontaram, cerca de 82% dos alunos da turma 3 e cerca de 74% dos alunos da turma 4 marcaram "5 (*great*/ótimo)" para esse quesito, logo, consideraram que as atividades estavam adequadas ao nível básico. As respostas dadas às questões dissertativas 5 e 6 reforçam o percentual de satisfação demonstrado nas respostas às perguntas objetivas.

Da mesma forma, podemos afirmar que os alunos consideraram as atividades de alta qualidade, como demonstrou a maioria das respostas dadas à pergunta 2: 86% dos alunos da turma 3 e cerca de 80% dos alunos da turma 4 avaliaram esse quesito com "5". As mensagens deixadas nas questões 5 e 6 revelam que muitos dos alunos se manisfestaram muito satisfeitos com o uso das ferramentas.

Em relação aos temas dos artigos, apenas um aluno fez menção, o aluno C da turma 3: "Gostei da maneira como os temas foram trazidos e principalmente o uso de diferentes ferramentas.". Associando esse dado aos outros já discutidos até agora, podemos considerar que o uso de um *corpus* com a temática dos ODS foi proveitoso, porém, essa é uma afirmação com base em uma generalização, pois para que pudéssemos colher a opinião dos alunos de forma mais específica, precisaríamos aplicar um novo questionário, planejado para a avaliação focada nas atividades e não no curso como um todo, como foi pensado o *Final Survey – 2024*, cujos resultados são apresentados na próxima seção.

4.4.3 Dados do *Final Survey* – 2024

O questionário *Final Survey* – 2024 foi criado com a intenção de colher informações a respeito da opinião dos alunos das novas turmas sobre a temática dos textos, os ODS, e sobre o uso das ferramentas de *corpora* empregadas nos exercícios de cada atividade, especificamente. Apesar de termos contado com a participação de mais estudantes em 2024 em comparação com 2023 (70), poucos alunos das novas turmas responderam ao novo formulário. Da turma 1-2024, colhemos 9 respostas, das turmas 2-2024, 3-2024 e 4-2024, colhemos 6 respostas em cada uma, o que totalizou 27 respostas.

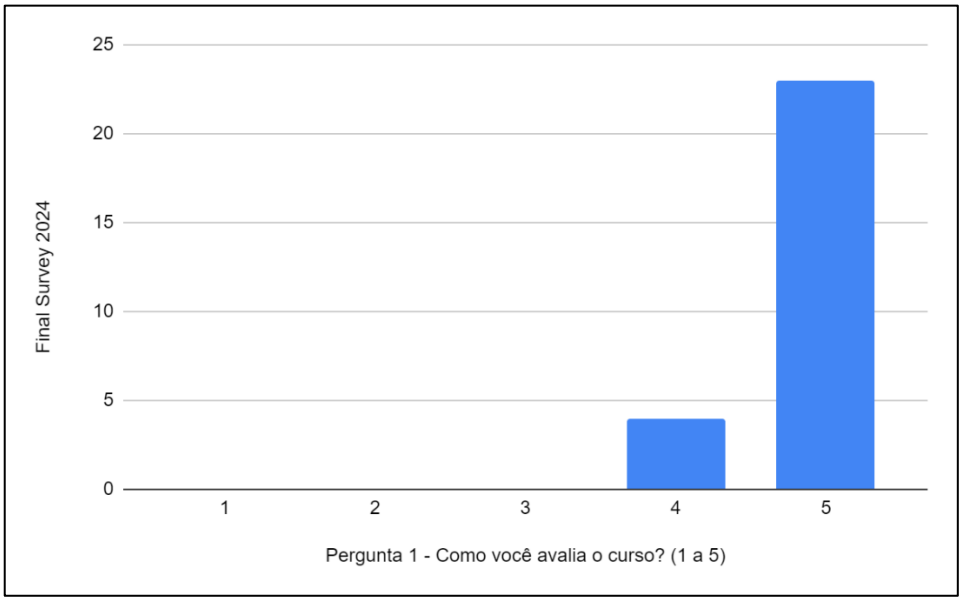
Acreditamos que o meio de contato feito com os participantes para divulgação do formulário tenha sido ineficiente – enviamos *e-mails* para aqueles que aceitaram participar da pesquisa. Em 2023, incluímos o questionário na seção de atividades do *Google Classroom*, ambiente em que todas os materiais e exercícios do curso eram postados. Entretanto, como as turmas eram semelhantes em certa medida, ou seja, compostas por pós-graduandos de áreas diferentes e de campi diferentes, ambos com o objetivo de desenvolver a leitura com propósitos acadêmicos, é possível considerar os resultados do *Final Survey* de 2023 como mostra prototípica.

A tentativa de colher informações de novas turmas nos interessa, principalmente, pelo distanciamento que há entre os estudantes e a pesquisadora, uma vez que esta não atuou como sua tutora, o que pode contribuir para uma avaliação mais imparcial das atividades desenvolvidas neste trabalho. Outro ponto importante que pode ser analisado a partir dos dados coletados no segundo questionário é o modo como as tutoras do ano de 2024 abordaram as ferramentas de *corpora* e as atividades movidas por dados, já que não tinham repertório de especialidade na DDL ou na LC e foram apresentadas às atividades elaboradas nesta pesquisa por meio de um vídeo explicativo.

Como obtivemos poucas respostas de cada uma das quatro turmas, optamos por juntar suas respostas ao questionário e analisá-las enquanto informações do grupo de 2024 como um todo. A seguir, debruçamo-nos sobre os dados colhidos pelo segundo formulário.

A questão 1, “Como você avalia o curso? (1 a 5)”, obteve os seguintes resultados:

Gráfico 9 – Respostas à pergunta 1 do *Final Survey* 2024

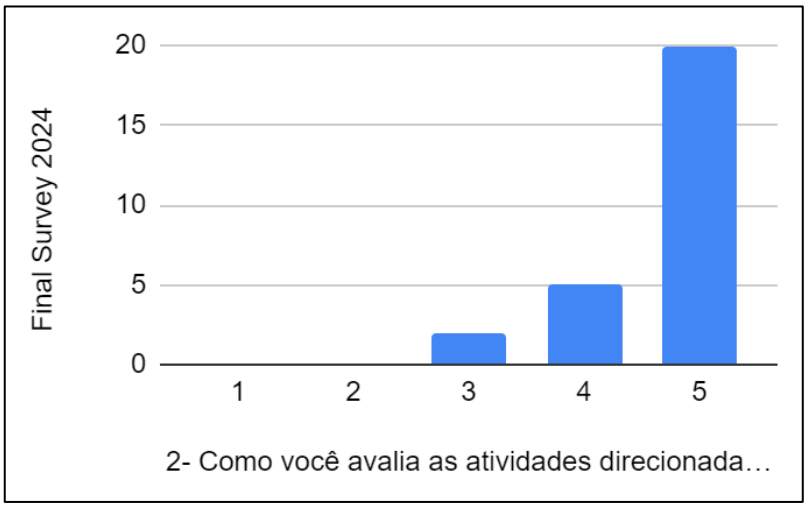


Fonte: *Final Survey* 2024.

Entre os alunos das turmas do primeiro semestre de 2024 que responderam ao questionário, 23 avaliaram o curso, na primeira pergunta, com o conceito máximo (5) e 4 avaliaram com conceito 4, ambos considerados boas avaliações.

Para a pergunta 2, “Como você avalia as atividades direcionadas por dados? (1 a 5)”, obtivemos as seguintes informações:

Gráfico 10 – Respostas à pergunta 2 do *Final Survey* 2024



Fonte: *Final Survey* 2024.

Na questão 2, em que deveriam avaliar as atividades direcionadas por dados, 20 alunos deram o conceito 5, 4 deram o conceito 4 e 2 deram o conceito 3. A maioria dos respondentes avaliou positivamente as atividades desenvolvidas na pesquisa. Na questão 3, em que se pedia uma justificativa para a resposta dada à pergunta 2, os alunos apontaram fatores como: i. a apresentação de ferramentas consideradas como "novidade" para a maioria dos estudantes; ii. a estruturação dos exercícios de modo a facilitar a compreensão dos textos. Apresentamos algumas respostas que ilustram tal constatação:

Aluno A-2024: "Foi muito interessante conhecer essas ferramentas, as quais eu nunca tinha ouvido falar e poder aprender sobre elas fornece muito mais robustez ao processo de aprendizagem."

Aluno B-2024: "As atividades foram de fácil compreensão, muito didáticas e interessantes."

Aluno C-2024: "Eram claras e seguiam uma lógica que facilitava a compreensão do texto"

Aluno D-2024: "Estou utilizando até na escrita do meu projeto para fazer a busca da sentença correta na base de dados."

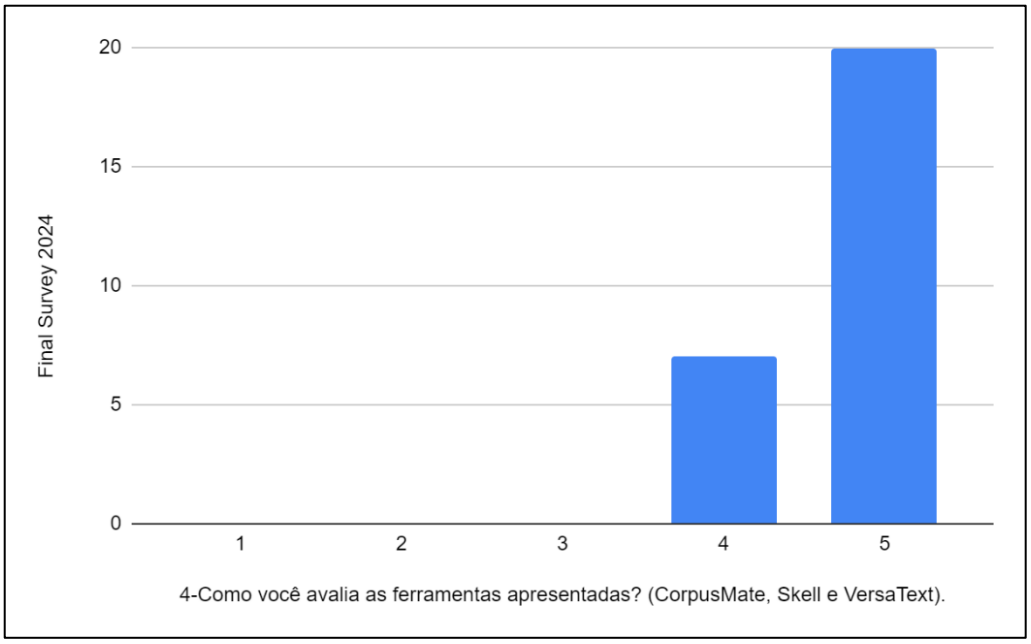
As justificativas dadas pelos alunos demonstram uma boa recepção das atividades, mas também podem refletir o modo como as tutoras as apresentaram a eles. A boa organização dos exercícios e a adequação das atividades para seu público-alvo é importante, mas o modo como os professores os utilizam também são, o que reforça o quão replicável as atividades podem ser consideradas e o quanto a DDL pode ser amplamente aproveitada por professores de línguas para além do escopo de pesquisadores em LC.

Entre respostas que possam ser consideradas negativas, observamos justificativas em que os estudantes mencionaram a extensão das atividades, muito longas na visão de um deles, e dificuldade de uso das ferramentas por parte de um deles. As atividades podem, sim, ser consideradas longas, pois contêm vários exercícios nas seções *pre-reading* e *post-reading*, algo que pode ser ajustado pelo professor a depender do tempo de que dispõe. Já a dificuldade de uso de ferramentas de *corpora* já é amplamente reconhecida como um dos desafios de implementação da DDL. Contudo, esse apontamento foi detectado em apenas uma das respostas

dadas, o que pode indicar que, de modo geral, o aproveitamento dos recursos de *corpus* foi positivo.

A pergunta 4, “Como você avalia as ferramentas apresentadas? (CorpusMate, Skell e VersaText)”, focava nas ferramentas de *corpora* e pedia para que os estudantes as avaliassem:

Gráfico 11 – Respostas à pergunta 4 do *Final Survey* 2024

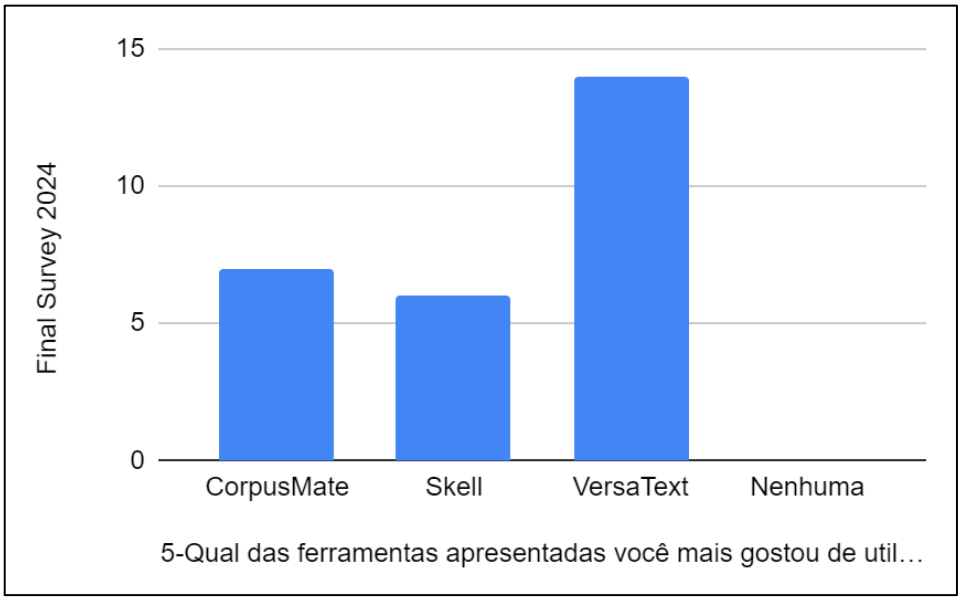


Fonte: *Final Survey* 2024.

Como se vê, as ferramentas foram muito bem avaliadas, recebendo 20 avaliações com conceito 5 e 7 com conceito 4. Isso pode ser relacionado a uma das dificuldades mencionadas por um dos estudantes na questão 3: apesar da dificuldade de uso, o aluno pode ter reconhecido valor nas ferramentas, o que fez com que não as avaliasse com baixos conceitos.

A pergunta 5, “Qual das ferramentas apresentadas você mais gostou de utilizar? (Pense na eficácia e na facilidade de uso da ferramenta para o estudo da língua)” revelou que a ferramenta mais bem recebida pelos alunos foi o *VersaText*, com 14 indicações. Em segundo lugar ficou o *CorpusMate*, com 7 indicações, e em terceiro, o *Skell*, com 6 indicações. É o que ilustra o gráfico seguinte:

Gráfico 12 – Respostas à pergunta 5 do *Final Survey* 2024



Fonte: *Final Survey* 2024.

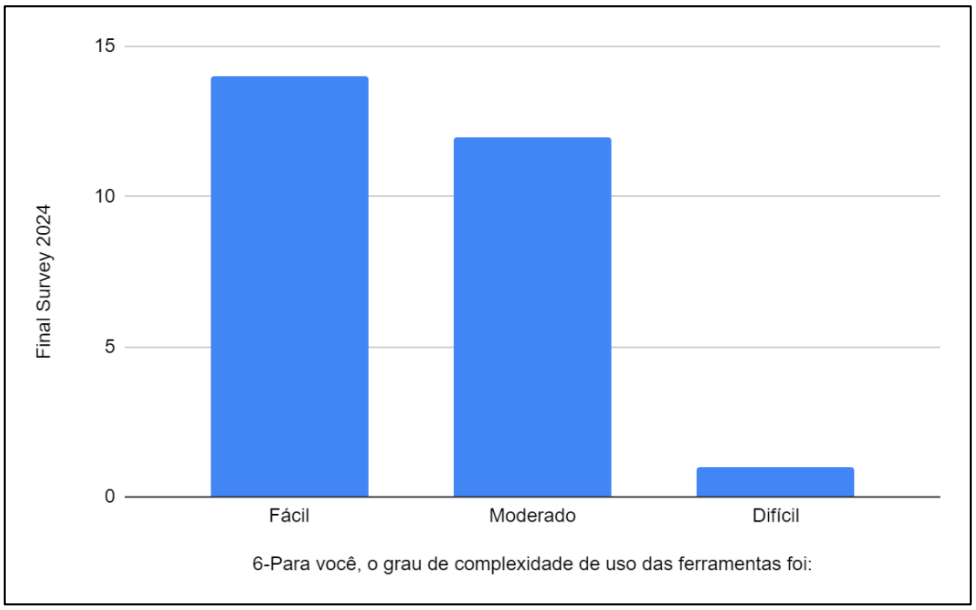
Todas as ferramentas foram empregadas, majoritariamente, nos exercícios de *pre-reading* com o objetivo de fazer com que os estudantes pesquisassem a respeito do vocabulário contido nos textos e os analisassem criticamente, seja por meio da observação de nuvem de palavras, da interpretação de expressões em linhas de concordância ou da busca por exemplos de expressões correlatas às investigadas em determinadas perguntas. Assim, buscou-se guiar os estudantes, por meio dos enunciados, para que utilizassem o *CorpusMate*, o *Skell* e o *VersaText* nos momentos de sondagem e de levantamento de conhecimento prévio para que se preparassem para a leitura dos textos.

Os exercícios de *pre-reading*, então, continham não somente perguntas nas quais os estudantes deveriam mobilizar seu *background* (conhecimento prévio) e realizar inferências (Brown, 2001; Khand, 2004; Leffa, 1996), mas também mobilizar estratégias cognitivas importantes para a leitura em LI por meio da exploração de dados reais da língua, como o reconhecimento de palavras-chave, de cognatos e falsos cognatos. Ficaram reservados, desse modo, os exercícios considerados de compreensão, em que são mobilizadas estratégias de *skimming* e de *scanning* e em que é feita a consolidação do conteúdo lido, como sugerem Khand (2004) e Tomitch (2009), para a seção de *post-reading*.

É possível afirmar que entre os alunos que responderam à questão, para a finalidade mencionada acima, a de acionar conhecimentos prévios e investigar a presença de palavras consideradas chave e cognatas, por exemplo, a maioria viu mais eficácia no *VersaText*.

A questão 6, “Para você, o grau de complexidade de uso das ferramentas foi: (Fácil; Moderado; Difícil)”, obteve os seguintes dados:

Gráfico 13 – Respostas à pergunta 6 do *Final Survey* 2024

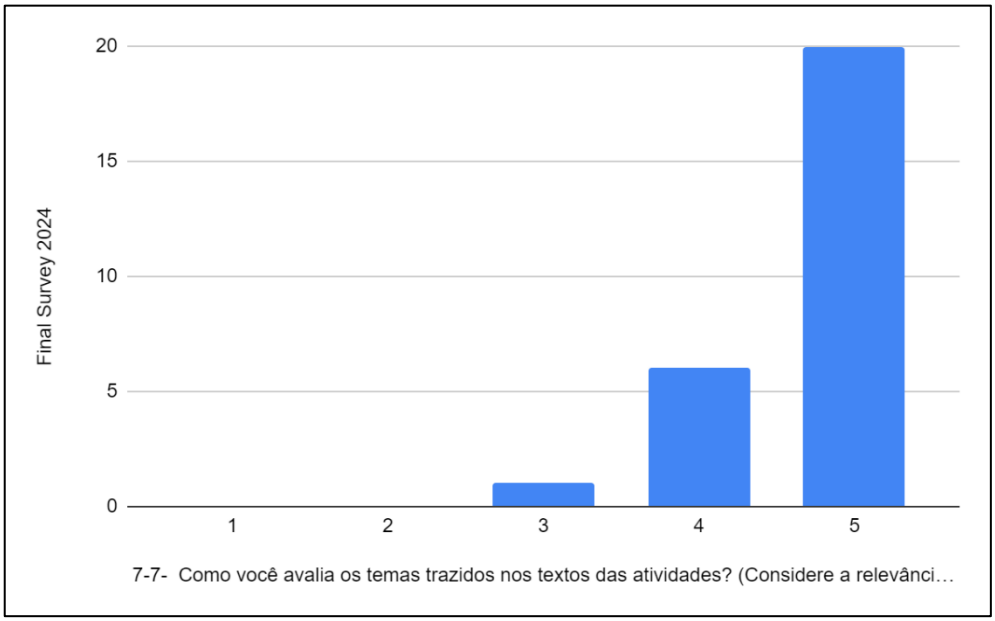


Fonte: *Final Survey* 2024.

Os dados refletem o nível de satisfação dos estudantes com as atividades, apontado nas questões anteriores, pois 14 indicaram grau de complexidade fácil para o uso das ferramentas, 12 apontaram grau moderado e apenas 1 apontou grau considerado difícil.

Com relação à temática dos textos, ou seja, temas relacionados aos ODS, os alunos responderam à questão 7, “Como você avalia os temas trazidos nos textos das atividades? (Considere a relevância dos assuntos abordados, seu nível de interesse pelos temas dos textos, a complexidade do vocabulário etc.)”, da seguinte forma:

Gráfico 14 – Respostas à pergunta 7 do *Final Survey* 2024



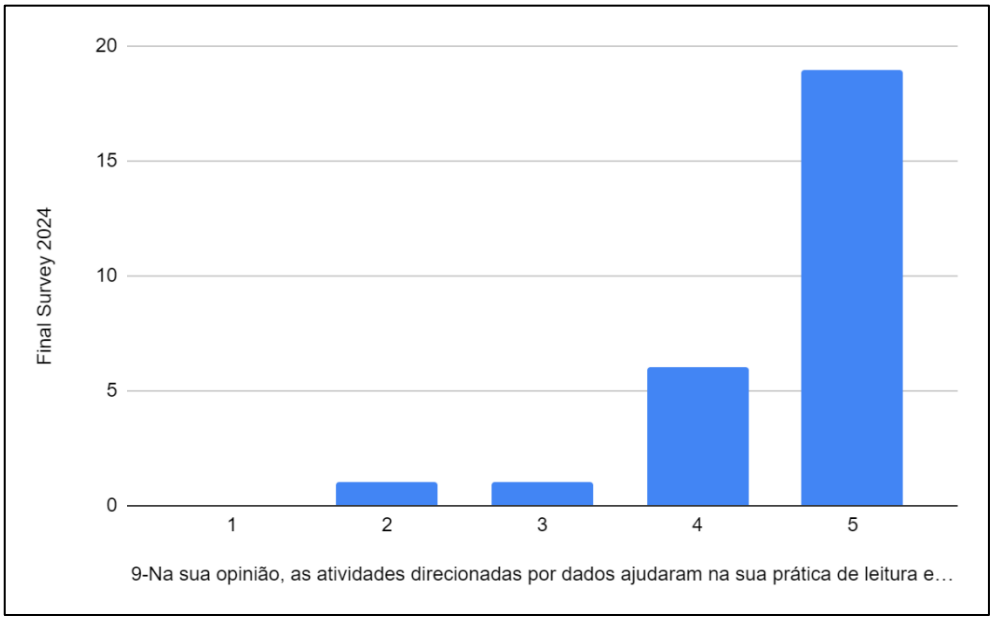
Fonte: *Final Survey* 2024.

A maioria dos alunos, 20, avaliou a temática dos textos com conceito 5, 6 alunos deram conceito 4 para esse quesito e apenas 1 aluno marcou conceito 3. O que se pode depreender desses dados é que, como apontam Bolster e Livrai (2017), os ODS cumprem sua função enquanto proposta de conteúdo multidisciplinar para cursos de IFA, pois, ainda que sejam poucas avaliações colhidas a esse respeito, nenhuma delas considerou a temática dos resumos de artigos lidos inadequada.

Para a questão 8, “Você pretende usar as ferramentas apresentadas nas atividades nos seus estudos futuros?”, 25 alunos responderam que sim e somente 2 alunos responderam que não. Como as respostas a essa pergunta não foram tão diversificadas quanto as anteriores, não julgamos necessário ilustrá-las em um gráfico. Não é possível saber se de fato os estudantes irão incorporar as ferramentas de *corpora* aos seus estudos de inglês, mas a maioria das respostas, afirmativas, podem demonstrar seu grau de interesse nos recursos apresentados.

A questão 9, “Na sua opinião, as atividades direcionadas por dados ajudaram na sua prática de leitura em inglês?”, obteve respostas também majoritariamente positivas:

Gráfico 15 – Respostas à pergunta 9 do *Final Survey* 2024

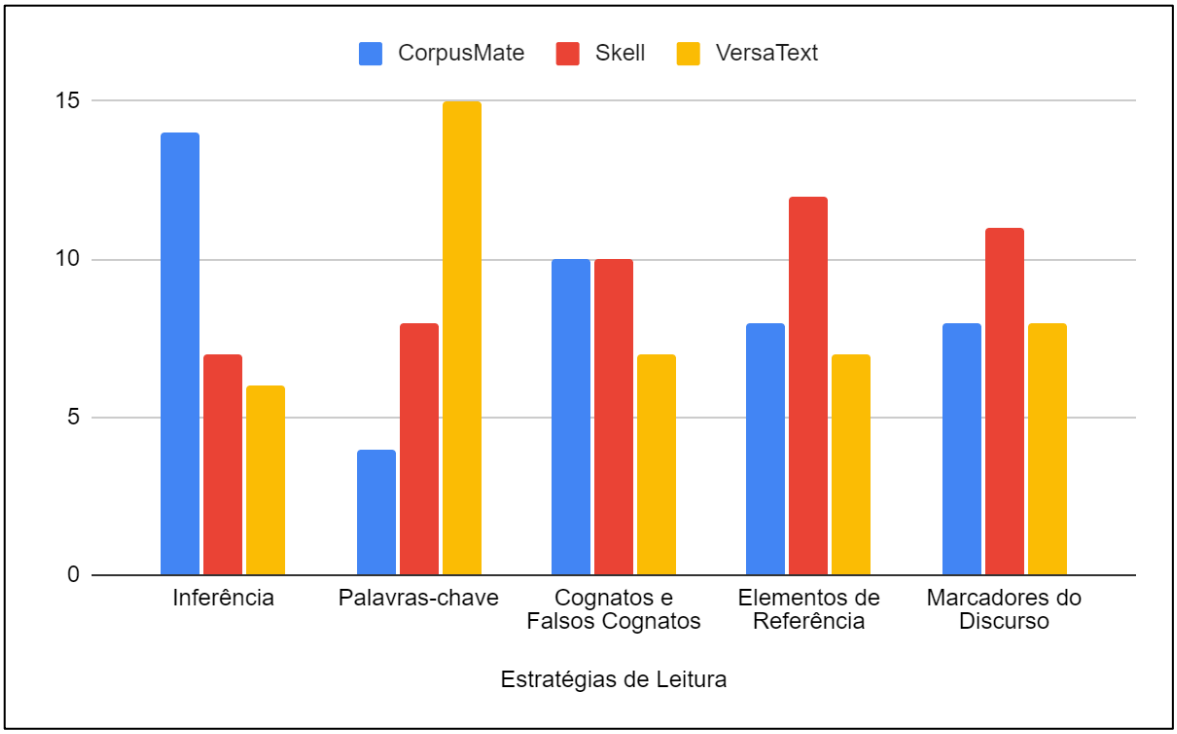


Fonte: *Final Survey* 2024.

Como é possível observar no gráfico, para o quesito relação eficaz entre as atividades movidas por dados e a prática de leitura, 19 alunos conferiram conceito 5, 6 alunos conferiram conceito 4, 1 aluno conferiu conceito 3 e 1 aluno conferiu conceito 2. A indicação de "3" e "2", consideradas mais baixas, podem ter relação com as respostas abertas em que identificamos 1 aluno que encontrou dificuldades no uso das ferramentas e 1 aluno que considerou as atividades muito longas.

Para saber se os estudantes consideraram o uso das ferramentas de modo compatível com o pensado no desenvolvimento das atividades, ou seja, se relacionaram o uso de cada uma às estratégias mobilizadas na seções *pre-reading*, recorremos aos dados obtidos na questão 10, "Nas atividades, vocês utilizaram o *CorpusMate*, o *Skell* e o *VersaText* para colocar em prática diferentes estratégias de leitura, lendo resumos de artigos científicos que tratavam de assuntos como saúde dos oceanos, mudança climática e saúde e bem-estar humano. Indique quais são as ferramentas que, na sua opinião, auxiliaram melhor a prática das estratégias estudadas ao longo do curso:". O gráfico a seguir mostra como os alunos relacionaram as estratégias às ferramentas:

Gráfico 16 – Respostas à pergunta 10 do *Final Survey* 2024



Fonte: *Final Survey* 2024.

As estratégias de leitura contidas no gráfico são aquelas mobilizadas na seção *pre-reading*. Para a estratégia de inferência, a ferramenta mais indicada pelos alunos foi o *CorpusMate* (14 votos), representada pela barra azul. Para a estratégia de reconhecimento de palavras-chaves, a ferramenta mais indicada foi o *VersaText* (15 votos), como se pode ver na barra em amarelo. Para a identificação de cognatos e falsos cognatos, a maioria dos estudantes indicou tanto o *CorpusMate* (10 votos) como o *Skell* (10 votos), na barra em vermelho. Para a estratégia de reconhecimento de elementos de referência, a ferramenta mais indicada foi o *Skell* (12 votos). Por fim, para a estratégia de identificação e análise dos marcadores do discurso, o *Skell* (11 votos) também foi a mais indicada.

Nos exercícios da atividade desenvolvida a partir do termo multipalavra *protected area*, as ferramentas utilizadas nos exercícios de *pre-reading* para o trabalho com as estratégias de inferência e reconhecimento de palavras-chave do texto foram o *CorpusMate* e o *VersaText*, assim como na atividade elaborada a partir do termo multipalavra *waterbird habitat*. Já na seção *pre-reading* da atividade feita a partir do termo multiplavra *ocean health*, para se trabalhar com a inferência e as palavras-chave foram usadas as ferramentas *Skell* e *VersaText*.

Na atividade elaborada a partir do termo multipalavra *human well-being*, para mobilizar a identificação de palavras-chave e de cognatos e o uso da inferência, foram empregados o *Skell*

e o *VersaText* na seção *pre-reading*. Nessa mesma atividade, utilizamos o *VersaText* em exercícios sobre marcadores do discurso e elementos de referência na seção *post-reading*. Os empregos das ferramentas nos exercícios foram descritos na seção de Materiais e Métodos e as atividades podem ser consultadas na íntegra nos apêndices. Para ilustrar de forma sucinta o emprego de cada ferramenta em cada uma das atividades, apresentamos o quadro a seguir:

Quadro 13 – Relação entre as ferramentas de *corpora* voltadas ao ensino e seu emprego nas atividades

Ferramentas de *corpora*: Emprego na seção *pre-reading*			
Estratégias de leitura abordadas no *pre-reading*	**Ferramentas**		
	CorpusMate	***Skell***	***VersaText***
Inferência	X	X	X
Palavras-chave			X
Cognatos e Falsos cognatos		X	X
Marcadores do discurso			X
Elementos de referência			X

Fonte: Elaboração própria.

É coerente a indicação feita pelos alunos do *CorpusMate* como ferramenta mais associada à inferência, já que esta ferramenta foi empregada com essa finalidade nos exercícios de duas das quatro atividades. Também é compreensível a indicação do *VersaText* como recurso mais associado ao reconhecimento de palavras-chave, pois em todas as atividades, nos exercícios em que se propunha esse fim, o *VersaText* foi a ferramenta mais utilizada.

O reconhecimento de cognatos foi associado às ferramentas *CorpusMate* e *Skell*, mas na maioria das atividades, o recurso usado com esse propósito foi o *Skell* juntamente com o *VersaText*, que também foi destinado aos exercícios com foco para marcadores do discurso e elementos de referência, como já foi mencionado. Dessa forma, é possível observar compatibilidades entre o uso proposto das ferramentas de *corpora* nas atividades e na percepção dos estudantes, mas também inconstâncias, que podem ser justificadas pelo uso que cada aluno fez de cada uma ao realizar as atividades. Isso porque, tendo sido apresentado a uma das ferramentas, o aprendiz pode utilizá-la nos momentos em que julgar necessário, sendo ela o instrumento foco de um exercício ou não.

Nesta seção, buscou-se analisar os dados obtidos a partir do questionário final aplicado aos alunos do curso oferecido em 2024. A seguir, tecemos algumas considerações que podem ser feitas a respeito da aplicabilidade das atividades.

4.5 A aplicabilidade das atividades

Podemos abordar o quesito "aplicabilidade" em pelo menos duas dimensões de análise: a da acessibilidade do emprego do *corpus* e das ferramentas de *corpora* voltadas ao ensino e a da eficácia. Ao falar da acessibilidade de uso de *corpus* e de ferramentas da LC, é necessário considerar o contexto em que estão inseridos os sujeitos envolvidos na prática de ensinagem, o professor e os alunos, assim como o local em que ela acontece, seja ele um espaço físico ou um ambiente virtual.

É preciso considerar que, no Brasil, o acesso à internet é um dos desafios relacionados à DDL (Rocha, 2017). Para que o ensino com base na abordagem aconteça, pelo menos quando se fala no emprego de atividades do tipo *hands on*, é necessário que tanto professor quanto aluno tenham acesso a tais recursos. No contexto desta pesquisa, esse quesito não representou um problema visto que o curso foi planejado para acontecer na modalidade on-line e as tutoras tinham acesso garantido à internet, assim como os alunos, que já estavam cientes de que para cursar a disciplina precisariam de um computador com uma boa conexão com a internet. Pouquíssimas vezes algum aluno perdia a conexão, geralmente quando chovia.

Com relação ao uso do *corpus SDG-Plos*, é importante ressaltar que ele foi criado por meio de um *software* aberto, o *AntCorGen*, mas foi compilado em um *software* pago, o *Sketch Engine*. Contudo, ainda que o *Sketch Engine* seja um instrumento de análise de *corpora* muito sofisticado, com ferramentas avançadas, ele não é o único *software* disponível para compilação. O *AntConc* 6 (Anthony, 2014) e o *WordSmith Tools* (Scott, 2016) são programas com versões abertas que oferecem todas as ferramentas utilizadas na pesquisa, como o *Keywords* e as linhas de concordância.

O *AntCorGen* possibilita a criação de *corpora* sobre diversas áreas e é possível encontrar muitos tutoriais de uso feitos por seu desenvolvedor na *web*. Os artigos sobre os ODS podem ser extraídos por meio desse *software* como também podem ser escolhidos por qualquer professor/pesquisador que deseje se debruçar sobre a temática. Já as ferramentas *CorpusMate*, *Skell* e *VersaText* são todas gratuitas e amigáveis, ou seja, de fácil acesso porque foram desenvolvidas especialmente para o ensino de LE.

A eficácia da DDL no ensino superior já é amplamente conhecida (Boulton; Cobb, 2017; Nesi, 2016). O que foi possível demonstrar neste estudo é que a DDL pode, sim, render bons resultados com alunos de nível básico, como observado na análise dos questionários feitos com os alunos, desde que sejam feitas as adaptações necessárias, conforme propuseram Mark e O'Keeffe (2022).

Neste trabalho, utilizamos atividades direcionadas por dados com foco para as estratégias de leitura para alunos de níveis A1 e A2, desse modo, foi necessária a investigação de quais itens lexicais poderiam servir para tal finalidade, os termos multipalavras. A criação das atividades levou tempo, tanto para a análise do *corpus* quanto para a elaboração dos exercícios em si, no entanto, conseguimos levar para alunos iniciantes os princípios da descoberta da língua e da autonomia no aprendizado, habilidades que podem ser fortalecidas a longo prazo, de um nível de proficiência para o outro.

5 CONSIDERAÇÕES FINAIS E ENCAMINHAMENTOS

A presente pesquisa apoiou-se na Linguística de *Corpus* e suas contribuições para o ensino de línguas, na abordagem *Data Driven Learning* e no Inglês Para Fins Acadêmicos para a elaboração de atividades de leitura direcionadas por dados com base em um *corpus* acadêmico de artigos sobre os Objetivos de Desenvolvimento Sustentável da ONU. Ao longo deste trabalho foram apresentadas a relevância do tema e o contexto de desenvolvimento da investigação, as bases teóricas nas quais se sustenta, os materiais e métodos empregados e as análises feitas a partir do que foi produzido.

Orientado pelo objetivo geral de elaborar atividades de leitura movidas por dados tematizadas pelos ODS e com foco para estudantes de nível básico, o desenvolvimento metodológico do trabalho estruturou-se em duas fases, a de extração dos dados do *corpus* de estudo e a de sua aplicação pedagógica.

Durante o percurso da pesquisa, os objetivos específicos, o levantamento de palavras-chave do *corpus*, a elaboração de atividades por meio das ferramentas *CorpusMate*, *Skell* e *VersaText*, com os temas dos ODS, e a avaliação de sua aplicabilidade, foram atingidos à medida em que o estudo avançava. Na fase inicial, foi priorizada a análise do *corpus* e do levantamento das palavras-chave para seu tratamento como termos multipalavras dos ODS, o que gerou o insumo para a produção das atividades.

Posteriormente, passamos para a investigação das ferramentas de *corpora* voltadas para o ensino e para a estruturação dos exercícios, que aconteceu juntamente com o planejamento das aulas do curso Estratégias de Leitura em Língua Inglesa – Nível Básico no segundo semestre de 2023. Durante o curso, as atividades foram aplicadas, assim como o questionário *Final Survey*, um dos instrumentos de coleta de dados utilizados.

Com a finalização do curso, foi possível avaliar as atividades feitas pelos alunos com mais atenção, bem como analisar as respostas dadas no questionário. A partir daí, deu-se o processo de análise da aplicabilidade do material e das ferramentas empregadas nas aulas e constatou-se uma das limitações do estudo, a falta de dados que servissem para a análise da temática do *corpus* e das ferramentas empregadas nos exercícios desenvolvidos.

Foi necessário, então, elaborar um novo questionário que contemplasse perguntas mais alinhadas aos propósitos da pesquisa e abranger os alunos das turmas do primeiro semestre de 2024 do curso como participantes para uma nova avaliação. Diante do trabalho realizado,

podemos retomar nossas perguntas de pesquisa e discuti-las a fim de pontuar o que foi possível desenvolver assim como o que se mostrou como obstáculos.

A primeira pergunta de pesquisa, "Quais itens lexicais podem servir para o desenvolvimento de atividades didáticas voltadas para a compreensão escrita?", nos conduziu à análise das palavras-chave enquanto termos multipalavras, ou *multiword technical terms* (Biber, 2009). Essas expressões fixas formadas por duas palavras lexicais foram propícias para o desenvolvimento de atividades de compreensão escrita porque, a partir delas, identificamos os textos que seriam levados para a leitura em aula. Pode-se dizer que os termos multipalavras caracterizam os artigos nos quais estão contidos.

Para Hyland (2008, p. 4), "Expressões multipalavras são um componente importante da produção linguística fluente e um fator chave no aprendizado de idiomas bem-sucedido"[37], assim, apresentar esses itens lexicais de modo estruturado e orientado a alunos em contexto acadêmico, como se faz no ensino de léxico-gramática orientado pela abordagem DDL, torna-os mais preparados para lidar com a linguagem comum àquela utilizada por seus pares.

Muitos outros tipos de exercícios além dos elaborados nesta pesquisa podem ser feitos explorando os termos multipalavras dos ODS. As atividades elaboradas podem se desdobrar em outras que exploram, a partir dos itens lexicais supracitados, não só as estratégias de leitura como outros conteúdos linguísticos tanto para a compreensão escrita como para as outras habilidades. Outra possibilidade é o desenvolvimento de atividades de leitura para os outros níveis de proficiência, já que existe o interesse por parte dos alunos de que o escopo do curso se estenda para além do A1 e do A2.

A segunda pergunta, "De que modo os ODS podem ser associados ao ensino de inglês em contexto acadêmico?" pode ser respondida com base no planejamento do curso e do resultado dos dados colhidos nos questionários. Os ODS se fizeram presentes nas atividades por meio dos resumos dos artigos e nos padrões lexicais explorados, ou seja, estavam no tema das atividades de forma implícita. Além disso, o *corpus* de artigos científicos contemplou a expectativa por parte dos estudantes em treinar a leitura de gêneros acadêmicos em inglês, constatada pela análise de necessidades conduzida com as turmas de 2023.

Com base nos resultados do *Final Survey*, é possível dizer que o *SDG-Plos* forneceu material de leitura proveitoso para os alunos do curso, ainda que fossem de PPGs diferentes, dada a forma como os alunos avaliaram as atividades. O questionário *Final Survey – 2024*,

[37] No original: *Multi-word expressions are an important component of fluent linguistic production and a key factor in successful language learning.* (Hyland, 2008, p. 4)

focado na avaliação do material e das ferramentas, revelou contentamento dos respondentes em relação aos temas dos textos, o que demonstrou o potencial dos ODS enquanto recurso didático multidisciplinar para o contexto de IFA.

No que tange à terceira pergunta, "Em que medida atividades didáticas direcionadas por dados podem influenciar a aprendizagem de leitura em LI de estudantes de nível básico?", pode-se dizer que as ferramentas utilizadas, o *CorpusMate*, o *Skell* e *VersaText* oferecem recursos de investigação da língua e de seus padrões, o que possibilita que os alunos ampliem seu vocabulário sem ter que lançar mão de uma ferramenta tradutória, como citou um dos participantes no formulário da turma 4 de 2023. Ademais, as atividades movidas por dados contemplaram as três etapadas de exercícios de leitura mencionados por Tomitch (2009), a pré-leitura, a leitura e a pós-leitura.

A DDL proporciona ao aprendiz o contato com materiais autênticos de uso da língua. Com a curadoria adequada do material, as ferramentas potencializaram as estratégias de leitura ensinadas aos alunos que, mesmo com conhecimentos básicos de inglês, foram capazes de realizar os exercícios propostos e, de maneira guiada, exploraram amostras reais da linguagem.

As atividades dirigidas por dados podem ainda estimular a autonomia na aprendizagem, característica requisitada aos estudantes de contexto acadêmico, como quando o aluno procura por palavras que geram combinações recorrentes em sua área de pesquisa, quando analisa uma nuvem de palavras e tenta identificar as classes morfológicas dos itens nela contidos, quando procura por exemplos reais de uso de determinada expressão e se depara com as linhas de concordância de um *corpus*, em que várias possibilidades de uso são expostas.

Por fim, reconhecemos que a segunda tentativa de aplicação das atividades na segunda oferta do curso no contexto do Portfólio de disciplinas para a pós-graduação da Unesp foi afetada pelo número baixo de participantes do *Final Survey – 2024* e que, por isso, não foi possível colher dados robustos como se esperava, em quantidade muito maior, para tecer considerações mais consistentes. No entanto, a quantidade obtida, serve como uma mostra prototípica em que, constatamos informações que demonstram que os estudantes das novas turmas também tiveram um contato produtivo com a abordagem DDL, assim como as tutoras que as aplicaram, visto que seu desempenho se deu com base em um vídeo explicativo sobre as atividades e reflete em certa medida a avaliação dada pelos alunos. Isso evidencia a contribuição deste trabalho com o desenvolvimento de materiais didáticos baseados em *corpora* reutilizáveis, fator importante para a disseminação da LC no ensino de línguas brasileiro

(Berber Sardinha, 2011), e, indiretamente, com a formação docente por meio da oportunidade de contato de professores de LE com as ferramentas *CorpusMate*, *Skell* e *VersaText*.

O desenvolvimento da proficiência linguística para a internacionalização em instituições de ensino superior, abordada na fundamentação teórica deste trabalho, pode contar com os recursos da LC e da DDL como impulsionadores da aprendizem de inglês no contexto acadêmico, especialmente nos cursos a distância, já que a LC "[...] por causa de suas raízes tecnológicas, integra-se nativamente com o EaD on-line" (Berber Sardinha, 2011, p 348) e se presta ao ensino de IFA tanto para os alunos considerados intermediários e avançados como para os iniciantes, como demonstrou-se neste estudo. A partir dos resultados da pesquisa, espera-se que o curso Estratégias de Leitura em Língua Inglesa crie um banco de materiais didáticos direcionados por *corpora* para o nível básico e amplie as ofertas das próximas edições para outros níveis.

REFERÊNCIAS

ABRAHÃO, M. H. V. Algumas reflexões sobre a abordagem comunicativa, o pós-método e a prática docente. *Revista EntreLínguas*. Araraquara, v. 1, n. 1, p. 25-41, 2015.

ANTHONY, L. *Antconc* (*Version 3.4.4*) [*Computer Software*]. Tokyo. Japan: Waseda University, 2011.

ANTHONY, L. *AntCorGen* (*Version 1.2.0*) [*Computer Software*]. Tóquio, Japão: Waseda University, 2022.

ANTHONY, L.; THOMPSON, P. *Automatic Creation and Discourse-Level Annotation of Individualized Discipline-Specific Corpora for the Data-Driven Learning (DDL) Classroom*. University of Birmingham, UK, 2017.

BAISA, V.; CROSTHWAITE, P. *CorpusMate*, 2023. Disponível em: https://corpusmate.com/. Acesso em: 29 jun. 2023.

BAISA, V.; SUCHOMEL, V. SkELL – Web Interface for English Language Learning. *In*: *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Tribun EU, 2014. p. 63-70.

BAFFI-BONVINO, M. A.; KANEKO-MARQUES, S. M. O processo de internacionalização e o ensino e aprendizagem de língua inglesa no contexto de uma universidade estadual. *Estudos Linguísticos*. São Paulo, v. 49, n. 1, p. 510-527, 2020.

BALDO, A. Recursos de inferência lexical em L2. *Letras de Hoje*. Porto Alegre, v. 44, n. 3, p. 60-69, jul./set. 2009.

BARROS, L. A. *Conhecimentos de terminologia geral para a prática tradutória*. São José do Rio Preto: Nova Graf, 2007.

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

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

BERBER SARDINHA, A. P. Como encontrar as palavras-chave mais importantes de um corpus com *Word Smith Tools*. *Revista D.E.L.T.A.* São Paulo, v. 21, n. 2, p. 237-250, 2005.

BERBER SARDINHA, A. P. *Como usar a Linguística de Corpus no ensino de língua estrangeira* - Por uma linguística de corpus educacional brasileira. *In*: VIANA, V; TAGNIN, S. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011. p. 301-356.

BERBER SARDINHA, A. P. *Computador, corpus e concordância no ensino do léxico* - gramática de língua estrangeira. *In*: *As palavras e sua companhia*: LEFFA, V. (Org.), O léxico na aprendizagem. Pelotas: Editora EDUCAT, 2000, p. 45–75.

BIBER, D. *A corpus driven-aproach to formulaic language in English*: Multi-word patterns in speech and writing. International Journal of Corpus Linguistics: John Benjamins Publishing Company, p. 275-311, 2009.

BOCORNY, A. E. P.; WELP, A. O desenho de tarefas pedagógicas para o ensino de Inglês para Fins Acadêmicos: conquistas e desafios da Linguística de Corpus. *Revista Estudos da Linguagem*. Belo Horizonte, v. 29, n. 2, p. 1529-1638, 2021.

BOLSTER, A.; LEVRAI, P. Using the Sustainable Development Goals in the English for Academic Purposes classroom. *In*: MALEY, A.; PEACHEY, N. (Org.). *Integrating global issues in the creative English language classroom:* With reference to the United Nations Sustainable Development Goals. Londres: British Council, 2017.

BONDI, M; SCOTT, M. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2010.

BOULTON, A. *Data-driven learning and language pedagogy*. *In*: THORNE, S.; MAY, S. (eds.). Language, Education and Technology: Encyclopedia of Language and Education. New York: Springer, 2017.

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.; COBB, T. *Corpus Use in Language Learning*: A Meta-Analysis. Language Learning. Ann Arbor, v. 67, n. 2, p. 348-396. 2017.

BRITISH COUNCIL. British Council Brasil, 2023. Quadro Comum Europeu de Referência para Línguas (CEFR). Disponível em: https://www.britishcouncil.org.br/quadro-comum-europeu-de-referencia-para-linguas-cefr. Acesso em: 25 de set. 2023.

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

CELANI, M. A. A. Introduction. *In*: CELANI, M. A. A. et al. (Org.). *ESP in Brazil:* 25 years of evolution and reflection. Campinas, SP: Mercado de Letras, 2005. p. 13-26.

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

CONSELHO DA EUROPA. *Quadro Comum Europeu de Referência para as Línguas* – Aprendizagem, Ensino e Avaliação. Portugal, Lisboa: ASA Editores, 2001.

CORPAS PASTOR, G. *Manual de Fraseología Española*. Editorial Gredos: Madrid, 1996.

CORPAS PASTOR, G.; MITKOV, R. Computational and Corpus-Based Phraseology - *Third International Conference, Europhras 2019*. Springer, 2019.

COSMIC Jesus. Intéprete: Black Alien; Forfun. Compositor: Danilo Cutrim; Gustavo Black Alien; Nicolas Fassano; Rodrigo Costa; Vitor Isensee. *In*: ALEGRIA compartilhada. Intérprete: Forfun. Rio de Janeiro: DeckDisc, 2011. 1 CD (43min4s), faixa 4.

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) – Pontíficia Universidade Católica de São Paulo – PUC. São Paulo, 2016.

FINARDI, K. R.; PORCINO, M. C. Tecnologia e Metodologia no Ensino de Inglês: Impactos da Globalização e da Internacionalização. *Revista Ilha do Desterro*. Florianópolis, n. 66, p. 239-282, jan/jun. 2014.

FINATTO, B. M. S; FROMM, G.; TAGNIN, S. E. O. Linguística de *Corpus*: conquistas e desafios. *Revista Estudos da Linguagem*. Belo Horizonte, v. 29, n. 2, p. 661-671, 2021.

GARCIA, D. N. M.; KANEKO-MARQUES, S. M. Implementação do Programa Idiomas sem Fronteiras na UNESP. *Estudos Linguísticos*. São Paulo, v. 48, n.1, p. 242-261, 2019.

GARCIA, D. N. M.; PINTO, P. T.; SANTOS, D. C. Ações de internacionalização para o ensino e a pesquisa na área de línguas. *Fórum Linguístico*. Florianópolis, v. 18, n. 1, p. 5618-5630, jan./mar. 2021.

HEEMANN, C. Inglês instrumental (ESP): o uso de estratégias de leitura em um curso on-line de leitura. *Signo*, Santa Cruz do Sul-RS, v. 34, n. 56, p. 137-156, jan. 2009.

HUTCHINSON, T. & WATERS, A. *English for Specific Purposes:* A Learning- Centred Approach. Cambridge University Press, 1987.

HYLAND, K. *As can be seen*: Lexical bundles and disciplinary variation. English for Specific Purposes, v. 27, p. 4-21, 2008.

HYLAND, K. *English for academic purposes: an advanced resource book*. Nova Iorque: Routledge, 2006.

HYLAND, K. English for Specific Purposes: What is it and where is it taking us?. *ESP today*, v. 10, n. 2, p. 202-2020, 2022.

JENKINS, J. *English as a lingua franca in the international university:* the politics of academic English language policy. Routledge, 2014.

JENKINS, J. *The internationalisation of higher education*. But what about its lingua franca?. *In*: KUMIKO MURATA (Org.), *Waseda Working Papers in ELF*, Volume 3. Waseda University, 2014.

JOHNS, A. M. *The history of English for Specific Purposes Research*. *In*: PALTRIDGE, B.; STARFIELD, S. *The Handbook of English for Specific Purposes*. Cambridge, 2013.

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

KHAND, Z. Teaching Reading Skills: Problems & Suggestions. *Journal of Research (Faculty of Languages & Islamic Studies)*, v. 5, p. 43-56, 2004.

KILGARRIFF, A. *et al*.: ten years on. Lexicography, v. 1, n. 1, p. 7-36, 2014.

LEFFA, V. *Aspectos da leitura:* uma perspectiva psicolinguística. Porto Alegre: Sagra - DC. Luzzatto Editores, 1996.

MARK, G.; O'KEEFFE, A. Principled pattern curation to guide data-driven learning design. *Applied Corpus Linguistics*, p. 1-15, 2022.

MCENERY, T.; XIAO, R. What corpora can offer in language teaching and learning. *In*: HINKEL, E.(ed.). *Handbook of Research in Second Language Teaching and Learning, vol. 2*, Routledge: London & New York, p. 364-380, 2010.

MILIZIA, D. Keywords and phrases in political speeches. *In*: BONDI; SCOTT (Org.). *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2010.

MULLER, P.; OHNHEISER, I.; OLSEN, S.; RAINER, F. *Word Formation* - An International Handbook of the Languages of Europe [HSK series]. Berlin: De Gruyter, 2011.

NADIN, O. L. Abordagem terminológico-discursiva: pelo resgate da Terminologia e da Terminografia no ensino e na aprendizagem de Línguas para Fins Específicos no contexto brasileiro. *Trabalhos de Linguística Aplicada*. Campinas, n. 61, v. 1, p. 97-108, 2022.

NAZZI-LARANJA, L. A. *O uso de agrupamentos lexicais em atividades de Língua Inglesa para o vestibular:* uma metodologia baseada em corpora. 2020. 148f. Dissertação (Mestrado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", São José do Rio Preto.

NESI, H. Corpus studies in EAP. *In*: HYLAND, K.; SHAW, P. (Org.). *The Routledge Hadbook of English for Academic Purposes*. Abingdon: Routledge, 2016.

O'BOYLE, A.; VIANA, V. *Corpus Linguistics for English for Academic Purposes*. Londres e Nova Iorque: Routledge, 2022.

ORENHA-OTTAIANO, A.; PINTO, P. T. Pedagogia do Léxico e da Tradução: Novas Práticas em Pesquisa. *In*: ROCHA, Nildicéia; RODRIGUES, A.; CAVALARI, S. (Org.). *Novas práticas em pesquisa sobre a linguagem:* rompendo fronteiras. 30.ed. Araraquara: Cultura Acadêmica, 2018, p. 127-144.

PINTO, P. T. Using Data Driven Learning Approach to Teach English for Academic Purposes and Improve Higher Education Internationalisation. *In*: CROSTHWAITE, P. (Org.).

Corpora for Language Learing: Bridging the Research-Practice Divide. Londres e Nova Iorque: Routledge, 2024, p. 227-239.

PINTO, P. T.; The Sustainable Development Goal (SDGS) words ‘Poverty’ and ‘Sustainability’ in the Brazilian research. *Cadernos de Linguística*. Campinas, n. 4, v. 2, p. 1-13, 2021.

PINTO, P. T.; SERPA, T.; SILVA, E. B. da; CAMARGO, D. C. de. Quando o léxico geral se torna terminologia no contexto social: Um estudo sobre os termos relacionados aos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas. *Letras & Letras*, Uberlândia, v. 39, p. 1–31, 2023.

PINTO, P. T.; BERBER SARDINHA, T.; DELFINO, M. C. N.; et al. The Academic Lexis in Research Papers on the Sustainable Development Goals: A Lexical Multidimensional Analysis (no prelo).

PINTO, P. T.; BERBER SARDINHA, T.; DELFINO, M. C. N.; OWA, M.; VIEIRA, S. R. Unveiling the academic vocabulary of sustainable development goal-research papers: a comprehensive lexical multidimensional analysis. *Revista Estudos da Linguagem* (no prelo).

RAMISCH, C; VILLAVICENCIO, A. *Chutando o balde ou batendo as botas?* Processamento de Linguagem Natural e Expressões Multipalavras na linguagem cotidiana e científica. *In*: LOPES PERNA, C. B.; DELGADO, H. O. K.; FINATTO, M. J.B. (Org.), *Linguagens especializadas em corpora* - Modos de dizer e interfaces de pesquisa. Porto Alegre: ediPUCRS, 2010, p. 29-49.

ROCHA, J. M. P. Abordagem orientada por dados: desafios e aplicações no ensino de LE. *Revista do GEL*. São Paulo, n. 3, v. 14, p.135-153, 2017.

SANTOS, E. M.; GOMES, R. B.; OLIVEIRA, H. L. *A internacionalização das IES e o ISF*-UFS – Uma análise de resultados. *In*: ABREU-E-LIMA, D. M., *et al*. (Org.), *Idiomas sem Fonteiras* - Internacionalização da educação superior e formação de professores de língua estrangeira. Belo Horizonte: Editora UFMG, 2021, p. 59-77.

SANTOS, L. V.; PINTO, P. T. Avaliação de um curso on-line de estratégias de leitura em língua inglesa ministrado no início da pandemia. *Revista EntreLínguas*. Araraquara, v. 8, n. 00, p. 1-22, 2022.

SCOTT, M. A. *Aprendizagem direcionada por dados*: uma homenagem a Tim Johns (1936 - 2009). *In*: VIANA, V.; TAGNIN, S. E. O. (Org.), *Corpora no ensino de línguas estrangeiras.* São Paulo: HUB Editorial, 2011, p. 25-95.

SCOTT, M. <2023>, *WordSmith Tools*. Stroud: Lexical Analysis Software.

SINCLAIR, J. M. *Corpus and Text* - Basic Principles. *In*: M. WYNNE (Org.), *Developing Linguistic Corpora*: a Guide to Good Practice. Oxford, 2005.

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

SIQUEIRA, S. *World Englishes, World English* – Inglês como Língua Internacional, Inglês como Língua Franca. *In*: LOGARES, X. C.; BAGNO, M. (Org.). *Políticas da norma e conflitos linguísticos*. São Paulo: Parábola Editorial, 2011, p.333-354.

SKETCH ENGINE. *Keywords and term extraction* — identifying typical words. Sketch Engine: 2024. Disponível em: https://www.sketchengine.eu/guide/keywords-and-term-extraction/. Acesso em 15 de jun. 2024.

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

THOMAS, J. *Two web-based tools for learning language from language*. Versatile: 2020. Disponível em: https://www.versatile.pub/uploads/8/1/6/3/81634112/two_web-based_tools_for_learning_language_from_language._thomas_2020_v2.pdf. Acesso em 15 de set. 2023.

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.

UNESP. Sumário Executivo PRINT e Alinhamento do Projeto de Internacionalização ao Plano de Internacionalização. São Paulo: 2018. Disponível em: https://www2.unesp.br/portal#!/print/general-information/print-project/. Acesso em: 23 de set. 2023.

UNESP. Centro de Línguas e Desenvolvimento de Professores, 2021. Programa de Língua Inglesa da Unesp. Disponível em: https://www.ibilce.unesp.br/#!/centro-de-linguas-1/idiomas-sem-fronteiras/. Acesso em 27 de set. 2023.

UNESP. Portifólio de Integração de Disciplinas da Pós-Graduação, 2023. Disponível em: https://www2.unesp.br/portal#!/propg/disciplinas-portfolio. Acesso em: 18 de novembro de 2023.

UNITED NATIONS. The 17 goals. Disponível em: https://sdgs.un.org/goals. Acesso em: 9 abr. 2023.

VIANA, V. Linguística de corpus: Conceitos, técnicas & análises. *In*: VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011, p. 25-95

ZANETTIN, F. Corpus-based Translation Activities for Language Learning, *The Interpreter and Translator Trainer*, n. 3, v. 2, p. 209-224.

APÊNDICES

APÊNDICE A – Atividade elaborada a partir do termo multipalavra *protected area*

Pre-reading

1- A frase a seguir é o título de um artigo cujo resumo leremos na aula de hoje: *Assessing climate change-robustness of protected area management plans—The case of Germany*. Há alguma palavra que você não conheça? Se sim, qual/quais?

_____
_____.

2- Procure as palavras que você desconhece na ferramenta *CorpusMate* (https://corpusmate.com). Ao clicar no botão "*Search*", a ferramenta irá gerar linhas de concordância com todos os resultados encontrados em seus *corpora*.

a) Clique em "*Pattern Finder*" e cole aqui um print com os resultados encontrados.

b) Em seguida, clique em "*Distribution by topic*". De acordo com os dados apresentados, a(s) palavra(s) que você pesquisou ocorre(m) mais vezes em quais áreas?
_____.

c) Em qual modalidade a(s) palavra(s) que você pesquisou parecem ser mais utilizadas, falada ou escrita?
_____.

d) O que as informações adquiridas nas questões anteriores permitem dizer sobre o vocabulário do texto que será lido?
_____
_____
_____

_____.

3- Acesse o artigo completo por meio do link para a revista onde foi publicado: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.018597. Cole apenas o resumo na ferramenta *VersaText* (https://versatext.versatile.pub/) e crie uma nuvem de palavras dessa amostra. Cole aqui um *print* da nuvem de palavras criada.

a) Qual a classe de palavras das palavras destacadas na nuvem (as que aparecem em fonte maior)?
_____.

b) Selecione cinco palavras que podem ser consideradas palavras-chave do texto.
_____.

c) Alguma das palavras do título aparece na nuvem de palavras? Qual/quais?
_____
_____.

4- O que é *robustness*? Você pode pesquisar o significado em um dicionário *online*.
_____
_____
_____.

Reading

- Click here to read text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0185972;

- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

Post-reading

1- Indique o *multiword term* adequado para as descrições abaixo.

a) climate change - b) protected areas - c) climate change-robustness

( ) O instrumento mais importante para a conservação da biodiversidade.
( ) Característica referente à alteração climática das áreas protegidas.
( ) Fenômeno para o qual as áreas protegidas precisam se preparar.

2- No estudo apresentado no artigo, os autores focam:

a) no aumento das áreas de conservação, principal meio de preservação da biodiversidade.
b) na robustez da mudança climática de 60 áreas de preservação na Áustria.
c) na avaliação dos planos de gestão de 60 áreas protegidas em toda a Alemanha com relação à sua robustez às mudanças climáticas.
d) nas áreas de conservação da rede Natura 2000, estabelecidas no abrigo da legislação de conservação da União Europeia.

3- Indique a ordem dos procedimentos metodológicos mencionados no resumo (1°, 2° e 3°).

( ) Avaliação do desempenho de planos de gestão individuais das áreas de conservação da Alemanha em relação à estrutura de robustez às mudanças climáticas.
( ) Definição da gestão de conservação robusta para mudanças climáticas usando 11 princípios e 44 critérios, que seguiram uma abordagem semelhante aos padrões de sustentabilidade.
( ) Descoberta do fato de a robustez da mudança climática das áreas protegidas quase não ter excedido 50% do seu desempenho potencial.

4- Substitua as palavras destacadas por sinônimos que exerçam a mesma função.

a) **First**, climate change-robust conservation management was defined using 11 principles and 44 criteria, which followed an approach similar to sustainability standards. _____.
b) We **then** evaluated the performance of individual management plans concerning the climate change-robustness framework. _____.
c) In general, the individual principles showed very different rates of accordance with our principles, **but** similarly low intensity. _____.

d) **Lastly**, we recommended the presented principles and criteria as essential guideposts that can be used as a checklist for working towards more climate change-robust planning.
e) Principles with generally higher performance values included holistic knowledge management, public accountability and acceptance **as well as** systemic and strategic coherence. _____.

5- Qual o referente dos elementos dos pronomes destacados no trecho a seguir?

_____.

Protected areas are arguably the most important instrument of biodiversity conservation. To keep ***them*** fit under climate change, ***their*** management needs to be adapted to address related direct and indirect changes.

6- Reflita: O artigo aborda a gestão das áreas de conservação na Alemanha. Como você acha que o Brasil administra suas áreas de conservação? Como o Brasil tem lidado com os problemas relacionados à mudança climática?

_____
_____
_____.

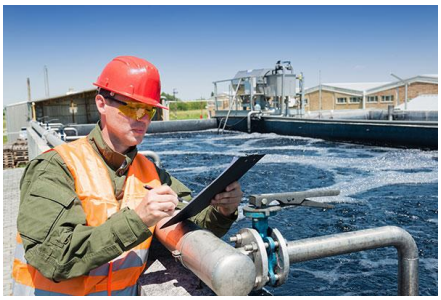
APÊNDICE B – Atividade elaborada a partir do termo multipalavra *waterbird habitat*

Pre-reading

1- Relacione os termos do quadro com as imagens a seguir:

Climate - Urbanization - Water management - Waterbird habitat

_____ _____



_____ _____

2- What could be considered a waterbird habitat?

_____.

3- What is water management? Explain it in your own words.

_____

_____

_____.

4- Procure por "*water management*" na ferramenta *Corpus Mate* (https://corpusmate.com) e diga quais são as áreas em que essa expressão tem mais ocorrência.

_____
_____.

5- Com base nas suas respostas aos exercícios anteriores e na observação da nuvem de palavras-chave do texto que leremos, diga qual pode ser o assunto a ser abordado. Você já ouviu falar sobre essa possível temática a ser tratada no texto?

_____
_____
_____.

Reading

- Click here to read text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169780;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

Post-reading

1- Read the passages and fill in the blanks with the appropriate expressions.

wetland restoration - water supply management - waterbird habitat - urban development

a) Future effects of urban development, changes in water supply management, and precipitation and air temperature related to global climate change on area of _____ in the Central Valley are uncertain, yet potentially substantial.

b) Results indicate that increased _____ and additional conservation and climate change adaptation strategies may be warranted to maintain habitat adequate to support waterbirds in the Central Valley.

c) Like many other mediterranean-climate ecosystems across the globe, the Central Valley has been subject to a burgeoning human population and expansion and intensification of agricultural and _____ that have impacted wildlife habitats.

d) The greatest reduction in waterbird habitat occurred in scenarios that combined warmer, drier climate and plausible _____ options affecting priority and delivery of water available for waterbird habitats.

2- Assinale a alternativa em que a *linking word* destacada exerça a mesma função que "despite" no trecho a seguir: *The Central Valley of California is one of the most important regions for wintering waterbirds in North America* ***despite*** *extensive anthropogenic landscape modification and decline of historical wetlands there.*

a) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **because of the** extensive anthropogenic landscape modification and decline of historical wetlands there.
b) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **as well as** extensive anthropogenic landscape modification and decline of historical wetlands there.
c) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **such as** extensive anthropogenic landscape modification and decline of historical wetlands there.
d) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **even with** extensive anthropogenic landscape modification and decline of historical wetlands there.

3- O *Central Valley*, ou Vale Central, em português, é um ecossistema

a) de clima equatorial, onde se encontra boa parte da vegetação nativa dos Estados Unidos.
b) de clima mediterrâneo, em que houve intensificação do desenvolvimento agrícola e urbano dos Estados Unidos.
c) de clima semiárido, que concentra boa parte do desenvolvimento urbano e agrícola dos Estados Unidos.
d) de clima mediterrâneo, onde se encontra boa parte da vegetação nativa dos Estados Unidos.

4- De acordo com o texto, são considerados efeitos futuros do desenvolvimento urbano no habitat de aves aquáticas no Vale Central:

a) manutenção dos modelos de gestão do abastecimento de água e alterações de temperatura relacionadas à mudança climática global.
b) manutenção dos modelos de gestão do abastecimento de água e alterações de temperatura e precipitação relacionadas à mudança climática global.

c) alterações na gestão do abastecimento de água e alterações de temperatura e precipitação relacionadas à mudança climática global.
d) alterações na gestão do abastecimento de água e manutenção da temperatura e dos índices de precipitação.

5- Qual é a proposta apresentada pelos autores no artigo, de acordo com o que foi exposto no resumo?

a) A restauração planejada de zonas úmidas até 2065 por meio de uma estrutura de modelagem de cenários e recursos hídricos.
b) A modelagem de cenários e recursos hídricos para a área de habitats de aves aquáticas para 17 cenários de clima, urbanização, gestão de abastecimento de água e restauração de áreas úmidas para os anos de 2006–2065.
c) A restauração planejada de zonas úmidas até 2065 por meio da gestão de abastecimento de água e restauração de áreas úmidas para os anos de 2006–2065.
d) A modelagem de cenários e recursos hídricos para a área de habitats de aves aquáticas para 17 cenários de clima, urbanização, gestão de abastecimento de água e restauração de áreas úmidas para os anos de 2006–2099.

6- O que os resultados da modelagem de cenário apresentada indicam?

_____

_____

_____

_____.

APÊNDICE C – Atividade elaborada a partir do termo multipalavra *ocean health*

Pre-reading

1- Take a look at the wordcloud below.

a) Can you identify any cognates?
b) What are the morphological classes represented by the different colors? Give one word (in the word cloud) for each morphological class identified as examples.
c) All of the words will appear in the text that we are going to read. What do you think the text is going to talk about?

2- The following chart brings the most important academic words from the text. Look for examples of usage of the words you don't know on Skell. Write down at least 4 sentences you found on Skell that helped you understand the meaning of the words.

Academic Word List

PoS	Items
noun	goals$_4$ assessment$_2$ index$_2$ data$_2$ areas$_2$ economies$_1$ resources$_1$ assessments$_1$ range$_1$ benefits$_1$ framework$_1$ refinements$_1$ input$_1$ formulations$_1$ variability$_1$ policy$_1$ focus$_1$
verb	derive$_1$ evaluate$_1$ focused$_1$ demonstrates$_1$ highlighting$_1$ identifying$_1$ enhancing$_1$ expanding$_1$ monitoring$_1$
adjective	fundamental$_1$ previous$_1$ global$_1$ regional$_1$
preposition	despite$_1$

i)_____
_____.

ii)_____
_____.

iii)_____
_____.

iv)_____

_____.

3- The chart below brings specific words that are found in the text. Are there any words unfamiliar to you? Look for examples of usage of the words you don't know on Skell. Write down at least 3 sentences you found on Skell that helped you understand the meaning of the words.

Text-specific words

PoS	Items
common nouns	coastlines$_1$ enjoyment$_1$ evaluates$_1$ oceans$_1$ congruent$_1$ mariculture$_1$ tourism$_1$ extractive$_1$ biodiversity$_1$ applicability$_1$ usefulness$_1$ fisheries$_1$ habitats$_1$
proper nouns	Brazil Ocean Integrated
adjective	coastal$_3$ marine$_3$ striking$_1$ relative$_1$

i)_____
_____.

ii)_____
_____.

iii)_____
_____.

4- Escolha duas palavras da tabela de palavras acadêmicas e duas da de palavras específicas do texto e tente explicar, com suas palavras, seus significados.

_____
_____
_____
_____
_____
_____
_____.

Reading

- Click here to read the text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092589;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

Post-reading

1- Qual o assunto abordado no texto?

_____
_____
_____.

2- Qual o gênero do texto? Quais as características típicas desse gênero?
_____
_____
_____.

3- Com base na leitura do texto, complete as frases a seguir com o marcador do discurso adequado.

i. The human use and enjoyment of coastal and marine resources is of fundamental importance to Brazil _____ the country has one of the largest and fastest growing economies and one of the largest coastlines in the world.

a) moreover.
b) hence.
c) because.
d) however.

ii. _____ refinements of input data and model formulations, the national score of 60 (out of 100) was highly congruent with the previous global assessment for Brazil of 62.

a) Therefore.
b) Although.
c) Besides.
d) Equally.

4- Match the words in order to create multi-word terms that we will find in the text. You need to fill the blanks with the words from the word cloud, presented in the following box:

protected - marine - national - coastal - regional - ocean

i. _____ level
ii. _____ habitats
iii. _____ scale
iv. _____ areas
v. _____ health
vi. _____ areas

5- Match the verbs with their possible complements and complete the sentences.

a) We **applied** …
b) This study **demonstrates**…
c) Here we **describe**…
d) … this study **suggests**…

( ) that future actions should focus on: enhancing fisheries management, expanding marine protected areas, and monitoring coastal habitats.
( ) the first such assessment for Brazil at both national and state levels.
( ) the Ocean Health Index framework, which evaluates ten public goals for healthy oceans.
( ) the applicability of the Ocean Health Index at a regional scale, and its usefulness in highlighting existing data and knowledge gaps and identifying key policy and management recommendations.

6- De acordo com o texto, as avaliações integradas da saúde dos oceanos são necessárias…

a) Porque evidenciam quais são os recursos marinhos ainda possíveis de serem explorados.
b) Para conhecer a variabilidade das possibilidades de vida marinha, amplamente preservadas pelo governo brasileiro.
c) Para fazer uma estimativa das práticas extrativistas permitidas, como a provisão de alimentos, e para avaliar quais são os recursos marinhos ainda possíveis de serem explorados.
d) Para entender a condição de vários benefícios que os seres humanos obtêm dos sistemas marinhos e para avaliar onde a atenção deve ser focada para melhorar a saúde desses sistemas.

7- O estudo descrito no artigo demonstra…

a) A variabilidade nas pontuações entre os estados costeiros e as principais políticas e recomendações de gestão.
b) A aplicabilidade do Índice de Saúde do Oceano em escala regional e sua utilidade em destacar os dados existentes e as lacunas de conhecimento, assim como em identificar as principais políticas e recomendações de gestão.
c) Avaliações integradas da saúde dos oceanos e ações futuras, que devem se concentrar em: melhorar a gestão da pesca, expandir as áreas marinhas protegidas e monitorar os habitats costeiros.
d) A variabilidade nas pontuações entre os estados costeiros e sua utilidade em destacar os dados existentes, as lacunas de conhecimento, assim como as principais políticas e recomendações de gestão.

8- Para melhorar a saúde dos oceanos no Brasil, o estudo sugere que…

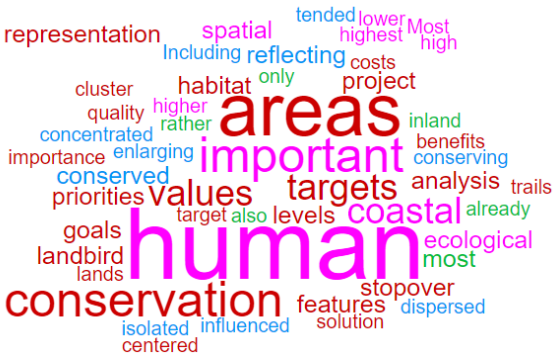
a) As ações futuras devem se concentrar em: melhorar a gestão das ações extrativistas, expandir as áreas de pesca e monitorar os habitats costeiros.
b) Avaliações integradas da saúde dos oceanos são necessárias para entender a condição de uma série de benefícios que os humanos obtêm dos sistemas marinhos.
c) As ações futuras devem se concentrar em: melhorar a gestão da pesca, expandir as áreas marinhas protegidas e monitorar os habitats costeiros.
d) As ações futuras devem se concentrar em: melhorar a gestão das ações extrativistas, expandir as áreas de pesca e monitorar os habitats costeiros para preservar a biodiversidade.

9- De que forma a saúde dos oceanos é afetada?

_____
_____
_____
_____.

APÊNDICE D – Atividade elaborada a partir do termo multipalavra *human well-being*

Pre-reading



1- Analyze the picture that follows in order to answer some questions.

a) What do you see in the picture?

_____.

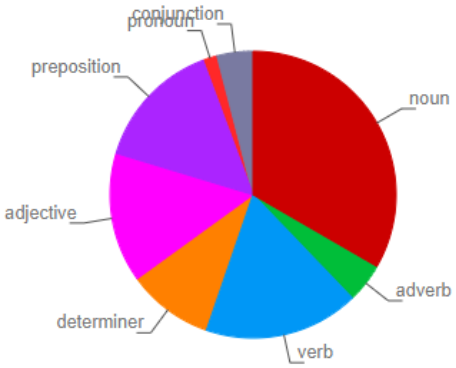
b) Do you know those words?

_____.

c) What are the words familiar to you?

_____.

2- Now, observe the following graph and find out the morphological class of the words.



3- Go to Skell, write the word that you don't know on the search bar and click on "Examples". Write down and say two sentences that helped you understand the meaning of the word.

_____
_____
_____
_____.

4- Create two sentences using some of the words from the word cloud.

_____
_____
_____
_____.

5- In the next word cloud that we are going to analyze we can find some verbs.

a) Do you know the meaning of these verbs?
_____.

b) Can you identify any cognates?
_____.

c) Choose at least three verbs and look for their synonyms on Skell. You will find them clicking on "Similar words".
_____
_____
_____
_____
_____.

6- We are going to read the text in which the words in the word cloud were selected. If the title of the paper were one of the sentences you have created, what could be the probable topic?

_____
_____.

7- The original title of the text is "Designing coastal conservation to deliver ecosystem and human well-being benefits".

a) What scientific area does this paper belong to?
_____.

b) What kind of researcher could have written it?
_____.

c) By looking at the title, what is the text about?
_____.

d) Are there any words that help you infer this information? Write them down.
_____.

Reading

- Click here to read text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172458;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

Post-reading

1- Complete the sentences with one word from the word cloud.

ecological - conservation - human - areas

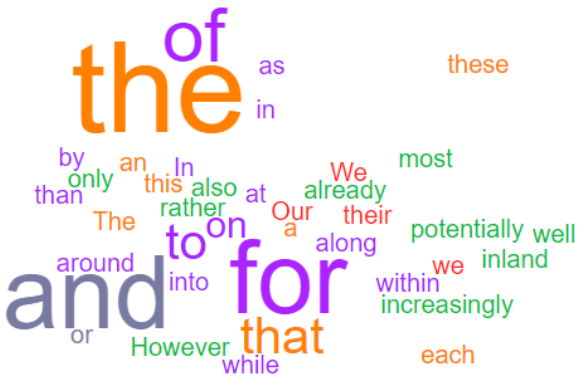
a) The influence that including human well-being had on project results was tested by running five scenarios and setting targets for _____ well-being at different levels in each scenario.

b) Our primary objective was to develop a spatially-explicit _____ plan that identifies the most efficient locations for conservation actions to meet _____ goals while sustaining or enhancing human well-being values within the coastal and nearshore areas of the western Lake Erie basin (WLEB).

c) Inland important _____ tended to cluster around trails and high quality inland landbird stopover habitat.

2- Take a look at the word cloud below. What kind of words are there?

_____.



a) Fill in the blanks with the words from the word cloud.

i. In this study we employ an approach for identifying regionally important human values and establishing specific spatial targets for ___ representation based on stakeholder outreach.

- What kind of discourse markers/reference devices did you find? What kind of ideas do they express?

_____
_____.

- The discourse marker/reference devices you have found is referring to what?

_____.

ii. _____, methods for defining quantitative targets for the spatial representation of human well-being priorities are less developed.

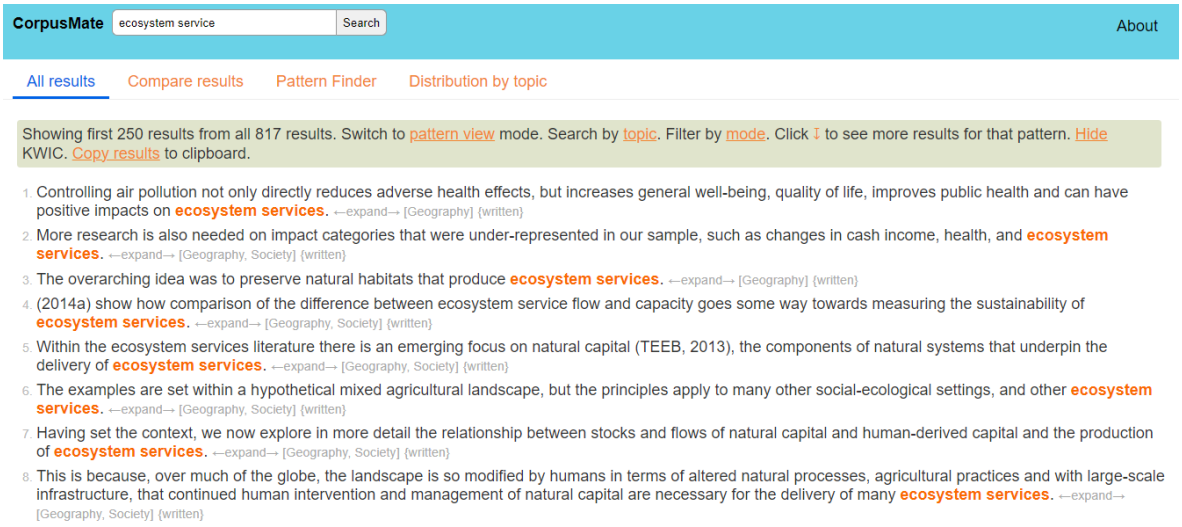
- What kind of discourse markers/reference devices did you find? What kind of ideas do they express?

_____.

APÊNDICE E – Atividade elaborada a partir do termo multipalavra *ecosystem service*

Pre-reading

1- a) What is *ecosystem service*? If you don't know it, take a look at the concordance lines that follow and try to infer a possible definition.



_____
_____
_____
_____
_____.

b) Search for a definition of *ecosystem service* and compare it with your answer given in the previous question.

2- Read the title of the text that we are going to read and identify its keywords (You can use the tool *VersaText* if you want): **Breaking the Link between Environmental Degradation and Oil Palm Expansion: A Method for Enabling Sustainable Oil Palm Expansion.**

_____
_____
_____
_____.

3- Paste the text on *VersaText* and write down some of the keywords suggested by the tool.

_____
_____
_____
_____
_____.

4- Are there any cognates in the word cloud created by *VersaText*? If so, write them down and give their meanings in Portuguese.

_____
_____
_____

_____
_____.

5- Create a word cloud, on *VersaText*, with only words which can work as discourse markers. Write down the discourse markers you have founded:

_____
_____
_____
_____.

6- Read the concordance lines given by *Versatext* in which the discourse markers appear. What kind of functions are they expressing?

_____
_____
_____
_____
_____.

Reading

- Click here to read text: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068610;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text (the abstract) in 2 or 3 minutes.

Post-reading

1- Qual o assunto abordado no texto?

_____
_____
_____
_____
_____.

2- Complete the sentences with the appropriate verb: **achieve - demonstrate - suggest - deliver - facilitate**.

a. To _____ the processes of avoided degradation and land rehabilitation, we have developed a methodology in which we have used international environmental and social sustainability standards to determine the suitability of lands for sustainable agricultural expansion.

b. Through visualizing options for sustainable expansion with our methodology, we _____ that the link between oil palm expansion and degradation can be broken.

c. Our methodology can serve as a template for designing such commodity and country specific tools and _____ such guidance.

d. Application of the methodology with criteria and thresholds similar to ours could help the Indonesian government and the industry to _____ its pro - growth, pro - job, pro-poor and pro-environment development goals.

e. The results _____ that substantial changes in current land-use planning are necessary for most new plantations to comply with international sustainability standards.

3- Match the multiword expressions according to what they can be related to in the text.

a. Land degradation. ( ) Sustainable expansion.
b. Avoided degradation. ( ) Global concern.
c. Land rehabilitation. ( ) International environmental and social
d. Oil palm expansion. sustainability standards.
e. Sustainable agricultural production. ( ) A major driver for deforestation in tropical regions globally.

4- Em áreas tropicais, de acordo com o texto, a degradação de terras está associada a quê?

a. À redução efetiva da degradação e ao uso de terras já degradadas para fins mais produtivos.
b. À conversão de florestas em terras não florestais e às perdas associadas de serviços ambientais.
c. Ao uso de terras já degradadas para fins mais produtivos e à conversão de florestas em terras não florestais.
d. Às mudanças substanciais no planejamento atual do uso da terra baseadas nos novos padrões internacionais de sustentabilidade.

5- De acordo com o texto, a metodologia proposta pelos autores para facilitar os processos de degradação evitada e reabilitação de terras se baseia no quê?
_____
_____
_____
_____.

6- Em que lugar no planeta foi testado o método desenvolvido pelos autores?
_____.

7- Qual tipo de plantação é considerado um dos principais impulsionadores do desmatamento em regiões tropicais no mundo?
_____.