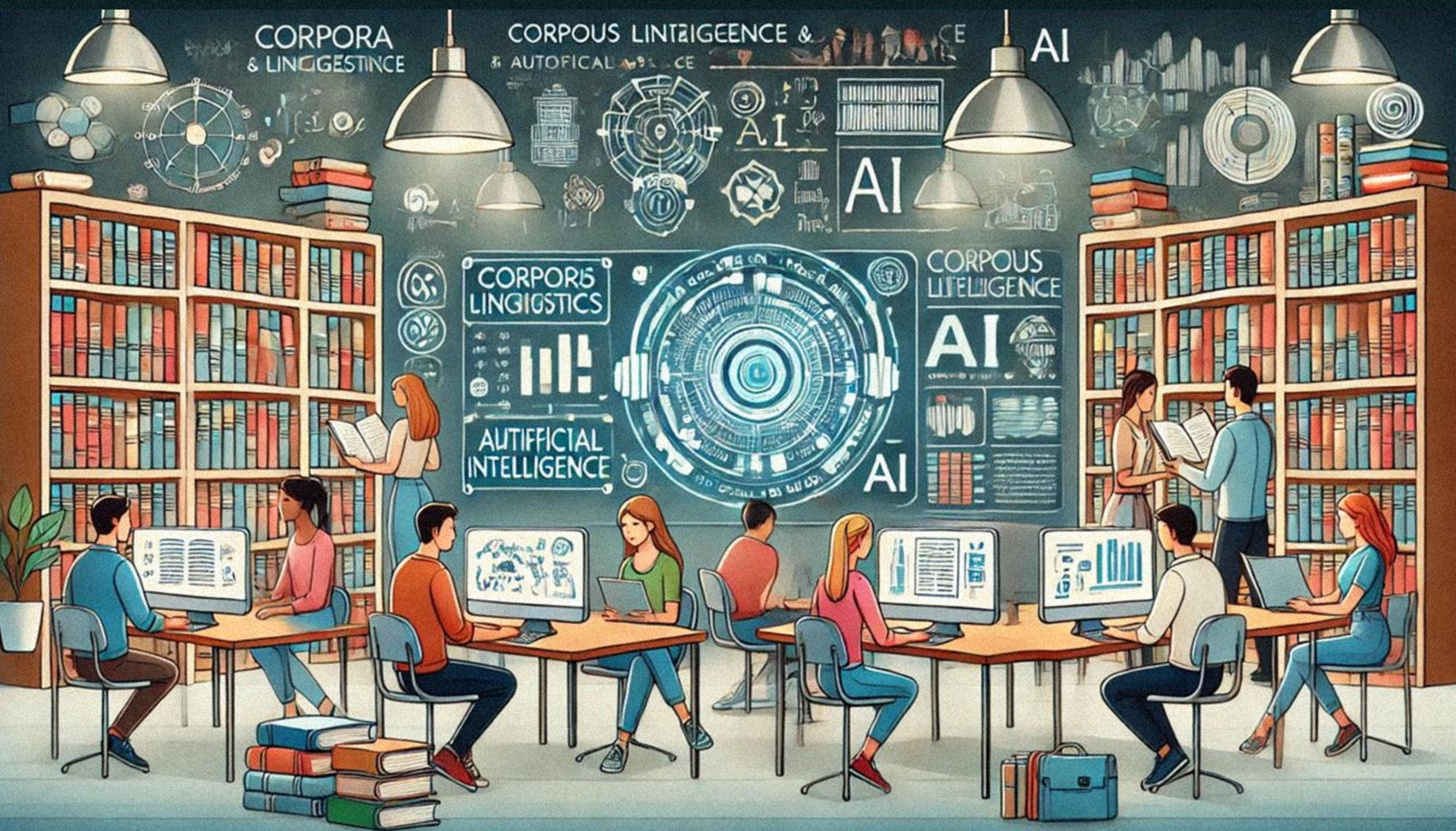


Corpora and Artificial Intelligence for Academic Students



Organizers:

Paula Tavares Pinto, Beatriz Curti Contessoto, Carolina Tavares de Carvalho, Cecília Nevack de Britto, Celso Fernando Rocha, Diva Cardoso de Camargo, Eduardo Batista da Silva, Emiliana Fernandes Bonalumi, Franciele Spinelli, Francine de Assis Silveira, Jeane Cardoso Costa, Luciano Franco da Silva, Liliane Mantovani, Luana Aparecida Nazzi Laranja, Luana Viana dos Santos, Maria Camila Moraes de Sousa, Mayra Aparecida dos Santos, Talita Serpa.

Paula Tavares Pinto, Beatriz Curti Contessoto, Carolina Tavares de Carvalho, Cecília Nevack de Britto, Celso Fernando Rocha, Diva Cardoso de Camargo, Eduardo Batista da Silva, Emiliana Fernandes Bonalumi, Franciele Spinelli, Francine de Assis Silveira, Jeane Cardoso Costa, Luciano Franco da Silva, Liliane Mantovani, Luana Aparecida Nazzi Laranja, Luana Viana dos Santos, Maria Camila Morais de Sousa, Mayra Aparecida dos Santos, Talita Serpa
(Organização)

Corpora and Artificial Intelligence for Academic Students

Araraquara
Letraria
2024

FICHA CATALOGRÁFICA

Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)

Corpora and artificial intelligence for academic students
[livro eletrônico]. 1. ed. Araraquara, SP : Letraria, 2024.

PDF.

Vários autores.

Vários organizadores.

Bibliografia.

ISBN 978-65-5434-106-6

1. Ensino superior 2. Inteligência artificial - Aplicações
educacionais 3. Língua inglesa - Estudo e ensino.

24-242176

CDD-371.334

Índices para catálogo sistemático:

1. Inteligência artificial : Educação 371.334

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

Todas as imagens das capas deste *e-book* foram produzidas pela
autora Carolina Tavares de Carvalho com uso de inteligência artificial.

Editorial committee

Camila Hölfling (UFSCar)

Sandra Mari Kaneko Marques (Unesp)

Contents

Acknowledgement/Agradecimentos	7
Presentation in English	8
Apresentação da obra em Português	13
Chapter 1 - First steps with Corpora - Tools: Skell, VersaText, and CorpusMate	18
Cecília Nevack de Britto	
Jeane Cardoso Costa	
Maria Camila Morais de Sousa	
Chapter 2 - AI-Powered Editing Techniques: Using CHATGPT-4o to Refine your Academic Work	32
Eduardo Batista da Silva	
Chapter 3 - Exploring academic vocabulary	50
Luana Aparecida Nazzi Laranja	
Luana Viana dos Santos	
Chapter 4 - Medical terminology: COVID-19, corpora tools, and AI in English for specific purposes	67
Talita Serpa	
Beatriz Curti Contessoto	
Francine de Assis Silveira	
Chapter 5 - AI tools to support your academic writing	77
Franciele Spinelli	
Chapter 6 - Exploring AI (ChatPDF)	89
Carolina Tavares de Carvalho	
Chapter 7 - Abstract	99
Liliane Mantovani	
Paula Tavares Pinto	

Chapter 8 - Research paper Introduction	111
Carolina Tavares de Carvalho Paula Tavares Pinto	
Chapter 9 - Literature Review	120
Emiliana Fernandes Bonalumi Diva Cardoso de Camargo Paula Tavares Pinto	
Chapter 10 - Materials and Methods	133
Paula Tavares Pinto	
Chapter 11 - Results and Conclusion	139
Mayra Aparecida dos Santos Paula Tavares Pinto	
Chapter 12 - Using the TED Corpus Search Engine to explore discourse organization expressions in Academic spoken language	147
Luciano Franco da Silva	
Answer Keys	155
Authors Bio	170

Acknowledgement/Agradecimentos

This e-book is originated from the research developed within the Postgraduate Program in Linguistic Studies (PPGEL) at São Paulo State University “Júlio de Mesquita Filho” (Unesp), São José do Rio Preto campus, through discussions from the Research Group En-Corpora: Corpus-Based and Corpus-Driven Studies (CNPq/Unesp), which receives financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES, processes 88882.317865/2019-01; 88887.507630/2020-00; 88887.596115/2020-00; 88882.180743/2007-01). Some of the work is also developed within two thematic projects of FAPESP (Process 2022/05908-0; Process 2022/05848-7).

The e-book was published with the support of the Program for Academic Excellence (PROEX-CAPES 0021/202).

Esta obra é oriunda das pesquisas desenvolvidas no âmbito do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), câmpus de São José do Rio Preto, por meio das discussões do Grupo de Pesquisa En-Corpora: Estudos Baseados e Dirigidos por Corpora (CNPq/Unesp), que conta com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal e Nível Superior (CAPES, processos 88882.317865/2019-01; 88887.507630/2020-00; 88887.596115/2020-00; 88882.180743/2007-01). Algumas atividades também foram desenvolvidas em dois projetos temáticos da FAPESP (Processo 2022/05908-0; Processo 2022/05848-7).

O e-book foi publicado com o apoio do Programa de Excelência Acadêmica (PROEX-CAPES 0021/202).

Presentation in English

The e-book **Corpora and Artificial Intelligence for Academic Students** offers us a guide of activities that integrates **digital corpora and artificial intelligence tools**, focusing on academic writing and listening. Interactivity is a central element in all the e-book's activities since the tools and platforms used are embedded in the lessons through specific links, providing teachers and students with easy and direct access, as well as active learning.

Boulton (2017) argues that the analysis of **digital corpora** can help students identify and understand the nuances of language, resulting in more contextualized and meaningful learning. In this sense, it is relevant to emphasize that the integration of digital corpora and artificial intelligence in learning environments has proven to be a significant strategy for improving students' linguistic understanding and the quality of their academic writing. The use of corpora allows students to explore authentic linguistic patterns, considering a real context of language use (Berber Sardinha, 2024).

Artificial Intelligence, consequently, allows systems to simulate intelligence based on patterns from exponential databases. Moreover, AI provides immediate feedback from prompts, enabling an adaptive learning process that meets students' individual needs. From this perspective, we understand the comprehension of AI and computational thinking as essential skills in today's education and knowledge society (Zapata-Ros; Pérez Paredes, 2019).

Teachers and students will find in each unit of this e-book a Theoretical Background that contextualizes the theme covered in each chapter and the tools to be used. Additionally, all activities are described in detail and contain the necessary links to the tools' websites and platforms, facilitating access and practical execution. The material also includes a specific section for Answer Keys, providing the reader with the opportunity to check their answers and track their progress.

In Chapter 1, Cecília Nevack de Britto, Jeane Cardoso Costa and Maria Camila Morais de Sousa introduce Skell, Linggle, and Versatext as the first tools and steps for using corpora and artificial intelligence in teaching and for developing initial skills in exploring linguistic data.

In Chapter 2, Eduardo Batista da Silva explores the use of the WordSmith Tools to develop students' vocabulary in academic writing in English, especially in reading and writing skills. The activity uses artificial intelligence to improve the clarity and appropriateness of academic texts.

Chapter 3, written by Luana Aparecida Nazzi Laranja and Luana Viana dos Santos, explores the use of the CorpusMate and Versatext tools to develop students' academic vocabulary, focusing on reading skills in English for academic purposes. The proposed activities are based on Corpus Linguistics and aim to facilitate the understanding of texts related to the United Nations' Sustainable Development Goals (SDGs) through a data-driven approach.

In Chapter 4, Talita Serpa, Beatriz Curti Contessoto and Francine de Assis Silveira discuss medical terminology related to COVID-19, using the Sketch Engine and Versa Text tools. The proposed activities include exercises with the COVID-19 corpus and culminate in a reading activity based on an AI-generated summary of a news article from FAPESP.

Next, in Chapter 5, Franciele Spinelli explores the use of Artificial Intelligence (AI) and Generative Artificial Intelligence (GenAI) tools in foreign language (L2) teaching, emphasizing their impact on academic writing. The activities propose programs and tools such as ChatGPT, Gemini, Claude, Scispace, ScholarAI, and Litmaps, which can help students develop writing, brainstorming, feedback, and translation skills.

Chapter 6, written by Carolina Tavares de Carvalho, addresses the use of AI tools, highlighting ChatPDF, which transforms the interaction with PDF documents in academic research. The suggested activities show how ChatPDF can improve efficiency and accessibility in academic writing, helping students develop critical skills such as analysis, synthesis, and understanding of complex content.

In Chapter 7, Liliane Mantovani and Paula Tavares Pinto explore the research paper abstracts as a discursive genre, based on Swales' (1990) definitions. For the analysis of recurring linguistic patterns, supporting the production of scientific research papers and the teaching of discursive genres, the activities present and suggest tools such as the Michigan Corpus of Upper-Level Student Papers (MICUSP), the Scientific Electronic Library Online (SciELO), the Manchester Phrasebank, AntConc, WordSmith Tools, Sketch Engine, and ChatGPT.

Chapter 8, written by Carolina Tavares de Carvalho and Paula Tavares Pinto, highlights the importance of integrating theoretical structures and corpora tools for academic writing. The text addresses the CARS (Create a Research Space) model by Swales and Feak (2009), which provides a strategic framework for drafting the introduction of scientific articles. The tools analyzed, such as the Manchester Academic Phrasebank, the Michigan Corpus of Upper-Level Student Papers (MICUSP), Lextutor, and Sketch Engine, are highlighted as resources that enhance the clarity and effectiveness of academic writing.

In Chapter 9, Emiliana Fernandes Bonalumi, Diva Cardoso de Camargo, and Paula Tavares Pinto explore the importance of the Literature Review as a fundamental section of scientific papers. Based on the University of Manchester's Academic Phrasebank, the authors offer various suggestions on how to start a Literature Review; furthermore, they present models for structuring it, such as Swales' CARS model (1981, 1990, 2004), Weissberg and Buker (1990), and the model by Aluísio and Oliveira (1996).

In Chapter 10, Paula Tavares Pinto analyzes the Materials and Methods section of a research paper, highlighting its importance in describing how the research was conducted. The author also mentions McCombes' (2019) perspective, which emphasizes that this section allows others to assess the reliability and validity of the research. Additionally, the author explores the different ways in which researchers describe the Methodology section, differentiating SHAPE (Social Sciences, Humanities, and Arts for People and the Economy) disciplines from STEM (Science, Technology, Engineering, and Mathematics) disciplines (Pinto *et al.*, 2024).

In Chapter 11, Mayra Aparecida dos Santos and Paula Tavares Pinto discuss the Results and Conclusion section in academic texts, emphasizing the importance of precision in each part of the work, which should be intelligently structured into introduction, methodology, results, and conclusion. The authors suggest using the Academic Phrasebank to highlight that conclusions serve two main purposes: to reflect and summarize the main topics covered in the text, as well as to present a final comment or judgment.

Finally, Chapter 12, written by Luciano Franco da Silva, analyzes and presents activities on the use of the TED Corpus Search Engine to explore expressions of discourse organization in spoken academic language. The author highlights the evolution of new genres aimed at academic popularization, which seek to democratize knowledge and facilitate the exchange of ideas between different audiences and disciplines (Gotti, 2014; Wang; Csomay, 2024).

Our hope is that this e-book serves as a practical and effective resource for English language teachers and students, so that, through all the proposed activities, they can develop linguistic skills that encompass the use of digital corpora and artificial intelligence tools—skills that are undoubtedly necessary in today’s educational context.

October/2024

Jeane Cardoso Costa e Carolina Tavares de Carvalho

References

ALUÍSIO, S. M.; OLIVEIRA Jr., O. N. A detailed schematic structure of research papers introductions: An application in support-writing tools. *Revista de La Sociedad Espanyola Para El Procesamiento Del Lenguaje Natural*, v. 19, p. 141–147, 1996.

BERBER SARDINHA, T. AI-generated vs. human-authored texts: a multidimensional comparison. *Applied Corpus Linguistics*, Amsterdam, v. 4, n. 1, p. 01–10, 2024.

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

GOTTI, M. Reformulation and recontextualization in popularization discourse. *Ibérica*, n. 27, p. 15–34, 2014.

PINTO, P. T.; SILVA, L. F.; SERPA, T.; CAMARGO, D. C. Do-It-Yourself Corpora to Support SHAPE and STEM Research Paper Writing. In: SARMENTO, S.; REBECHI, R., MATTE M. L. (org.). *English for Academic purposes: reflections, description e pedagogy*. 1. ed. Porto Alegre: Zouk, 2024. v. 01, p. 97–126.

SWALES, J. M. *Aspects of article introductions*. Language Studies Unit, University of Aston in Birmingham, 1981.

SWALES, J. M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.

SWALES, J. M. *Research genres: Explorations and applications*. Ernst Klett Sprachen, 2004.

WANG, W.; CSOMAY, E. Constructing proximity in popularization discourse: Evidence from lexical bundles in TED talks. *English for Specific Purposes*, v. 73, p. 95–109, 2024. Available at: <https://doi.org/10.1016/j.esp.2023.10.003>. Access on: Nov. 25, 2024.

WEISSBERG, R.; BUKER, S. *Writing up research*. Prentice Hall Englewood Cliffs, NJ, 1990.

ZAPATA-ROS, M.; PÉREZ-PAREDES, P. El pensamiento computacional. *Análisis de una competencia clave*, 2019.

Apresentação da obra em Português

A obra **Corpora and artificial intelligence for academic students** nos oferece um guia de atividades que **integra corpora digitais e ferramentas de inteligência artificial**, com o foco em escrita e em apresentações acadêmicas. A interatividade é um elemento central em todas as atividades do *e-book*, uma vez que as ferramentas e plataformas utilizadas estão incorporadas às lições por meio de *links* específicos, o que permite ao leitor (professor/aluno) um acesso facilitado e direto, além de uma aprendizagem ativa.

Boulton (2017) argumenta que a análise de **corpora digitais** pode ajudar os estudantes a identificarem e a compreenderem as nuances da linguagem, o que resulta em uma aprendizagem mais contextualizada e significativa. Nesse sentido, é relevante enfatizar que a integração de *corpora* digitais e de inteligência artificial nos ambientes de aprendizagem tem se revelado uma estratégia significativa para o aprimoramento da compreensão linguística e da qualidade da escrita acadêmica dos alunos. O uso de *corpora* permite aos alunos a exploração de padrões linguísticos autênticos, considerando um contexto real de uso da linguagem (Berber Sardinha, 2004).

Já a **Inteligência Artificial**, por conseguinte, permite que sistemas simulem uma inteligência baseada em padrões de exponenciais bancos de dados. Ademais, a IA oferece *feedback* imediato a partir de *prompts*, o que possibilita um processo de aprendizagem adaptativo que atende às necessidades individuais dos alunos. Sob esse viés, entendemos a compreensão acerca da IA e o pensamento computacional como competências essenciais no contexto atual da educação e da sociedade do conhecimento (Zapata-Ros; Pérez Paredes, 2019).

Professores e alunos encontrarão em cada unidade deste *e-book* um *Theoretical Background* que contextualiza a temática abordada em cada capítulo e as ferramentas a serem utilizadas. Além disso, todas as atividades estão descritas detalhadamente e possuem os *links* necessários para os *sites* das ferramentas e para as plataformas necessárias, facilitando o acesso e a execução prática. O material também inclui uma seção específica para as *Answer Keys*, proporcionando ao leitor a oportunidade de verificar suas respostas e de acompanhar o seu progresso.

No Capítulo 1, as autoras Cecília Nevack de Britto, Jeane Cardoso Costa e Maria Camila Morais de Sousa introduzem as ferramentas *Skell*, *Linggle* e *Versatext* como primeiros passos para o uso de *corpora* e de inteligência artificial no ensino e para o desenvolvimento de habilidades iniciais na exploração de dados linguísticos.

No Capítulo 2, o autor Eduardo Batista da Silva explora o uso da ferramenta *WordSmith Tools* para desenvolver o vocabulário dos alunos na escrita acadêmica em inglês, especialmente em habilidades de leitura e de redação. A atividade utiliza inteligência artificial para melhorar a clareza e a adequação dos textos acadêmicos.

O Capítulo 3, elaborado pelas autoras Luana Aparecida Nazzi Laranja e Luana Viana dos Santos, explora o uso das ferramentas CorpusMate e Versatext para desenvolver o vocabulário acadêmico dos alunos, focando as habilidades de leitura em inglês para fins acadêmicos. As atividades propostas são baseadas na Linguística de Corpus e visam facilitar a compreensão de textos relacionados aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, por meio de uma abordagem orientada por dados.

No Capítulo 4, as autoras Talita Serpa, Beatriz Curti Contessoto e Francine de Assis Silveira discutem a terminologia médica relacionada à COVID-19, utilizando as ferramentas Sketch Engine e o Versa Text. As atividades propostas incluem exercícios com o *corpus* COVID-19 e culminam em uma atividade de leitura baseada em um resumo gerado por inteligência artificial de uma notícia da FAPESP.

Na sequência, no Capítulo 5, Franciele Spinelli explora o uso de ferramentas de Inteligência Artificial (IA) e Inteligência Artificial Generativa (GenAI) no ensino de línguas estrangeiras (L2), enfatizando seu impacto na escrita acadêmica. As atividades propõem ferramentas como ChatGPT, Gemini, Claude, Scispace, ScholarAI e Litmaps que podem auxiliar os alunos a desenvolverem habilidades de escrita, *brainstorming*, *feedback* e de tradução.

O Capítulo 6, escrito por Carolina Tavares de Carvalho, aborda a utilização de ferramentas de Inteligência Artificial (IA), destacando o ChatPDF, que transforma a interação com documentos PDF na pesquisa acadêmica. As atividades sugeridas evidenciam como o ChatPDF pode melhorar a eficiência e a acessibilidade na escrita acadêmica, auxiliando os alunos a desenvolverem habilidades críticas, como análise, síntese e compreensão de conteúdos complexos.

No Capítulo 7, Liliane Mantovani e Paula Tavares Pinto exploram acerca do resumo de artigos científicos como um gênero discursivo, fundamentando-se nas definições de Swales (1990). Para a análise de padrões linguísticos recorrentes, como suporte à produção de artigos científicos e ao ensino de gêneros discursivos, as atividades apresentam e sugerem ferramentas como o Michigan Corpus of Upper-Level Student Papers (MICUSP), a Scientific Electronic Library Online (SciELO), o Manchester Phrasebank, o AntConc, o WordSmith Tools, o Sketch Engine e o ChatGPT.

O Capítulo 8, por sua vez, escrito por Carolina Tavares de Carvalho e Paula Tavares Pinto, evidencia a importância da integração de estruturas teóricas e ferramentas de *corpora* para a escrita acadêmica. O texto aborda o modelo CARS (*Create a Research Space*) de Swales e Feak (2009), que fornece um *framework* estratégico para a elaboração da introdução de artigos científicos. As ferramentas analisadas, como o Manchester Academic Phrasebank, o Michigan Corpus of Upper-Level Student Papers (MICUSP), o Lextutor e o Sketch Engine, são destacadas como recursos que melhoram a clareza e a eficácia da escrita acadêmica.

No Capítulo 9, Emiliana Fernandes Bonalumi, Diva Cardoso de Camargo e Paula Tavares Pinto exploram a importância da Revisão de Literatura como uma seção fundamental em trabalhos científicos. Com base no Academic Phrasebank da Universidade de Manchester, as autoras oferecem diversas sugestões de como iniciar uma Revisão de Literatura; ademais, apresentam modelos para a sua estruturação, como o modelo CARS de Swales (1981, 1990, 2004), Weissberg e Buker (1990), e o modelo de Aluísio e Oliveira (1996).

Já no Capítulo 10, Paula Tavares Pinto analisa a seção de Materiais e Métodos de um trabalho de pesquisa, destacando sua importância na descrição de como a pesquisa foi conduzida. A autora também menciona a visão de McCombes (2019), que reforça que essa seção permite que outros avaliem a confiabilidade e a validade da pesquisa. Além disso, a autora explora as diferentes maneiras pelas quais os pesquisadores descrevem a seção de Metodologia, diferenciando as disciplinas SHAPE (Ciências Sociais, Humanidades e Artes para as Pessoas e a Economia) das disciplinas STEM (Ciências, Tecnologia, Engenharia e Matemática) (Pinto *et al.*, 2024).

No Capítulo 11, Mayra Aparecida dos Santos e Paula Tavares Pinto discutem a seção de Resultados e Conclusão em textos acadêmicos, enfatizando a importância da precisão em cada parte do trabalho, que deve ser inteligentemente estruturado em introdução, metodologia, resultados e conclusão. As autoras sugerem o recurso Academic Phrasebank para enfatizar que as conclusões têm duas finalidades principais: refletir e resumir os tópicos principais abordados no texto, além de apresentar um comentário final ou julgamento.

Finalmente, o capítulo 12, escrito por Luciano Franco da Silva, analisa e traz atividades acerca do uso do TED Corpus Search Engine para explorar expressões de organização do discurso na linguagem acadêmica falada. O autor destaca a evolução de novos gêneros voltados para a popularização acadêmica, que visam democratizar o conhecimento e facilitar a troca de ideias entre diferentes públicos e disciplinas (Gotti, 2014; Wang; Csomay, 2024).

O nosso desejo é que este *e-book* sirva como um recurso aplicável e eficaz para professores e para alunos de Língua Inglesa, de tal modo que haja, por meio de todas as atividades propostas, o desenvolvimento de habilidades linguísticas que contemplem a utilização de *corpora* digitais e de ferramentas de inteligência artificial; habilidades essas, indubitavelmente, necessárias no contexto educacional vigente.

São José do Rio Preto, outubro de 2024.

Jeane Cardoso Costa e Carolina Tavares de Carvalho

Referências

ALUÍSIO, S. M.; OLIVEIRA Jr., O. N. A detailed schematic structure of research papers introductions: An application in support-writing tools. *Revista de La Sociedad Espanyola Para El Procesamiento Del Lenguaje Natural*, v. 19, p. 141-147, 1996.

BERBER SARDINHA, T. AI-generated vs. human-authored texts: a multidimensional comparison. *Applied Corpus Linguistics*, Amsterdam, v. 4, n. 1, p. 01-10, 2024.

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

GOTTI, M. Reformulation and recontextualization in popularization discourse. *Ibérica*, n. 27, p. 15-34, 2014.

PINTO, P. T.; SILVA, L. F.; SERPA, T.; CAMARGO, D. C. Do-It-Yourself Corpora to Support SHAPE and STEM Research Paper Writing. In: SARMENTO, S.; REBECHI, R., MATTE M. L. (org.). *English for Academic purposes: reflections, description e pedagogy*. 1. ed. Porto Alegre: Zouk, 2024. v. 01, p. 97-126.

SWALES, J. M. *Aspects of article introductions*. Language Studies Unit, University of Aston in Birmingham, 1981.

SWALES, J. M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.

SWALES, J. M. *Research genres: Explorations and applications*. Ernst Klett Sprachen, 2004.

WANG, W.; CSOMAY, E. Constructing proximity in popularization discourse: Evidence from lexical bundles in TED talks. *English for Specific Purposes*, v. 73, p. 95–109, 2024. Available at: <https://doi.org/10.1016/j.esp.2023.10.003>. Access on: Nov. 25, 2024.

WEISSBERG, R.; BUKER, S. *Writing up research*. Prentice Hall Englewood Cliffs, NJ, 1990.

ZAPATA-ROS, M.; PÉREZ-PAREDES, P. El pensamiento computacional. *Análisis de una competencia clave*. 2019.

Chapter 1

First steps with Corpora – Tools: Skell, VersaText, and CorpusMate

Cecília Nevack de Britto
Jeane Cardoso Costa
Maria Camila Morais de Sousa



Theoretical Background

This chapter will address some introductory notions about the use of corpora. According to Sanchez (1995, p. 9), corpora is “a comprehensive and detailed set of linguistic data, which may originate from speech or writing, or even both, organized according to specific criteria.” By analyzing carefully collected structured sets of written or spoken texts (corpora), researchers can examine linguistic patterns, test hypotheses, and develop theories based on empirical data and evidence from the language. The main objective of a corpus is to provide a representative database for the analysis of linguistic patterns. According to McEnery and Wilson (2001, p. 104), the construction and analysis of corpora are fundamental for developing linguistic theories based on empirical evidence. From this, we will briefly mention some applications of corpora in linguistic research. We have different types of corpora: **general ones** (with a wide variety of texts from different genres and registers); **specialized ones** (composed of texts from a specific domain. They are helpful for terminology studies); **comparative ones** (used to compare language use in different linguistic varieties); **and learner corpora** (containing texts written or spoken by language learners. They are valuable for second language acquisition research and language teaching). Regarding the applications of corpora in linguistic studies, we encounter a range of possibilities. Let’s consider some examples:

Analysis of Lexical Patterns

According to Sinclair (1991, p. 115), collocation analysis reveals important aspects of a language’s lexical structure. In this context, corpora can identify lexical patterns, such as collocations. It is possible to investigate which words frequently occur together and how certain expressions are used in different contexts.

Frequency Studies

Another application of corpora in linguistic studies concerns the analysis of word and phrase frequencies, which is crucial for understanding the relative importance of different linguistic elements. As Stubbs (2003, p. 12) points out, frequency is a key indicator for linguistic usage analysis. Frequency studies help identify the most common and relevant words and grammatical structures in specific corpora.

Analysis of Linguistic Variation

Corpora can also study linguistic variations among dialects, registers, and historical periods. This is done through comparative corpora, which can be synchronous or diachronic.

Translation Studies and Contrastive Studies

There are bilingual or multilingual corpora, which are used to compare translations and to study differences between languages. Baker (1993, p. 238) states that a bilingual corpus analysis can reveal important insights into translation practices, which is valuable for translators and linguists interested in equivalence and language transfer issues.

Development of Teaching Materials

An essential application of corpora use is related to developing teaching materials based on real language use. According to McCarthy and O’Keeffe (2010), using corpora to develop teaching materials contributes to teaching authenticity, ensuring that learners are exposed to authentic and relevant examples of language use.

Error Analysis

According to Granger (1998, p. 18), corpus-based error analysis provides an empirical basis for correction and feedback. Learner corpora allow for analyzing common errors in learner productions, enabling teachers to identify problematic areas and develop more effective and intentional teaching strategies.

In summary, the uses and applications of corpora in linguistic research and teaching offer a data-driven approach to understanding and teaching language. Corpus analysis tools allow for the exploration of complex linguistic patterns and the application of insights gained for the development of teaching materials and teaching strategies. As Biber *et al.* (1998) assert, Corpus Linguistics will continue to be a vital area for linguistic research and pedagogical practice.

Based on the considerations mentioned above, we understand that using corpora has been undeniably essential and fruitful. In this practical part of this chapter, we will also present some activities to be carried out with three specific corpus analysis tools,

namely Skell, VersaText, and CorpusMate. Let's see below a brief description of each tool before we start the proposed activities:

SKELL

The first tool is SKELL (Sketch Engine for Language Learning). It is an advanced corpus analysis tool designed to meet the needs of English students and teachers. We can highlight its accessibility and ease of use. The tool uses advanced algorithms and natural language processing technologies to automatically extract relevant examples, collocations, and synonyms for entered words and phrases based on a vast linguistic database with billions of textual occurrences.

VERSATEXT

The second tool is VersaText, an online tool designed to assist English learners in detailed language analysis of specific texts. The tool generates word clouds, allowing users to overview the most common words and separate them into different grammatical classes. It also features a profiler, bringing important information about the inserted text. Combines features from two online text tools, Concordance and Vocabulary Profiler. Concordance creates a word cloud from a text and then shows how individual words are used in that text in concordance lines. Vocabulary Profiler provides frequency statistics about the words in the text. Moreover, VersaText includes many additional features that build upon both resources.

CORPUSMATE

This corpus platform (CorpusMate) provides an enhanced and streamlined experience for linguistic data analysis for students, integrating the best functionalities of existing tools, **such as word and phrase search, concordance view (KWIC), filtering and result hiding (pattern finder and distribution), in the same digital environment.** The CorpusMate **Platform** presents a variety of written and oral corpora that have been carefully filtered with the Python program so that they can only present educational and scientific contexts in twenty school subjects. It is a particularly suitable tool for both undergraduate and postgraduate students in English for Academic Purposes (EAP) courses, **with its content selected from** reputable sources such as BAWE, TED Talks, Simple English Wikipedia, BBC Teach, Elsevier, and BNC 2014 Spoken. The selected texts were normalized, tokenized and

PoS tagged. The texts were checked for profanity with a special tool and all offending sentences were filtered out from the corpus, and only articles containing preselected keywords were processed.

Moving forward, we will present some practical exercises using the tools mentioned earlier.

Exercises:

- ▶ Ø The activities below will allow you to develop skills in corpus analysis, using the **SkELL** tool to explore linguistic patterns, word frequency, and usage contexts of the word “**research**” among other academic words.
- ▶ Ø To begin the suggested activities, **access the SkELL tool** (skell.sketchengine.eu/) in your web browser.

→ Activity 1

Read each statement about the SkELL tool below. Determine if the statement is *true (T)* or *false (F)* based on your knowledge or a quick research. List your corresponding answers.

Statements:

- Ø () **SkELL** is a simple tool for students and teachers of English to easily check whether or how real speakers of English use a particular phrase or a word.
- Ø () The “Word Sketch” function in SkELL summarizes how a word is used grammatically and lexically.
- Ø () SkELL only offers examples in English, not supporting other languages.
- Ø () The SkELL tool is paid and requires a subscription for full access

→ Activity 2

SkELL (Sketch Engine for Language Learning) is an online tool developed to help language learners and linguists explore and analyse linguistic patterns. Have you heard about this tool? **(If not, at the end of this activity, return to this question)**. List some of its features below.

Features:

--

→ Activity 3

Now, **type the word “research”, then click in “Word Sketch”, a function in SkELL** to generate a list of contexts where the word “research” appears. List at least five collocations. Then, analyse the different contexts and discuss this word’s various meanings and uses.

Concordance:

--

→ Activity 4

Choose three words of your preference (**words you like, related to the academic context**) and use the **“Examples”** button to display examples of that word in context. List some of these examples below.

Words:

Examples:

--	--

→ Activity 5

Now, use the “Concordance” button. Then, based on the analysis of concordance and the list of frequent words, create sentences or short paragraphs using the word “research” in different contexts.

Paragraph:

Corpus Linguistics Activity Using the VersaText Tool

Activities– Analyzing an academic article based on its data (Data analysis)

- **Pre-reading activity.** To start, **access** the Versatext tool (<https://versatext.versatile.pub/>) in your web browser, and follow the steps below.

1. Choose an academic article from your field of interest, and copy and paste its **title and abstract** on the Versatext input feature.
2. Analyses the highlighted words in the word cloud feature. Which ones are you interested in learning more about? Write them down below:

Analyzed word	Word class (noun, verb, adjective, adverb)
cried	verb

3. Click on the words that you choose and explore their **concordance** lines, then answer the following questions on the board:

- A. Based on your interpretation of the context provided, what do these words mean?
- B. Research their meaning in a dictionary of your preference;
- C. Reflect on whether your interpretation matches the words' dictionary meaning and write down your thoughts.

Word	Interpretation	Dictionary meaning	Notes
cried	spoke loudly	past simple and past participle of cry	they are different: speak in a sad voice and weep

4. In the profiler feature, try to obtain more information about the words you studied and report their importance for the analyzed text, try to fit them into the categories:

Word	Categories- (Function word; versatile academic; text-specific; other	Notes
cried	text-specific	It is a content word rather than a function word.

5. Did the reading activities with VersaTex help you understand the main features of the academic text selected? Produce a short paragraph explaining your answer.

Corpus Linguistics Activity Using the CorpusMate Tool

- The activities below will allow you to develop corpus analysis skills using the **CorpusMate**. To begin the suggested activities, **access** (<https://corpusmate.com/>) in your web browser. The examples are going to be highlighted.

→ Exercise 1– Concordance Analysis

This exercise aims to raise awareness of how different terms influence specific words when combined. It will enhance your ability to conduct detailed linguistic analysis and improve your skills in corpus-based research.

Follow the steps below:

Step 1- Use the search function on CorpusMate to find the term “**development.**” Explore the different contexts in which the word is used and observe the variations in meaning. Then, write the types of texts in which the term appears below and determine the occurrence of the term in each text.

Types of texts	Occurrences
Law text	Development of the children

Step 2- Now, identify the words that most frequently occur before or after “development” using the collocation analysis tool CorpusMate.

Before	After
its	of

Step 3- After, explore how these words influence the meaning of “development”. Write about this:



Its development- evolving of something.

Developing of- it’s probably about the advance of somebody or something.

Step 4- Finally, analyses if specific collocation patterns are associated with “development” in different fields of study or text types. If there are, write them below. Use the functionality “Distribution by Topic” to help:



“Developing of” is very common in Engineering and Psychology areas.

→ Activity 2- Verbs and concordance

Using both VersaText and CorpusMate, this activity is designed to build proficiency in corpus analysis, enhance linguistic awareness, and provide practical experience in analyzing language use in different contexts. It includes identifying verb usage patterns, analyzing collocations, gaining insights into text characteristics, and reflecting on linguistic implications.

Follow the steps below:

Step 1- Choose an academic text from your field of interest, and copy and paste its **title and abstract** on the Versatext input feature;

Step 2- On the **word cloud feature**, use only the option “verbs” in the content words;

Step 3- Upon the most used verbs in the text, select five of them and list them (**for example- the verb “run”**):

Verbs
Run

Step 4- Use the concordance feature in Versatext to analyse the contexts in which these verbs appear and try to explain in your own words in what do they mean in this context:

Verbs	Meaning
Run	to walk fast

Step 5- Take each of the top five verbs and search for them on the CorpusMate tool. Look for examples in which the verb has the same meaning as the ones from the previous step. Then, identify the words that most frequently occur after the verbs on the Pattern Finder:

Verbs	Words that are common after them
Run	for, away, up, out, by

Step 6- Now, writing time! Observing what was reported in the previous steps, write new sentences using the combination of the verbs and one of the words that occur after them. Do as many combinations as possible.



She is running for city councilor. We ran out of time.

Step 7- **Based on the previous activities**, reflect on the implications of your findings. Consider how the context of your text (e.g., academic, literary, technical) might influence the frequency and usage of these verbs. Discuss any insights gained about the language and style of the text.



References

BAKER, M. Corpus Linguistics and translation studies: Implications and applications. *In: BAKER, M. et al. (ed.). Text and technology: In honor of John Sinclair.* Philadelphia/ Amsterdam: John Benjamins, 1993.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press, 1998.

CORPUSMATE. Available at: <https://corpusmate.com/>. Access on: May 25, 2024.

GRANGER, S. *Learner English on computer.* Longman, 1998.

MCENERY, T.; WILSON, A. *Corpus Linguistics.* 2. ed. Edinburgh: Edingurgh University Press, 2001.

O'KEEFFE, A.; MCCARTHY, M. (ed.). *The Routledge handbook of corpus linguistics.* London: Routledge, 2010.

SANCHEZ, A. Definicion e historia de los corpus. *In: SANCHEZ, A. et al. (org.). CUMBRE – Corpus Linguistico de Espanol Contemporaneo.* Madrid: SGEL, 1995.

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

SKELL. Available at: <https://skell.sketchengine.eu/>. Access on: May 25, 2024.

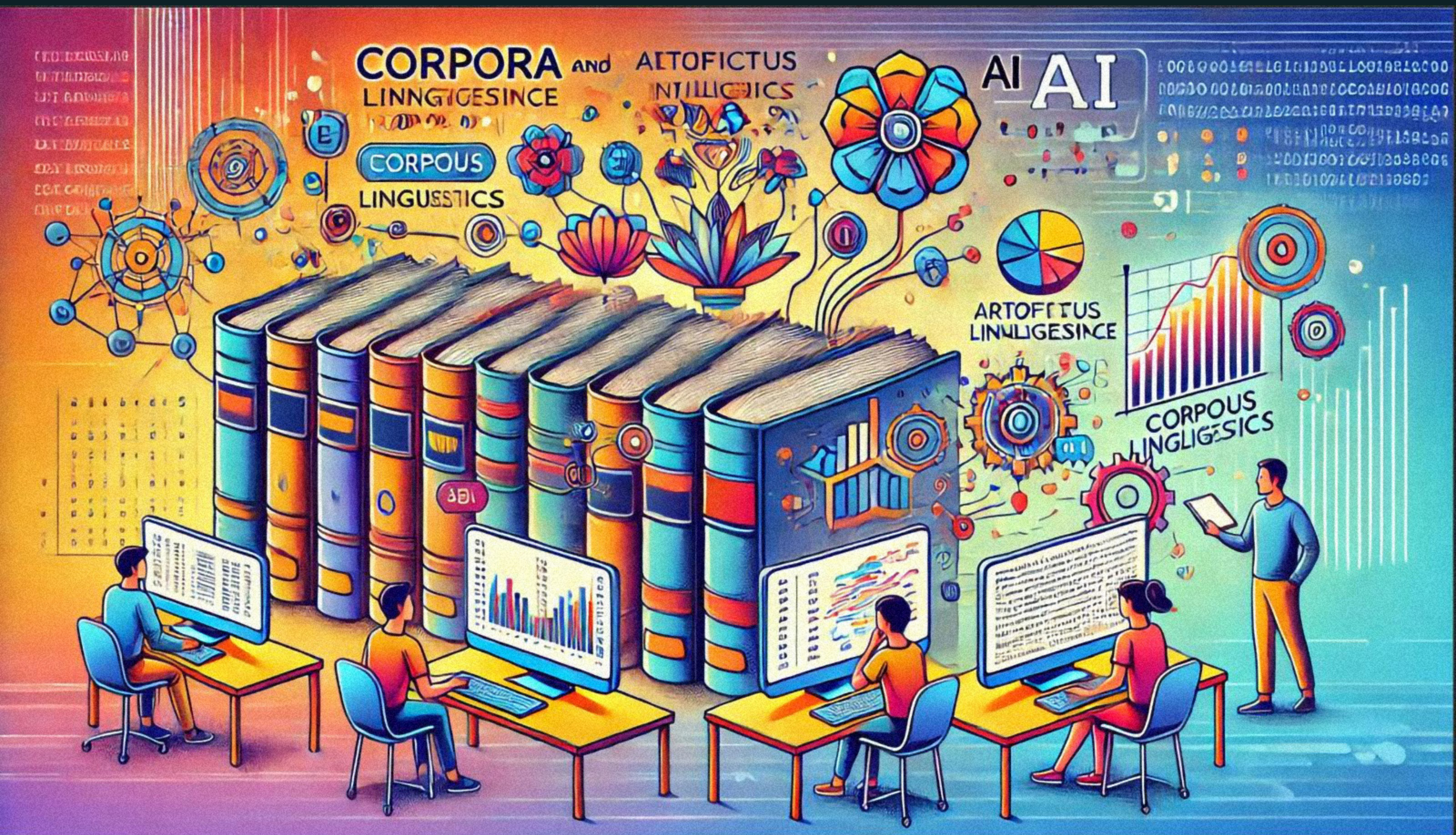
STUBBS, M. Conrad in the computer: examples of quantitative stylistic methods. *In*: Conrad, Concordance, Collocation: Heart of Darkness or light at the end of the tunnel? *Language and Literature*. Trier, Germany: University of Birmingham, 2003. p. 5-24.

VERSATEXT. Available at: <https://versatext.versatile.pub/>. Access on: May 25, 2024.

Chapter 2

AI-Powered Editing Techniques: Using CHATGPT-4o to Refine your Academic Work

Eduardo Batista da Silva



Theoretical Background

In recent years, various higher education demands in Brazil have triggered the ‘publish or perish’ production model, which entails Portuguese academic skills and English written proficiency, mostly for reading and writing. Moreover, lecturing and attending conference meetings have become necessary for those who delve into the research realm.

Aware of the linguistic issues involved, this chapter will provide some editing techniques to use ChatGPT-4 Omni (also known as ChatGPT-4o), an Artificial Intelligence (AI) model launched in mid-May, 2024. With it, Brazilian graduate and undergraduate non-native English-speaking students are likely to enhance the quality of their theses, dissertations and articles. These pivotal registers in academia deserve attention for their situational characteristics, purposes of communication, setting, participants involved and pervasive linguistic features, as discussed in previous investigations by Biber and Conrad (2019).

Biber and Egbert (2023) explored the importance of understanding register as a fundamental aspect of language that shapes how linguistic features are used in different contexts. Biber and Egbert argue that registers influence language choices, which in turn shape the linguistic features of texts within those registers. The study suggests that understanding register variation is crucial for evaluating the effectiveness of AI-generated texts, as it impacts the ability of AI to mimic human-like textual variation and meet specific communicative goals. Overall, Biber and Egbert’s research underscores the importance of register in linguistic studies and highlights the complexity of accounting for both linguistic and situational variations in textual analysis.

Berber Sardinha (2024) compared AI-generated and human-authored texts across different registers to understand if and how they differ linguistically. AI-generated academic texts, particularly in applied linguistics, closely mirror human-authored counterparts, though they are 1.6 times less likely to use explicit references. The similarity between AI-generated and human-authored academic texts highlights AI’s reliance on formal, information-dense language.

There can be an intersection between register studies and computational processing. ChatGPT-4o provides advanced text analysis for academic research since it can analyze large volumes of text, identifying patterns, trends, and anomalies that may not be immediately evident through manual analysis. It leverages deep learning algorithms to model language usage and predict word sequences, aiding in the study of language structures and syntax. It can process and analyze data quickly, making it possible to handle

extensive corpora and perform complex analyses in a fraction of the time compared to traditional methods. It can be used for various tasks, including sentiment analysis, semantic analysis, and translation studies, making it a versatile tool in linguistic research.

We strongly support the ethical use of this tool. We endorse its use to enhance your writing and promote learning and skill development rather than doing the work for you. That being said, you will be introduced to some procedures to make your text more academic. The suggestions offered here do not mean to be exhaustive at all. On the contrary, they might help interested research authors become acquainted with an AI-powered model useful for language processing. Let us begin!

For the activities, you will need the following material:

1 ChatGPT-4o (**Warning:** At the time of this publication, access to ChatGPT-4 may require a paid subscription. While a free version is available, it comes with limitations on the number of interactions. For unlimited access and additional features, a paid plan like ChatGPT Plus is required.)

2 Academic Corpus of English (ACE) Wordlist, freely available on https://1drv.ms/f/s!Ak9g7VEirac5p8NHOxf_fYvPiAEj6Q?e=JGGOrS. (for Activity 4)

3 Sample texts (pieces of writing of yours and an abstract)

The activities will introduce students to straightforward procedures regarding specialized techniques such as proofreading, text conciseness, organizational sections, rhetorical moves and academic vocabulary.

ACTIVITY 1

Proofreading

Running a spell check is an easy task for word editors. However, vocabulary and grammar correction effectiveness will hinge on many rules that are not always available in software packages. The quality of your texts can be improved through different prompts that will enhance your writing. As you become familiar with the output, the prompts can be modified several times and used in combination so as to suit your needs.

Suppose you just wrote a paragraph in English. It is late at night, and you are exhausted and hungry. You might be running out of patience to go any further. At this point, you

cannot identify any mistakes in your paragraph. There is no need to worry. Let us check a sample text:

Corpus Linguistic, a metod wide utilized in linguistical research, involve the analyses of the large collections of texts, known like corpora, to study language patterns and usage. Inspired for the works of Douglas Biber, this field employs quantitative techniques to examine variations between different registers. Biber's multidimensional analysis revelation how linguistic features vary according with situational contexts, offering insights into language complexity and use. Her research highlights the importance of empiric datas in understand language dynamics, challenging traditional notions of gramar and sintax. Corpus Linguistics thus provides a comprehensive approach to linguistic analysis, bridging theoretical and applied linguistics.

You can have it proofread with ChatGPT-4o in seconds and eventually learn all the corrections made. For that, the following prompt will do:

Prompt	Proofread this paragraph for spelling, vocabulary, grammar and punctuation errors. Write the corrections in bold: [insert text].
--------	--

The output text can be seen next:



Here are the corrections:

Corpus Linguistics, a **method** widely utilized in **linguistic** research, **involves** the **analysis** of **large** collections of texts, known **as** corpora, to study language patterns and usage. Inspired **by** the works of Douglas Biber, this field employs quantitative techniques to examine variations between different registers. Biber's multidimensional analysis **reveals** how linguistic features vary **according to** situational contexts, offering insights into language complexity and use. **His** research highlights the importance of **empirical data** in **understanding** language dynamics, **challenging** traditional notions of **grammar** and **syntax**. Corpus Linguistics thus provides a comprehensive approach to linguistic analysis, bridging theoretical and applied linguistics.



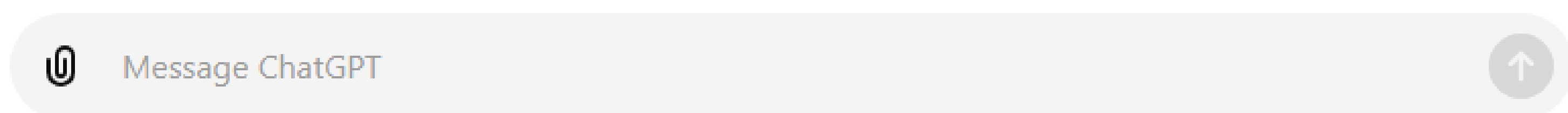
Depending on your needs, you can test the effectiveness of other prompts, adjust them until you obtain the desired textual outcome and save your customized prompt for future reference.

Now, it is your turn to work with your very own piece of academic writing. Choose one or more of these prompts:

Prompts	<p>Please review this section to ensure it maintains an academic tone: [insert text].</p> <p>Check this document for consistency in terminology and style: [insert text].</p> <p>Suggest synonyms for the word 'important'</p> <p>Paraphrase this sentence: The results of the experiment were significant.</p> <p>Can you combine these two sentences into one complex sentence: The study was conducted in 2023. It focused on climate change.</p> <p>Rewrite this paragraph to be more concise and clear: [insert text].</p> <p>How can I improve the cohesion between these two paragraphs: [insert text].</p> <p>Can you revise this paragraph to maintain an academic tone: [insert text].</p>
---------	--

The next step is:

Write one or more of these prompts and paste a paragraph of yours. Then message ChatGPT.



After reading the edited paragraph, how satisfied are you with the edited text?

If you think it can be somewhat improved, consider modifying the prompt(s) used.

ACTIVITY 2

Editing Abstracts

As writing an effective abstract is crucial for academic papers, often times you create a long abstract and due to space constraints the abstract ought to be shortened, either for institutional thesis/dissertation formatting guidelines or journal stylesheet demands. If you are failing to decide what to leave aside, AI can assist you.

What you need is to shorten your abstract text, which requires conciseness. Yet, it is desirable to maintain it brief while informative. The following prompt can be of good use:

Prompt Write a concise abstract: [insert text].

Write a concise abstract:

Understanding and utilizing academic vocabulary is crucial for reading and writing scientific articles and reference works in English. This vocabulary exhibits high frequency and

In the right column, the concise abstract can be seen.

Original abstract

Understanding and utilizing academic vocabulary is crucial for reading and writing scientific articles and reference works in English. This vocabulary exhibits high frequency and consistent distribution across eight specialized fields (Formal and Earth Sciences, Life Sciences, Engineering, Health Sciences, Agricultural Sciences, Applied Social Sciences, Humanities, Linguistics, Language Arts, and Arts), sorted into 69 subfields. The present study aims to describe and analyze academic vocabulary in English across these fields. The specific objectives are to: 1) constitute a specialized corpus in English; 2) propose a methodology to identify and retrieve academic vocabulary; 3) identify fundamental academic vocabulary; 4) establish equivalence in Portuguese; 5) develop a terminological dictionary of fundamental academic vocabulary in English with Portuguese equivalents; and 6) revise the Academic Word List and the Academic Vocabulary List. The theoretical framework is based on significant contributions from Terminology (Barbosa, 1999, 2009; Barros, 2004; Cabré, 1993, 1999), Corpus Linguistics (Berber Sardinha, 2004; Sinclair, 2004), and Lexical Statistics (Larson & Farber, 2012; Oakes, 1998; Butler, 1985). Regarding the methodology, we created an academic corpus in English containing 113,337,773 tokens. WordSmith Tools, version 5, was used as the linguistic-statistical software to process the corpus and retrieve terms. To identify the terms, we used the use coefficient (resulting from multiplying Juilland's dispersion coefficient by the normalized frequency), followed by expert validation of the eligible terms. The results highlight a core set of 505 terms, predominantly nouns (e.g., function, structure, process, chapter, method), verbs (e.g., vary, analyze, require, occur, involve), and adjectives (e.g., individual, similar, major, available, specific). Overall, academic vocabulary use in the eight knowledge areas is more intense than in general language, as indicated by the Wilcoxon Rank Sum Test results with 95% confidence. The academic vocabulary identified in our investigation shows high occurrence and marked distribution across different corpora, indicating it does not belong to a specific specialized domain. Therefore, a systematic study of this vocabulary is necessary to properly evaluate and improve both the understanding and production of scientific texts by Brazilians.

Concise abstract

Understanding and utilizing academic vocabulary is crucial for reading and writing scientific articles and reference works in English. This vocabulary exhibits high frequency and consistent distribution across eight specialized fields, covering 69 subfields. This study aims to describe and analyze academic vocabulary in English within these fields. Specific objectives include: constituting a specialized corpus in English, proposing a methodology to identify and retrieve academic vocabulary, identifying fundamental academic vocabulary, establishing equivalence in Portuguese, developing a terminological dictionary with Portuguese equivalents, and revising the Academic Word List and Academic Vocabulary List. Theoretical foundations are drawn from Terminology, Corpus Linguistics, and Lexical Statistics. Methodologically, an English academic corpus of 113,337,773 tokens was created and analyzed using WordSmith Tools, version 5. Terms were identified using the use coefficient and validated by experts. Results highlight 505 core terms, predominantly nouns, verbs, and adjectives. Academic vocabulary usage in these fields is more intense than in general language, as confirmed by the Wilcoxon Rank Sum Test with 95% confidence. The findings suggest a need for systematic study to improve understanding and production of scientific texts by Brazilians.

The original abstract has 1,221 running words, whereas the concise abstract has 181 words. This time, there was no command to reduce the text to a certain number of words. Alternatively, a more specific prompt can be created such as:

Prompt	Based on my abstract, write a 100-word abstract: [insert text].
--------	---

Graduate students are aware that their thesis/dissertation will have to fit the 150-500 words count. When they write a research article, there is a little change and the count becomes 100-250, according to the Brazilian Technical Standards Association (ABNT, 2021).

Now use an abstract you have written in English and run any of the prompts provided so far.

Another useful editing task is to visualize the sections present in the abstract. It is quite common to read abstracts that lack goals, material and methods, or even results, undoubtedly fundamental parts. You can also check your abstract overall structure, including the sections recommended by your advisor.

Prompt	Review my abstract and highlight background, research goals, theoretical framework, material and methods, results, and implications: [insert text].
--------	---



Here is your abstract with the background, research goals, theoretical framework, material and methods, results, and implications highlighted:

Abstract:

Background:

Reading and writing of scientific articles and reference works in the English language rely on the knowledge of the academic vocabulary – a terminological set which shows high frequency and regular distribution in eight specialized areas (Formal and Earth Sciences, Life Sciences, Engineering, Health Sciences, Agricultural Sciences, Applied Social Sciences, Humanities, Linguistics, Language Arts and Arts), sorted in 69 subfields.

Research Goals:

The present study has the general objective of describing and analyzing the academic vocabulary in the English language, occurring in eight fields of knowledge. Regarding the specific objectives, the research intends to:

1. constitute a specialized corpus in the English language;
2. propose a methodology to identify and retrieve the academic vocabulary;
3. identify the fundamental academic vocabulary;
4. establish equivalence in the Portuguese language;
5. develop a terminological dictionary of the fundamental academic vocabulary in the English language with equivalents in the Portuguese language; and
6. revise the Academic Word List and the Academic Vocabulary List.



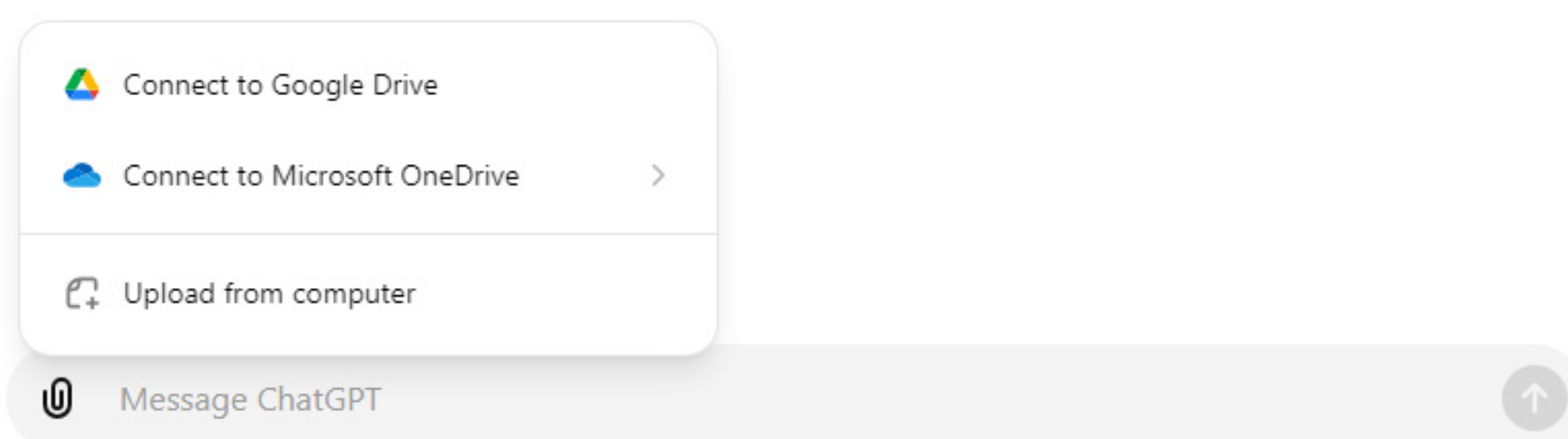
Having the opportunity to check the important sections at a glance will help you enhance your abstract. The missing parts can be written and included after the consultation is finished.

ACTIVITY 3

Writing an effective introduction section

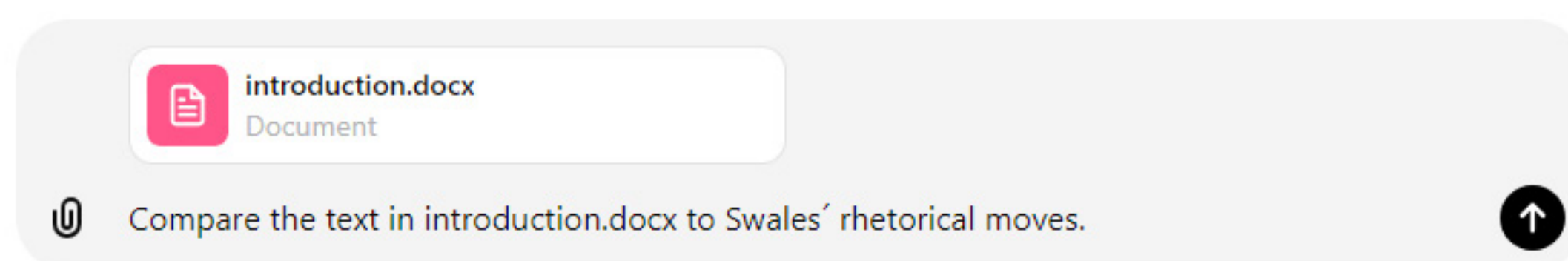
Writing the introduction section in research articles with Swales' rhetorical moves (2011) will surely help authors. The "Creating a Research Space" model has been used for decades and still remains as an authoritative reference. Using ChatGPT-4o to compare your introduction against Swales' rhetorical moves is a practical and effective way to enhance your academic writing. It offers objective and comprehensive feedback that can help you improve the clarity, coherence and persuasiveness of your research paper.

Along with ChatGPT-4o, an innovative resource was launched, enabling researchers to upload a given file and have it analyzed.



We uploaded an MS Word file in which we added part of an article: a 3-page long introduction section, published in 2024 in an Applied Linguistics journal.

Prompt	Compare the text in introduction.docx to Swales' rhetorical moves.
--------	--



The goal is to make sure that the introduction is clear and consistent with academic conventions, making it more effective in communicating research goals and contributions.

Within seconds, the analysis provided feedback and excerpts retrieved from the uploaded file on the main moves, namely "Establishing a Territory", "Establishing a Niche" and "Occupying the Niche". At the end, a summary was given, as can be seen:

Announcing Principal Findings (more common in abstracts):

- While this move is not explicitly present in the introduction, it sets up the framework for presenting findings in the subsequent sections.

Indicating Structure:

- The introduction does not explicitly indicate the structure of the paper, but it sets up a clear research question and the approach to be taken.

Summary

Overall, the introduction aligns well with Swales' CARS model. It effectively establishes the importance of the topic, identifies gaps in the existing research, and outlines the purpose and approach of the current study. This alignment enhances the clarity and persuasiveness of the introduction, making it a strong foundation for the research paper.



Comparing your introduction against Swales' model can help you identify areas where it aligns well with established academic standards and areas that may need improvement.

Now upload your introduction. Run the prompt.

How well does your introduction meet the criteria of each of Swales' moves?

Refine your introduction. After the initial comparison, make revisions based on the feedback and then use ChatGPT again to assess the updated version. Identify specific sentences or sections that correspond to each move and write possible enhancements.

ACTIVITY 4

Checking for academic vocabulary. The following text discusses the importance of using academic vocabulary. Read it before doing the activities.

Silva (2015) developed the file to be used for Activity 4 during his doctoral research. His academic wordlist comprises 505 word families commonly present in eight specialized areas (Formal and Earth Sciences, Life Sciences, Engineering, Health Sciences, Agricultural Sciences, Applied Social Sciences, Humanities, Linguistics, Language Arts, and Arts), sorted in 69 subfields, based on a comprehensive academic corpus of 113,337,773 words.

Before starting the activity, remember that writing research papers in English relies on the knowledge of specific vocabulary items, namely academic vocabulary – a non-domain specific set of words, thus, spreading throughout different research fields. Hence, there is a need for a systematic study on this vocabulary so as to properly evaluate and consequently improve both the understanding and production of scientific texts by Brazilians. The academic vocabulary set found in specialized texts serves as a proxy for producing and promoting scientific information in several subject fields. The academic vocabulary shown here has a statistically significant use when compared to the so-called general English.

In the past years, there have been paramount investigations regarding academic words. Coxhead (2000) created the Academic Word List (AWL). Her list has been used in software programs and dictionaries. Gardner and Davies (2014) published the Academic Vocabulary List and sought to stand as a reference list. In this context, Dayrell (2010) investigated the differences between scientific article abstracts written in English by Brazilian graduate students and abstracts originally published in English. The study reveals significant differences in both lexical and grammatical choices of non-native students compared to native and/or experienced speakers.

In previous research, Silva and Babini (2012) compared the presence/absence of 60 academic words in papers written in English by two groups: native English speakers and Brazilians with proficient English. They found that 23% of the words were overused, 60% were underused, and only 17% were normally used. Such findings indicate that authors should be aware of academic content in their writing.

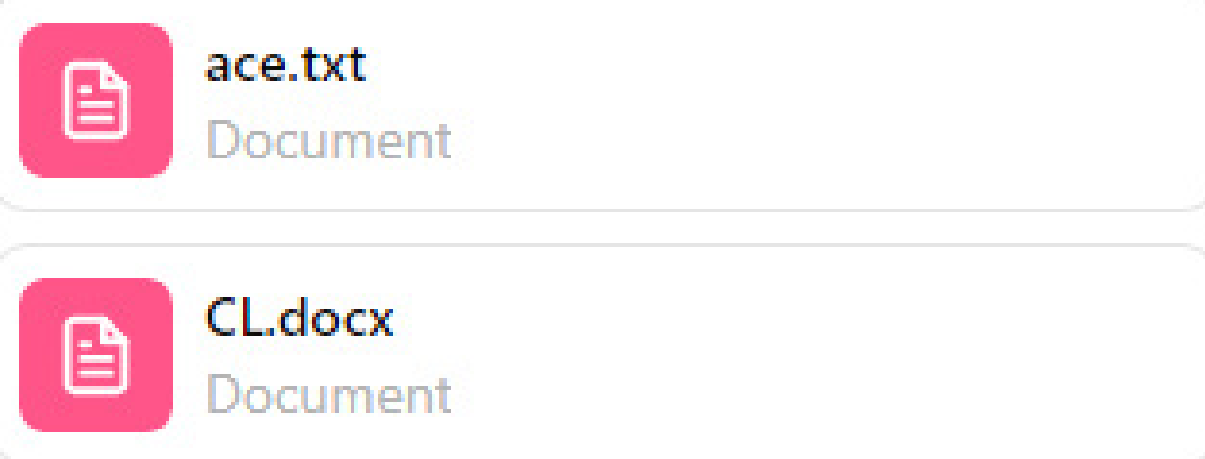
To have students make progress in terms of academic vocabulary, it is beneficial to explore its presence in academic texts.

For this activity, you will need to download the Academic Corpus of English (ACE) Wordlist. It's freely available on https://1drv.ms/f/s!Ak9g7VEirac5p8NH0xf_fYvPiAEj6Q?e=JGGOrS.

Activity 4 differs from the previous activities because it requires a specific file to be run with ChatGPT-4o. Even though AI can perform automated tasks, it needs to be fed with data. Some pieces of information are not in its data bank and its neural networks, for obvious reasons, cannot learn or offer suitable answers. A hassle-free solution is to use data obtained from your research findings.

Download the ACE wordlist. As study material, you can use a Word file with your text. Following the same procedure in Activity 3, upload the ACE wordlist and your Word file.


For our demonstration, we used a sample text on Corpus Linguistics. The purpose is to show how to work with multiple files and ask for uncommon tasks, since ChatGPT-4o is far from being only a search-engine.



Create a new Word file. Copy the full text in CL.docx and highlight in red all the words that are also found in ace.txt.

Prompt	Create a new Word file. Copy the full text in CL.docx and highlight in red all the words that are also found in ace.txt.
--------	--

A new Word file was created and the message shown was as can be seen:

 I have created the Word file with the full text from "doingCL.docx" and highlighted the common words in red. You can download it using the link below:

[DoingCL_Highlighted_Red_Common_Words.docx \[-\]](#)


After downloading the file and opening it to check the output, this is what was rendered:

years, the field of corpus linguistics has garnered **significant** attention from **scholars** and researchers across various linguistic disciplines. This burgeoning interest can be attributed to the field's **unique** ability to provide quantitative **data** on language use, which allows for more **objective** and replicable analyses. The **empirical** nature of corpus linguistics sets it apart from other linguistic approaches that may **rely** more heavily on introspection and

In an attempt to grasp a quantitative perspective in terms of academic words, a second prompt was given:

Prompt	Count the number of words in CL.docx. Count the number of words in red. What's the percentage of words in red?
--------	--

The answer was:

 The document "CL.docx" contains a total of 664 words. Out of these, 62 words are highlighted in red. This means that approximately 9.34% of the words are highlighted in red. [-]



In our sample text, 9.34% of the words belonged to the academic wordlist. Nation (2022) demonstrates that such a relation might vary between 8% and 12%. In this regard, the sample text contains a percentage of academic words that might be considered normal – and expected. Awareness and proper use of academic words typically employed in academic registers will undoubtedly enrich Brazilian authors’ written quality.

Were you familiar with pervasive academic vocabulary in the English-speaking world research papers?

What is the percentage of academic vocabulary in your text?

From now on, you can learn the mean academic vocabulary profile among different authors, self-monitor your very own academic vocabulary profile and enhance your academic vocabulary profile.

Final Remarks

Using these strategies and prompts, non-native English-speaking students can significantly enhance their academic writing and reading skills with the help of ChatGPT-4o, an accessible tool for both novice and experienced writers.

As shown here, researchers can tailor prompts to specific research questions, allowing for precise and targeted investigations into particular linguistic phenomena, and fine-tuning the model to better suit specific needs.

Even though the activities shown here were in English, ChatGPT-4o works with commands in Portuguese as well.

Appendix

If you enjoyed the idea of having your text put to the test, you can make use of specific prompts to thoroughly scan your written work and eventually gain more confidence.

If you upload a file, there is no need to add a colon and the linguistic content to be checked. Then you will not use this sequence “: [insert text]”. Simply use the words before the colon.

Here you are suggestions for editing, proofreading and reviewing:

Analyze the discourse markers used in this paragraph: [insert paragraph].

Analyze the rhetorical devices used in this paragraph: [insert paragraph].

Analyze the use of conjunctions for logical relationships: [insert paragraph].

Analyze the use of direct and indirect speech in this paragraph: [insert paragraph].

Assess the clarity of the thesis statement: [insert thesis statement].

Assess the overall structure of this section: [insert section].

Assess the relevance of this paragraph to the main topic: [insert paragraph].

Check for alignment between the objectives and conclusions: [insert text].

Check for appropriate use of abbreviations and acronyms: [insert text].

Check for appropriate use of headings and subheadings: [insert text].

Check for appropriate use of linking words and phrases: [insert text].

Check for consistency in citation style (eg, APA, MLA, Chicago): [insert text].

Check for correct apostrophe usage in this paragraph: [insert paragraph].

Check for correct preposition usage in this paragraph: [insert paragraph].

Check for proper use of semicolons in this text: [insert text].

Check if the argument in this paragraph is logically sound: [insert paragraph].

Check if the main idea of this paragraph is clear: [insert paragraph].

Check if this paragraph effectively addresses counterarguments: [insert paragraph].

Check the paragraph for appropriate use of emphasis and focus: [insert paragraph].

Check the use of discourse markers for clarity: [insert text].

Check this paragraph for correct article usage: [insert paragraph].

Check this paragraph for correct verb tense usage: [insert paragraph].

Check this paragraph for punctuation mistakes: [insert paragraph].

Combine these two sentences into one complex sentence: [insert sentences].

Conduct a final readthrough to ensure overall readability and polish: [insert text].

Confirm logical flow and coherence between paragraphs: [insert text].

Confirm that all in-text citations match the reference list: [insert text].

Correct any grammar errors in this paragraph: [insert paragraph].

Edit this paragraph for clarity and conciseness: [insert paragraph].

Edit this paragraph to enhance its academic tone: [insert paragraph].

Edit this paragraph to ensure it is free of jargon: [insert paragraph].

Edit this paragraph to ensure it stays on topic: [insert paragraph].

Edit this paragraph to make it more engaging: [insert paragraph].

Edit this sentence to make it more concise: [insert sentence].

Enhance the argument in this paragraph with additional evidence: [insert paragraph].

Enhance the transitions between sentences in this paragraph: [insert paragraph].

Ensure all sentences in this paragraph are complete: [insert paragraph].

Ensure proper use of commas in this text: [insert text].

Ensure that all appendices are clearly referenced in the text: [insert text].

Ensure that all quotations in this paragraph are properly punctuated: [insert paragraph].

Ensure that all tables and figures are correctly labeled: [insert text].

Ensure that any limitations of the study are acknowledged: [insert text].

Ensure that bullet points are used consistently and correctly: [insert text].

Ensure that each paragraph has a clear main idea: [insert paragraph].

Ensure that keywords are relevant and accurate: [insert text].

Ensure that quotations are properly integrated and cited: [insert text].

Ensure that results are presented clearly and objectively: [insert text].

Ensure that sentences are complete and not fragments: [insert text].

Ensure that the conclusion summarizes key points and reiterates the thesis: [insert conclusion].

Ensure that the evidence provided in this paragraph is up-to-date: [insert paragraph].

Ensure that the paragraph maintains a formal register: [insert paragraph].

Ensure that the tone of this paragraph is appropriate for an academic audience: [insert paragraph].

Ensure that this paragraph maintains an academic tone: [insert paragraph].

Ensure the page numbers are correct and formatted properly: [insert text].

Evaluate the conclusion for a strong summary of the main points: [insert conclusion].

Evaluate the effectiveness of the examples used in this paragraph: [insert paragraph].

Evaluate the strength of the argument in this paragraph: [insert paragraph].

Evaluate the use of modality in this paragraph: [insert paragraph].

Identify and correct any typographical errors in this text: [insert text].

Improve the cohesion between these two paragraphs: [insert paragraphs].

Improve the overall readability of this paragraph: [insert paragraph].

Improve the sentence structure in this paragraph: [insert paragraph].

Improve the word choice in this paragraph: [insert paragraph].

Make sure sentences are complete and not fragments: [insert text].

Proofread this paragraph for correct hyphenation: [insert paragraph].

Proofread this paragraph for spelling errors: [insert paragraph].

Proofread this sentence for proper punctuation: [insert sentence].

Proofread this text for consistent use of tenses: [insert text].

Proofread this text for proper noun capitalization: [insert text].

Remove any unnecessary words from this sentence: [insert sentence].

Reorganize the ideas in this paragraph for better structure: [insert paragraph].

Reorganize this paragraph to improve the flow of ideas: [insert paragraph].

Replace colloquial expressions with formal language in this text: [insert text].

Replace vague words with specific terms in this sentence: [insert sentence].

Review the consistency of the argument throughout this section: [insert section].

Review the introduction to ensure it effectively sets up the topic: [insert introduction].

Review the paragraph for appropriate use of passive and active voice: [insert paragraph].

Review the use of evaluative language in this paragraph: [insert paragraph].

Review the use of hedging language in this paragraph: [insert paragraph].

Review the use of nominalization in this paragraph: [insert paragraph].

Review the use of synonyms and antonyms for variety: [insert paragraph].

Review the use of transitions between paragraphs: [insert text].

Review this paragraph for any logical fallacies: [insert paragraph].

Review this paragraph for any potential plagiarism: [insert paragraph].

Review this paragraph for bias and objectivity: [insert paragraph].

Review this paragraph for logical consistency: [insert paragraph].

Review this text for coherence and cohesion: [insert text].

Rewrite this paragraph to avoid redundancy: [insert paragraph].

Rewrite this paragraph to improve its coherence: [insert paragraph].

Rewrite this sentence to avoid ambiguity: [insert sentence].

Rewrite this sentence to eliminate passive voice: [insert sentence].

Rewrite this sentence to improve its readability: [insert sentence].

Shorten this paragraph without losing its main ideas: [insert paragraph].

Verify consistent tense usage: [insert text].

Verify that acknowledgments, if included, are appropriate: [insert text].

Verify that all sources are correctly cited: [insert text].

Verify that any footnotes or endnotes are used correctly: [insert text].

Verify that figures and tables are referenced in the text: [insert text].

Verify the correct usage of homophones in this text: [insert text].

Verify the correct use of conjunctions in this paragraph: [insert paragraph].

References

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 6028: informação e documentação: resumo: apresentação*. Rio de Janeiro, 2021.

BERBER SARDINHA, T. Como usar a Linguística de Corpus no ensino de língua estrangeira: por uma Linguística de Corpus educacional brasileira. In: VIANA, V.; TAGNIN, S. (org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 293-348.

BERBER SARDINHA, T. AI-generated vs. human-authored texts: a multidimensional comparison. *Applied Corpus Linguistics*, Amsterdam, v. 4, n. 1, p. 01-10, 2024.

BIBER, D.; EGBERT, J. What is a register?: accounting for linguistic and situational variation within – and outside of – textual varieties. *Register studies*, Amsterdam, v. 5, n. 1, p. 01-22, 2023.

BIBER, D.; CONRAD, S. *Register, genre, and style*. 2. ed. Cambridge: Cambridge University Press, 2019.

COXHEAD, A. A New Academic Word List. *TESOL quarterly*, Hoboken, v. 34, n. 2, p. 213-238, 2000.

DAYRELL, C. Corpora no ensino do inglês acadêmico: padrões léxico-gramaticais em abstracts de pós-graduandos brasileiros. In: VIANA, V.; TAGNIN, S. (org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 137-171.

GARDNER, D.; DAVIES, M. A New Academic Vocabulary List. *Applied Linguistics*, v. 35, n. 3, p. 305-327, jul. 2014.

NATION, I. S. P. *Learning vocabulary in another language*. 3rd. ed. Cambridge: Cambridge University Press, 2022. (Kindle edition).

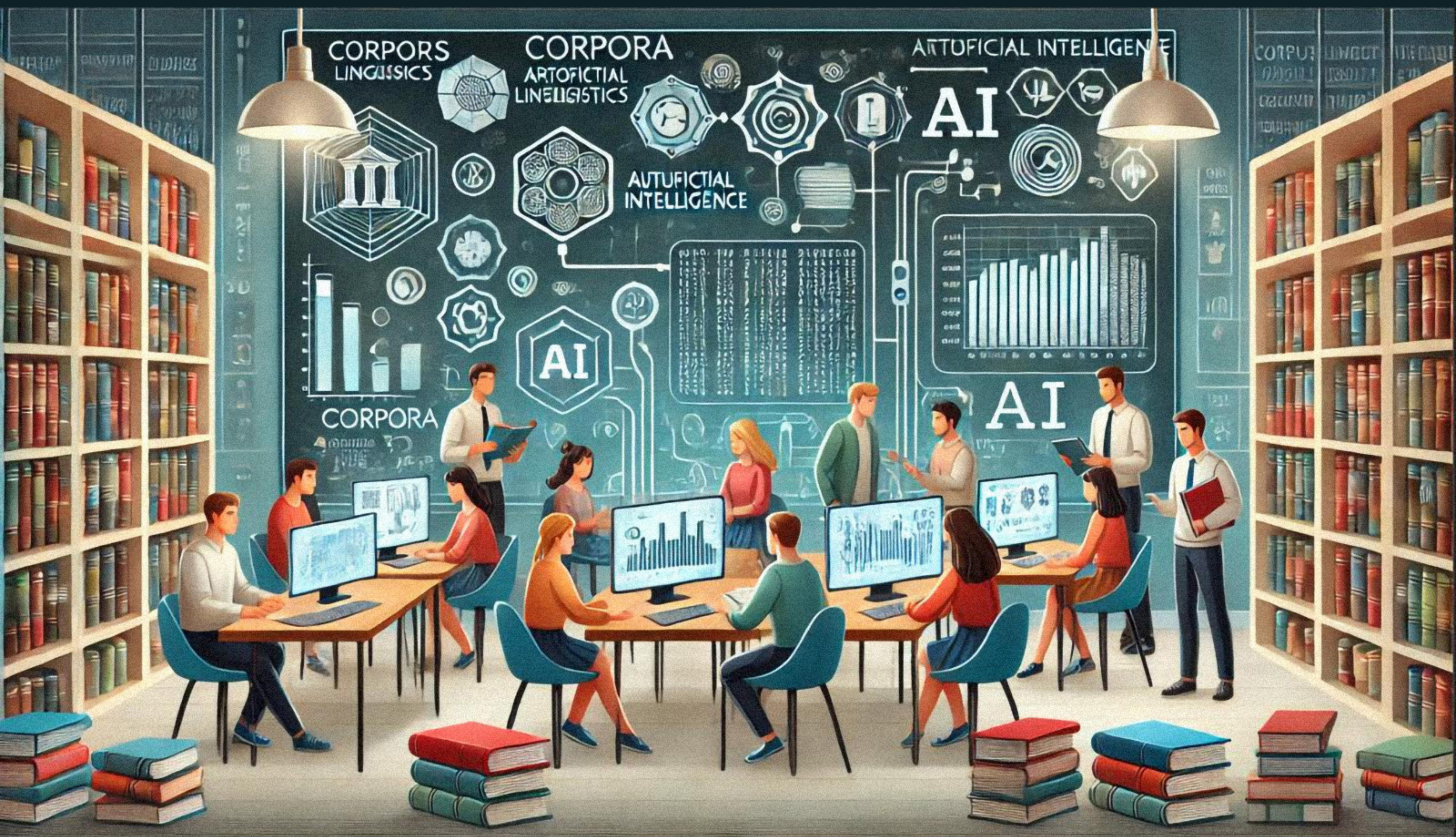
SILVA, E. B.; BABINI, M.; A terminologia acadêmica nos textos científicos em língua inglesa: uma abordagem baseada em corpus. *In*: ISQUERDO, A. N.; SEABRA, M. C. T. C. (org.). *As ciências do léxico: Lexicologia, Lexicografia, Terminologia*. v. 6. Campo Grande: Ed. UFMS, 2012. p. 415-427.

SILVA, E. B. *Identificação e análise do vocabulário acadêmico em língua inglesa presente em textos acadêmico-científicos*. 2015. 289 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2015.

Chapter 3

Exploring academic vocabulary

Luana Aparecida Nazzi Laranja
Luana Viana dos Santos



Theoretical Background

It is known that language is already a barrier to accessing information, as proficiency in English has represented and continues to represent a challenge for the internationalization of Brazilian universities (Baffi-Bonvino; Kaneko-Marques, 2020; Santos; Gomes; Oliveira, 2021; Pinto *et al.*, 2021). International partnerships require intermediate or advanced knowledge of English, according to Pinto *et al.* (2021, p. 5623), however, the results of a survey carried out by IsF revealed that between 2017 and 2019, about 40% of undergraduate students at Unesp campuses demonstrated proficiency levels A1 and A2, referring to the basic level according to the Common European Framework of Reference for Languages (CEFR) (Pinto *et al.*, 2021).

Directing attention to beginner students is crucial, given that this student profile may not have a wide vocabulary repertoire, a factor that influences their reading comprehension (RC) (Baldo, 2009). It is necessary to turn our attention to entry-level undergraduate and graduate students who are already immersed in the academic context and have to deal with the necessary demands for research in today's world.

Therefore, the RC activities developed in this chapter were designed through Corpus Linguistics (CL), for students at levels A1 and A2 in the context of teaching English for Academic Purposes (EAP) (Anthony, 2018; Hyland, 2006; O'Boyle; Viana, 2022), the keywords are from a corpus of scientific articles produced by authors that address the Sustainable Development Goals (SDGs) of the United Nations (UN), established and signed in the 2030 Agenda.

Vocabulary and Corpus Linguistics

It is known that successful reading in foreign languages is related to the process of vocabulary acquisition (Baldo, 2009), mainly contextual, so the knowledge of syntactic structures would have less influence on reading proficiency, as this is influenced by the repertoire of information guaranteed by the readers' lexical collection. Hirvela (2013, p. 84) states, "Vocabulary occupies an important place in both the native language and second/ foreign language reading scholarship, since the words on the page are the starting point for reading".

Corpora can be incorporated into vocabulary teaching because they allow information to be extracted from authentic texts of the language. In addition, this information

(keywords, lexical groupings, concordance lines) is statistically analyzed, providing data that, in most cases, are not noticed by the human eye.

In other words, when the teacher/researcher observes a text in search of keywords, for example, they will select them by intuition because there is no certainty, statistically speaking, that those words are pivotal in the text in question. On the other hand, when professionals use a corpus in their research/material, the concordant does this statistical work for them, including additional information, such as the texts in which this information is most found, which texts do not have certain groupings, what is the percentage of this occurrence, among other details that can help and improve the research/work of the researcher.

As Frankenberg-Garcia *et al.* (2019) state, “using corpora, you are not asking for one person, but for several people together.” Furthermore, Coxhead (2013, p. 118) explains that “Corpus studies have contributed a great deal to our quest to identify and understand more about specialized vocabulary”. Due to corpora relevance in the teaching and learning context, approaches have been created to develop and guide activities in the classroom, such as Data Driven Learning (DDL).

DDL is an innovative proposal for teaching a Foreign Language (FL) first introduced by Tim Johns in the late 1980s (Viana; Tagnin, 2011). DDL is based on the exposure of learners to the language (through the use of corpora) and its exploration in an organized and guided way, so that they have access to collections of texts of their interests and to the frequencies and distribution of occurrences of lexical-grammatical items in a contextualized way (Boulton, 2017). This approach aims to help learners develop a particular ability to see patterns in the data presented, assuming that the students will generalize about the target language.

In the proposed activities, we develop hands-on and hands-off approaches. Hands-on activities can be applied in both face-to-face and online classroom contexts, offering easy accessibility for students. These activities provide students with direct interaction with LC tools. On the other hand, the hands-off activities are designed to engage students without direct contact with the LC tools. Both types of activities are described and explained as follows.

Data-driven activities for vocabulary learning

In this chapter, we present data-driven exercises focused on reading comprehension skills to exemplify how corpora tools can be implemented in vocabulary lessons. The corpora tools used in the exercises are CorpusMate (Baisa; Crosthwaite 2023), Skell (Kilgarriff *et al.*, 2014), and Versatext (Thomas, 2020).

Developed by the researchers Peter Crosthwaite and Vít Baisa (2022), CorpusMate is a platform designed for language teaching based and directed by corpora to basic education students (Baisa; Crosthwaite 2022). It shelters five major corpora, according to the authors, the *BAWE corpus*, the *TED Talks corpus*, *Simple English Wikipedia*, the *BBC Tech corpus*, the *Elsevier Corpus* and the *BNC 2014 Spoken*, which provides a range of data on the written and spoken modality of the language.

Among the resources offered by the platform, we can mention statistics related to the search expression, patterns found with the searched word and its distribution across the areas in which it occurs. It is a very useful platform as it is designed for basic education students, that is, outside the Higher Education context (Baisa; Crosthwaite, 2023). Because of this characteristic, the software is simpler than traditional corpus analysis concordancers, making it easier for students between levels A1 and A2 - basic levels according to the Common European Framework of Reference for Languages (2001).

The second tool is Skell (Kilgarriff *et al.*, 2014), a user-friendly tool designed for students and teachers of English to quickly check how a particular word or phrase is used by real speakers. It draws from corpora and offers several key resources: (i) **Examples**, which present sentences from the corpora where the search word appears, helping users see it in context; (ii) **Word Sketch**, which displays common collocations, showing how the search word functions as a subject, object, or modifier; and (iii) **Similar Words**, which provides synonyms to expand vocabulary (Kilgarriff *et al.*, 2014).

The third one is VersaText, an online tool for English learners to explore the language of a single text. With this tool, students can paste text samples and investigate vocabulary through the **Word Cloud** feature, perform statistical and morphological analysis of the lexicon using **Profiler**, or explore the text with the **Concordance** feature, which highlights each word class with a different color, making it easier to understand combinatorial patterns.

To create reading comprehension activities for students in the context of English for Academic Purposes (EAP), Santos and Pinto (2024) worked with an academic corpus of scientific articles written by international researchers who addressed the Sustainable Development Goals (SDGs) (United Nations 2024) in their work, the corpus SDG-PLOS (Pinto et al, in press). To demonstrate how teachers can start using DDL tasks focused on vocabulary with the assistance of CL, we present exercises from the activities developed in a previous study.

In the exercises presented here, the texts were chosen based on the analysis of some of the corpus keywords. The corpus keywords representing the SDGs were selected to identify the texts that would be the best fit for the activities (Santos; Pinto, 2024). The exercises of each activity are separated into three sections: pre-reading, reading, and post-reading.

The pre-reading exercises are good examples of how the vocabulary of a text can be explored to activate students' previous knowledge the previous knowledge of students. The subsequent exercises of the activities are the reading itself and questions about the content of texts that put into practice reading skills such as skimming and scanning. Post-reading exercises can also take advantage of corpora tools, as well as all kinds of exercises - focused on reading, writing, listening and speaking because learners are always in contact with new vocabulary. These steps set the foundation for the following practical activities, where students will apply these strategies in real-time interactions with texts and language data.

The complete activities are presented below:

Activity 1 – “Waterbird habitat”

Pre-reading

1- Match the words in the box with the following images:

Climate - Urbanization - Water management - Waterbird habitat

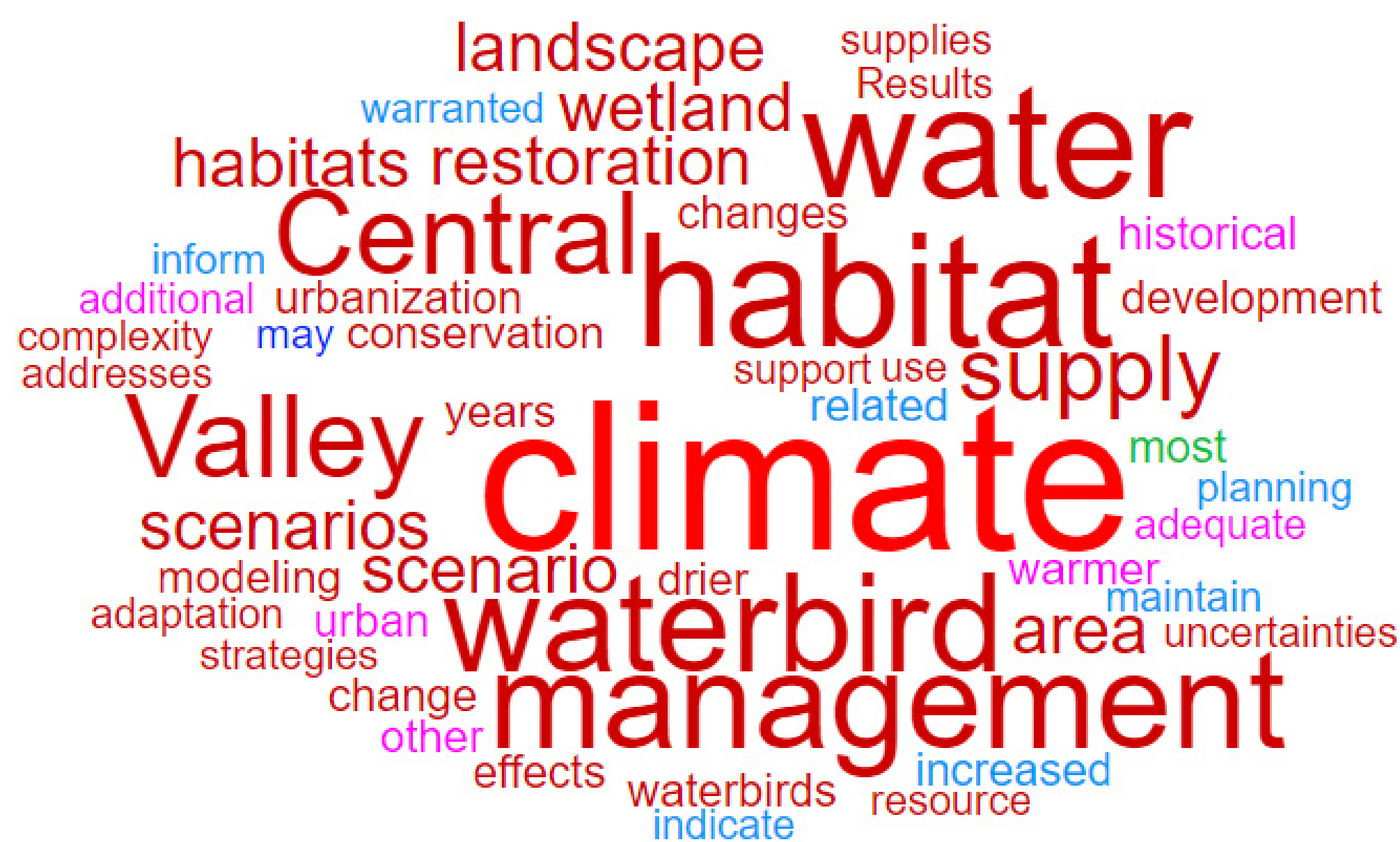


2- What could be considered a waterbird habitat?

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

4- Search for “water management” in the Corpus Mate tool (<https://corpusmate.com>) and tell us which areas this expression is most common.

5- Based on your answers to the previous questions and in the observation of the keywords of the word cloud below, say what could be the topic covered in the text.



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- Mark the option in which the linking word highlights the same function of “despite”: *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- The Central Valley is an ecosystem

- a) with an equatorial climate, where much of the native vegetation of the United States is found.
- b) with a mediterranean climate, in which agricultural and urban development in the United States intensified.
- c) with a semi-arid climate, which concentrates much of the urban and agricultural development in the United States.
- d) with a mediterranean climate, where much of the native vegetation of the United States is found.

4- According to the text, future effects of urban development on waterbird habitat in the Central Valley are considered:

- a) maintenance of water supply management models and temperature changes related to global climate change.
- b) maintenance of water supply management models and changes in temperature and precipitation related to global climate change.
- c) changes in water supply management and changes in temperature and precipitation related to global climate change.
- d) changes in water supply management and maintenance of temperature and precipitation rates.

5- What is the proposal presented by the authors in the article, according to what is exposed in the abstract?

- a) Planned wetland restoration until 2065 through a water resources and scenario modeling framework.

- b) Waterbird habitat and water resource modeling for 17 climate, urbanization, water supply management, and wetland restoration scenarios for the years 2006–2065.
- c) Planned wetland restoration until 2065 through water supply management and wetland restoration for the years 2006–2065.
- d) Waterbird habitat scenario and water resource modeling for 17 climate, urbanization, water supply management, and wetland restoration scenarios for the years 2006–2065.
- e) Planned wetland restoration until 2065 through water supply management and wetland restoration for the years 2006–2099.

Activity 2 – “Ocean health”

Pre-reading



1- Take a look at the word cloud 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 the words will appear in the text that we are going to read. What topic do you think will be covered in the text?

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.

i) _____

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

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- Choose two words from the chart of academic words and two from the chart of specific words in the text and try to explain, in your own words, their meanings.

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- What subject is covered in the text?

2- What is the genre of the text? What are the typical characteristics of this genre?

3- Based on your reading of the text, complete the following sentences with the appropriate discourse marker.

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 in 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- According to the text, integrated assessments of ocean health are necessary...

- a) Because they highlight which marine resources are still possible to be explored.
- b) To learn about the variability of marine life possibilities, largely preserved by the Brazilian government.
- c) To estimate permitted extractive practices, such as food provision, and to assess which marine resources are still possible to be exploited.
- d) To understand the status of various benefits that humans derive from marine systems and to assess where attention should be focused to improve the health of these systems.

7- To improve the health of the oceans in Brazil, the study suggests that...

- a) Future actions should focus on: improving the management of extractive actions, expanding fishing areas and monitoring coastal habitats.
- b) Integrated assessments of ocean health are needed to understand the status of a range of benefits that humans derive from marine systems.

- c) Future actions should focus on: improving fisheries management, expanding marine protected areas and monitoring coastal habitats.
- d) Future actions should focus on: improving the management of extractive actions, expanding fishing areas and monitoring coastal habitats to preserve biodiversity.

References

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. Available at: <https://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/5765>. Access in: May. 22, 2023.

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

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.

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

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

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

CROSTHWAITE, P.; BAISA, V. *CorpusMate*. Available at: <https://corpusmate.com/>. Access in: May.25, 2024.

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

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.

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

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

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

MCENERY, T.; WILSON, A. *Corpus Linguistics*. 2. ed. Edinburgh: Edinburgh University Press, 2001.

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

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* (in press).

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 Fronteiras – Internacionalização da educação superior e formação de professores de língua estrangeira*. Belo Horizonte: Editora UFMG, 2021. p. 59-77.

SANTOS, L. V. dos; PINTO, P. T. Data Driven Learning para aprendizes iniciantes de inglês: uma experiência em um curso de estratégias de leitura. *Revista Horizontes de Linguística Aplicada*, v. 23, n. 1, p. AG3-AG3, 2024.

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.

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. Available at: https://www.versatile.pub/uploads/8/1/6/3/81634112/two_web-based_tools_for_learning_language_from_language._thomas_2020_v2.pdf. Access in: May. 25, 2024.

UNITED NATIONS. The 17 goals. Available at: <https://sdgs.un.org/goals>. Access in: May. 26, 2024.

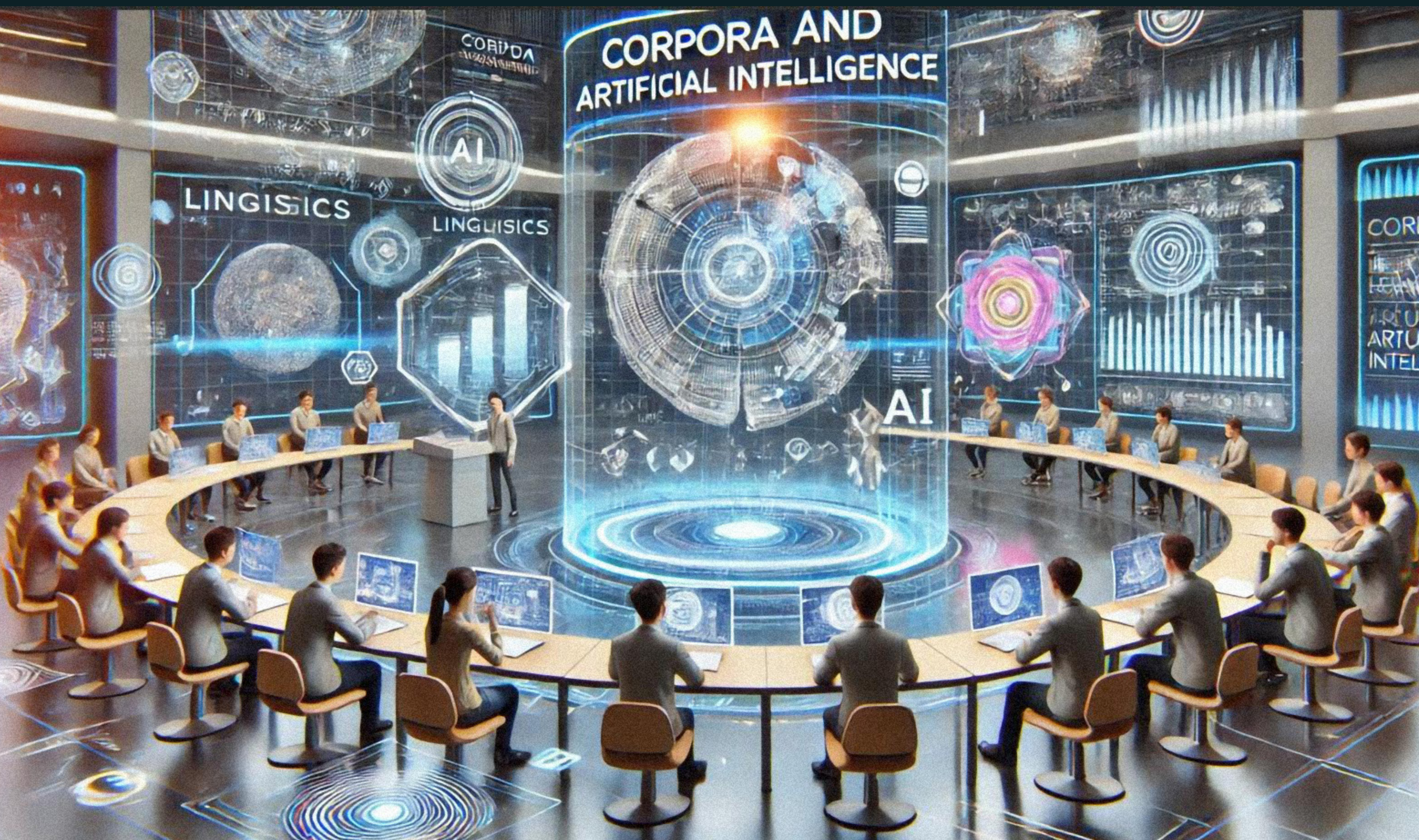
Chapter 4

Medical terminology: COVID-19, corpora tools, and AI in English for specific purposes

Talita Serpa

Beatriz Curti Contessoto

Francine de Assis Silveira



Theoretical Background

In academic communication, the precise use of language is paramount. It is imperative not only to write and speak with formality and conciseness but also to employ specialized terminology essential for accurately disseminating scientific information. In this context, corpora and artificial intelligence provide substantial benefits. Leveraging a corpus of scientific papers within a specific field can facilitate the verification of terminologies *in vivo*. Moreover, corpora are invaluable resources for examining and comprehending how a particular scientific discipline articulates its concepts, methodologies, and analyses.

In the preceding paragraph, several interconnected concepts are examined. The foremost concept pertains to terminology. Terminology is often used interchangeably with vocabulary and generally refers to a collection of specialized words. For instance, terms frequently encountered in scientific articles, such as “purpose,” “data collection,” and “assumptions,” are considered scientific terminology. However, in the context of our discussion, the term “terminology” is employed with a distinct meaning.

As a scientific discipline, Terminology, in a more restrictive sense, studies the specific terms of a particular subject area (e.g., a research field or technical area) and their main characteristics and interrelations. Accordingly, a term is considered a lexical entity that can express concepts, objects, or processes, possessing a specialized meaning within specialized languages (Krieger; Finatto, 2004, p. 79). A collection of such terms may also be referred to as terminology, which is why expressions like “terminology of botany” or “terminology of medicine” are common. To distinguish between the concepts, we typically capitalize the T when referring to Terminology as “the language discipline dedicated to the scientific study of the concepts and terms used in specialized languages” (Pavel; Nolet, 2001, p. 17).

Although terminology and specialized vocabulary are closely related and commonly understood, it is essential to carefully consider the semantic relationship inferred between terms, words, and vocabulary. According to Barros (2004, p. 40, our translation), a word can be understood as “a lexical unit, that is, a linguistic sign composed of expression and content, which belongs to one of the major lexical classes (noun, verb, adjective, or adverb).” In this context, terms and words can be classified as lexical units. The semantic feature that differentiates a ‘term’ from a ‘word’ becomes apparent in its application within a text. A term is specific to a specialized field, whereas a word is a lexicon unit encompassing a given language’s entire set of words. As Kocourek (1991, p. 105, our

translation) notes, words or lexical units are characterized as terms “when they are defined and used in specialty texts.” In other words, the specific status of a term is granted by the pragmatic conditions that adapt it to a particular type of communication (Cabr , 1999). This specialized communication is primarily facilitated through terms that linguistically represent the specific concepts of a field of expertise.

This chapter focuses on medical terminology, precisely terms related to COVID-19. Medical terminology has a rich history rooted in Latin and Greek (Yorova; Iqbal, 2023). Research on compounds and derivatives has garnered significant attention in English (Lieber; tekauer, 2014) and other languages (Pavol *et al.*, 2015). In Brazilian Portuguese, studies have also been conducted (Alves, 2022), including investigations into syntagmatic compositions, such as the types name + adjective and name + prepositional syntagma. The activities proposed in this chapter aim to highlight various recurring syntagmatic constructions in English formed from noun terms. Additionally, we strive to utilize corpora to identify the verbs most commonly used in this context.

All these patterns of linguistic usage are found in textual corpora, which, in this case, are sets “of selected texts that serve as the basis for terminological analysis” (Pavel; Nolet, 2003). These corpora function as a natural habitat for terms (Cabr ; Cunha, 2022). Various criteria for composition and typology must be considered to ensure the corpora produced are suitable and tailored to the specific task at hand.

At this point, another scientific discipline within linguistics becomes relevant alongside terminology: Corpus Linguistics. Its purpose is to examine the frequency and keyness of lexical items as primary data sources. As Berber Sardinha (2010, p. 304, our translation) highlights, these characteristics are critical components of the language system, accounting for vital elements of discourse, such as “the way ‘things’ are said and how all the constructions of a language are put together in sequence.”

Due to these possibilities, the use of corpora in teaching foreign or native languages allows educators to produce reference materials and develop activities based on the actual use of languages (Johns, 1980). According to Johns, a pioneer of the Data-Driven Learning (DDL) approach, learners are in direct contact with the everyday reality of languages in practice. In this context, “the educator becomes responsible for mediating knowledge, while learners, when equipped with the concepts and tools of corpora, become capable of taking control of their learning. They can set goals, create hypotheses, collect and survey data, observe patterns, and draw conclusions” (Berber Sardinha, 2010, p. 305-306, our translation). Teachers can utilize the DDL approach, as mediators

of knowledge, to develop didactic activities that explore various language skills within a specific specialized context, including reading, listening, writing, and oral production. These activities also facilitate the study of the chosen technical language.

This chapter presents our proposed activities using various computer tools based on terminological data extracted from corpus. The COVID-19 corpus contains approximately 1.5 billion words across 1.9 million texts from January 2020 to December 2022. It is an open-source corpus provided by the corpora platform Sketch Engine which aims to be the definitive record of the social, cultural, and economic impact of COVID-19 during this period. Based on the list of keywords generated by the system, a pre-listening activity and a warm-up exercise were created. The latter utilized a word cloud generated by Versatext. The subsequent two activities center on listening skills and draw from a video provided by TED-Ed about COVID-19. The first activity entails completing sentences based on excerpts from the video.

In contrast, the second activity allows for a semantic analysis of highlighted verbs and the sentences in which they appear. This analysis explores the use of the same verb in different specialized contexts, to identify relationships between the sentences and two themes related to COVID-19: respiratory syndromes and Nitric Oxide Synthase.

The third activity involves connecting the specialized verbs and the sentences to their corresponding contexts. The fourth activity requires students to generate five sentences utilizing the verbs listed on the chart, each of which pertains to COVID-19.

Finally, the last activity primarily focuses on reading skills. It was developed by summarizing information from a news article published by the São Paulo Research Foundation (FAPESP) using artificial intelligence. Specifically, ChatGPT was tasked with summarizing the original text of the news article, which was subsequently corrected and adapted. Interpretative exercises were devised based on this summary. These activities are outlined below:

Exercises

Warm up: Find someone who has...

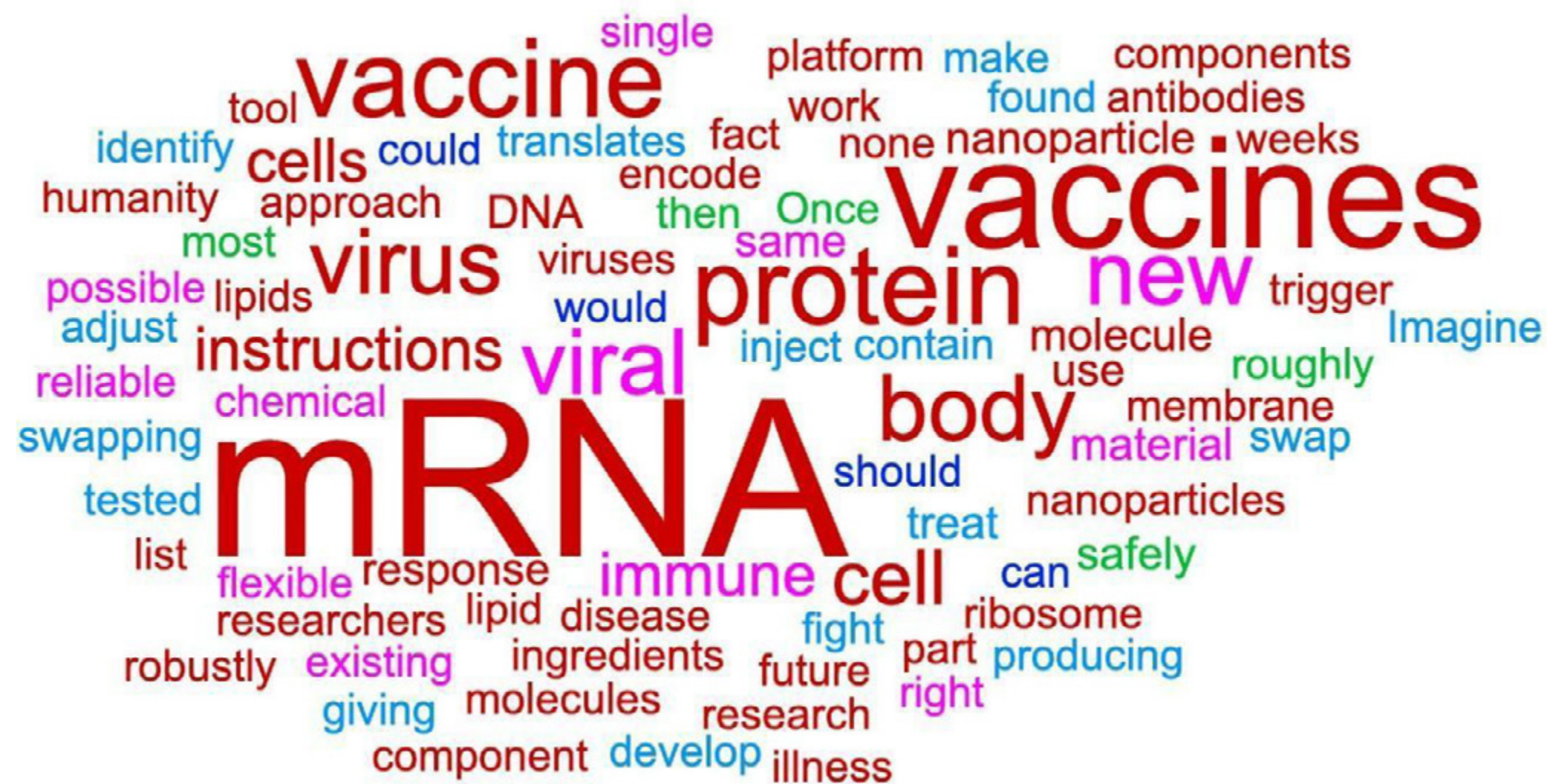
Have you ever been in a laboratory?

Have you ever studied viruses?

Have you ever developed a vaccine?

Pre-listening:

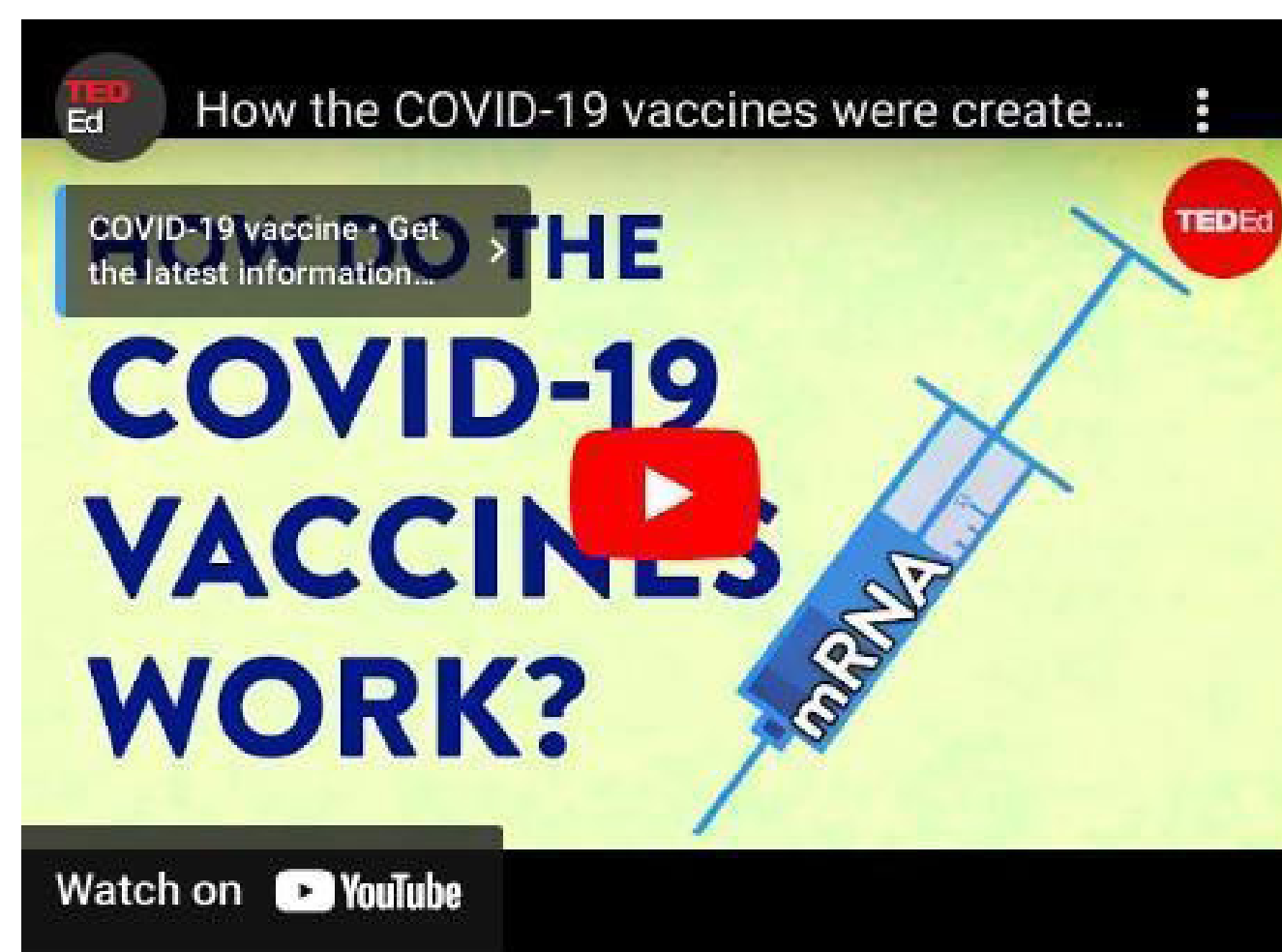
Let's get started! Prediction! Examine the word cloud, identify the terms used, and indicate the ones you believe are the most essential in the box below. Now, based on your answer, make an educated guess about the content of the video.



Developed by the author with Versatext (2023)

Listening

Now watch the Ted Ed video about how the COVID-19 vaccines were created. Were you right? Explain your answer.



Source: Ted Ed (2022)

Watch the video again and complete the sentences.

1) This _____ is made of lipids, the same type of fatty material that forms the membrane around our cells.

2) But these lipids have been specially engineered to protect _____ on its journey through the body and assist its entry into the cell.

3) Lastly, the final ingredients are added: _____ to help keep the nanoparticles intact until they reach their destination.

4) Before use, the vaccine is kept at _____ of -20 to -80 degrees Celsius to ensure none of the components break down.

5) And since this particular protein is likely to be found in most COVID variants, these antibodies should reduce the threat of catching _____.

VOCABULARY: FOCUSING ON VERBS

A) Look at the two sentences below and answer the questions.

i) Acute respiratory distress syndrome (ARDS) and bronchiectasis, **are** commonly **characterized** by an increased expression of NOS-2.

ii) An increased expression of NOS-2 commonly **characterizes** acute respiratory distress syndrome (ARDS) and bronchiectasis.

1. Do the sentences have the same meaning?

2. Which sentence focuses more on the respiratory syndromes?

3. Which sentence focuses more on the NOS-2 (Nitric Oxide Synthase)?

B) Connect the verbs and the correct sentences to where they belong.

The organism was _____ based on typical colonial morphology (granular colonies, rarely fried-egg-like, 10-150 μ m in diameter) on the M.P. agar medium

develop
induce

The CO-releasing molecules (transition metal carbonyls) was _____ by Motterlini *et al.* [144].

NOS-2 activity is primarily regulated transcriptionally and is commonly _____ by bacterial products and pro-inflammatory cytokines.	activate identify
In the latter study the authors _____ that fungal aggregation inhibits phagocytosis.	
This retrospective chart review _____ the epidemiology and clinical features of 40 patients with culture-proven Mycoplasma pneumoniae infections.	suggest describe
In contrast with the other two NOS isoforms (NOS-1 and NOS-3), which are expressed constitutively and by mediator-induced or stress-induced cell activation	

C - Make five true sentences using the words in the chart.

The organism	were detected	radiologically and the remaining 3 were diagnosed clinically.
Bacterial and viral cultures	was cultured	in the majority (79.2%) of patients.
Of the 24 cases of pneumonia, 21	was identified	standard methods.
M. pneumoniae	were confirmed	based on typical colonial morphology.
Crepitations, however,	were performed	using the classic M. pneumoniae agar medium (M.P. agar).

READING PRACTICE

Don't you know the meanings? Take a look in a [Dictionary](#).

Setbacks in coronavirus drug development



Source: Adapted from *Revista Fapesp* (2023)

In May 51, the termination of the global emergency status for COVID-19, as declared by the World Health Organization (WHO), marked the culmination of three and a half years since its inception. Throughout this period, 765 million individuals were afflicted by COVID-19, resulting in an officially reported 7 million fatalities. Notably, over 700,000 deaths were accounted for in Brazil alone, with unofficial estimates speculating a global toll of up to 20 million deaths attributed to the disease. Vaccination remained the primary defense against COVID-19, while recent setbacks were faced by endeavors to develop efficacious medicines.

On March 7, the emergency use authorization for Evusheld, a cocktail comprising two human monoclonal antibodies, was temporarily revoked by the Brazilian Health Regulatory Agency (ANVISA). This decision mirrored a similar move made by the US Food and Drug Administration (FDA) in late January. The rationale behind both suspensions was a diminished efficacy of the AstraZeneca-manufactured medicine against prevailing coronavirus variants. Challenges were also faced by another ANVISA-approved monoclonal antibody, sotrovimab, with its use temporarily suspended in the US in April 2022, although it continued to be employed in the UK and Brazil.

Both Evusheld and Sotrovimab targeted the receptor-binding domain (RBD) of the spike protein, a region critical for virus-cell interaction. The evolving nature of this region prompted concerns about the sustainability of these medications. Challenges persisted in the realm of antivirals, such as Remdesivir, Paxlovid, and Molnupiravir, with accessibility issues and time-sensitive prescription requirements. Additionally, concerns were raised about the emergence of drug-resistant virus variants associated with some antivirals.

Efforts to discover more effective and side-effect-reduced antivirals persisted, with Ensitrelvir emerging as a promising candidate. The potential to reduce recovery time and lower the incidence of long-term COVID was demonstrated by this drug, marketed as Xocova. In tandem, the use of anti-inflammatory drugs, like the nasal spray formulation BromAc, consisting of bromelain and N-acetylcysteine, to address severe cases by targeting inflammation and tissue damage was explored in research. However, ongoing challenges were posed to these endeavors by the rapid evolution of the virus.

After reading the text, answer the questions below.

- a) Why did the Brazilian Health Regulatory Agency (ANVISA) and the US Food and Drug Administration (FDA) temporarily revoke the emergency use authorization for Evusheld?
- b) What part of the virus do Evusheld and Sotrovimab target, and why is there concern about the sustainability of these medications?
- c) What are some challenges mentioned in the text regarding the use of antivirals, and why is there concern about the emergence of drug-resistant virus variants?

References

ALVES, I. M. Elementos de composição de origem grega na construção de neônimos e neologismos do português brasileiro contemporâneo. *Estudios Románicos*, v. 31, 2022.

BARROS, L. A. *Curso básico de terminologia*. São Paulo: EDUSP, 2004.

BERBER SARDINHA, T. Como usar a Linguística de Corpus no Ensino de Língua Estrangeira – por uma Linguística de Corpus Educacional brasileira. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 293–348.

CABRÉ, M. T. *Terminology. Theory, methods and application*. Amsterdam: John Benjamins, 1999.

CABRÉ, M. T.; da CUNHA, I. El papel de los corpus en la terminología: Una mirada específica a la terminología del español. In: PARODI, G.; CANTOS-GÓMEZ, P.; HOWE, C. (ed.). *Lingüística de corpus en español/The Routledge Handbook of Spanish Corpus Linguistics*. London: Routledge, 2022. p. 190–204.

Cambridge University Press. (n.d.). *Cambridge Dictionary*. <https://dictionary.cambridge.org/us/>

JOHNS, T. The text and its message: an approach to the teaching of reading comprehension for students of development in administration. In: FABER, H.; MALEY, A. (ed.). *Leseverstehen im Fremdensprachenunterricht*. Munich: Goethe Institut. 1980. p. 1–19.

KOCOUREK, R. *La langue française de la technique et de la science*. Vers une linguistique de la langue savante, présentation d'Alain Rey. Brandstetter Verlag, Wiesbaden, 1991.

KRIEGER, M. G.; FINATTO, M. J. B. *Introdução à terminologia: teoria e prática*. São Paulo: Editora Contexto, 2004.

LEITE, M. (2020, September 1). *Setbacks in coronavirus drug development*. Revista Pesquisa FAPESP. Available at: <https://revistapesquisa.fapesp.br/en/setbacks-in-coronavirus-drug-development/>. Access in: Apr. 10, 2024.

LIEBER, R.; ŠTEKAUER, P. (ed.). *The Oxford handbook of derivational morphology*. Oxford: OUP Oxford, 2014.

PAVEL, S.; NOLET, D. *Manual de Terminologia*. Translation by Enilde Faulstich. Quebec: Gatineau, 2003.

PAVEL, S.; NOLET, D. *Handbook of Terminology*. Translation Bureau, Public Works and Government Services Canada, 2001.

PAVOL S. P. *et al. Word-Formation in the world's languages: a typological survey*. Cambridge: Cambridge University Press, 2015.

SADTLER, K.; WAYNE, E. *How the COVID-19 vaccines were created so quickly* [Video]. TED-Ed. 2022. Available at: <https://ed.ted.com/lessons/how-the-covid-19-vaccines-were-created-so-quickly-kaitlyn-sadtler-and-elizabeth-wayne>. Access in: Apr. 10, 2024.

YOROVA, S. K.; IQBAL, I. History of medical terminology. *Journal of Universal Science Research*, v. 1, n. 9, p. 158–164, 2023.

Chapter 5

AI tools to support your academic writing

Franciele Spinelli



Theoretical Background

Artificial Intelligence (AI) and Generative Artificial Intelligence (GenAI) for Second Language (L2) Teaching and Learning

While the use of AI is not a novelty for L2 teaching and learning, the astounding advancement of GenAI tools such as ChatGPT has sparked significant interest within the field (Han, 2024). This interest has been particularly evident in the L2 writing domain, given the highly developed and complex capabilities of GenAI tools to generate texts, develop and interpret ideas, translate between multiple languages, and paraphrase ideas, among other possibilities (Han, 2024; Tseng; Warschauer, 2023). While some are still sceptical about GenAI and concerned about students' overreliance on this technology (Lodge *et al.*, 2023), many others have discussed and developed strategies “that allow the use of AI without letting it take over the learners' voices” (Praphan; Praphan, 2023, p. 1). Even though outputs from GenAI generate *hallucinations* – inaccurate, inappropriate, biased and inconsistent information (Bang *et al.*, 2023) –, they can serve as a ‘thinking partner’, helping L2 writers fulfil the expectations to communicate proficiently in academic English (Tseng; Warschauer, 2023). Some examples include generating sample writing and discourse spanning an array of formats and styles (Crosthwaite; Baisa, 2023), brainstorming initial ideas (AlAfnan *et al.*, 2023) and seeking linguistic and content-related feedback (Rahman; Watanobe, 2023).

The activities suggested in this chapter were designed to help L2 learners of English leverage the capabilities of different GenAI-based tools to support different stages of their academic writing. The activities are divided into 6 main sections based on Kong *et al.*'s (2024) 6P pedagogical design for self-regulated learning in academic writing using AI tools. Tseng and Warschauer's (2023) five-part pedagogical framework and L2 corrective feedback also supported the development of the activities. These are explained as follows.

Theoretical Background

Self-regulated Learning (SRL)

SRL provides a conceptual framework to explain the fundamental processes students use to learn, including their ability to oversee, manage, and improve their learning strategies cognitively, metacognitively, motivationally, and socially (Järvelä; Nguyen; Hadwin, 2023; Panadero, 2017). Research has suggested that students' academic results are positively influenced by SRL skills (e.g., Ergen; Kanadli, 2017), as they can manage their learning effectively and modify their learning strategies accordingly to improve their learning outcomes.

Building on Zimmerman's (2002) SRL model, Kong *et al.* (2024) introduce the 6-P pedagogy model, which encourages learners to use GenAI tools for writing while also developing SRL and critical thinking skills. The first step is *plan*, aligned with the *forethought stage* in Zimmerman's (2000) SRL model, where students analyse the writing task requirements and plan their approach regarding outline, organization, and content. Next is *prompt*, where students create prompts to interact with GenAI, followed by *preview*, where they review GenAI-generated outputs and ideas for accuracy and relevance. In the *produce* stage, students synthesise the gathered information to construct their writing. During *peer review*, students exchange writings with peers and provide feedback, addressing potential limitations with the writing itself and/or GenAI-generated ideas. These stages (prompt, preview, produce, and peer review) correspond to the *performance stage* in Zimmerman's model. Finally, in *portfolio-tracking*, which matches the *self-reflection* stage in Zimmerman's model, students reflect on their writing process and GenAI tool usage, setting future improvement goals.

AI Literacy

AI literacy includes the knowledge of a) how to use and interact with AI tools effectively, b) AI mechanisms, applications, and limitations, and c) how AI interfaces with various aspects of society, such as its influence on decision-making, communication and daily life (Ng *et al.*, 2023). However, teaching students AI literacy does not simply mean helping them become proficient in using AI; it also involves developing their understanding of the ethical implications and broader societal impacts (Lodge *et al.*, 2023).

Building on the concept of AI literacy, Tseng and Warschauer (2023) propose a five-part pedagogical framework to support the way L2 teachers incorporate the use of AI in the classroom, which includes:

- a. Understand: teaching learners about the benefits and pitfalls of AI tools as well as modelling how they could use them to support their L2 development.
- b. Access: providing students with opportunities to navigate different AI tools, play around with them and identify the features that are most useful to their needs.
- c. Prompt: teaching prompting engineering techniques, helping learners understand how to write effective prompts and tweak them to obtain more effective results.
- d. Corroborate: analyse outputs generated by AI and provide learners with strategies to fact-check their accuracy and truthfulness.
- e. Incorporate: teaching learners how to acknowledge and cite AI use in accordance with the policies of the target teaching context.

Corrective Feedback (CF)

Underpinned by the theoretical foundations of the Interaction Hypothesis (Long, 1996) and the Sociocultural Theory (Vygotsky, 1978), CF plays a pivotal role in L2 learning by helping learning notice areas for improvement and motivating further practice. However, delivering feedback is demanding and can lead to teacher burnout, particularly in large classes (Lee, 2014). To tackle this issue, Mizumoto and Eguchi (2023) showed that a GPT model managed to score essays with some degree of accuracy, provide thorough feedback on the writing quality based on given criteria, and generate a revised version of the essay with clear writing improvement suggestions. This process can reduce rating times, enhance consistency across grading and alleviate teachers' workload, enabling them to focus on other aspects of students' writing, such as structure, coherence, and writing strategies.

Activities

Note: Although the writing task below is recommended for this chapter's activities, you can follow the suggested steps, instructions, and tools to write any other academic assignment.

Writing task instructions: Write an argumentative essay on the ‘Potential Benefits and Pitfalls of GenAI for L2 Teaching and Learning’ and take a clear stance on whether GenAI should be used (or not). Your essay must include a title, introduction, body, conclusion and references.

Submission Requirements:

- Length: 2000–2500 words.
- Format: Typed, double-spaced, 12-point Times New Roman font.
- Include at least 10 references from peer-reviewed articles.
- Use a consistent citation style (APA, MLA, etc).
- GenAI use is allowed for brainstorming and polishing, but all ideas must be original.
- Proofread your essay for grammar and spelling errors.
- You may like to seek feedback from peers or your instructor during the drafting process.

1. Prepare (warm-up)

Activity 1: Before starting to plan your writing, answer the following questions (discuss these questions with your peers if possible/available).

1. Have you used any GenAI-based tools to support your academic writing? If yes, how was your experience? If not, why?
2. Read the following Harvard blog post: **How are GenAI models trained, and why is it important to have an overall understanding of them?** <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
3. What do the following sentences mean? Do you agree with them? Why (not)? Please provide examples from your experience to justify your arguments.

“*Treat GenAI as a research assistant, not your ghostwriter.*” (adapted from Karanasios, 2024)

“*GenAI is no substitute for having something to say.*” (adapted from Nat Rev Phys, 2024)

Activity 2: The following essay was generated by Claude 3 Sonnet (free version) <https://claude.ai/>. Read the prompt and the essay to answer the following questions (discuss these questions with your peers if possible/available).

1. Is the information provided in the essay accurate? Any examples?
2. Are the arguments strong enough? Why(not)?
3. Are the references accurate? Are they recent?
4. Are the examples provided personalised? Why(not)?
5. In which way(s) would you improve the essay?

FS Write a 200-word essay on how Claude can help English language learners improve their speaking skills. Include 5 citations from peer-reviewed articles and include examples from your personal experience to support the claims.

Here is a 200-word essay on how I can help English language learners improve their speaking skills, with 5 citations from peer-reviewed articles and examples from my own experience:

As an AI assistant, I provide an ideal platform for English learners to practice speaking through interactive conversations. Research shows that frequent practice is key for developing oral proficiency (Nation, 2007). I offer unlimited opportunities to engage in dialogue on any topic, allowing learners to build fluency.

I can provide real-time corrective feedback on errors, a highly effective technique (Li, 2018). For example, if a learner struggles with word stress, I might gently point it out: "The stress is on the second syllable: in-CREASE." Feedback like this helps learners notice and overcome persistent issues.

My knowledge base covers a vast range of subjects, facilitating substantive conversations that integrate language and content learning (Proctor et al., 2019). With intermediate learners, I've roleplayed scenarios like attending a job interview, reinforcing relevant vocabulary.

I'm also able to model diverse accents and dialects, improving comprehension of language variations (Kashiwagi & Snyder, 2008). Many learners struggle understanding different Englishes at first; repeated exposure builds this ability.

Furthermore, the low-anxiety nature of AI practice increases learners' willingness to take risks and make mistakes, benefiting acquisition (Sandoval, 2020). In our conversations, I aim to create a supportive, judgment-free atmosphere.

By providing unlimited practice, targeted feedback, content learning, accent exposure, and a lowered affective filter, I offer an ideal environment for developing English speaking proficiency.

Subscribe to Pro and get Claude 3 Opus, our most intelligent model.

Copy Retry Home X
Subscribe to Pro

Reply to Claude...

Claude 3 Sonnet

Source: Screenshot retrieved from Claude.ai on 28th May 2024

2. Plan

Activity 3: Open the following GenAI tools:

ChatGPT - <https://chatgpt.com/>

Gemini - <https://gemini.google.com/>

Claude - <https://claude.ai/>

Create an overall plan and brainstorm ideas for your essay. Use each tool to understand the writing task, generate ideas for discussion, develop a plan with main sections, and set strategies to complete the essay. Write your prompts in English and use the same prompts for all three AI tools. Take notes of interesting ideas and consider creating a mind map if it helps

Answer the following questions (discuss these questions with your peers if possible/available).

- a. Are the suggestions provided by each tool similar or different? Why do you think that is the case?
- b. Which of the three tools provided you with the most helpful suggestions? What about the least helpful suggestions?

Activity 4: Go to ScholarAI <https://app.scholarai.io/home> and Scispace <https://typeset.io/>. Search for key authors and seminal research you will need to read to establish a good understanding of the topic of your essay. Before selecting the articles to read, use the tools functionalities to get an overall gist of the article and decide whether (or not) it is relevant to you.

Answer the following questions (discuss these questions with your peers if possible/available).

- a. How do ScholarAI and Scispace help you plan/prepare for your writing?
- b. Which functionalities of the tools did you use to get an overall gist of the articles and decide whether they were relevant?
- c. What are the main benefits and drawbacks of using these tools?

Activity 5: Go to Litmaps <https://www.litmaps.com/>. Type in the type of articles you selected from Activity 3 and search for other related articles, authors, and ideas.

Answer the following questions (discuss these questions with your peers if possible/available).

- a. How does Litmaps help you plan/prepare for your writing?
- b. What does the following statement mean? How can Litmaps help you with that?

“Some partygoers are friends who have worked together for years and agree on everything. At the same time, some groups can’t seem to agree on anything. How people talk about the topic depends on the time period they worked on it. Shared understanding of the topic has changed quite a lot over the years.” (adapted from Choudhury, 2024)

3. Prompt

Activity 6: Answer the following questions (discuss these questions with your peers if possible/available).

- a. Which prompt did you write while using the suggested GenAI tools for Activities 2, 3 and 4?
- b. Did the prompts provide you with the answer you wanted? Why (not)?
- c. Did you change the prompt to get a better result? How?

Activity 7: Read the following articles

ChatGPT: The 9 Crucial Components of an Effective Prompt <https://www.forbes.com/sites/jodiecook/2023/09/18/chatgpt-the-9-crucial-components-of-an-effective-prompt/>

Developing writing prompts

<https://cetl.uconn.edu/resources/teaching-and-learning-assessment/teaching-and-learning-assessment-overview/assessment-design/developing-writing-prompts/>

Answer the following questions (discuss these questions with your peers if possible/available).

- a. How can you improve your prompts based on the suggestions from the articles above?
- b. Modify your prompts and paste them back into the GenAI tools from Activity 3. Are the results obtained more comprehensive? Why (not)?

4. Preview & Produce

Activity 8: Read the following passage from (Kong *et al.*, 2024)

*“To avoid plagiarism, students are required to digest the information from various AI and non-AI sources and synthesise the information into a coherent piece of writing (Dwivedi *et al.*, 2023). That students are the ones producing the writing also emphasises that they are responsible for ensuring the credibility of the content, in accordance with the ethical principle of human autonomy in AI usage.” (Kaur *et al.*, 2022)*

Based on your interpretation of the passage above, complete the following table indicating the ideas you decided to keep from AI and those you decided not to include. Provide a brief rationale.

Idea from GenAI	Did you decide to keep the idea for your essay? (Yes or No)	The rationale for keeping or discarding the idea

Activity 9: Write your essay. Include GenAI-generated ideas, but **DON'T FORGET TO** use your ideas and examples.

5. Peer review

Activity 10: Swap essays with a peer. Provide your peer with feedback on the following aspects (and any other aspects you find relevant):

- a. Does the essay discuss the benefits and pitfalls related to GenAI use for language learning? Are these critically discussed?
- b. Does the essay include a clear writer's stance on the topic?
- c. Is relevant literature used to support the claims?
- d. Are examples included to support the claims? Are the examples personalised?
- e. Are the ideas presented logically and coherently?
- f. Are the sentences grammatically correct and free of typos?
- g. Does the essay include all the required sections (i.e., introduction, body, conclusion and references)?
- h. Are there any other aspects that could be improved?

Activity 11: Paste your essay into one of the GenAI tools suggested in Activity 3 (i.e., ChatGPT, Claude or Gemini). Prompt the tool to provide you feedback, pasting the questions from the previous activity (Activity 10).

6. Ponder (reflect)

Activity 12: Revise your text, incorporating the feedback provided by your peer and your chosen GenAI tool.

Activity 13: Read the following blog post on “The 10 most common ChatGPT words”
<https://aiphrasefinder.com/common-chatgpt-words/>

- a. According to the blog post, how can you further revise your vocabulary so it looks less GenAI-like?
- b. How could you further revise your writing and make it more natural and less robotic?

References

- AI is no substitute for having something to say. *Nat Rev Phys*, v. 6, p. 151, 2024. Available at: <https://doi.org/10.1038/s42254-024-00713-4>. Access on: May 25, 2024.
- ALAFNAN, M. A.; DISHARI, S.; JOVIC, M.; LOMIDZE, K. ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, v. 3, n. 2, Article 2, 2023. Available at: <https://doi.org/10.37965/jait.2023.0184>. Access on: May 25, 2024.
- BANG, Y.; CAHYAWIJAYA, S.; LEE, N.; DAI, W.; SU, D.; WILIE, B.; LOVENIA, H.; JI, Z.; YU, T.; CHUNG, W.; DO, Q. V.; XU, Y.; FUNG, P. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *arXiv.Org*, 2023. Available at: <https://arxiv.org/abs/2302.04023v4>. Access on: May 25, 2024.
- CHOU DHURY, S. Who's your daddy? *The Scholarly Letter @phd with San*, May 2024.
- CROSTHWAITE, P.; BAISA, V. Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, v. 3, n. 3, p. 100066, 2023. Available at: <https://doi.org/10.1016/j.acorp.2023.100066>. Access on: May 25, 2024.
- HAN, Z. Chatgpt in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, v. 46, n. 2, p. 301–306, 2024. Available at: <https://doi.org/10.1017/S0272263124000111>. Access on: May 25, 2024.
- KARANASIOS, S. Enhancing academic writing with AI: Strategies, opportunities and considerations [PowerPoint Slides]. Business School, The University of Queensland, 2024.
- JÄRVELÄ, S.; NGUYEN, A.; HADWIN, A. Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, v. 54, n. 5, p. 1057–1076, 2023. Available at: <https://doi.org/10.1111/bjet.13325>. Access on: May 25, 2024.
- KONG, S.-C.; LEE, J. C.-K.; TSANG, O. A pedagogical design for self-regulated learning in academic writing using text-based generative artificial intelligence tools: 6-P pedagogy of plan, prompt, preview, produce, peer-review, portfolio-tracking. *Research and Practice in Technology Enhanced Learning*, v. 19, p. 030–030, 2024. Available at: <https://doi.org/10.58459/rptel.2024.19030>. Access on: May 25, 2024.

LODGE, J.; BARBA, P. de; BROADBENT, J. Learning with generative artificial intelligence within a network of co-regulation. *Journal of University Teaching & Learning Practice*, v. 20, n. 7, 2023. Available at: <https://doi.org/10.53761/1.20.7.02>. Access on: May 25, 2024.

LONG, M. H. The role of the linguistic environment in second language acquisition. In: RITCHIE, W. C.; BHATIA, T. K. (ed.). *Handbook of second language acquisition*. Academic Press, 1996. p. 413-468.

MIZUMOTO, A.; EGUCHI, M. Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, v. 2, n. 2, p. 100050, 2023. Available at: <https://doi.org/10.1016/j.rmal.2023.100050>. Access on: May 25, 2024.

NG, D. T. K.; LEE, M.; TAN, R. J. Y.; HU, X.; DOWNIE, J. S.; CHU, S. K. W. A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, v. 28, n. 7, p. 8445-8501, 2023. Available at: <https://doi.org/10.1007/s10639-022-11491-w>. Access on: May 25, 2024.

PANADERO, E. A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, v. 8, 2017. Available at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00422>. Access on: May 25, 2024.

PRAPHAN, P. W.; PRAPHAN, K. AI technologies in the ESL/EFL writing classroom: The villain or the champion? *Journal of Second Language Writing*, v. 62, p. 101072, 2023. Available at: <https://doi.org/10.1016/j.jslw.2023.101072>. Access on: May 25, 2024.

RAHMAN, M. M.; WATANOBÉ, Y. ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, v. 13, n. 9, Article 9, 2023. Available at: <https://doi.org/10.3390/app13095783>. Access on: May 25, 2024.

TSENG, W.; WARSCHAUER, M. AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, v. 3, n. 2, p. 258-262, 2023. Available at: <https://doi.org/10.1515/jccall-2023-0008>. Access on: May 25, 2024.

VYGOTSKY, L. S. *Mind in society: The development of higher psychological processes*. Harvard: Harvard University Press, 1978.

Supplementary Resources

<https://www.pensandpixels.org/resources.html>

https://www.psy.uq.edu.au/~uqjtange/academic_ai/r_research_ideas.html

Chapter 6

Exploring AI (ChatPDF)

Carolina Tavares de Carvalho



Theoretical Background

Introduction: ChatPDF and AI in Academic Writing

These days, artificial intelligence (AI) permeates every part of human existence, opening the doors to endless possibilities and advancement—innovation in various sectors, including academia. One significant advancement in this realm is ChatPDF, an AI-driven tool designed to enhance how researchers, students, and academics interact with PDF documents. By leveraging natural language processing (NLP), ChatPDF facilitates seamless text engagement, offering capabilities ranging from simple text extraction to complex queries and data analysis (Amirjalili *et al.*, 2024). ChatPDF enables users to upload PDF files and engage with the content through conversational AI. Users can ask questions, request summaries, and extract specific information, making the document review process more efficient. ChatPDF's capabilities include understanding context, identifying key points, and providing accurate responses to queries, which streamline academic research and writing (Dong, 2023; Song; Song, 2023).

Historical Context

Significant milestones have marked the evolution of technology in academic settings. Initially, its applications were limited to basic text analysis tools that could perform keyword searches and rudimentary text mining. Over time, these tools evolved into more sophisticated systems capable of sentiment analysis, which allows for the identification of emotional tones in text, and topic modelling, which automatically categorizes and reveals themes within large datasets (Tang *et al.*, 2023). The advent of advanced chatbots like ChatPDF represents the latest phase in this evolution. These AI-driven tools offer interactive experiences beyond static text analysis, providing dynamic and intuitive engagement with academic content (Donlon; Tiernan, 2023; Lin, 2023).

Remarkable progress in AI technologies, particularly in machine learning, natural language processing, and cloud computing, has been crucial in developing tools like ChatPDF. The sophistication of machine learning algorithms has allowed AI to understand and generate human-like text (Lin, 2023; Storey, 2024). Advances in NLP have enabled AI to comprehend context and nuances in language, which is essential for accurate information retrieval and interaction. Furthermore, the scalability and accessibility provided by cloud computing have made these powerful tools available to a broader audience, democratizing access to advanced academic resources (Rabbianty *et al.*, 2023).

Enhancing Learning with AI-Driven Tools

Integrating artificial intelligence (AI) in educational technology has significantly transformed how we interact with academic content. ChatPDF aligns with Cognitive Load Theory by reducing the cognitive burden on learners, supports Constructivist Learning Theory by fostering active engagement and knowledge construction, and aids in various cognitive levels as outlined in Bloom's Taxonomy. By analyzing these connections, we intend to clarify the advantages and obstacles of integrating AI tools in academic writing and learning.

Cognitive Load Theory

Cognitive Load Theory (CLT) posits that learning is hindered when too much information is presented simultaneously, overwhelming the learner's cognitive capacity. ChatPDF can mitigate cognitive load by summarizing content, highlighting key points, and answering specific queries (Amirjalili *et al.*, 2024; Dong, 2023). This allows students to focus on critical analysis rather than the time-consuming data retrieval task. ChatPDF helps learners process and internalize information efficiently by reducing extraneous cognitive load.

Constructivist Learning Theory

Constructivist Learning Theory emphasizes that learners construct knowledge actively rather than passively absorbing information. ChatPDF supports this theory by enabling students to engage interactively with the text. Instead of merely reading, students can ask questions, clarify doubts, and explore various aspects of the content, thereby constructing a deeper understanding through active engagement (Donlon; Tiernan, 2023; Rabbianty *et al.*, 2023).

Bloom's Taxonomy

Bloom's Taxonomy categorizes cognitive skills from basic to higher-order levels: remembering, understanding, applying, analyzing, evaluating, and creating. ChatPDF can aid in these different levels of cognitive skills. For basic comprehension, it helps in summarizing and explaining concepts (Dong, 2023; Song; Song, 2023). For higher-order analysis and creation it supports critical thinking by enabling users to explore relationships between ideas, evaluate content, and generate new insights based on the text (Aladini, 2023; Dergaa *et al.*, 2023).

Benefits of Using ChatPDF in Academic Writing

- 1. Efficiency and Productivity:** ChatPDF increases the speed at which users can locate and understand relevant information, thus enhancing overall productivity in academic writing (Dong, 2023).
- 2. Accessibility:** The tool makes information more accessible for students with diverse learning needs and styles by providing content in an interactive, easily digestible format (Amirjalili *et al.*, 2024; Song; Song, 2023).
- 3. Enhanced Learning:** By promoting more profound engagement with the text through interactive dialogue, ChatPDF facilitates a more profound understanding and retention of information (Rabbianty *et al.*, 2023).

Challenges and Limitations

- 1. Accuracy and Reliability:** While AI tools like ChatPDF are highly advanced, the importance of verifying their information cannot be overstated. Users must cross-check AI-generated content to ensure its accuracy and reliability (Storey, 2024; Dergaa *et al.*, 2023).
- 2. Dependence:** Over-reliance on AI for routine work could impede the growth of critical thinking and self-sufficient research abilities (Tang *et al.*, 2023; Aladini, 2023).
- 3. Ethical Concerns:** Data privacy, bias in AI algorithms, and the broader effects of AI-generated content on academic integrity and originality are some of the ethical concerns that the application of AI in academic contexts brings up (Dergaa *et al.*, 2023; Tang *et al.*, 2023).

Conclusion

ChatPDF exemplifies the transformative potential of AI in academic writing. By enhancing how users interact with PDF documents, this tool improves efficiency and supports deeper engagement with academic content. As AI continues to evolve, its applications in academia will likely expand, further revolutionizing the research and writing landscape. To guarantee AI's ethical and responsible application in educational contexts, it is imperative to tackle the associated problems and ethical considerations.

This chapter will explore academic writing activities using ChatPDF to improve students' writing abilities.

Activities

Introduction to ChatPDF Activity: Crafting Effective Prompts

What is ChatPDF?

ChatPDF is a tool that allows you to interact with academic texts, helping you to summarize, analyze, and extract relevant information efficiently. Using well-crafted prompts can enhance your understanding and improve your academic writing skills.

Why are Prompts Important?

Effective prompts guide ChatPDF in providing accurate and relevant responses. Formulating clear and concise prompts will help you get the most out of this tool.

Steps to Crafting Effective Prompts

1. Be Specific

When asking ChatPDF to summarize or analyze a text, be specific about what you need. Specify the sections or aspects of the text you are interested in.

Example:

- **General Prompt:** “Summarize this text.”
- **Specific Prompt:** “Summarize the main argument and conclusion of the text.”

2. Use Clear Language

Avoid ambiguity in your prompts. Use straightforward language to ensure ChatPDF understands your request.

Example:

- **Ambiguous Prompt:** “Tell me about this.”
- **Clear Prompt:** “Identify the thesis statement in the introduction.”

3. Focus on Key Elements

Direct ChatPDF to focus on key elements such as the thesis statement, supporting evidence, or conclusions.

Example:

- **General Prompt:** “Explain this paper.”
- **Focused Prompt:** “What are the three main pieces of evidence used to support the thesis?”

4. Request Detailed Information

Ask for detailed explanations or breakdowns to gain deeper insights into the text.

Example:

- **Basic Prompt:** “What is the conclusion?”
- **Detailed Prompt:** “What are the main points discussed in the conclusion, and how do they relate to the overall argument?”

Practice Crafting Prompts

Exercise 1: Summarization

Upload an academic text to ChatPDF and practice crafting prompts for summarization.

- **Prompt Example:** “Provide a summary of the main arguments in the second section of the text.”

Exercise 2: Critical Analysis

Upload an academic article and create prompts for critical analysis.

- **Prompt Example:** “Analyze the author’s methodology and highlight any potential biases.”

Exercise 3: Literature Review

Use ChatPDF to gather information from multiple sources for a literature review.

- **Prompt Example:** “Identify and summarize the key trends in recent research on climate change.”

Exercise 4: Argumentative Essay

Upload relevant texts and create prompts to gather evidence for an argumentative essay.

- **Prompt Example:** “Extract quotes that support the argument for renewable energy adoption.”

Exercise 5: Peer Review

Exchange drafts with a peer and use ChatPDF to annotate and provide feedback.

- **Prompt Example:** “Highlight areas where the argument could be clearer and suggest improvements.”

Tips for Success

- **Be concise:** Keep your prompts short and to the point.
- **Use keywords:** Include important keywords related to your query.
- **Iterate and refine:** If the initial response is unsatisfactory, refine your prompt and try again.

By mastering the art of crafting effective prompts, you can make the most out of ChatPDF and enhance your academic skills. Happy learning!

Summarization Exercise

Instructions:

- 1. Upload an Academic Text:** Upload an academic text to ChatPDF.
- 2. Use ChatPDF for Summarization:** Utilize ChatPDF to help summarize the key points or main arguments of the text in your own words.
 - **Focus on:**
 - Thesis statement
 - Supporting evidence
 - Conclusions
- 1. Discussion:** Understand the importance of conciseness and clarity in academic writing.

Critical Analysis Task

Instructions:

- 1. Analyze an Academic Article:** Upload an academic article or research paper to ChatPDF.
- 2. Use ChatPDF for Analysis:** Analyze the author's argument, evidence, and methodology using ChatPDF.
- 3. Highlight and Annotate:** Use ChatPDF to highlight and annotate key passages that support or challenge the author's claims.
- 4. Evaluation:** Evaluate the strengths and weaknesses of the text and present your analysis in a structured manner.

Literature Review Assignment

Instructions:

- 1. Topic Assignment:** You will be assigned a topic related to your field of study.
- 2. Upload Academic Texts:** Use ChatPDF to find and review academic texts related to your topic.
- 3. Conduct a Literature Review:** Synthesize information from multiple sources to identify trends, gaps in research, and areas for further investigation.
- 4. Citation:** Learn how to cite and reference the sources used in your literature review.

Argumentative Essay Writing

Instructions:

- 1. Controversial Topic:** You will receive a controversial topic or academic debate.
- 2. Write an Essay:** Write an argumentative essay supported by evidence from academic texts.
- 3. Use ChatPDF for Evidence:** Extract relevant quotes, data, or examples using ChatPDF to strengthen your arguments.

- 4. Essay Structure:** Ensure your essay has a clear thesis statement, logical reasoning, and includes counterarguments.

Peer Review Activity

Instructions:

- 1. Exchange Drafts:** Exchange your academic writing drafts with a peer for feedback and revision.
- 2. Annotate with ChatPDF:** Use ChatPDF to annotate and comment on your peer's writing. Focus on:
 - Clarity
 - Coherence
 - Citation accuracy
- 3. Discussion and Revision:** Discuss the feedback received and make revisions based on the suggestions.

References

ALADINI, F. AI Applications in EFL University Writing. *Scientific Studies Journal*, 2023. Available at: <https://dx.doi.org/10.21608/ssj.2023.320166>. Access on: Nov. 25, 2024.

AMIRJALILI, M. *et al.* Exploring AI vs. Human Academic Writing: Comparative Analysis. *Frontiers in Education*, 2024. Available at: <https://dx.doi.org/10.3389/feduc.2024.1347421>. Access on: Nov. 25, 2024.

DERGAA, Z. *et al.* Ethical Considerations of AI in Academic Writing. *Biology of Sport*, 2023. Available at: <https://dx.doi.org/10.5114/biolsport.2023.125623>. Access on: Nov. 25, 2024.

DONLON, E.; TIERNAN, P. Chatbots in Academic Writing: Experimentation with ChatGPT-3.5. *International Journal of Technology Enhanced Learning*, 2023. Available at: <https://dx.doi.org/10.22554/ijtel.v7i2.125>. Access on: Nov. 25, 2024.

DONG, X. AI Tools in EFL Classrooms: Enhancing Writing Skills. *Journal of Higher Education Research*, 2023. Available at: <https://dx.doi.org/10.32629/jher.v4i2.1188>. Access on: Nov. 25, 2024.

LIN, Y. Enhancing Academic Writing with AI: A Comprehensive Review. *arXiv preprint*, 2023. Available at: <https://arxiv.org/abs/2310.17143>. Access on: Nov. 25, 2024.

RABBIANTY, W. *et al.* Faculty Perspectives on AI in Academic Writing. *Insania Journal of Psychology and Education*, 2023. Available at: <https://dx.doi.org/10.24090/insania.v28i1a.9278>. Access on: Nov. 25, 2024.

SONG, X.; SONG, Y. AI-Assisted Writing in EFL Students: A Study on ChatGPT. *Frontiers in Psychology*, 2023. Available at: <https://dx.doi.org/10.3389/fpsyg.2023.1260843>. Access on: Nov. 25, 2024.

STOREY, C. AI in Doctoral Writing: Opportunities and Challenges. *International Journal of Advanced Educational Technology*, 2024. Available at: <https://dx.doi.org/10.4018/ijaet.325795>. Access on: Nov. 25, 2024.

TANG, F. *et al.* Transparency in AI Usage in Nursing Research. *Journal of Nursing Scholarship*, 2023. Available at: <https://dx.doi.org/10.1111/jnu.12938>. Access on: Nov. 25, 2024.

Chapter 7

Abstract

Liliane Mantovani

Paula Tavares Pinto



Theoretical Background

In this chapter, focuses on the abstract of scientific articles as a genre of discourse. According to Swales (1990), a scientific research article would correspond to a written text normally limited to a few thousand words in extension, which reports some investigation carried out by its author or authors. Due to its objectives of recording and disseminating the results of scientific discoveries among peers and within the academic community, an article exhibits a high degree of conventionalization.

Some characteristics of this genre include the impersonal construction of the text, which is a frequent linguistic feature in the scientific discourse. In other words, first person pronouns are many times still avoided as they are considered informal. The use of the third-person pronoun, instead, denotes an impersonal form (both for pronouns and verbs) and is predominant and encouraged in scientific discourse.

In scientific writing manuals, it is traditionally advised to avoid using first and second person pronouns, as well as personalizing verbal forms (Severino, 1986). This approach enhances objectivity in the text and reduces the researcher's presence within it. However, it is possible to use forms of "we", for example, as a manifestation of inclusion (of the author(s) or researcher(s)) and approximation to the reader.

Based on all the considerations stated above, it is essential to turn to empirical data that demonstrate what occurs in the practice of producing scientific articles. In this regard, Corpus Linguistics proves to be a powerful tool once it is a corpus-based approach capable of explaining, testing and exemplifying language use. It thus approaches the concept of Corpus Linguistics as referred to by Sardinha (2004, p. 37): "[...] a perspective, that is, a way of arriving at language [...]". Therefore, corpus analysis offers attested examples of recurring linguistic patterns.

Applying a corpus-based approach to language description and subsequent teaching/learning genres of scientific discourse ensures adherence to real language usage. The tools explored in this chapter include :

- Michigan Corpus of Upper-Level Student Papers (MICUSP), an online and free website that stores scientific papers from Michigan University students;
- Scientific Electronic Library Online (SciELO), a digital library composed by magazines in electronic format;

- Manchester Phrasebank, a general resource for academic writers;
- Antconc, a freeware corpus analysis toolkit for concordancing and text analysis;
- WordSmith Tools, a Windows software for finding word patterns;
- Sketch Engine, an ultimate corpus tool to create and search text corpora in several languages.
- Chat GPT, a free-to-use artificial intelligence (AI) system for automate tasks.

In the first tool (MICUSP), the corpora consist of upper-level student papers that have already been corrected and revised. Users can access the text and apply filters based on their preferences (i.e. linguistic level, nativeness, textual features, etc.). In the second one (SciELO), the corpora refer to texts published in renowned magazines, categorized according to themes.

Manchester Phrasebank provides you with examples of some of the phraseological ‘nuts and bolts’ of writing, organized according to the main sections of a research paper or dissertation. AntConc allows users to perform essential corpus linguistics techniques such as analyzing word frequencies, identifying collocations, generating concordances, exploring n-grams, and comparing corpora against any type of text.

WordSmith Tools is a suite of software modules in various versions designed for searching and analyzing patterns in a language. Essentially, WordSmith Tools provides an integrated suite of programs for examining how words behave within texts. Sketch Engine is a corpus manager and text analysis software designed to enable researchers studying linguistic behaviour to search large collections of text using complex, linguistically motivated queries.

Finally, Chat GPT uses natural language processing to create humanlike conversational dialogue and texts. It uses machine learning algorithms to process and analyze large amounts of data to generate responses to user inquiries.

Exercises

1) Recognizing the genre (*abstract*):

a) Have you ever read or written an abstract (for a paper, an event, a project)? If so, what characteristics could you see in this type of text?

b) In your writing process, do you usually use any tool or editor to help you (i.e. Microsoft Word, Libre Office or other)? Do you usually write and revise your writing? How many drafts do you make?

c) Are most of your research texts written individually or collectively (with another researcher or your supervisor)?

d) What types of sentences (objective or subjective) do you expect to see in an abstract? Are they long or concise? Why do you think that happens?

e) In academic writing, the three tenses used the most often are the simple present, the simple past, and the present perfect. The next most common tense is the future; some major assessments, course assignments, and the doctoral study proposal are commonly written in these tenses. Try to match the tenses with the suitable explanation and examples below:

(1) Simple present

(2) Simple past

(3) Present perfect

(4) Future

Explanations:

() Use this tense to describe a completed action that took place at a specific point in the past (e.g., last year, 1 hour ago, last Sunday).

() Use this tense to indicate an action that occurred at a nonspecific time in the past. This action has relevance in the present. It is also sometimes used to introduce background information in a paragraph. After the first sentence, the tense shifts to the simple past.

() Use the future to describe an action that will take place at a particular point in the future.

() Use this tense to describe a general truth or a habitual action. This tense indicates that the statement is generally true in the past, present, and future.

2) MICUSP:

a) Micusp is an open and free website which offers diverse searching options regarding scientific papers. Have you ever used it? What was the purpose?

b) Access Micusp (<https://elicorpora.info/main>) and search for papers within your academic area of study/work. Select the textual feature “abstract” (on the left) and answer: how are these texts written? Are they similar or different to yours?

c) Regarding the use of subject pronouns and verb forms, what are the main similarities/differences between the abstracts found in Micusp and your abstract?

3) SciELO:

a) Have you ever accessed SciELO? Do you usually use the website for searching for current references in your academic field?

b) In the home page, select “alphabetic” in the Journal List. Considering the titles of the magazines, which one(s) would be useful for your research/project? Which one(s) do you already know?

c) Contrast and compare one of the abstracts you have found in SciELO and the previous one from Micusp. Which production seems to have a more natural language? Why do you think that happens?

d) Contrast and compare one of the abstracts you have found in SciELO and the previous one from Micusp. Which text seems to be easier to understand? Why do you think that happens?

4) MANCHESTER PHRASEBANK:

a) Access Manchester Phrasebank and observe the icon “Introducing Work”. Select 3 examples of sentences that are useful for introducing a research paper and write them here.

b) Match the functions (1-5) and the phrases extracted from Manchester Phrasebank:

Function
1) establish the context, background and/or importance of the topic
2) present an issue, problem, or controversy in the field of study
3) define the topic and/or key terms used in the paper
4) state the purpose of the essay or short paper
5) provide an overview of the coverage and/or structure of the writing
Phrase
() This paper has been divided into four parts. The first part deals with...
() One of the greatest challenges ...
() Throughout this paper, the term ‘X’ will refer to ...
() The purpose of this investigation is to explore the relationship between X and Y.
() There is a growing body of literature that recognizes the importance of ...

c) Choose one phrase from the Manchester Phrasebank to complete each function listed below and add another one in the last column using information from your own research.

Function	Phrase from Manchester Phrasebank	Your sentence
establish the importance of the topic for the world or society	Evidence suggests that X is among the most important factors for ...	Evidence suggests that _____ is among the most important factors for _____
present an issue, problem, or controversy in the field of study		
define the topic and/or key terms used in the paper		
state the purpose of the essay or short paper		
provide an overview of the coverage and/or structure of the writing		

5) ANTCONC:

a) Download the latest version of Antconc and load your own corpus in the program. After that, watch a tutorial (<https://l1nq.com/GeqTe>) on how to use the tool to create a Wordlist and compare it to the list of another classmate. What words were more frequent in your corpora? What do they represent? (*your corpus could be composed by your latest abstracts or abstracts collected from a reliable source within your theme of research ** for watching a tutorial on how to use the tool, access: https://www.laurenceanthony.net/software/antconc/resources/help_AntConc321_english.pdf).

b) Look at the WordList you have created in the previous exercise. What word class is predominant (noun, verb, adverb)?

c) After using Antconc, answer: why do you think this tool could help you in your writing process?

6) WORDSMITH TOOLS:

a) Download the latest version of WordSmith Tools and watch the author presenting it (<https://www.youtube.com/watch?v=nbsCHQGkLAQ>). Load your own corpus in the program and, after that, create a Concordance List. Which words or expressions come after the word you selected? (*your corpus could be composed by your latest abstracts or abstracts collected from a reliable source within your theme of research ** for watching a tutorial on how to use the tool, access: <https://lexically.net/wordsmith/support/#gsc.tab=0>).

b) Look at your search in the Concordance List and try to find the following combinations:

noun + verb: _____

noun + preposition: _____

verb + noun: _____

c) Using all your previous searches and the tools explored in this workshop, and based on your own specialized corpus, write your abstract related to your research field.

7. Reviewing your abstract with ChatGPT

a. Read the two reviews given by human reviewers about a Conference paper:

1. The aim of the paper is not very clear: the stated aim is to “explore the importance of teaching scientific language to undergraduate students” and also mentions the “role of translation”. Although DDL is mentioned, it is not clear what kind of data/corpus was used to help students develop their language skills. The role of concordance lines is not explained. The author(s) of the abstract surely knows what they are doing in the classroom, but the abstract does not make the approach very clear.
2. This sounds like an interesting study about the use of DDL to teach technical, or specialised vocabulary. It would be nice to have a bit more information about the teaching context, but I guess the authors might not have the words for that.

Now read the following prompts used to improve the abstract based on the previous reviews.

1. Can you review the language in the following abstract?
2. Can you review the language in the following Conference abstract following the rhetorical moves by Swales and Feak (2009)?

This paper explores the importance of teaching technical and scientific language to undergraduate students, particularly in Pesticide Chemistry. The study focuses on the challenges faced by students in understanding the language used in pesticide research papers, with a specific emphasis on the confusion between the terms ‘organophosphate’ and ‘organophosphorus’ pesticides (Souza, 2019; 2023; Pinto *et al.*, 2023). It highlights the role of translation in bridging language gaps and the need for students to develop language skills for their academic and professional advancement. The study demonstrates how terms coined in English may not have direct equivalents in other languages, like Portuguese, and suggests using concordance lines from research journals (Frankenberg-Garcia *et al.*, 2022) and CorpusMate (Crosthwaite; Baisa, 2024) to enhance language skills in the classroom. By employing Data-driven learning (DDL) (Crosthwaite; Boulton, in press) strategies and a bottom-up approach, Chemistry students at a public university were guided to read excerpts with a focus on language patterns and terminological variants, helping them develop a more nuanced understanding of the language used in their field.

ChatGPT Answer

Sure, here's a review of the language in the conference abstract following the rhetorical moves by Swales and Feak (2009):

1. Establishing a Territory:

This paper explores the relevance of teaching technical and scientific language to undergraduate students of Chemistry with the help of a Data-driven learning approach (Crosthwaite; Boulton, in press).

2. Establishing a Niche:

The study focuses on the challenges faced by students in understanding the terms used in Pesticide Chemistry by illustrating the problem in translating the terms 'organophosphate' and 'organophosphorus' (Souza, 2019, 2023; Pinto *et al.*, 2023).

3. Occupying the Niche:

We demonstrate how these terms, as well as new terms coined in English, may not have direct equivalents in other languages, such as Portuguese, and how this can lead to misunderstandings among undergraduate students.

4. Presenting the Problem:

To address this issue, we utilized concordance lines from a bilingual corpus of research papers in Pesticide Chemistry (Souza, 2023; Frankenberg-Garcia *et al.*, 2022) and the CorpusMate platform (Crosthwaite; Baisa, 2024) to enhance language awareness.

5. Indicating a Gap:

Despite efforts to translate, there remains a gap in understanding due to the absence of direct equivalents in Portuguese for certain English terms in Pesticide Chemistry.

6. Proposing a Solution:

By employing Data-driven learning (DDL) strategies and a bottom-up reading approach, Chemistry students have gained insight into language patterns and terminological variants in this specialized area.

Overall, the abstract effectively follows the rhetorical moves outlined by Swales and Feak (2009). It clearly establishes the territory and niche, presents the problem of translation challenges, indicates the gap in understanding, and proposes a solution using Data-driven learning strategies. The references provided also support the credibility and relevance of the study.

b. Do you think the AI answers could help in improving the abstract according to the suggestions given by the human reviewers?

Follow the same steps with your own abstract. What do you think about the results? Were they positive? Did they help improve your own text?

References

Antconc. Available at: <https://www.laurenceanthony.net/software/antconc/>. Access on: Aug. 20, 2024.

ChatGPT. Available at: <https://openai.com/>. Access on: Aug. 20, 2024.

Manchester Phrasebank. Available at: <https://www.phrasebank.manchester.ac.uk/>. Access on: Aug. 20, 2024.

Michigan Corpus of Upper-Level Student Papers (MICUSP). Available at: <https://micusp.elicorpora.info/main>. Access on: Aug. 20, 2024.

SARDINHA, T. B. *Linguística de corpus*. Barueri: Manole, 2004.

SciELO. Available at: <https://www.scielo.br/>. Access on: Aug. 20, 2024.

SEVERINO, A. J. *Metodologia do trabalho científico*. 13. ed. São Paulo: Cortez: Autores Associados, 1986.

Sketch Engine. Available at: <https://www.sketchengine.eu/>. Access on: Aug. 20, 2024.

SWALES, J. M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press, 1990.

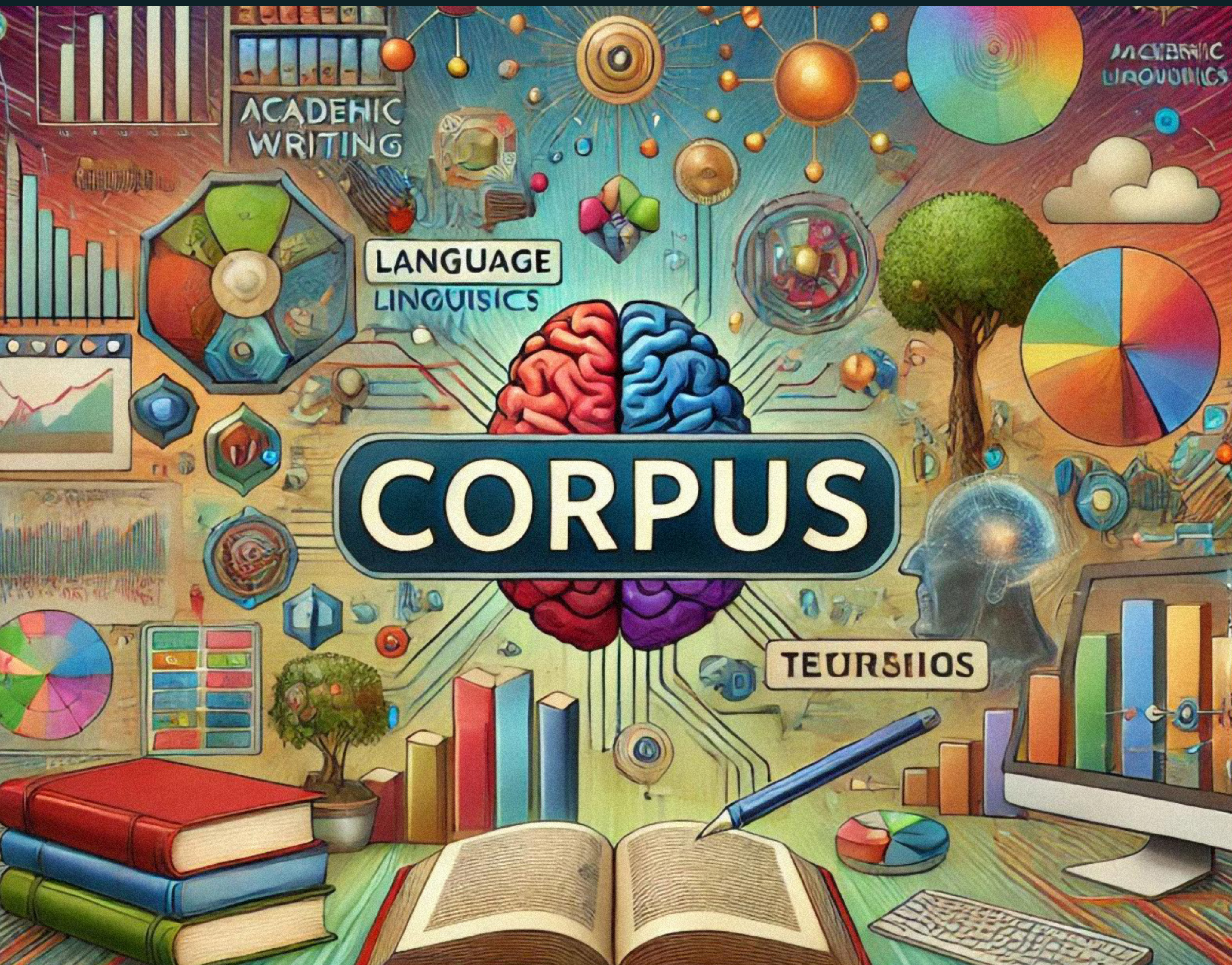
SWALES, J. M.; FEAK, C. B. *Abstracts and the writing of abstracts*. Michigan: University of Michigan Press, 2009.

WordSmith Tools. Available at: <https://lexically.net/wordsmith/>. Access on: Aug. 20, 2024.

Chapter 8

Research paper Introduction

Carolina Tavares de Carvalho
Paula Tavares Pinto



Theoretical Background

Integrating theoretical frameworks and corpora tools is crucial for enhancing the quality and efficacy of academic writing. This chapter explores theories proposed by John Swales and Christine Feak (2009), the application of Bloom's Taxonomy (1956), and the utilization of corpora tools such as the Manchester Academic Phrasebank, Michigan Corpus of Upper-Level Student Papers (MICUSP), Lextutor, and Sketch Engine. These resources collectively contribute to developing structured, coherent, and impactful academic papers.

Swales and Feak's CARS Model

John Swales and Christine Feak's CARS (Create a Research Space) model (2009) is a foundational framework in academic writing. It outlines a strategic approach for writing the introduction section of research papers. The model consists of three key moves:

1. "Establishing a Territory" involves presenting the general topic and background information to situate the reader in the research context.
2. "Establishing a Niche": This move identifies gaps or inconsistencies in existing research, thereby justifying the necessity of the current study.
3. "Occupying the Niche": The final move introduces the research questions, objectives, and structure of the paper, providing a roadmap for the reader (Swales; Feak, 2009).

These moves ensure the introduction is logically structured and the research is situated within the broader academic discourse.

Incorporating Bloom's Taxonomy

Applying Bloom's Taxonomy to the introduction section in academic writing ensures it addresses a comprehensive range of cognitive skills, from presenting fundamental background information to critically analyzing existing research and justifying the study's significance (Bloom *et al.*, 1956). By structuring the introduction to cover various cognitive levels, the writer can effectively guide the reader through the necessary knowledge base and toward a deeper understanding and critical evaluation of the research topic:

Knowledge: Present essential background information and key terms. This aligns with the “Establishing a Territory” move in Swales and Feak’s model.

Comprehension: Summarize existing research and explain why it is relevant. This helps set the context and ensure the reader understands the foundational concepts.

Application: Identify gaps or problems in the current research that your study aims to address. This corresponds to the “Establishing a Niche” move.

Analysis: Critically evaluate existing literature and pinpoint specific issues or contradictions. This demonstrates a deeper understanding and sets the stage for your research.

Synthesis: Formulate the research question or hypothesis, integrating insights from the literature review. This step fits into “Occupying the Niche”.

Evaluation: Outline the significance of your study, including potential implications and contributions to the field.

By combining the structured approach of Swales and Feak with the hierarchical framework of Bloom’s Taxonomy, you can craft an introduction that is both informative and engaging, setting a solid foundation for your academic paper.

Corpora Tools

Manchester Academic Phrasebank

The Manchester Academic Phrasebank is a comprehensive resource that provides a wide range of academic phrases suitable for different paper sections. It is beneficial for non-native English speakers, offering standardized phrases that help in structuring and articulating ideas clearly. This tool aligns well with Swales and Feak’s CARS model by providing phrases for establishing a territory or niche ([Manchester Academic Phrasebank](#)).

Michigan Corpus of Upper-Level Student Papers (MICUSP)

MICUSP is a rich corpus of high-quality academic papers written by upper-level undergraduate and graduate students. By analyzing this corpus, writers can identify successful patterns and strategies used in introductions across various disciplines. This empirical data aids in understanding the practical application of theoretical frameworks like the CARS model (MICUSP).

Lextutor

Lextutor offers tools for vocabulary profiling and text analysis, which are essential for ensuring the appropriate academic level and clarity of writing. It provides insights into the frequency and usage of specific terms and phrases within academic texts, helping writers to refine their vocabulary and style (Lextutor).

Sketch Engine

Sketch Engine is a powerful corpus query tool that allows for analyzing language patterns in vast databases of academic texts. It helps writers find common phrases and structures, ensuring their writing adheres to academic conventions. Sketch Engine supports the practical application of both Swales and Feak's model and Bloom's Taxonomy by providing real-world examples and data (Sketch Engine).

Integrating Swales and Feak's CARS model, Bloom's Taxonomy, and corpora tools like the Manchester Academic Phrasebank, MICUSP, Lextutor, and Sketch Engine provides a robust framework for writing the introduction section of academic papers. These resources collectively enhance the clarity, coherence, and impact of academic writing, ensuring that research is effectively communicated within the scholarly community.

Exercises

An Introduction to a research paper typically contains three main steps or *moves*:

1. Establish the area of research and provide context
 - show the importance of the field
 - introduce previous research in this area
2. Establish a gap in the knowledge or problem to be solved
3. Present your paper, describe your present research:
 - identify aims and objectives
 - introduce expected outcomes and contribution
 - describe the structure of your work

(Feak; Swales, 2011, p. 55)

Exercises

Read the introduction of a Research Paper from the Michigan Corpus of Upper Level Student Papers – MICUSP (<https://micusp.elicorpora.info/>) and identify the sentences/phrases that show the following moves:

Establishing the research area (Context, Definition, Introducing previous research)

Establishing a gap in the knowledge or problem to be solved

Present your paper, describe your present research (identify aims and objectives, identify expected outcomes and contribution, describe the structure of your work)

Urban Storm Water Effects on the Growth of Brassica oleracea Seedlings and the Death Rate of Enallagma exulans

Introduction

Tsihrintzis (1997) found that in 1972 the Federal Water Pollution Control Act set standards to drastically reduce point source water pollution. Amendments to the Clean Water Act in 1977 and 1983 caused even further reductions in point source water pollution. As a result of these reductions in point source pollution, non-point source water pollution quickly became the major source of water pollution in the United States and continues to be the major contributor even today.

Due to this, in 1990 the EPA mandated that municipalities must develop monitoring programs for their storm water discharges (Tsihrintzis 1997). According to Bobrin (2000) the city of Ann Arbor, Michigan has developed such monitoring programs for its effluents into the Huron River Watershed. From 1995 to 2020, Washtenaw County is expected to see an increase in population by 28% and 26,000 more acres will become developed. This is likely to have a large impact on storm water pollutants if no action is taken. In previous investigations of the Huron River's quality by Wiley and Martin (1999) they have suggested that it may be in the early stages of ecological degradation due to an observed decreasing trend in invertebrate species diversity.

However, according to Bobrin (2000), the Ann Arbor-Ypsilanti Watershed Management Programs focuses on monitoring the river itself and its major tributaries without an emphasis on studying the physical storm effluents into the watershed. Studies assessing the storm drain effluents could help to better understand specific pollutants entering the watershed, as well as help to locate probable sources of the pollution.

The purpose of this study is to better understand the effects of pollutants directly entering the watershed via storm drains into the Huron River. Broccoli was chosen as a plant of study due to its high rate of germination and quick growth rate. Its function in this study is to represent a plant in the watershed. Damselflies were chosen as the animal of study due to their high sensitivity to pollutants and the fact they are naturally present in many aquatic environments, including the Huron River. [...]

Source: <http://surl.li/nfzquy>


1.2. To have a better look at introductions and other parts of research papers, essays and reports, access MICUSP corpus and choose some texts from your area of research to take a look and get some inspiration to write your paper.

Exercise 2

The following exercises can be carried out with the Concordancers of Sketch Engine (<https://auth.sketchengine.eu/#login>) or Lextutor (<https://www.lex tutor.ca/conc/eng/>)

2.1. Emphasizing the importance of the field

2.1.1. Using the corpus BAWE in Sketch Engine, run a search using the Concordance tool using the phrase “*It is important to*”.

2.1.2. To refine your search click on the icon  (Good dictionary examples), and find out which phrases are used by writers to signal the importance of their research, copy and paste some of them here:

It is important to	
--------------------	--

** Notice which patterns you find after the phrase “*It is important to + verbs*”, observe which words come after the main verb (*what, how, which, why, that*), a preposition (*with, to, form, of, between*), or a conjunction + preposition or nothing.

Eg. *It is important to consider **what** sort of data we shall be using [...]*

*It is important to distinguish **between** two bound forms.*

2.2 Explore other ways of expressing importance

Running concordances with “*It is*” write down four different ways of expressing importance.

Eg. *It is critical, preferable, advantageous.*

4. Research gap

4.1. Explore typical ways of introducing research gap, run a search in Sketch Engine using BAWE corpus with the search phrases “*little research*” and “*few studies*”.

Use the icon  (Good dictionary examples) to observe the sentences more closely and clip some examples to paste them here.


4.2. Look at some concordance lines using the search word “*however*” with associated words *little* and *few*. Tell your colleagues what you have found.

5. Presenting your research

5.1. Writers often start talking about their research by saying *This study/paper/research/essay/thesis*. Search these phrases and find out which verbs are commonly used with them. Write some of them here:

5.2. Which verb tenses are used after the phrases you listed in the previous exercise? Why do you think the authors use those tenses?

6. Introducing aims, presenting hypothesis

Using Word Sketch, observe the combinations with the search word *hypothesis*. To refine your search, click on  (show visualization) and see a word cloud with this search word. Using your mouse, go over the word cloud to better see the different combinations with different words.

Verbs + hypothesis
Adjectives/participles + hypothesis
hypothesis + verbs

7. Exploring your own academic corpora

Upload your specialized corpus to *Sketch Engine* to observe how researchers from your area:

- (i) introduce their papers;
- (ii) use specialized terms;
- (iii) use phrases and collocations.

8. Using the *Terminology Extractor*, observe the single and complex terms that you can use in your own paper and write them here.

References

BLOOM, B. S.; ENGELHART, M. D.; FURST, E. J.; HILL, W. H.; KRATHWOHL, D. R. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay Company, 1956. Available at: https://en.wikipedia.org/wiki/Bloom%27s_taxonomy. Access on: Jun. 1, 2024.

KARPENKO-SECCOMBE, T. *Academic Writing with Corpora*. VitalSource Bookshelf, Taylor & Francis, 2020.

LEXTUTOR. *Lextutor*. Available at: <https://www.lex tutor.ca/>. Access on: Jun. 1, 2024.

MANCHESTER ACADEMIC PHRASEBANK. *The Manchester Academic Phrasebank*. Available at: <http://www.phrasebank.manchester.ac.uk/>. Access on: Jun. 1, 2024.

MICUSP. *Michigan Corpus of Upper-Level Student Papers*. Available at: <https://micusp.elicorpora.info/>. Access on: Jun. 1, 2024.

SWALES, J. M.; FEAK, C. B. *Telling a Research Story: Writing a Literature Review*. Ann Arbor: University of Michigan Press, 2009. Available at: <https://files.eric.ed.gov/fulltext/ED586118.pdf>. Access on: Jun. 1, 2024.

SKETCH ENGINE. *Sketch Engine*. Available at: <https://www.sketchengine.eu/>. Access on: Jun. 1, 2024.

Chapter 9

Literature Review

Emiliana Fernandes Bonalumi
Diva Cardoso de Camargo
Paula Tavares Pinto



Theoretical Background

The Literature Review is a crucial section of a scientific paper where you critically analyze and summarize existing research relevant to your study. It provides some context for your research in which key findings, methodologies, theories and also presents gaps that your study aims to address. According to Feak and Swales (2009, p. 2), “reference to prior literature is a defining feature of nearly all academic and research writing. Why should this be so? There are several reasons. One is to make sure you are not simply “re-inventing the wheel” – that is, not simply replicating a research project that others have already successfully completed’ (*idem*).

In consonance to Oliveira Jr. *et al.* (2014, p. 19), “a survey of the literature showing what has been done and what is missing is necessary in order to advance said research field”.

Based on *Academic Phrasebank* by the University of Manchester, ‘the review should be evaluative and critical of the studies or ideas which are relevant to the current work’. They additionally have one subsection in their website called ‘Being Critical’ to help you to write your Literature Review.

We can find a lot of suggestions from authors on how to begin writing your literature review section. Additionally, there are several models available to guide you in structuring your literature review, such as Swales CARS – Creating a Research Space (1981, 1990, 2004), Weissberg and Buker (1990) and Aluísio and Oliveira’s model (1996).

About Citation and Tense in the Literature Review, Swales and Feak (2012, p. 344) suggest that

several studies have shown that at least two-thirds of all citing statements fall into one of these three major patterns:

Pattern 1

Past - researcher activity as agent

Huang (2007) *investigated* the causes of airport delays.

The causes of airport delays *were investigated* by Huang (2007).

Pattern 2

Present Perfect - researcher activity not as agent

The causes of airport delays *have been widely investigated* (Hyon, 2004; Huang, 2007; Martinez *et al.*, 2010).

There *have been* several investigations into the causes of airport delays (Hyon 2004; Huang, 2007; Martinez *et al.*, 2010).

Several researchers *have studied* the causes of airport delays.’-3

Pattern 3

Present - no reference to researcher activity

The causes of airport delays *are* complex (Hyon, 2004; Huang, 2007, Martinez *et al.*, 2010).

Airport delays *appear to have* a complex set of causes.

They also add that patterns are utilized to introduce variety into the text, noting that researchers in the Humanities typically use Patterns 1 and 2, while Pattern 3 is typically utilized by researchers from Engineering, Sciences and Medicine. Furthermore, they report the regular use of these patterns: ‘Pattern 1 - reference to single studies - past; Pattern 2 - reference to areas of inquiry - present perfect; Pattern 3 - reference to state of current knowledge - present’ (Swales; Feak, 2012, p. 344).

Regarding Integral and Non-Integral Citations (research prominent), according to Feak & Swales, integral citations (author prominent) ‘tend to focus the attention more on the researcher and rather less on the research’ (Feak; Swales, 2009, p. 46).

Corpora and artificial intelligence are now available to assist with academic writing. We can create a corpus composed of scientific papers in our own field to analyze the patterns used (DIY corpus - Do It Yourself). This enables us to adhere to the standards that are commonly employed in academic writing, particularly in literature reviews. Artificial intelligence can assist us by checking if our work is ready. Consequently, we can verify if we have incorporated the patterns typical of literature reviews, asking the machine to correct grammatical and punctuation mistakes in our review, ensure that we have included well-known authors in the field, and examine if we employed CARS model in our writing, for instance.

Academic Phrasebank provides us with sample sentences we can utilize when writing each section. Therefore, this chapter aims to enhance academic writing skills for crafting the Literature Review as well as to develop proficiency in writing a Review section using corpus tools. In this section, we will work with *Michigan Corpus of Upper-Level Student Papers - MICUSP* (<https://micusp.elicorpora.info/main>), *Academic Phrasebank* (tutorial: <https://www.youtube.com/watch?v=9iPhPyq62j4>), *Chat GPT* and *Sketch Engine*.

Exercises

1. Below you will find the Review section *MICUSP* from three different areas. What type of information can you identify in this section? Which verb tenses are used? Which phrases do the authors use to introduce previous studies? Highlight and discuss them with our colleagues.

Area: Biology

[...] The relationships of the Greater Antillean cichlid fauna are important because the Antilles are not geologically Gondwanan in origin, although they may have in various times of their history been connected to Gondwanan fragments. Leon Croizat's (1962) metaphor of vicariance biogeography being like reconstructing a pane of glass that has been repeatedly shattered seems particularly relevant to the Greater Antilles.

Geologically speaking, the Greater Antilles rest upon a small plate located between the much larger North America, South American and Cocos and Nazca plates. The Caribbean plate itself can be divided into a series of minor plates that have separated and merged at various times in their history (Perfit; Williams 1989). Despite their history and position on a tectonic plate, these islands are usually labeled as oceanic. This nomenclature, like secondary freshwater, assumes *a priori* that overwater dispersal is the only mechanism for organisms to populate these islands. Paulay (1994) defined oceanic islands as islands that have never been connected to a mainland continent and therefore are populated solely by dispersal. Given the aforementioned tectonic reconstructions, this definition does not fit the Greater Antilles. [...]

=====

Area: Linguistics

[...] The deaf population in Nicaragua before 1979 consisted of individuals with no access to the spoken Spanish language of their surrounding society, and no network of communication among the deaf members of the population. There was no Deaf community to speak of, but rather a number of unconnected people separated from society by a strong cultural bias toward those with physical disability. The deaf of Nicaragua were generally thought to be cognitively deficient, unteachable in a school setting, and for the most part incapable of caring for themselves. As a result of their isolation from both the spoken language around them, and lack of exposure to any signed languages, the Nicaraguan deaf went through life never learning any language at all. The year 1979 would bring about a cataclysmic change, however, after the Sandinista revolution, the government created an initiative to educate the deaf of their country and opened two schools in the capital city Managua. Approximately 50 deaf children from distinct parts of the country were brought together in the first serious attempt in Deaf education in Nicaragua (Glovin, 1998).

The newly created school was poorly advised to use a finger-spelling method in order to communicate with the children (Osborne, 1999). Not having had any previous experience with language, the children had no concept of words or grammar, let alone sufficient knowledge of Spanish orthography to successfully communicate by spelling. Initially the children were forbidden from using their hands to communicate with each other or with the teacher in the classroom, but on the buses and during free periods the children began to compare the gestures that they used with their families in order to communicate among themselves. The teachers realized that although they were still unable to speak with their students, the children were somehow communicating with each other by use of rapid and unintelligible hand signals. Mystified and eager to understand what was happening in the school, the government recruited an American linguist, Judy Kegl, to decode the hand signals and explain how the children had been able to spontaneously create their own complex system of communication (Osborne, 1999). [...]

Area: Mechanical Engineering

[...] Common sense informs us that one can reduce the number of fall-related injuries in two ways: (a) reduce the number of falls by intervening on the risk factors for falls (for example, Weerdesteyn *et al.*, 2006), and (b) when a fall happens, reduce the risk of injury from that fall (for example, DeGoede; Ashton-Miller, 2002, 2003, Groen *et al.*, 2005, Lo, 2006). As far as the first approach is concerned, many extrinsic and intrinsic risk factors for falls have been identified (for example, Tinetti *et al.*, 1994) and several types of interventions have been developed to address the most important of these factors. However, the number of falls each year remains high and the probability of ever reducing the number of falls to zero is non-existent. Furthermore, sooner or later everyone falls unintentionally, regardless of their age. So, the second approach leads us to want to find out what people need to know *a priori* and what they need to do during a fall so that they can avoid injury when they do suffer the inevitable fall.

It is known that when a fall results in the greater trochanter directly impacting the ground, then the risk of hip injury is 30 times higher than when the greater trochanter does not strike the ground (Nevitt *et al.*, 1993, Hayes *et al.*, 1993, Schwartz *et al.*, 1998). The underlying message from these studies is that if an elderly person falls, don't, ever, land on your greater trochanter. [...]

1. After reading the excerpts and accessing the complete texts, you will notice that authors present the Review Literature in different sections and subsections. How do authors in your area discuss these sections?
2. Below you will find the Review section *MICUSP* from three different areas. According to Citation and Tense in the Literature Review by Swales and Feak (2012), can you identify the three different patterns in the texts below? Which verb tenses and sentences do you recognize? Please Highlight and discuss them with our colleagues.

Area: Education

Race is a socially constructed way of dividing the human population by phenotypes that has been politically reproduced since the colonial era. Race was created to justify social inequality, and became a marker of inequality, specifically, the subordination of those with darker who are not White skin to the domination of those who are White with lighter skin. Race, or the color line (DuBois, 1903), has been the basis for both individual prejudice and governmentally sponsored subjugation, including denial of personhood, rights, citizenship, and countless other manifestations of freedom. DuBois called it the color line, which is to say the separation of some from others on the basis of skin color.

Genders are the socially constructed meanings attached to the sexes, including the functional norms that dominate people's understanding of masculinity and femininity. In the last ten or so years, dialogue on gender has begun to open up to include not only traditionally female and male orientations toward self and the world, but also categories such as transgender (i.e., one's physical sex does not match gender identity/role), pangender (i.e., one cannot be labelled as female or male), and genderqueer (i.e., one takes a both/and approach to female and male) identities and roles. Like race, gender has been used as the basis for systematic domination of one portion of the population (in this case, males) over another (females).

From these definitions, it is clear that race and gender, like all socially constructed categories, are dynamic in meaning and laden with values. One's status—and I believe it is fair to say, one's freedom—is shaped by the salience of one's race, class, gender, national origin, sexuality, and religion in a particular temporal, geographic, and social context. More importantly, the ways that an individual's race, gender, and other dimensions of identity play out independently, however, are the ways that they intersect to shape experience on an individual level. Poor, uneducated Irish immigrants, for example, may initially have been excluded from the white racial category despite their light skin color and European origins, but within a few generations had assimilated the values of privilege and the discourses and behaviors of domination that permitted them to claim status within the hierarchy (Marable, 2000, p. 245).

Source: <https://micusp.elicorpora.info/view?pid=EDU.G1.03.1>

=====

Area: Sociology

[...] Because of the grave social and painful individual implications of suicide, it has been the subject of much study among scholars. Many scholars believe that a persistent feeling of hopelessness is the driving force behind suicide in adults. They define hopelessness more specifically than its colloquial usage. Minkoff, in his work, describes it as, a feeling that nothing will turn out right, nothing will succeed, important goals are unattainable, and worst problems will never be solved (Handout, 2005). Weishaar and Beck also focus on the role of hopelessness in suicide. Many other scholars, including Beck, Pearson, Brown, Conaghan, Davidson, Heisel, Hill, and Uncapher further emphasize the role of hopelessness in causing suicide among specific populations, such as psychiatric patients and the elderly.

Related to theories centering on hopelessness is what are called the escape theory and the meaning-centered approach. Both of these theoretical approaches discuss a particular manifestation of hopelessness as center to suicidality. Escape theorists such as Baumeister and Hewitt posit that suicidal behavior results when a person perceives herself as failing or not meeting perfectionist goals. This state of negative self-awareness leads to cognitive deconstruction (Handout, 2005) in which the person's judgment and inhibition falters and they choose suicide to escape their painful state of disappointment and self-perceived failure. The meaning-centered approach similarly focuses on an individual's perception of her life. Frankl and Linehan in particular propose that a person's will to live depends on their ability to find meaning in new life events. Those who can find meaning even in negative events or suffering are thus less prone to suicide.

The medical profession has even gotten involved in the debate. Some scientific research indicates that predisposition to suicide may be a genetically inherited trait. However, psychiatrists and psychologists, such Americans Karl Menninger and Edwin Schneidman, still tend to focus on the role of hopelessness and inability to see solutions to problems as driving factors behind suicide (Handout, 2005).

All of the above literature focuses on traits of individuals that may cause suicide. There, however, a substantial body of literature that claims the individual factors causing suicide are negligible when compared to the social forces influencing an individual's perceptions and actions. To understand such a seemingly counterintuitive position, we must first review the sociological literature on individual agency and social structure.

Two basic schools of thought dominate the structuralist camp. Many sociologists, including prominent scholars Emile Durkheim and Karl Marx, argue that social structures are so powerful that they completely control an individual, and leave no room for individual's actions or characteristics to determine her life. More recently, Berger and Luckmann have argued that Durkheim and Marx's ideal is too restrictive and that, in the beginning, individuals have agency in giving meaning to the world they see around them. After this first act of free choice, however, these meaning become institutionalized and then form the social structure which goes on to limit the future agency of the individual. In discussing suicide, structural sociologists would look at the influence of social institutions and practices on an individual's decision to commit suicide.

On the other side of the coin are sociologists who believe that individuals have considerable power in shaping their world and perceptions. These scholars, known as symbolic-interactionists, do not ignore the influence of society. They recognize the power of social forces such as stigma, situation, and peer groups and the limitations placed by society on individuals. They do believe, however, that individuals retain some agency and ability to effect change in their part of the world. Proponents of this theory include George Herbert Mead and others of the Chicago school.

Source: <https://micusp.elicorpora.info/view?pid=SOC.G0.03.1>

=====

Area: Nursing

A study from the University of Minnesota found that time spent performing physical activity significantly decreased from early adolescents to late adolescents (University of Minnesota, 2007). Decreasing physical activity becomes a habit, and a person will develop a sedentary lifestyle. This lifestyle leads to increase risk for obesity and other related illnesses (University *et al.*, 2007). In today's world it is often harder to be active than it is to be not active. The priority of convenience and time saving habits leads to less leisure time, and less activity. People depend on cars more for transportation to decrease effort and save time, work leads to a decrease time to participate in healthy activities.

A study published by The Journal of Adolescent Health verifies the high prevalence of minimal activity in adolescent males and females. Sixty percent of females, and forty-three percent of males did not meet the national guidelines for activity. Compliance was defined as being physically active for 60 minutes five times a week. This study also concluded that race or region of residency did not have an impact of amount of activity. Girls were more likely than boys to decrease activity as age increased (Butcher; Sallis; Mayer; Woodruff, 2008). This article supports efforts to increase the amount of physical activity beginning at or before adolescents. This will prevent the onset of a sedentary lifestyle as well as the physical and emotional complications form inactivity. (Butcher *et al.*, 2008).

Drexel University studied the contributing factors of childhood obesity by observing the physical and social environment changes. The study recognized the physical environment decreasing opportunity for activity, and the social environment promoting food high in fat content and calories. The study concluded the need for a shift of assuming individual responsibility for obesity, and recognition of the environment as the primary determinant of obesity (Budd; Haymann, 2008). While there may need to be a larger focus on the contributing factors to limited activity among the population, an individual must do what s/he is in control of. Drexel recommended the use of nurses to promote and support individual change while setting an example for the community (Budd *et al.*, 2008). Until America can accept responsibility and admit its faults, individuals must do what is in their power to decrease their risk of the complications of inadequate activity.

Source: <https://micusp.elicorpora.info/view?pid=NUR.G0.15.2>.

4. Access an Academic Corpus (*BAWE*, *OCAE*) in *Sketch Engine* or *Lextutor* and read some parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. Do you think the previous excerpts could be improved in any way? How? Write four examples below.

1.
2.
3.
4.

5. Do Your Own Corpus in *Sketch Engine* and read some parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. What can you say about the findings? Are they similar to the ones you found at *BAWE* or *OCAE*? Can you identify the patterns by Swales and Feak (2012)? Discuss with our colleagues.
6. After your discussion with our colleagues, select an excerpt from Do Your Own Corpus and *BAWE* or *OCAE* and put in *Chat GPT*, asking it to identify the patterns, according to Swales and Feak (2012).
7. Below you will find the Review section *MICUSP* from three different areas. According to Integral (author prominent) and Non-Integral (research prominent) Citations in the Literature Review by Feak and Swales (2009), can you identify them in the texts below? Which areas can you find Integral Citation? Which areas can you find Non-Integral Citations? Highlight and discuss them with our colleagues.

Area: Natural Resources and Environment

Urban development in areas surrounding coastal ecosystems has been accompanied by changes in ecosystem structure. A commonly reported estuarine malady is an increase in the frequency and severity of seasonal hypoxic and anoxic conditions, associated with increased levels of algae and phytoplankton in the system (Jackson *et al.*, 2001; Brawley *et al.*, 2000). These events are often explained as being caused proximately by increased anthropogenic contributions to estuarine nutrient loads and have triggered significant effort into the modeling of nutrient inputs to estuaries from groundwater, point and non-point source runoff into estuarine feeders. The goal of such work is to determine estimates of critical nutrient load, or loading targets to guide agricultural and land-use policy (Brawley *et al.*, 2000; Cerco *et al.*, 1995). However, it is important to recognize that increased runoff and nutrient input is not the only anthropogenic influence on coastal ecosystems. Habitat destruction and fishing practices have affected the ability of macrofauna, like the oysters, to exert control over the structure of their ecosystems. The oyster reefs of the Chesapeake Bay were once capable of filtering the entire water column in only a few days, and it is only since exhaustive dredging of the bay led to the collapse of the oyster fishery that hypoxia and anoxia begin to be observed (Jackson *et al.*, 2001). Jackson *et al.* discuss this and other correlations of ecosystem decline with destructive fishing practices, making it clear that comprehensive attempts to curb hypoxic and anoxic events via policy should consider the potential for both bottom-up and top-down effects to be significant in a target system.

Source: <https://micusp.elicorpora.info/view?pid=NRE.G3.01.1>

=====

Area: Political Science

However, partisan politics has not always led to the complete destruction of beneficial initiatives for Americans. In fact, it can be argued that it is precisely such partisanship that promotes healthy democratic debates over issues. With myriad opinions from both sides being contributed through suggested amendments even on the basis of party lines, the final bill can be further refined and improved before it goes to the floor, increasing the probability of it being passed into law. This can be seen in the example of Medicare where in 1995, for the first time in 30 years, a highly public, partisan and ideologically divisive debate occurred (Oberlander, 2003). The 1995 Republican Medicare reform bill championed achieving a balanced budget through large cuts in program spending by introducing a political cap on Medicare expenditures (Oberlander, 2003). Democrats were opposed to such a change because they believed that this would have led to a shortfall in program finances and crippled the program, thus leading then President Bill Clinton to veto the bill, while congressional Democrats joined in solidarity to criticize the Republican Medicare proposal as abandoning its social contract with the people (Oberlander, 2003). However, the 1995 defeat of the bill created the gateway for further political compromises, and finally in 1997, Medicare reforms were passed in a bipartisan manner with the hard cap on Medicare spending that would have activated automatic spending cuts in the program being removed (Oberlander, 2003).

This was a key concession by the Republicans as Democrats had identified this particular cap as threatening the ability of seniors to access quality medical care and were thus strongly opposed to it (Oberlander, 2003). Spending cuts were also less harsh this time around. In hindsight, these political compromises have been recognized as essential in preventing traditional Medicare from being devastated through cuts and restructuring, allowing it to continue serving its target population (Oberlander, 2003). Thus in this scenario, we are able to observe how partisan politics might have served the American people.

Source: <https://micusp.elicorpora.info/view?pid=POL.G0.30.1>

=====

Area: Civil & Environmental Engineering

Constructed stormwater wetlands are water treatment wetlands designed to improve the water quality of urban stormwater while increasing on-site detention to mitigate the effects of storm surges on water treatment plants or riverways [6]. Constructed wetlands can also be used to treat municipal wastewater prior to discharge [7], or for treating industrial wastewater such as wood waste or landfill leachate [8].

Constructed wetlands are considered to be favorable control mechanisms for storm water management because they reduce stormwater contaminant loading, provide increased detention capacity on site, and increase recreational opportunities and wildlife habitat. Although not always suitable as a sole stormwater management control mechanism, constructed wetlands have gained increased attention as an option for controlling stormwater within a diverse stormwater management infrastructure.

Source: <https://micusp.elicorpora.info/view?pid=CEE.G3.03.1>

8. Access an Academic Corpus (*BAWE, OCAE*) in *Sketch Engine* or *Lextutor* and read other parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. Could you find more Integral or Non-Integral Citations? Write four examples below and discuss with your colleagues.

- 1.
- 2.
- 3.
- 4.

9. With your Do Your Own Corpus in *Sketch Engine*, read other parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. What can you say about the findings? Could you find more Integral or Non-Integral Citations? And which is the reason for your findings? Discuss with our colleagues.
10. Start planning your literature review by consulting academic corpora and your own corpus. Write one example of each function.

Referring to Sources	Phrases used in this section taken from the Manchester Phrasebank	Phrases used in this section taken from your specialized corpus
General comments on the relevant literature		
Previous research: a historical perspective		
Previous research: methodological approaches taken		
Previous research: area investigated		
Previous research: What has been established or proposed		
Stating what is currently known about the topic		
Reference to a previous investigation: researcher prominent		
Reference to a previous investigation: time prominent		
Reference to a previous investigation: investigation prominent		
Reference to a previous investigation: topic prominent		

Reference to what other writers do in their published work		
Reference to another writer's idea or position		
Synthesizing material: supporting evidence or ideas		
Synthesizing material: contrast evidence or ideas		
Some ways of introducing quotations		
Summarizing the review or parts of the review		

References

Academic Phrasebank. Available at: <https://www.phrasebank.manchester.ac.uk/referring-to-sources/>. Access on: Jul. 23, 2021.

ALUÍSIO, S. M.; OLIVEIRA Jr., O. N. A detailed schematic structure of research papers introductions: An application in support-writing tools. *Revista de La Sociedad Espanyola Para El Procesamiento Del Lenguaje Natural*, v. 19, p. 141–147, 1996.

Chat GPT. Available at: <https://chatbotapp.ai/>. Access on: Jul. 23, 2021.

FEAK, C. B.; SWALES, J. M. *Telling a Research Story. Writing a Literature Review*. 3rd Edition: Michigan ELT, 2009.

MICUSP. Available at: <https://micusp.elicorpora.info/>. Access on: Jul. 23, 2021.

OLIVEIRA, Jr. O. N.; SCHUSTER, E.; LEVKOWITZ, H.; ZUCOLOTTO, V. The Fundamentals of Scientific Writing. In: SCHUSTER, E.; LEVKOWITZ, H.; OLIVEIRA Jr., O. N. *Writing Scientific Papers in English Successfully*. Andover & São Carlos, 2014.

Sketch Engine. Available at: <https://auth.sketchengine.eu/#login>. Access on: Sep. 8, 2021.

Lextutor: Web Concordance English. Available at: <https://www.lex tutor.ca/conc/eng/>. Access on: Sep. 8, 2021.

SWALES, J. M. *Aspects of article introductions*. Language Studies Unit, University of Aston in Birmingham, 1981.

SWALES, J. M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.

SWALES, J. M. *Research genres: Explorations and applications*. Ernst Klett Sprachen, 2004.

SWALES, J. M.; FEAK, C. B. *Academic Writing for Graduate Students*, 3rd Edition: Michigan ELT, 2012.

WEISSBERG, R.; BUKER, S. *Writing up research*. Prentice Hall Englewood Cliffs, NJ, 1990.

Chapter 10

Materials and Methods

Paula Tavares Pinto



Theoretical Background

According to Karpenko-Seccombe (2020, p. 127), the Methods (or Methods and Methodology) section of a research paper provides an account of how the research was conducted, justifies the choice of specific methods, outlines the study's design, and describes the exact steps and procedures used to answer the research questions, test hypotheses, and achieve the research aims and objectives. Similarly, as summarized by McCombes (2019), the methodology section details what the researcher did and how they did it, allowing others to evaluate the reliability and validity of the research. This section includes discussions on the type of research carried out, the data collection and analysis processes, and the tools and materials used. Typically, this section is written in the past tense. In this chapter, the authors will explore different ways in which researchers describe the Methodology section, whether they belong to SHAPE disciplines (Social Sciences, Humanities, and the Arts for People and the Economy) or STEM disciplines (Science, Technology, Engineering, and Mathematics) (Pinto *et al.*, 2024).

Exercises

1. The words **Method** and **Methodology** are sometimes confused by graduate students. What is the difference between both? Think about it, and briefly discuss with your colleagues how you would define each term.

Find a definition for each one of the words from a reliable online dictionary and write them here:

2. Using both words, *Method* and *Methodology*, run a search in the British Academic Written English Corpus - BAWE from Lextutor (<https://www.lex tutor.ca/conc/eng/>) or Sketch Engine (<https://auth.sketchengine.eu/#login>). If you have your own Specialized Corpus compiled from journals in your research area, you can use the concordance lines from it. If you are not aware on how to use these tools, you will find a tutorial online.

Then, **carefully read the concordance lines** and pay close attention to the words that appear before and after *Method* and *Methodology* in each context. As you reflect on the usage of these terms, consider the following:

- **Context:** In what type of academic writing is each term used? Is it part of a broader explanation or a specific description of a research approach?
- **Word Class:** What types of words (e.g., adjectives, nouns, or verbs) commonly appear before or after *Method* and *Methodology*? Do you notice any patterns, such as *Method* being modified by specific adjectives or *Methodology* being followed by certain verbs?
- **Function:** How do *Method* and *Methodology* function in the sentence? Does one seem to refer more to practical steps, while the other addresses theoretical concepts?

Suggestions: Try to select lines where *Method* and *Methodology* are used in clearly different ways. For example, you might find *Method* used in a procedural context (e.g., “a method to collect data”) and *Methodology* in a theoretical context (e.g., “the methodology behind the study”).

Concordance lines from _____ with the query METHODS

(Adapted from Karpemco–Seccombe, 2020)

Concordance lines from _____ with the query METHODOLOGY

2.1. After reading the lines or the expanded paragraphs, what have you noticed about the language used in these sections? Which verbs are used? In which tenses? Are there phrases you could use in your own research papers? Think about it and use them in your paper template for future reference.

3. Now, explore the collocates of both *Method* and *Methodology* using either the **WordSketch** tool from Sketch Engine or the **KWIC (Key Word In Context)** tool in Lextutor.

If you are using the **KWIC tool in Lextutor**, follow these steps:

1. Enter *Method* or *Methodology* in the search field.
2. Select **Collocates** from the top menu.
3. In the **Part of Speech (POS)** drop-down menu, choose:
 - **adj.ALL** to find collocates with adjectives.
 - **verb.ALL** to find collocates with verbs.

Next, if *Method* and *Methodology* are not terms commonly used in your area of research, think about which terms are. For instance, you might use *Approach* or another term. In that case, replace *Method* and *Methodology* with the relevant term and perform the same search. Record the combinations you find below.

COLLOCATES OF METHODOLOGY

verbs (<i>methodology</i> as an object):
adjectives:
verbs (<i>methodology</i> as a subject):
for+V.ing
for + noun

COLLOCATES OF METHOD

verbs (<i>method</i> as an object):
adjectives:
verbs (<i>method</i> as a subject):
prepositions and prepositional phrases

4. Access the MICUSP corpus, find some excerpts that describe the functions below, and discuss your findings with a classmate.

- to identify a gap in the methodology
- to provide support for the claim
- to justify the choice of methods
- to present a critical analysis of existing methods

(Adapted from Karpenko–Seccombe, 2020)

1. _____
2. _____
3. _____
4. _____

5. Using all your previous searches, write five sentences about the methods/ methodology/ approach you are using in your research and, from that point on, expand your section adding more details:

References

KARPENKO-SECCOMBE, T. *Academic Writing with Corpora*. VitalSource Bookshelf, Taylor & Francis, 2020.

LEXTUTOR. Web Concordance English. Available at: <https://www.lex tutor.ca/conc/eng/>. Access on: Nov. 25, 2024.

McCOMBES, S. *Understanding Different Sampling Methods*. Scribbr. 2019. Available at: <https://www.scribbr.com/methodology/sampling-methods/>.

MICUSP. Available at: <https://micusp.elicorpora.info/>. Access on: July 23rd, 2021.

PINTO, P. T.; SILVA, L. F.; SERPA, T.; CAMARGO, D. C. Do-It-Yourself Corpora to Support SHAPE and STEM Research Paper Writing. *In*: SARMENTO, S.; REBECHI, R., MATTE M. L. (org.). *English for Academic purposes: reflections, description e pedagogy*. 1. ed. Porto Alegre: Zouk, 2024. v. 01, p. 97-126.

SKETCH ENGINE. Available at: <https://auth.sketchengine.eu/#login>. Access on: Sep. 8, 2021.

Chapter 11

Results and Conclusion

Mayra Aparecida dos Santos
Paula Tavares Pinto



Theoretical Background

Whenever we write a report, a paper, a book, or any kind of academic text, it is important to be precise in each and every part of the text. The text must be intelligently divided into parts such as introduction, methodology, results, and, finally, conclusion. Although that is the regular sequence for the work when it is done, it is certainly not the common path for the writer. It is more likely to write a precise abstract and conclusion once the intermediate chapters are done.

Considering that the author is fluent and has a good proficiency in the target language, it will be easier to write which doesn't mean that the text will be similar to one written by a native speaker of English. In order to do so, Corpora and AI tools can be very helpful and provide a rich repertoire of phrases that are common in academic texts and provide a better reading experience.

According to the website Academic Phrasebank, conclusions have mostly two main purposes in academic texts: look back and summarize the principal topics covered in the writing and indicate a final comment or judgement. It might be “optional in research articles where consolidation of the study and general implications are covered in the Discussion section” (Academic Phrasebank website), but it is highly expected in most kinds of papers, once the reader is able to see the author's whole idea.

Conclusions not only wrap up the text, but also may point out ways to continue the research or to branch out ideas for new findings in the scientific field. For that, writers have the option of suggesting implications, recognize and acknowledge limitations of the studies, summarize or restate the aims and the findings, state and make recommendations and also, state practices and policies.

Exercises

1. Which sentence below would you choose to begin your Discussion Section? Why? Highlight the words/ phrases that helped you decide which one.

- A. Our findings suggest that patients who had partially blocked arteries and were treated with a stent as the first line of intervention had no better outcomes than patients who were not given any surgical treatments.

B. Our findings noted that patients who received stents demonstrated slightly higher rates of asthma than those who did not. In addition, the placement of a stent did not impact their rates of cardiac events in a statistically significant way.

(Enago Academy, 2021).

2. In the Discussion and Conclusions sections, writers discuss and interpret results. Look at the following concordance lines and highlight examples in which authors:

- a. highlight the significance of the results
- b. compare their results with previous research
- c. emphasise the novelty and contribution of their research
- d. suggest treating results with caution

<s>Ultimately societal morphogenesis or morphostasis is the result of the interaction between third order emergent properties (the results of the results of **the results** of social interaction) (Archer, 1995; 327).</s>

<s>Therefore **the results** would most likely yield conclusions that are incongruent with Parsons' implication that the political structure is a direct reflection of society's underlying collective goals, as the particular form a structure takes is the result of interaction between competing interest groups.</s>

<s>Table 2 illustrates that there are a number of significant correlations in **the results** obtained, consisting of both inter-correlations between sub-scales of the Ways of Coping Questionnaire and correlations between scores on particular sub-scale with AUDIT scores.</s>

<s>It would generally be expected that participants scoring highly on problem-focused coping would be unlikely to also score highly on emotion-focused coping, which is what **the results** indicate; as planful problem solving score goes up, distancing score goes down.</s>

<s> **The results** obtained support the hypothesis posited at the beginning of the study, which stated that those with higher levels of alcohol consumption would exhibit more avoidance coping.</s>

<s> **The results** therefore imply that, when considered in relation to academic workload in particular, a higher level of alcohol consumption is associated with a tendency to exhibit more avoidant coping strategies, and a smaller likelihood of going to others for help, whether this is emotional help in the form of seeking sympathy and understanding, or practical help in the form of asking a lecturer or tutor for guidance with work.</s>

<s>This is highlighted by the significant negative correlation obtained in **the results** between the ‘distancing’ and ‘seeking social support’ sub-scales, illustrating that individuals who are high in social support seeking are low in avoidant coping strategies, and resultantly less likely to show high levels of alcohol consumption.</s>

<s> **The results** partially confirm the original hypothesis - there was a significant main effect of speed.</s>

<s> **Th results** broadly support previous research.</s>

2. *Conclude* is probably the most obvious word used in the last section of a research paper. Understanding how this word is used in context can help clarify how academic writing summarizes key findings and emphasizes the research’s impact.

So, let’s look at how *Conclude* is used in academic writing to express conclusions and contributions. Follow these steps:

- **Step 1:** Search for the word *Conclude* in your specialized corpus or in an academic corpus (such as BAWE or another available corpus). This will give you examples of how *Conclude* is used in real academic texts.
- **Step 2:** Examine the expanded concordance lines that include this word, paying close attention to how it introduces conclusions and reflects the study’s contributions.
- **Step 3:** Select two expanded concordance lines that, in your opinion, most clearly show the study’s conclusions and contributions. Write it down in the space below.
- **Step 4:** Underline the key phrases in each line that helped you decide they were effective at presenting the study’s closure and outcomes.

3. Run a search with the query word *Implications* in Wordsketch of SKell. Which are the most common verbs used with it? If you are using Sketch Engine, take a picture of the word cloud

COLLOCATES OF *IMPLICATIONS*

verbs (*methodology* as an object):

adjectives:

verbs (*methodology* as a subject):

Source: Adapted from Karpenko-Seccombe (2020)

4. Considering your findings in the exercise above, choose two collocates of *implications* to expand the concordance lines and select similar words for them, in order to improve your vocabulary knowledge.

verbs (*methodology* as an object)

- 1.
- 2.
- 3.
- 4.
- 5.

verbs (*methodology* as a subject):

- 1.
- 2.
- 3.
- 4.
- 5.

Adjectives:

- 1.
- 2.
- 3.
- 4.
- 5.

Source: Adapted from Karpenko–Seccombe (2020)

5. The word *observations* is regularly used in the conclusion section of papers. Find the five most common collocates of *observations* with verbs in Wordsketch and add 3 similar words to each of the collocates.

COLLOCATES OF *OBSERVATION*

- | |
|----|
| 1. |
| 2. |
| 3. |
| 4. |
| 5. |

Source: Adapted from Karpenko–Seccombe (2020)

6. The previous activity presented a substantial amount of phrases and vocabulary that helps vary the options when the author of a paper needs to write the conclusion paragraphs or sections. Using the phrases listed above, choose three examples to develop paragraphs related to the conclusion of a paper you have written or are still in progress.

1 _____

2 _____

3 _____

7. Considering that you have just finished a research, and you have concluded that your studies found some limitations. Choose two of the options below to point out these limitations and develop a paragraph.

- a. A limitation of this study is that ...
- b. The most important limitation lies in the fact that ...
- c. The generalisability of these results is subject to certain limitations. For instance, ...
- d. This current study is limited by the absence of ...

1 _____

2 _____

References

ACADEMIC PHRASEBANK. Available at: <https://www.phrasebank.manchester.ac.uk/writing-conclusions/>. Access on: Jun. 8, 2024.

ENAGO ACADEMY. *Discussion Vs. Conclusion*: know the difference before drafting manuscripts.

KARPENKO-SECCOMBE, T. *Academic Writing with Corpora*. VitalSource Bookshelf, Taylor & Francis, 2020.

LEXTUTOR. Web Concordance English. Available at: <https://www.lex tutor.ca/conc/eng/>. Access on: Nov. 25, 2024.

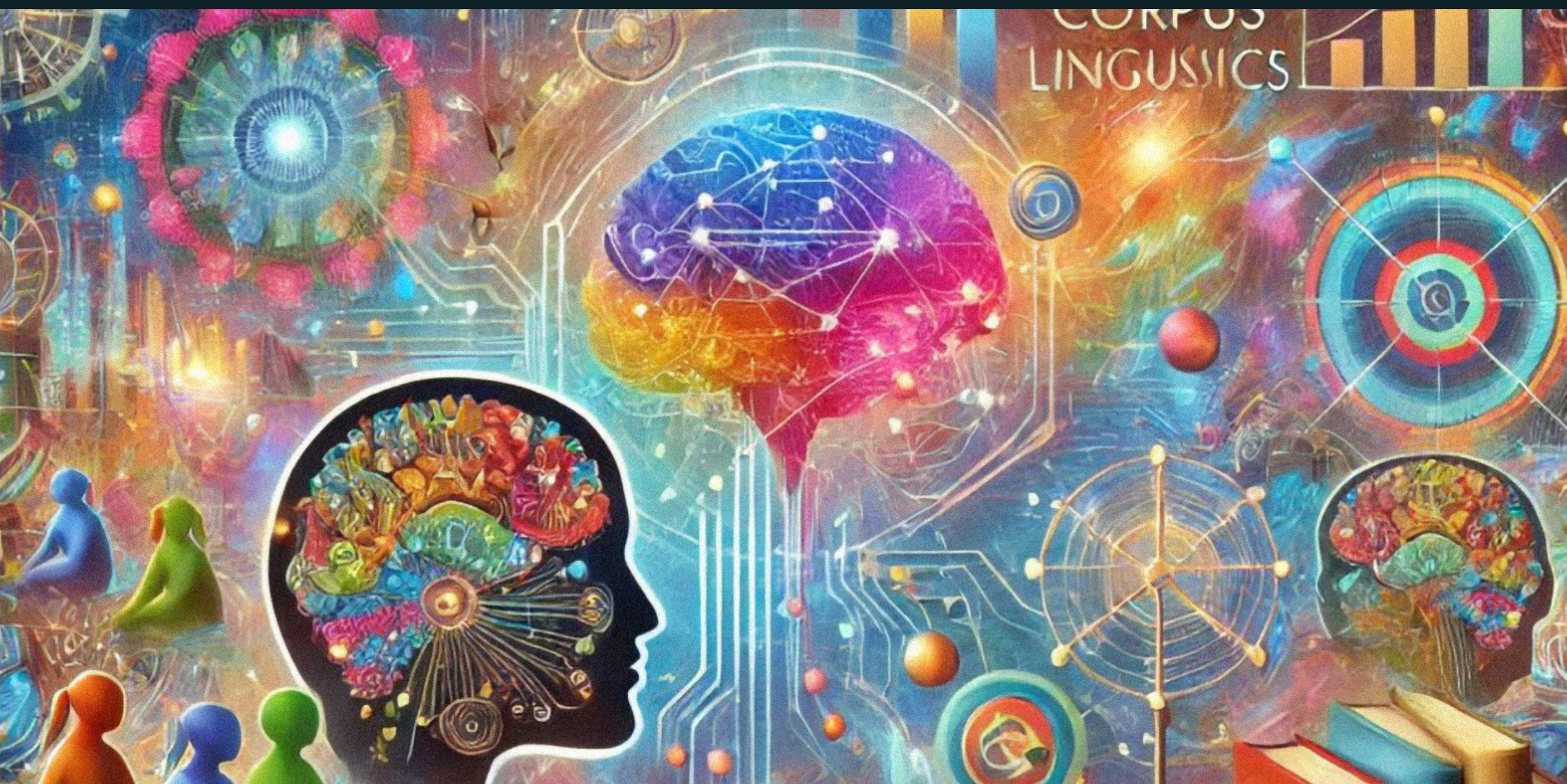
MICUSP. Available at: <https://micusp.elicorpora.info/>. Access on: Jul. 23, 2021.

SKETCH ENGINE. Available at: <https://auth.sketchengine.eu/#login>. Access on: Sep. 8, 2021.

Chapter 12

Using the TED Corpus Search Engine to explore discourse organization expressions in Academic spoken language¹

Luciano Franco da Silva



¹ **Note to the Reader:** Chapter 12, *Using the TED Corpus Search Engine to Explore Discourse Organization Expressions in Academic Spoken Language* is dedicated to guiding teachers in implementing the activities in the classroom. It offers practical strategies for educators to help students engage with academic spoken language through corpus-based tasks. This chapter enhances the instructional experience by providing educators with structured tools and insights, making it an invaluable resource for both teaching and learning.

Theoretical Background

In recent years, we have witnessed the emergence and evolution of new genres aimed at academic popularization. Unlike traditional approaches that simplify specialized knowledge for non-experts, academic popularization serves as a vital medium of scientific communication, fostering the democratization of knowledge and facilitating the exchange of ideas across diverse audiences and disciplines (Gotti, 2014; Wang; Csomay, 2024). These genres are produced by academics and experts from various fields for audiences who do not share the same scientific background. Consequently, academics must recontextualize their communication methods to ensure their research findings are appealing and comprehensible to a broad audience. Researching academic popularization genres involves examining how academics use language to engage society, promote their research, assert their authority as experts, and disseminate scientific practices into the public domain (Hyland, 2009).

One well-known type of academic popularization discourse is TED Talks (Caliendo, 2014; Compagnone, 2016; Wang Csomay, 2024). These freely available recorded presentations are delivered by experts from various fields to a general audience. They exemplify the intersection of academic expertise and public accessibility, as experts from a wide range of fields bring forward to a lay audience what is novel, exciting, and groundbreaking in their respective research areas. TED Talks can be described as a hybrid web-mediated genre of academic popularization discourse (Caliendo, 2014). This hybrid nature is evident in their content and presentation style, as TED speakers blend academic discourse features with entertainment elements to present complex ideas engagingly. As a result, TED Talks often deviate from traditional academic presentations by avoiding formal and technical language, evolving into a digital phenomenon that transcends conventional academic and scientific communication boundaries (Caliendo, 2014; Compagnone, 2016). Over the years, they have leveraged the internet's power to amplify their reach and impact, transforming from a specialized local conference to a globally recognized platform for sharing ideas and projects.

Due to its empirical nature, corpus linguistics (CL) offers numerous potential benefits for English for academic purposes (EAP) settings. Corpus analysis enables EAP professionals to examine academic language use from both broad and specific perspectives. They can investigate general academic discourse or focus on specific genres or disciplines, gaining nuanced insights into language requirements and conventions within different academic fields. Analyzing large and diverse text collections allows EAP practitioners to identify

patterns, structures, and features intrinsic to academic language, informing pedagogical practices and helping students develop more accurate and effective language skills.

The utility of CL in exploring register variation in language use is crucial in EAP (Hyland, 2004; Viana; O’Boyle, 2022), and this importance extends to the investigation of academic popularization genres. A significant area of interest is how TED Talk presenters recontextualize specialized discourse to make it accessible to a target lay audience. The following pedagogical activities explore the organization and presentation of relevant information in spoken academic discourse, focusing on two key features: *appeal* and *framing*. Appeal refers to strategies used by presenters to capture the audience’s interest or establish their expertise. Many TED speakers begin their presentations by establishing common ground with the audience, presenting topics as matters of shared concern or daily life interest, thereby engaging the audience and setting the tone for the talk. Framing, on the other hand, involves presenting technological or scientific breakthroughs as unique, relevant, and significant. The use of words like ‘first’ or phrases such as ‘for the first time’ is particularly significant in highlighting the groundbreaking nature of the topic. By emphasizing novelty, speakers create a compelling narrative that maintains the audience’s interest throughout the talk.

The activities presented in this lesson will be valuable to future CL-EAP researchers interested in academic discourse. As the democratization and globalization of knowledge gradually narrow the gap between academia and the general public (Calsamiglia, 2003), this type of discourse is likely to gain further popularity, especially in online mediums (see Luzón; Pérez-Llantada, 2019; Pérez-Llantada, 2021).

Exploring the expression *a little bit about* for discourse organization

- 1.** Begin the class by asking students to share the types of academic events they have participated in or expect to participate in during their undergraduate or graduate studies. Encourage them to list as many examples as possible. Some responses might include expository classes, seminars, group presentations, poster presentations, 3-minute thesis presentations, and webinars.
- 2.** Ask students to suggest words or phrases they would use to highlight their academic expertise and background, direct the audience’s attention, and emphasize information as relevant and groundbreaking. Write their responses on the whiteboard.

3. Divide the students into small groups. In their groups, have students discuss and decide which phrases they consider most appropriate for spoken academic registers and why.
4. Walk around the classroom, providing feedback to individual groups as needed. Once the group discussions are over, invite students to share their answers with the entire class.
5. Emphasize the importance of appeal and framing strategies in spoken academic registers, noting that it is crucial for academics to effectively communicate key messages to the general public.
6. Draw a three-column table on the board with the headings: *Stance*, *Engagement*, and *Topic Elaboration*.
7. Invite students to complete the table with phrases they can remember. Suggested answers are provided below, but students may add other phrases.

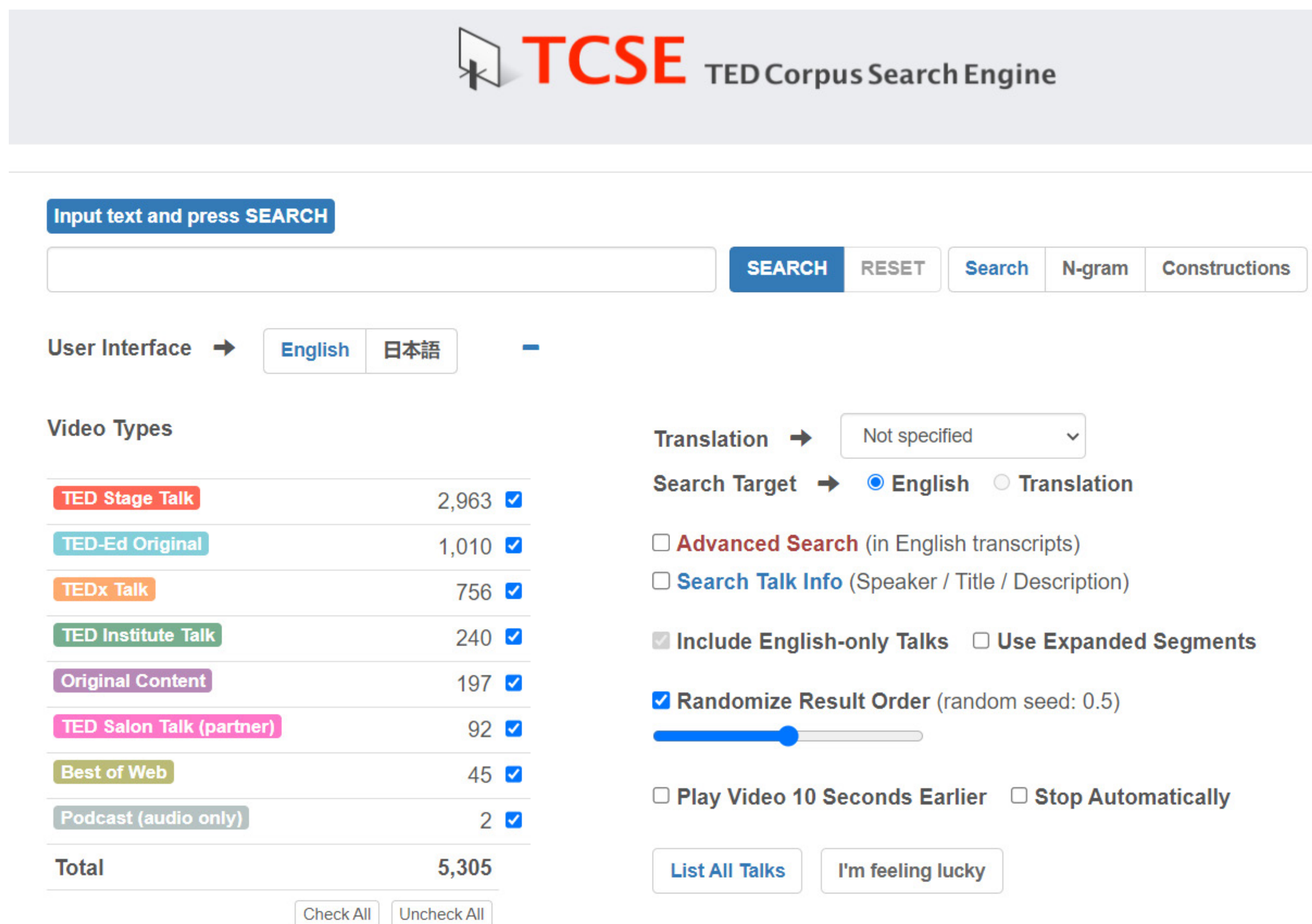
Table 1 – Suggested answers for stance, engagement, and topic elaboration expressions

Stance	Engagement	Topic elaboration
and I think that I can tell you I don't know if I don't know what I think this is might be able to so that we can to be able to we were able to	(I'm) + going to tell you and you can see as you can see I want to show + (you) I want you to if you look at if you want to let me show you let me tell you take a look at those of you who when you look at you can see here	a little bit about all of a sudden in the first place it turns out that let me give you + (an example on the other hand one of the first one of the most one of the reasons the problem is that this is a picture this is the first

8. Ensure students understand the expressions listed on the board by inviting individual students to explain the meanings of these combinations.
9. Ask the class if they are familiar with TED Talks, an initiative aimed at communicating scientific ideas to the general public. If necessary, explain that TED Talks are organized by a non-profit organization focusing on “ideas worth spreading” and are delivered by experts to a general audience. Numerous video-recorded TED Talks are freely available at <https://www.ted.com/>.

10. Familiarize students with the TED Corpus Search Engine (TCSE), a specialized tool for exploring TED Talks transcripts and aligning concordance lines with their original video sources² (see Hasebe, 2015).

Figure 1 – TCSE search page³



11. Show students how they can use some of the TCSE facilities⁴ to explore word sequences that express stance and discourse organization.

12. Ask students to search for the expression *a little bit about* in the TCSE to access its concordance lines. These lines contain short fragments from the original TED Talks where the sequence was used by the speaker. Provide a hands-off example using Table 2, which shows a few concordance lines extracted from the TCSE.

13. Encourage groups to read the concordance lines to explore the meaning of the expression *a little bit about* in TED Talks. Students should notice that *a little bit about* is

² Regarding to data availability and source, it is important to highlight that TED Talks are accessible for non-commercial use under the Creative Commons license (CC BY-NC-ND).

³ Available at <https://yohasebe.com/tcse>.

⁴ For richer contextual information on the TCSE see: <https://tcse.gitbook.io/doc>.

mostly used to indicate that the imposition on the audience’s time and concentration will not be significant (e.g., Examples 02, 03, 05). It is also used to signal that the information provided is only a small part of the broader subject (e.g., Examples 07, 10). Students should also observe that in spoken academic registers like TED Talks, speakers use *a little bit about* more often as a discourse organizer than as a quantifier.

14. Have the groups access the *Constructions* tab in the TCSE (see Figure 1) and explore the meaning of any expressions they are unfamiliar with. The goal here is to focus on a group of pre-selected formulaic expressions, analyze their use, and expand students’ lexical knowledge.

15. Ask each group to choose three expressions from the *Constructions* section of the TCSE and come up with ideas about the functions these expressions serve in the discourse. Encourage each group to present their suggestions to the class. The other students should ask questions about the usage of the expressions and whether there are other expressions that serve the same function.

16. The TCSE can be used to explore various other multi-word sequences (e.g., N-grams⁵) or many other pre-selected multi-word combinations (e.g., *about/high time + V-ed, in a nutshell, and on the other hand*).

Table 2 – Concordance lines for *a little bit about* in the TCSE (adapted).

01	Knowing just	a little bit about	the natural world can help you perceive the full spectrum around you all the time.
02	I thought I’d talk	a little bit about	some just big ideas about this, and then get John back out here so we can talk interactively a little bit more and think and ask questions.
03	I’m gonna talk	a little bit about	open-source security, because we’ve got to get better at security in this 21st century.
04	Somehow, 13 billion years ago there was a Big Bang, and you’ve heard	a little bit about	, you know, the origin of the universe.
05	And I’m going to use them to tell you	a little bit about	my favorite subject: deception.
06	I think I’ll start out and just talk	a little bit about	what exactly autism is.

⁵ See Franco and Viana (2022) for a more detailed example on how to use the TCSE, and the N-gram feature for pedagogical activities.

07	Now, I'm a neuroscientist by training and I know	a little bit about	how the brain performs under stress.
08	I'm going to talk to you	a little bit about	a few of these different companies.
09	I'm going to talk	a little bit about	where technology's going.
10	Much has been said of the Latino vote in this election, which is something I know	a little bit about	, having been working obsessively over it for the last 16 years.

References

CALINDO, G. The popularization of science in web-based genres. *In*: BONGO, G.; CALIENDO, G. (ed.). *The Language of Popularization - Die Sprache der Popularisierung*. Peter Lang, 2014. p. 111-138. Available at: <https://doi.org/10.3726/978-3-0351-0670-1/15>. Access on: Nov. 25, 2024. CALSAMIGLIA, H. Popularization discourse. *Discourse Studies*, v. 5, n. 2, p. 139-146, 2003.

COMPAGNONE, A. The pragmatics of spoken academic discourse in the framework of TED Talks: A case study. *In*: BIANCHI, F.; GESUATO, S. (ed.). *Pragmatic Issues in Specialized Communicative Contexts*, Brill Rodopi, v. 29, p. 71-91, 2016. Available at: https://doi.org/10.1163/9789004323902_006. Access on: Nov. 25, 2024.

FRANCO, L.; VIANA, V. Research findings for all: Popular science communication on global challenges. *In*: VIANA, V. (ed.). *Teaching English with corpora: A resource book*. 1st ed., Routledge, 2022. p. 300-306.

GOTTI, M. Reformulation and recontextualization in popularization discourse. *Ibérica*, n. 27, p. 15-34, 2014.

HASEBE, Y. Design and Implementation of an Online Corpus of Presentation Transcripts of TED Talks. *Procedia - Social and Behavioral Sciences*, v. 198, p. 174-182, 2015. Available at: <https://doi.org/10.1016/j.sbspro.2015.07.434>. Access on: Nov. 25, 2024.

HYLAND, K. *Disciplinary Discourses: Social Interactions in Academic Writing*. The University of Michigan Press, 2004.

HYLAND, K. *Academic Discourse: English in a Global Context*. Continuum, 2009.

LUZÓN, M. J.; PÉREZ-LLANTADA, C. *Science communication on the internet: Old genres meet new genres*. John Benjamins Publishing Company, 2019.

PÉREZ-LLANTADA, C. Genres and languages in science communication: The multiple dimensions of the science-policy interface. *Language and Communication*, v. 78, p. 65–76, 2021. Available at: <https://doi.org/10.1016/j.langcom.2021.02.004>. Access on: Nov. 25, 2024.

VIANA, V.; O'BOYLE, A. *Corpus Linguistics for English for Academic Purposes*. Routledge, 2022.

WANG, W.; CSOMAY, E. Constructing proximity in popularization discourse: Evidence from lexical bundles in TED talks. *English for Specific Purposes*, v. 73, p. 95–109, 2024. Available at: <https://doi.org/10.1016/j.esp.2023.10.003>. Access on: Nov. 25, 2024.

Answer Keys

Chapter 1: Answer Key

→ Activity 1

SkELL (Sketch Engine for Language Learning) is an online tool developed to help language learners and linguists explore and analyze linguistic patterns. Have you heard about this tool? List some of its features below.

Features:

Word Sketch
Concordance
Thesaurus
Translations
Word Frequency
Example Frequency
Search Filters

→ Activity 2

Now, access the tool and use the “*Word Sketch*” function in SkELL to generate a list of contexts where the word “research” appears. List at least five collocations and their concordance. Then, analyze the different contexts and discuss this word’s various meanings and uses.

Concordance:

verbs with **research** as the subject
research suggests that
verbs with **research** as the object
research conducted
adjectives with **research**
as a **research** assistant to
modifiers of **research**
scientific **research**
nouns modified by **research**
research project

→ Activity 3

Choose three words of your preference (related to the academic context) and use the “**Examples**” button to display examples of that word in context. List some of these examples below.

<i>Words:</i>	<i>Examples:</i>
Theory Methodology Analysis	The second alternative is state autonomy theory. Their conclusion and methodology were widely criticized. The current analysis includes data from 2005.

→ Activity 4

Now, based on the analysis of concordance and the list of frequent words, **create** sentences or short paragraphs using the word “research” in different contexts.

Paragraph:

Research suggests that corpus studies play a pivotal role in understanding language usage patterns and linguistic phenomena. By analyzing large collections of authentic texts, linguists can uncover insights into vocabulary, grammar, and discourse structures across different contexts and genres. Moreover, corpus research provides empirical evidence to support language teaching methodologies, translation practices, and language policy decisions. With advances in technology, corpus linguistics continues to evolve, offering invaluable resources for both theoretical linguistics and practical applications in fields such as natural language processing and computational linguistics.”

Paragraph:

Scientific research in corpus linguistics has revolutionized our understanding of language structure and usage. By systematically collecting and analyzing vast amounts of linguistic data from diverse sources, corpus linguists gain valuable insights into language patterns, variations, and evolution. This empirical approach allows researchers to explore language phenomena across different contexts, genres, and time periods with unprecedented accuracy and depth.

→ Activity 5

Read each statement about the SkELL tool below. Determine if the statement is *true* (T) or *false* (F) based on your knowledge or a quick research. List your corresponding answers.

Statements:

Ø (V) **SKELL** is a simple tool for students and teachers of English to easily check whether or how real speakers of English use a particular phrase or a word.

Ø (V) The «Word Sketch» function in SkELL summarizes how a word is used grammatically and lexically.

Ø (F) SkELL only offers examples in English, not supporting other languages.

Ø (V) The SkELL tool is paid and requires a subscription for full access.

Tool: Versatext

Activities– Analysing an academic article on its data (Data analysis)

Pre–reading activity– follow the steps below

1. The objective is to help the student feel more involved with the investigation process by choosing by themselves the subject, the text, and the data analyzed. In the given context, the activity is to search for an article and pre-analyze its title and abstract with the corpora-based tool, but other academic text genres can also replace the article.
2. The goal is to learn more academic vocabulary by studying its usage, in this step, the student will have to decide which words they will try to understand better in the context presented.
3. **A.** The student has to extract some meaning from the analyzed words by their context of usage and their previous knowledge;

4. **B.** Confirm if the interpretation agrees with its dictionary meaning or if it is used in a different context;
5. **C.** If the meanings diverge or not, it is important to understand why this occurs in the given context.
6. To better analyze the words and the academic context in which they are used, the student has to explore the **profiler** feature of the tool and write down some of their reflections on the importance of those academic specific and general words and terms.
7. Finally, the students have to reflect on why they are engaging in this activity and how this systematization of knowledge can help them in a more productive reading.

Tool: CorpusMate

The results of activities 1 and 2 are not set and cannot be fully predicted because they will depend on the subjectiveness of the user. So, to make sure the activities are correct, follow the instructions and the examples that are given in the activities.

Chapter 2: Note on Chapter 2

Chapter 2 does not include an answer key because the activities' results depend on the user's chosen data. As each data may vary, the answers will differ accordingly.

Chapter 3: Answer Key

Exploring academic vocabulary - Exercises: Answer Key

Activity 1 – “Waterbird habitat”

Pre-reading:

1. i) Urbanization; ii) Waterbird habitat; iii) Climate; iv) Water management.
2. Possible answer: wetlands such as rivers, lakes, and the ocean.
3. Personal answer. The students are supposed to use their background knowledge.
Possible answer: The effort of monitoring waterbirds and supporting actions to preserve their habitat.

4. Geography, Agriculture and Environment, Society.
5. Personal answer. The students are supposed to use their background knowledge. Possible answers: The preservation of waterbird habitats; The impacts of climate change on waterbird habitats; Actions to preserve waterbirds and natural resources.

Post-reading:

1. a) waterbird habitat; b) wetland restoration ; c) urban development; d) water supply management.
2. Alternative D.
3. Alternative B.
4. Alternative C.
5. Alternative E.

Activity 2 – “Ocean health”

Pre-reading:

1. a) Personal answer. Possible answers: Yes. Some examples are the words “Brazil”, “ocean”, “regional”, “actions”, “systems”, “marine”, etc.
b) Red - nouns (e.g. “ocean”); pink - adjectives (e. g. “national”); blue - verbs (e.g. “demonstrates”).
c) Personal answer. The students are supposed to use their background knowledge. Possible answers: Policies to support and improve ocean health in Brazil; The impacts of the management and assessment of ocean health on biodiversity.
2. i) Personal answer; ii) Personal answer; iii) Personal answer; iv) Personal answer.
3. i) Personal answer; ii) Personal answer; iii) Personal answer.
4. Personal answer.

Post-reading

1. The ocean health in Brazilian costlines.

2. The text is an abstract. It presents the aims, the theoretical background, the methods and the results of a research/study.
3. i) Alternative C; ii) Alternative B.
4. i) national level; ii) coastal habitats; iii) regional scale; iv) protect areas; v) ocean health; vi) marine areas.
5. (d), (c), (a), (b).
6. Alternative D.
7. Alternative C.

Chapter 4: Answer Key

Exercises

Warm up: free answers

Pre-listening:

Examples:

membrane – vaccine – nanoparticles – chemical – mRNA

ribosome – lipids – viral – viruses – molecule

Listening

1. This nanoparticle is made of lipids, the same type of fatty material that forms the membrane around our cells.
2. But these lipids have been specially engineered to protect the mRNA on its journey through the body and assist its entry into the cell.
3. Lastly, the final ingredients are added: sugars and salt to help keep the nanoparticles intact until they reach their destination.
4. Before use, the vaccine is kept at a temperature of -20 to -80 degrees Celsius to ensure none of the components break down.

5. And since this particular protein is likely to be found in most COVID variants, these antibodies should reduce the threat of catching new strains.

Vocabulary

A – Questions:

1. Same meaning, different emphasis.
2. Sentence i – passive voice
3. Sentence ii – active voice

B – Connecting verbs to sentences

- The organism was identified based on typical colonial morphology (granular colonies, rarely fried-egg-like, 10-150 μ in diameter) on the M.P. agar medium
- The CO-releasing molecules (transition metal carbonyls) was developed by Motterlini *et al.* [144].
- NOS-2 activity is primarily regulated transcriptionally and is commonly induced by bacterial products and pro-inflammatory cytokines.
- In contrast with the other two NOS isoforms (NOS-1 and NOS-3), which are expressed constitutively and activated by mediator-induced or stress-induced cell activation.
- In the latter study the authors suggested that fungal aggregation inhibits phagocytosis.
- This retrospective chart review describes the epidemiology and clinical features of 40 patients with culture-proven *Mycoplasma pneumoniae* infections.

C- Make true sentences

Bacterial and viral cultures were performed using standard methods.

Of the 24 cases of pneumonia, 21 were confirmed radiologically and the remaining 3 were diagnosed clinically.

M. pneumoniae was cultured using the classic *M. pneumoniae* agar medium (M.P. agar) and the Pneumofast tray (Pneumofast[®], International Microbio, Signes, France).

Crepitations, however, were detected in the majority (79.2%) of patients. Pleuritic chest pain and pleural effusion were rare.

The organism was identified based on typical colonial morphology (granular colonies, rarely fried-egg-like).

READING PRACTICE

Questions

a) The emergency use authorization for Evusheld was temporarily revoked by ANVISA and the FDA due to a reduced effectiveness of the AstraZeneca-manufactured medicine against the prevalent coronavirus variants.

b) Evusheld and sotrovimab target the receptor-binding domain (RBD) of the spike protein, a critical part of the virus. Concerns about the sustainability of these medications arise because the RBD is constantly evolving.

c) Challenges with antivirals include issues with accessibility and time-sensitive prescription requirements. There is concern about the emergence of drug-resistant virus variants associated with some antivirals like remdesivir and molnupiravir.

Chapter 5: Note on Chapter 5

Chapter 5 does not include an answer key because the activities' results depend on the user's chosen data. As each data may vary, the answers will differ accordingly.

Chapter 6: Note on Chapter 6

Chapter 6 does not include an answer key because the activities' results depend on the user's chosen data. As each data may vary, the answers will differ accordingly.

Chapter 7: Answer Key

Exercises

1 - a) Open answer (yes/no). Possible answers: linguistic characteristics (past form verbs, objective sentences, formal vocabulary, etc.) and extra linguistic characteristics (small text, 1 paragraph construction, etc.).

b) Open answer.

c) Open answer.

d) Objective and concise sentences. This happens because abstracts are characterised by short sentences (only 1 paragraph) and, therefore, the text must be brief and direct.

e) 2, 3, 4, 1.

2 - a) Open answer. Possible answers: the website offers high-level native speakers written productions which could be used as a model / example for students when they are writing their papers.

b) Open answer.

c) Open answer.

3 - a) Open answer.

b) Open answer.

c) Open answer. Possible answers: when comparing the two websites, students will be able to find a more natural writing style in MICUSP once they can see the range of vocabulary employed in the texts of native speakers (in comparison to Brazilian publications at SciELO).

d) Open answer.

4 - a) Open answer.

b) 5, 2, 3, 4, 1.

c) Open answer.

5 - a) Open answer. Possible answers: after articles, prepositions and numbers, verbs and nouns will probably be the most frequent words in the corpora. A verb represents an action; and a noun represents a person, a thing, a concept or place.

b) Open answer.

c) Open answer.

6 - a) Open answer.

b) Open answer.

c) Open answer.

Chapter 8: Note on Chapter 8

Chapter 8 does not include an answer key because the activities' results depend on the user's chosen data. As each data may vary, the answers will differ accordingly.

Chapter 9: Answer Key

Exercises

1. Below you will find the Review section *MICUSP* from three different areas. What type of information can you identify in this section? Literature Review. Which verb tenses are used see? simple present; present perfect; simple past; past continuous; past perfect. Which phrases do the authors use to introduce previous studies? Given the aforementioned...; Despite their...; Common sense informs us...; several types of interventions...; Highlight and discuss them with our colleagues. Personal answers.
2. After reading the excerpts, and accessing the complete texts, you will notice that authors present the Review Literature in different sections and with different subsections. How do authors of your area discuss these sections? Personal answer
3. Below you will find the Review section *MICUSP* from three different areas. According to Citation and Tense in the Literature Review by Swales and Feak (2012), can you identify the three different patterns in the texts below? Which verb tenses and sentences do you recognize ? Please Highlight and discuss them with our colleagues.

Pattern 1 Past - researcher activity as agent - “Girls were more likely than boys to decrease activity as age increased” (Butcher, Sallis, Mayer, Woodruff, 2008); “The study concluded the need for a shift of assuming individual responsibility for obesity, and recognition of the environment as the primary determinant of obesity” (Budd, Haymann, 2008); “While there may need to be a larger focus on the contributing factors to limited activity among the population, an individual must do what s/he is in control of. Drexel recommended the use of nurses to promote and support individual change while setting an example for the community” (Budd *et al.*, 2008).

Pattern 3 Present - no reference to researcher activity- “important goals are unattainable, ...” (Handout, 2005); “This state of negative self-awareness leads to cognitive deconstruction” (Handout, 2005); “to see solutions to problems as driving factors behind suicide” (Handout, 2005); “This lifestyle leads to increase risk for obesity and other related illnesses” (University *et al.*, 2007);

4. Access an Academic Corpus (*BAWE, OCAE*) in *Sketch Engine* or *Lextutor* and read some parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. Do you think the previous excerpts could be improved in any way? How? Write four examples below. Personal Answers
5. Do Your Own Corpus in *Sketch Engine* and read some parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. What can you say about the findings? Are they similar to the ones you found at *BAWE* or *OCAE*? Can you identify the patterns by Swales and Feak (2012)? Discuss with our colleagues. Personal answers.
6. After your discussion with our colleagues, select an excerpt from Do Your Own Corpus and *BAWE* or *OCAE* and put in *Chat GPT*, asking it to identify the patterns, according to Swales and Feak (2012). Personal answer
7. Below you will find the Review section *MICUSP* from three different areas. According to Integral (author prominent) and Non-Integral (research prominent) Citations in the Literature Review by Feak and Swales (2014), can you identify them in the texts below? **Integral (author prominent) Citations** - “Jackson *et al.* discuss this and other correlations of ecosystem decline with destructive fishing practices, making it clear that comprehensive attempts to curb hypoxic and anoxic events via policy should consider the potential for both bottom-up and top-down effects to be significant in a target system”. **Non-Integral (research prominent) Citations** - “Urban development in areas surrounding coastal ecosystems has been

accompanied by changes in ecosystem structure”; “Thus in this scenario, we are able to observe how partisan politics might have served the American people”; “Constructed wetlands are considered to be favorable control mechanisms for storm water management because they reduce stormwater contaminant loading, provide increased detention capacity on site, and increase recreational opportunities and wildlife habitat”. Which areas can you find Integral Citation? **Natural Resources and Environment**. Which areas can you find Non-Integral Citations? **Natural Resources and Environment; Political Science; Civil & Environmental Engineering**. Highlight and discuss them with our colleagues.

8. Access an Academic Corpus (*BAWE, OCAE*) in *Sketch Engine* or *Lextutor* and read other parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. Could you find more Integral or Non-Integral Citations? Write four examples below and discuss with our colleagues. Personal answers.
9. With your Do Your Own Corpus in *Sketch Engine*, read other parts of Literature Review by using specific queries such as “Literature Review”, “Previous studies” and “Recent studies”. What can you say about the findings? Could you find more Integral or Non-Integral Citations? And which is the reason for your findings? Discuss with our colleagues. Personal answers.
10. Start planning your literature review consulting academic corpora and your own corpus. Write one example of each. Personal answers.

Chapter 10: Note on Chapter 10

Chapter 10 does not include an answer key because the activities’ results depend on the user’s chosen data. As each data may vary, the answers will differ accordingly.

Chapter 11: Note on Chapter 11

Chapter 11 does not include an answer key because the activities’ results depend on the user’s chosen data. As each data may vary, the answers will differ accordingly.

Chapter 12: Answer Key

USING THE TED CORPUS SEARCH ENGINE TO EXPLORE DISCOURSE ORGANIZATION EXPRESSIONS IN ACADEMIC SPOKEN LANGUAGE

Exercises

Table 1 – Suggested answers for stance, engagement, and topic elaboration expressions

Stance	Engagement	Topic elaboration
and I think that I can tell you I don't know if I don't know what I think this is might be able to so that we can to be able to we were able to	(I'm) + going to tell you and you can see as you can see I want to show + (you) I want you to if you look at if you want to let me show you let me tell you take a look at those of you who when you look at you can see here	a little bit about all of a sudden in the first place it turns out that let me give you + (an example on the other hand one of the first one of the most one of the reasons the problem is that this is a picture this is the first

Table 2 – Concordance lines for *a little bit about* in the TCSE (adapted).

01	a little bit about	the natural world can help you perceive the full spectrum around you all the time.
02	a little bit about	some just big ideas about this, and then get John back out here so we can talk interactively a little bit more and think and ask questions.
03	a little bit about	open-source security, because we've got to get better at security in this 21st century.
04	a little bit about	, you know, the origin of the universe.
05	a little bit about	my favorite subject: deception.
06	a little bit about	what exactly autism is.

- | | | |
|-----------|---------------------------|--|
| 07 | a little bit about | how the brain performs under stress. |
| 08 | a little bit about | a few of these different companies. |
| 09 | a little bit about | where technology's going. |
| 10 | a little bit about | , having been working obsessively over it for the last 16 years. |

Authors Bio

Paula Tavares Pinto is an Associate Professor in the Department of Modern Languages at São Paulo State University (Unesp, Brazil). She teaches courses for the B.A. in Languages and Linguistics and is a member of the postgraduate program in Linguistic Studies. Her research focuses on Data-Driven Learning for Academic and Specific Purposes, Translation Studies, and Terminology.

Beatriz Curti Contessoto is an associate member of CeRLA at Université Lumière Lyon 2, France. Her research interests include intralinguistic equivalences in Lusophone comparative law, the influence of diachrony on legal translation between Portuguese and French, the socio-cultural and historical dimensions of legal terminology, and improving terminological accessibility in both legal and medical contexts.

Carolina Tavares de Carvalho is a Ph.D. student in Applied Linguistics at Unesp/Ibilce, with a Master's in Applied Linguistics, in the research line of Pedagogy of the Lexicon and Translation Based on Corpora. She studies Data-Driven Learning (DDL), focusing on its application in developing countries contexts. She is a member of translation and corpora research groups, is involved in SDG-related projects, and holds a degree in Civil Engineering.

Cecília Nevack de Britto graduated in Law from Centro Universitário Toledo (2020) and is a student of Literature at Unesp/Ibilce, since 2022. She has experience as a foreign language teacher in private courses and as an assistant teacher in a bilingual school. During her academic journey, she conducted a research project funded by a PIBITI scholarship, focusing on English language teaching strategies. She currently works in English language teaching, specializing in reading strategies and the use of corpora tools to enhance students' learning.

Celso Fernando Rocha is a Professor in the Department of Modern Languages at São Paulo State University (Unesp, Brazil). He teaches courses for the B.A. and Spanish Language for the Teacher's Training Program. His research focuses on Translation Studies, Legal Terminology and Lexicon and Pedagogy.

Diva Cardoso de Camargo is a Professor of Translation Studies at the University of the State of São Paulo, Brazil. She lectures on translation theory and literary translation, and supervises a number of research students in the area of translation studies and literature. Her main area of research interest is the use of corpora as a resource for studying features of translation and the distinctive styles of individual translators.

Eduardo Batista da Silva holds a PhD in Linguistic Studies from Unesp and is currently a professor of English at the Universidade Estadual de Goiás (UEG). He also teaches in UEG's postgraduate program in Language, Literature, and Interculturality. His research focuses on English teaching, Lexicology, Corpus Linguistics, and Lexical Statistics. He is a member of several academic organizations, including the Linguistic Society of America, AAAL, and TESOL International Association.

Emiliana Fernandes Bonalumi is an Associate Professor at the Federal University of Rondonópolis, Brazil. She has experience in education and literature, with an emphasis on translation, and she works mainly on the following topics: linguistic education practices, foreign language teaching, and corpus-based translation studies.

Franciele Spinelli is a PhD candidate at the University of Queensland in Brisbane Australia, where she also works as a casual academic, learning designer and research assistant. She holds a Master Degree in Applied Linguistics from the same university and a Bachelor Degree in Languages and Education. Her research is currently focused on the use of GenAI to support the academic language needs of international students.

Francine de Assis Silveira is an Associate Professor in the Institute of Literature and Linguistic - ILEEL at the Federal University of Uberlândia-UFU (Brazil). She teaches courses for the Undergraduate Translation course (English/Portuguese - Bachelor's degree). Her research interests include Terminology, Translation Studies, Specialized Translation and Translator Training. Currently, she is a member of academic groups, including En-Corpora: Corpora-Based and Corpora-Driven Teaching (CNPq).

Jeane Cardoso Costa is a university professor and a Ph.D. candidate in the Linguistic Studies program at São Paulo State University (Unesp), São José do Rio Preto. She is a member of the En-Corpora Research Group. Her research focuses on Corpus Linguistics, Natural Language Processing, and lexical analysis of essays written in Portuguese.

Luciano Franco da Silva is a university professor at the Federal Institute of Education, Science, and Technology of Paraná (IFPR). He holds a Ph.D. in Linguistic Studies from the São Paulo State University (Unesp), and his research/teaching areas include corpus linguistics, English for specific purposes and academic popularization discourse.

Liliane Mantovani is an Assistant Professor in the Department of Modern Languages at Maringá State University (UEM, Brazil). She teaches courses for the B.A. in Languages, Linguistics and Translation. She is a member of the research group En-corpora (Corpus-Based and Corpus-Driven Teaching). Her research interests are Translation Studies, Corpus Linguistics and Academic Writing.

Luana Aparecida Nazzi Laranja holds a Master's degree in Linguistic Studies from the São Paulo State University (Unesp). Her research focuses on corpus linguistics, data-driven learning and teaching reading comprehension skills.

Luana Viana dos Santos Graduated in Languages with a qualification in Portuguese and English from the São Paulo State University “Júlio de Mesquita Filho” (Unesp). Holds a Master's degree in Linguistic Studies from Unesp, specializing in Applied Linguistics, with a research focus on Lexicon Pedagogy and Corpus-Based Translation. Works as an English teacher in public elementary and secondary education.

Maria Camila Morais de Sousa, Undergraduate in Languages with a qualification in Portuguese and English at São Paulo State University (Unesp, Brazil). She is developing corpus-based research in data-driven learning for teaching English for academic purposes (EAP).

Mayra Aparecida dos Santos is an English teacher with 16 years of experience. She holds a Master's degree in Linguistic Studies from Unesp and a Bachelor's degree in Translation and Interpretation from Unilago. Her research focuses on English teaching, particularly writing production, and she has experience in academic coordination and school administration.

Talita Serpa is a faculty member at Paula Souza Center and a collaborator professor in the Postgraduate Program in Linguistic Studies at São Paulo State University. Her primary interests are corpus-based translation studies, terminology, law translation, and social sciences. Currently, she is a member of various academic groups, including the Translation, Terminology, and Corpora Groups, and En-Corpora: Corpora-Based and Corpora-Driven Teaching (CNPq).

Publique com a gente e
compartilhe o conhecimento



www.lettraria.net

 Letraria®