

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

ROGÉRIO YAMADA

FERRAMENTAS BASEADAS EM INTELIGÊNCIA ARTIFICIAL COMO
ASSISTENTES DE ESCRITA ACADÊMICA EM INGLÊS PARA AUTORES EM
INGLÊS COMO SEGUNDA LÍNGUA: UMA ANÁLISE MULTI-DIMENSIONAL

MESTRADO EM LINGUÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

SÃO PAULO

2025

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

ROGÉRIO YAMADA

FERRAMENTAS BASEADAS EM INTELIGÊNCIA ARTIFICIAL COMO
ASSISTENTES DE ESCRITA ACADÊMICA EM INGLÊS PARA AUTORES EM
INGLÊS COMO SEGUNDA LÍNGUA: UMA ANÁLISE MULTI-DIMENSIONAL

Dissertação apresentada à banca examinadora da
Pontifícia Universidade Católica de São Paulo como
exigência parcial para obtenção do título de Mestre
em Linguística Aplicada e Estudos da Linguagem,
sob orientação do Prof. Dr. Antonio Paulo Berber
Sardinha.

SÃO PAULO

2025

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação de mestrado por processos fotocopiadores ou eletrônicos.

Assinatura: _____

Data: [work in progress]

E-mail: eyamrog@gmail.com

Curriculo Lattes: <http://lattes.cnpq.br/9204618373258699>

[Informação bibliográfica – a ser incluída após a defesa da dissertação]
--

Rogério Yamada

Ferramentas Baseadas em Inteligência Artificial como Assistentes de Escrita Acadêmica em Inglês para Autores em Inglês como Segunda Língua: Uma Análise Multi-Dimensional

Aprovada em: [work in progress]

Dissertação apresentada à banca examinadora da Pontifícia Universidade Católica de São Paulo como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem, sob orientação do Prof. Dr. Antonio Paulo Berber Sardinha.

Banca Examinadora

Prof^a. Dr^a. Marilisa Shimazumi

Prof^a. Dr^a. Simone Sarmiento

A Letícia

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Processo 131464/2023-0.

AGRADECIMENTOS

Gostaria de expressar minha mais profunda gratidão a todos os professores do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem (LAEL) por atuarem como meus mentores nesta área.

Um agradecimento especial ao Professor Tony Berber Sardinha, meu orientador neste projeto de pesquisa, que sempre foi para mim um modelo inspirador.

Também desejo agradecer à Professora Maria Cláudia Nunes Delfino e ao Professor Carlos Henrique Kauffmann pelo incentivo.

Por fim, gostaria de expressar meus sinceros agradecimentos a todos os colegas e membros da equipe que me acompanharam e apoiaram ao longo da minha trajetória.

'[R]egistros' são categorias culturais, não categorias científicas. Essas categorias podem ser descritas por suas características situacionais e linguísticas típicas. Mas elas não são definidas nesses termos. Na verdade, registros não possuem definições em termos de características necessárias. Em vez disso, culturas e línguas evoluem naturalmente em termos de tais organizações categóricas, sem qualquer fundamento científico.

FERRAMENTAS BASEADAS EM IA COMO ASSISTENTES DE ESCRITA
ACADÊMICA EM LÍNGUA INGLESA PARA AUTORES EM INGLÊS COMO
SEGUNDA LÍNGUA: UMA ANÁLISE MULTIDIMENSIONAL

Rogério Yamada

2025

Resumo

Pesquisadores que não têm o inglês como língua materna frequentemente precisam escrever em inglês, pois a predominância do inglês na academia o estabeleceu como o principal meio de publicação em periódicos acadêmicos de qualidade (Baumvol, Sarmento, & Da Luz Fontes, 2021; Belcher, 2007; Cargill & Burgess, 2008; Flowerdew, 2012). No entanto, o desafio reside em utilizar uma linguagem idiomática que atenda às expectativas de registro do inglês acadêmico. O registro acadêmico em inglês baseia-se em padrões retóricos e lexicogramaticais específicos (Ädel & Erman, 2012; Biber & Barbieri, 2007), que diferem de muitos dos presentes na formação acadêmica dos autores (Ädel & Erman, 2012; Pang, 2010; Wray, 2019). Partindo do pressuposto de que Modelos Amplos de Linguagem, como o ChatGPT, foram extensivamente treinados no inglês acadêmico, este estudo buscou verificar até que ponto ferramentas baseadas em IA são capazes de ajudar escritores em inglês como L2 a atender às necessidades lexicogramaticais e retóricas do inglês acadêmico. Pesquisas anteriores sobre o inglês acadêmico gerado por IA sugerem que tais ferramentas não necessariamente reproduzem as escolhas retóricas ou lexicogramaticais dos autores humanos (Berber Sardinha, 2024). Para aprofundar nessa questão, o estudo empregou um corpus de textos produzidos por falantes de inglês como L2 (English-as-L2 Authored Papers – EL2AP), compilado a partir do repositório SciELO Pré-prints (*SciELO Preprints*, n.d.), e um corpus de textos publicados em periódicos de qualidade (Quality Journals Published Papers – QJPP), refletindo as mesmas disciplinas do EL2AP. Foram incluídos apenas artigos submetidos antes do advento do ChatGPT. Os textos do EL2AP foram revisados com o ChatGPT e reunidos no corpus AI-EL2AP. Para avaliar as similaridades e diferenças entre textos produzidos por humanos e por IA, foi realizada uma Análise Multidimensional (Biber, 1988, 1995) por meio de uma análise aditiva (Berber Sardinha, Pinto, Mayer, Zuppari, & Kauffmann, 2019). De modo geral, os resultados indicaram que a redação acadêmica assistida por IA diverge dos padrões humanos ao recorrer a padrões não típicos do inglês acadêmico.

Palavras-chave: Análise Multidimensional; Autores em inglês como segunda língua; Linguagem gerada por IA; Registro inglês acadêmico

AI-POWERED TOOLS AS ACADEMIC ENGLISH WRITING ASSISTANTS
TO ENGLISH-AS-L2 AUTHORS: A MULTI-DIMENSIONAL ANALYSIS

Rogério Yamada

2025

Abstract

Non-English-speaking scholars are often required to write in English, as the predominance of English in academia has established it as the primary medium for quality academic journals (Baumvol et al., 2021; Belcher, 2007; Cargill & Burgess, 2008; Flowerdew, 2012). However, the challenge lies in writing idiomatic language that meets the register expectations of academic English. The academic register in English relies on specific rhetorical and lexicogrammatical patterns (Ädel & Erman, 2012; Biber & Barbieri, 2007) that differ from many of those in their authors' academic background (Ädel & Erman, 2012; Pang, 2010; Wray, 2019). Assuming that Large Language Models such as ChatGPT have been extensively trained on academic English, this study aimed to verify the degree to which AI-powered tools are capable of assisting English-as-L2 writers in meeting the lexicogrammatical and rhetorical needs of academic English. Previous research in AI-generated academic English suggests that AI tools do not necessarily reproduce the rhetorical or lexicogrammatical choices of human authors (Berber Sardinha, 2024). To further explore this, the study employed an English-as-L2-Authored Papers (EL2AP) corpus compiled from the SciELO Preprints archive (*SciELO Preprints*, n.d.) and a Quality Journals Published Papers (QJPP) corpus reflecting the same disciplines as EL2AP. Only articles submitted before the advent of ChatGPT were included. EL2AP texts were revised with ChatGPT and compiled into the AI-EL2AP corpus. To gauge the similarities and differences between human- and AI-generated texts, a Multi-Dimensional Analysis (Biber, 1988, 1995) was carried out through an additive analysis (Berber Sardinha et al., 2019). Overall, the results indicated that AI-assisted academic writing diverges from human standards by relying on non-typical patterns of academic English.

Keywords: AI-generated language; English academic register; English-as-L2 authors; Multi-Dimensional Analysis

Lista de Figuras

Lista de Tabelas

Listings

Sumário

1	Introduction	16
1	Appendices	19
1.1	work in progress	19

1 Introduction

Referências Bibliográficas

- Ädel, A., & Erman, B. (2012, April). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92. doi: 10.1016/j.esp.2011.08.004
- Baumvol, L., Sarmiento, S., & Da Luz Fontes, A. B. A. (2021, August). Scholarly publication of Brazilian researchers across disciplinary communities. *Journal of English for Research Publication Purposes*, 2(1), 5–29. doi: 10.1075/jerpp.20012.bau
- Belcher, D. D. (2007, March). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1), 1–22. doi: 10.1016/j.jslw.2006.12.001
- Berber Sardinha, T. (2024, April). AI-generated versus human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. doi: 10.1016/j.acorp.2023.100083
- Berber Sardinha, T., Pinto, M. V., Mayer, C., Zuppari, M. C., & Kauffmann, C. H. (2019). Adding Registers to a Previous Multi-Dimensional Analysis. In T. Berber Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (1st ed., pp. 165–186). London: Bloomsbury Academic.
- Biber, D. (1988). *Variation across speech and writing* (1st ed.). Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge; New York: Cambridge University Press.
- Biber, D., & Barbieri, F. (2007, January). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. doi: 10.1016/j.esp.2006.08.003
- Biber, D., & Egbert, J. (2023, June). What is a register?: Accounting for linguistic and situational variation within – and outside of – textual varieties. *Register Studies*, 5(1), 1–22. doi: 10.1075/rs.00004.bib
- Cargill, M., & Burgess, S. (2008, April). Introduction to the Special Issue: English for Research Publication Purposes. *Journal of English for Academic Purposes*, 7(2), 75–76. doi: 10.1016/j.jeap.2008.02.006
- Flowerdew, J. (2012, November). English for Research Publication Purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (1st ed., pp. 301–321). Wiley. doi: 10.1002/9781118339855.ch16
- Pang, W. (2010, October). Lexical Bundles and the Construction of an Academic Voice: A Pedagogical Perspective. *The Asian EFL Journal*, 47.
- SciELO Preprints. (n.d.). <https://preprints.scielo.org/index.php/scielo>.
- Wray, A. (2019). Concluding question: Why don't second language learners more proactively target

formulaic sequences? In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 248–269). New York London: Routledge, Taylor & Francis Group. doi: 10.4324/9781315206615

1 Appendices

1.1 work in progress