

UNIVERSITATEA DE VEST DIN TIMIȘOARA
ȘCOALA DOCTORALĂ DE ȘTIINȚE UMANISTE

DOMENIUL FILOLOGIE

TEZĂ DE DOCTORAT

Conducător științific:

CS II Habil. Dr. Mădălina CHITEZ

Comisie de îndrumare:

Prof. Univ. Dr. Habil. Loredana PUNGĂ, Universitatea de Vest din Timișoara

Prof. Univ. Dr. Habil. Codruța GOȘA, Universitatea de Vest din Timișoara

Conv. Univ. Dr. Irina Diana MĂDROANE, Universitatea de Vest din Timișoara

Lect. Univ. Dr. Claudia Ioana DOROHOLSCHI, Universitatea de Vest din Timișoara

Doctorand:

Andreea Dincă

Timișoara

2025

UNIVERSITATEA DE VEST DIN TIMIȘOARA
ȘCOALA DOCTORALĂ DE ȘTIINȚE UMANISTE
DOMENIUL FILOLOGIE

REZUMATUL TEZEI DE DOCTORAT

***Academic Phraseology in Romanian Learner English: a
Corpus-Based Applicative Study***

Conducător științific:

CS II Habil. Dr. Mădălina CHITEZ

Comisie de îndrumare:

Prof. Univ. Dr. Habil. Loredana PUNGĂ, Universitatea de Vest din Timișoara

Prof. Univ. Dr. Habil. Codruța GOȘA, Universitatea de Vest din Timișoara

Conv. Univ. Dr. Irina Diana MĂDROANE, Universitatea de Vest din Timișoara

Lect. Univ. Dr. Claudia Ioana DOROHOLSCHI, Universitatea de Vest din Timișoara

Doctorand:

Andreea Dincă

Timișoara

2025

Rezumat

Pe măsură ce limba engleză se impune ca *lingua franca* a mediului academic la nivel global, numeroase universități oferă cursuri de limba engleză în scopuri academice (EAP – *English for Academic Purposes*), în cadrul cărora studenții învață să scrie texte academice. Dincolo de activitatea didactică, domeniul EAP include și o componentă de cercetare menită să sprijine procesul de predare. Una dintre abordările de cercetare utilizate este lingvistica bazată pe corpus, unde un corpus lingvistic este analizat cu ajutorul unor instrumente digitale. Cercetările EAP asistate de corpus au demonstrat că scrierea academică are o natură intrinsec frazeologică (vezi, de exemplu, Simpson-Vlach & Ellis, 2010). Astfel, competența comunicativă a unui autor poate depinde de capacitatea sa de a stăpâni expresiile-cheie utilizate în domeniul respectiv. Totuși, această abilitate nu se dezvoltă în mod natural la cei care se află la început de drum în scrierea academică, în special la cei care sunt și vorbitori de limba engleză ca limbă străină, care se confruntă frecvent cu dificultăți în utilizarea combinațiilor lexicale recurente ce îndeplinesc funcții discursive specifice (Flowerdew, 2015, p. 469).

În ultimii ani, sprijinul pentru dezvoltarea scrierii academice a fost tot mai des oferit prin intermediul unor instrumente inteligente dezvoltate atât în medii academice, cât și comerciale. Aceste instrumente variază de la corpusuri digitale care conțin exemple de texte academice de nivel expert până la platforme ce oferă feedback instantaneu unui text. Recent, disponibilitatea pe scară largă a Inteligenței Artificiale Generative (GenAI) a adus atât oportunități noi de învățare, cât și preocupări legate de onestitatea academică, determinând universitățile și profesorii de scriere academică să elaboreze ghiduri privind utilizarea etică a acestora. Deși instrumentele digitale sprijină procesul de învățare, ele sunt în mare parte dezvoltate pornind de la modele de tip expert ale scrierii, fără a lua în considerare nevoile lingvistice individuale ale celor care învață o limbă străină. Rezultatele empirice tot mai numeroase din domeniul *Learner Corpus Research (LCR)* oferă perspective diferite asupra nevoilor și dificultăților specifice ale studenților non-nativi care urmează cursuri EAP. Din păcate, aceste rezultate sunt rareori luate în considerare în proiectarea platformelor digitale de sprijin pentru scrierea academică.

Această teză își propune să acopere această lacună prin oferirea unor modele de suport frazeologic pentru scrierea academică în limba engleză, adaptate nevoilor specifice ale studenților vorbitori nativi de limba română. Studiul investighează mai întâi nevoile existente

prin analiza a două concepte frazeologice-cheie: lexical bundles (grupurile lexicale) și complexitatea frazeologică. Pe baza rezultatelor acestei analize, sunt propuse trei aplicații practice pentru sprijinirea scrierii academice. Prima este o colecție de activități de tip data-driven learning (DDL), care pot fi utilizate fie cu instrumente digitale pe bază de corpus, fie sub formă de fișe de lucru. A doua este proiectarea conceptuală a unui instrument digital denumit PhraseAPP, dezvoltat pentru a oferi sprijin frazeologic în scrierea academică. Un punct forte al PhraseAPP ar fi concentrarea sa pe nevoile specifice ale studenților români care scriu în limba engleză la nivel universitar. În al treilea rând, studiul își propune să evalueze potențialul GenAI ca instrument de îmbunătățire a acurateții frazeologice în scrierea studenților și reflectează asupra integrării etice a acestuia în predarea scrierii academice la nivel universitar.

Metodologia principală de cercetare utilizată este lingvistica bazată pe corpus, care permite atât tehnici cantitative, cât și calitative de analiză. Studiile de caz efectuate includ analize ale *lexical bundles*, analizarea complexității frazeologice și o evaluare calitativă a feedbackului generat de ChatGPT (OpenAI, 2024). Instrumentele utilizate variază de la platforme clasice de interogare a corpusurilor, cum ar fi LanCSBox X (Brezina & Platt, 2024), la limbajul de programare Python (Python Software Foundation, n.d.). Cercetarea utilizează trei corpusuri lingvistice. Corpusul principal analizat este *The Romanian Genre Corpus* (ROGER) (Chitez et al., 2021), un corpus bilingv, comparabil, care conține texte academice scrise de studenți de la mai multe universități românești. Celelalte două corpusuri utilizate pentru comparație sunt *Corpus of Expert Writing in Romanian and English* (EXPRES) (Chitez et al., 2022), un corpus bilingv și multidisciplinar care include texte academice scrise de experți, și *The British Academic Written English Corpus* (BAWE) (Nesi & Gardner, 2012), o colecție de texte scrise de studenți competenți din învățământul superior britanic.

Rezultatele principale ale studiului evidențiază mai multe dificultăți frazeologice întâmpinate de studenții români, vorbitori nativi de limba română, care scriu în limba engleză. Una dintre problemele principale este utilizarea frecventă a expresiilor autoreferențiale și gama limitată de strategii de exprimare a precauției (*hedging*), ceea ce duce la un ton subiectiv, în contrast cu formulările mai obiective din scrierea academică de tip expert. De asemenea, studenții se bazează excesiv pe un set restrâns de lexical bundles familiare pentru structurarea textelor, recurgând adesea la expresii informale care afectează stilul academic. Analiza pe domenii a arătat că fiecare disciplină (Științe Umaniste, Științe Politice, Economie și

Tehnologia Informației) are propriile convenții frazeologice, subliniind nevoia unui sprijin adaptat specificului disciplinar.

Analiza complexității frazeologice a arătat un nivel scăzut de sofisticare și diversitate în scrierea în limba engleză ca L2, în special în combinațiile verb-substantiv, în timp ce textele în limba română (L1) au prezentat un nivel mai ridicat de sofisticare, dar limitări similare în ceea ce privește variația. În cele din urmă, constatările calitative sugerează existența unui transfer din limba română în limba engleză, incluzând traduceri directe și utilizarea de „false friends”, care afectează suplimentar acuratețea frazeologică.

Această teză propune soluții concrete de suport frazeologic al scrierii academice în limba engleză a studenților români, vorbitori nativi de limba română. În primul rând, sunt prezentate exerciții DDL centrate pe ariile-cheie de îmbunătățire identificate în cadrul studiului (de exemplu, exprimarea atitudinii, stilul impersonal, sofisticarea frazeologică). În al doilea rând, este introdusă proiectarea conceptuală a unui instrument digital pentru frazeologie academică (PhraseAPP) care urmează să fie integrat în platforma web ce găzduiește corpul de scriere expertă EXPRES. Un avantaj al PhraseAPP este focusarea sa pe nevoile specifice ale studenților români care scriu în limba engleză. Platforma va include un modul cu materiale de învățare și exerciții bazate pe corpus, precum și o funcționalitate dedicată oferirii de feedback automat asupra utilizării frazeologiei de către studenți. Ultima soluție propusă este o serie de seminare dedicate utilizării responsabile a ChatGPT în contextul predării scrierii academice.

Studiul de față poate fi relevant atât pentru cercetători, cât și pentru profesorii care predau cursuri de scriere academică. Pe de o parte, studiul contribuie la literatura de specialitate privind scrierea studenților români în limba engleză și oferă un punct de plecare pentru cercetări viitoare în frazeologie. Conceptualizarea bazată pe cercetare a PhraseAPP oferă, de asemenea, o bază solidă pentru atragerea de fonduri în vederea dezvoltării platformei în cadrul unui proiect de cercetare. Pe de altă parte, soluțiile concrete și fundamentate empiric, prezentate sub forma exercițiilor DDL, pot fi implementate direct în activitatea de predare a scrierii academice.

Abstract

As English has established itself as the lingua franca of academia worldwide, many universities offer training for students in the form of English for Academic Purposes (EAP) classes, where students learn academic writing. Apart from teaching, the EAP field is also concerned with conducting research to better inform those classes. One such research approach is corpus linguistics, where a language corpus is analyzed using digital tools. Corpus-assisted EAP research has shown that academic writing is inherently phraseological (see, e.g., Simpson-Vlach & Ellis, 2010), and that a writer's communicative competence may depend on their ability to master the key phrases used in their field. However, this skill may not develop naturally in academic novices, particularly in those who are also L2 learners of English, who often face challenges in using recurrent word combinations that serve specific discursive functions (Flowerdew, 2015, p. 469).

In recent years, academic writing support has increasingly been offered through intelligent tools developed in both academic and commercial contexts. These tools, ranging from digital corpora with samples of successful academic writing to platforms offering instant feedback on drafts, can complement in-class teaching. More recently, the widespread availability of Generative Artificial Intelligence (GenAI) has brought both new opportunities for learning and raised concerns about academic dishonesty, prompting universities and writing instructors to develop guidelines for its ethical use. While digital tools show promise as academic support, they have largely been developed based on expert models of writing, often overlooking insights from actual learner language. Yet, the growing body of empirical findings from Learner Corpus Research (LCR) offers valuable perspectives on the specific needs and challenges of diverse learner populations in EAP. Unfortunately, these findings are rarely taken into account when designing writing support platforms.

The present study seeks to address this gap by proposing phraseological support models for academic writing in English, tailored specifically to the needs of first-language Romanian university students. It first investigates these needs by analyzing two key phraseological constructs: lexical bundles and phraseological complexity. Based on these findings, the study proposes three practical applications to support academic writing. First, it offers a collection of data-driven learning (DDL) activities, which can be used either with digital corpus tools or in paper-based worksheets. Second, it presents the conceptual design of a digital tool called PhraseAPP, developed to offer phraseological support for academic writing. A central strength

of PhraseAPP is its focus on the specific phraseological needs of L1 Romanian students writing in English at university level. Third, the study evaluates GenAI as a potential tool for improving phraseological accuracy in student writing and reflects on its ethical integration into university-level academic writing instruction.

The primary research methodology employed is Corpus Linguistics, which enables both quantitative and qualitative corpus analysis techniques. The analyses performed include lexical bundle analysis, phraseological complexity analysis, and a qualitative evaluation of ChatGPT (OpenAI, 2024)-generated feedback. The tools used in the analysis range from traditional corpus query platforms, such as LancsBox X (Brezina & Platt, 2024), to the programming language Python (Python Software Foundation, n.d.). This investigation uses three language corpora. The primary corpus under investigation is the Romanian Genre Corpus (ROGER) (Chitez et al., 2021), a bilingual, comparable learner corpus containing academic writing by students enrolled in several Romanian universities. The other two corpora used for comparison with ROGER are the Corpus of Expert Writing in Romanian and English (EXPRES) (Chitez et al., 2022), which is a bilingual, multidisciplinary corpus showcasing expert academic writing in English and Romanian, and The British Academic Written English Corpus (BAWE) (Nesi & Gardner, 2012), which is a collection of proficient student writing from British higher education.

The key findings from the study highlight several phraseological challenges faced by first-language Romanian university students writing in English. One major issue is the frequent use of self-referential stance expressions and a limited range of hedging strategies, which results in a subjective tone that contrasts with the more objective phrasing observed in expert academic writing. Students also rely heavily on a narrow set of familiar lexical bundles for structuring their texts, often resorting to informal expressions that undermine academic style. An analysis across disciplines revealed that each field (Philology, Political Science, Economics, and Engineering) has its own distinct phraseological conventions, emphasizing the need for discipline-sensitive writing support.

Phraseological complexity analysis showed low sophistication and diversity in English L2 writing, especially in verb-object combinations, while Romanian L1 texts displayed higher sophistication but similar limitations in variation. Finally, qualitative findings suggest a transfer from Romanian to English in collocation use, including direct translations and the use of false friends, which further hinders phraseological accuracy.

This thesis proposes solutions for phraseological support in academic writing in English for L1 Romanian university students. It first presents a collection of DDL exercises that focus on the key areas for improvement identified by this study (e.g., expression of stance, impersonal style, phraseological sophistication). Second, it introduces the conceptual design of a digital tool for academic phraseology, PhraseAPP, which is intended to be integrated into the web-based platform containing the expert writing corpus, EXPRES. A key strength of PhraseAPP is its focus on the specific needs of first-language Romanian students writing in English at university level. The platform will consist of a module containing learning materials and corpus-integrated exercises, alongside a functionality designed to provide automatic feedback on students' use of phraseology. The final solution proposed is a seminar series on the responsible use of ChatGPT in academic writing classroom contexts.

This study can be of value to both researchers and writing instructors. On the one hand, it contributes to the body of research on L1 Romanian student writing in English and offers a starting point for further investigation into phraseology. The research-based conceptualization of PhraseAPP also provides a solid foundation for applying to funding agencies that could support the development of this platform within a research project. On the other hand, the ready-made, research-informed solutions presented in the form of DDL exercises can be readily implemented in academic writing instruction.

References

- Brezina, V., & Platt, W. (2024). *#LancsBox X* [Software]. Lancaster University.
<http://lancsbox.lancs.ac.uk>
- Chitez, M., Bercuci, L., Dincă, A., Rogobete, R., & Csürös, K. (2021). *Corpus of Romanian Academic Genres (ROGER)* [Data set]. West University of Timișoara. <https://roger-corpus.org/>
- Chitez, M., Mureșan, V., Rogobete, R., & Dincă, A. (2022). *Corpus of Expert Writing in Romanian and English (EXPRES)* [Data set]. West University of Timișoara. <https://expres-corpus.org/>
- Flowerdew, L. (2015). Learner corpora and language for academic and specific purposes. In F. Meunier, G. Gilquin, & S. Granger (Eds.), *The Cambridge handbook of learner corpus research* (pp. 465–484). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139649414.021>
- OpenAI. (2024). *GPT-4 Omni* [Large language model]. OpenAI. <https://openai.com/research/gpt-4o>
- Python Software Foundation. (n.d.). *Python* [Programming language]. <https://www.python.org/>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>