



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

REZUMATUL TEZEI DE DOCTORAT

ENGLISH LANGUAGE TEACHING MEETS CORPUS LINGUISTICS:
a case study of upper-secondary students' argumentative essay writing in L2
English in Romania

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

Lect. Univ. Dr. Valentina MUREȘAN, Universitatea de Vest din Timișoara

Doctorand:

Ionela Tatiana MANDA (n. RACOLȚA)

Timișoara

2025

Table of Contents

Declarație	i
Acknowledgements	ii
List of figures.....	iii
List of tables.....	vi
List of acronyms and abbreviations	vii
Abstract.....	ix
Rezumat	xiv
Introduction.....	1
Argument.....	1
Aims and objectives	2
Thesis outline	4
Chapter 1. Theoretical framework: an overview of the key concepts	7
1.1. EFL in Romanian pre-university settings.....	7
1.1.1. Theoretical embedding.....	7
1.1.1.1. Key competences for lifelong learning	7
1.1.1.2. Foreign Language Acquisition.....	9
1.1.2. The evolution of the EFL curriculum in Romania.....	9
1.1.2.1. General considerations.....	9
1.1.2.2. The EFL Romanian school syllabus	11
1.1.2.3. Unique textbook vs. Alternative textbooks	12
1.1.3. Issues in EFL writing in Romanian pre-university settings.....	15
1.1.3.1. Instruction and assessment.....	15
1.1.3.2. Essay writing in national and international examinations	16
1.2. Learner Corpus Linguistics	22
1.2.1. Why corpus linguistics?.....	22
1.2.2. General overview of learner corpus linguistics	23
1.2.2.1. The international context	24
1.2.2.2. The Romanian context	26
1.2.3. Learner corpora in language education.....	26
1.3. Corpus based linguistic analyses.....	30
1.3.1. Lexical diversity in learner essays	30
1.3.2. Writing style in learner corpus essays	32
1.3.3. Coherence and cohesion in L2 argumentative writing	35

1.3.4. Lexical bundles in learner corpora.....	37
1.3.5. Interlanguage analysis.....	40
1.3.6. Writing with AI in educational settings	42
Chapter 2. Corpus-based approaches to teaching EFL at pre-university level	45
2.1. Learner English corpus research: the AGORA Project.....	45
2.1.1. Purpose.....	46
2.1.2. Design criteria	47
2.1.3. Representativeness	53
2.1.4. Processing	57
2.1.5. Biases, challenges and limitations	66
2.2. Methodology	68
2.3. Didactic avenues for AGORA.....	73
Chapter 3. Key linguistic features of AGORA essays	74
3.1. Corpus datasets.....	74
3.2. Contrastive Interlanguage Analysis	78
3.2.1. Lexico-grammatical features	80
3.2.1.1. Lexical diversity.....	80
3.2.1.2. Key word analysis.....	83
3.2.1.3. Pronoun-related features	88
3.2.1.4. Adjective-related features	92
3.2.1.5. Noun-related features.....	97
3.2.1.6. Verb-related features	100
3.2.1.7. Lexical bundles	102
3.2.2. Discoursal features: the use of cohesive devices	106
3.3. Lexical patterns in AGORA: Descriptive analysis	112
3.3.1. Idiomatic expressions.....	113
3.3.2. Collocations	119
3.4. Key results.....	138
Chapter 4. AGORA error analysis.....	142
4.1. Error analysis approaches.....	142
4.2. Analysis and patterns	143
4.2.1. Punctuation errors	146
4.2.2. Spelling errors.....	147
4.2.3. Style errors	147

4.2.3.1. Weak adjectives.....	148
4.2.3.2. Use of the sequencing discourse marker “first of all”	150
4.2.4. Grammar errors	152
4.2.4.1. Preposition errors	153
4.2.4.2. Article errors	158
4.2.4.3. Conditional clauses	159
4.2.5. Miscellaneous errors	164
4.2.5.1. Word order errors	165
4.2.5.2. Compound-word errors	167
4.2.6. Confused words	169
4.3. Pedagogical insights	171
Chapter 5. How does AI influence students’ essays?.....	173
5.1. Comparative analysis	174
5.1.1. Case study #1	174
5.1.2. Case study #2	176
5.1.3. Case study #3	179
5.1.4. Case study #4	182
5.1.5. Case study #5	185
5.2. The use of AI: pedagogical potentials and constraints.....	191
Chapter 6. Conclusions.....	197
6.1. Summary of findings.....	197
6.2. Pedagogical implications.....	198
6.3. Contributions to knowledge	201
6.4. Limitations	203
6.5. Further research.....	203
References.....	206
Webpage references.....	222
Appendices.....	225
Appendix 1	225
Appendix 2	231
Appendix 3	238
Appendix 4	240
Appendix 5	258

Abstract

The increasing demand for English proficiency in a globalised world implies a deeper understanding of the challenges faced by English as a Foreign Language (EFL) learners. While corpus linguistics offers valuable tools for analysing language use, its integration into EFL instruction, particularly in contexts like the pre-university system in Romania, remains limited. This PhD thesis addresses this gap by exploring the potential of corpus linguistics to inform and enhance argumentative essay writing instruction for upper-secondary Romanian EFL students. It contributes to the existing body of knowledge by providing empirical evidence and practical recommendations for pedagogical practices, ultimately aiming to improve the writing skills of non-native English learners in Romania.

The research is motivated by the disparity between the potential of corpus linguistics and its underuse in EFL classrooms. Despite the availability of various corpora, only a small fraction of educators actively engages with corpus research. This discrepancy calls for greater integration of corpus-informed methodologies in language education, extending beyond quantitative analyses to encompass the insights offered by qualitative approaches. By focusing on argumentative essay writing, a crucial skill evaluated in national and international English exams, this thesis directly addresses a practical need within the Romanian educational system.

The **purpose** of this thesis is to explore how corpus linguistics analysis can be used to inform and improve English language teaching in Romanian high schools. This central aim is achieved through the investigation of a newly compiled learner corpus and a comparative analysis with native speaker data, culminating in practical pedagogical recommendations. The thesis sets out to answer the following **research questions**:

1. How do L2 English high school students in Romania use grammatical categories, i.e. adjectives, nouns, verbs, and pronouns differently from native English speakers?
2. How does the lexical diversity in argumentative essays written by L2 English high school students in Romania compare to that of native English speakers?
3. What are the main similarities and differences in the use of cohesive devices between argumentative essays written by L2 English high school learners in Romania and those written by native English speakers?
4. What are the most recurrent lexical, grammatical, and cohesion-related errors in the argumentative essays written by L2 English high school students in Romania, and how does interlanguage influence (L1 Romanian vs. L1 Hungarian) contribute to these errors?

5. How do the writing styles of L2 English high school students in Romania differ based on their L1 background (Romanian vs. Hungarian), and what insights does this provide for enhancing English language teaching in Romanian schools?
6. How does the integration of AI impact the structural organisation and coherence of argumentative essays produced by L2 English high school students in Romania?
7. How can insights from the lexical, grammatical, and cohesion-related challenges in L2 English high school students' argumentative essay writing in Romania be applied to enhance instructional strategies for teaching writing in L2 English?

To address these research questions, the thesis employs a hybrid **methodology** that integrates quantitative and qualitative corpus analysis techniques. At the heart of this approach lies the AGORA (ArGumentative & Opinion Romanian learner essAys) corpus, a self-compiled learner corpus of 1,083 L2 English argumentative and opinion essays written by upper-secondary students (grades 9th-12th) in Romania between 2021 and 2023. Essays encompass proficiency levels B1, B2, and C1, and metadata includes demographic information, academic details, textual attributes, and administrative data. The data was collected with the support of 104 English teachers from 74 schools nationwide, ensuring an 81% coverage of Romanian counties. The AGORA corpus was compared to a sample from the Learner Corpus of Native English Essays (LOCNESS), used as a benchmark corpus. Tools like AntConc and LancsBox X were used for generating wordlists, frequency lists, and key word analyses. Further qualitative analysis involved the KWIC (Key Words in Context) tool, followed by error analysis using the LanguageTool API to automate error detection, which was then manually refined. Lastly, several essays were rewritten using ChatGPT and compared to the original texts.

The **thesis outline** reflects a logical progression through the research process:

- Chapter 1 provides the theoretical framework for understanding EFL education in Romania, focusing on curriculum development, writing instruction, and standardised assessments.
- Chapter 2 details the development and analysis of the AGORA corpus, outlining the design criteria, representativeness, ethical considerations, and processing steps.
- Chapter 3 presents a contrastive interlanguage analysis of the AGORA corpus essays, authored by Romanian and Hungarian L1 speakers, using the native-speaker LOCNESS corpus as a benchmark to analyse writing styles and lexico-grammatical and discoursal features. In addition to this contrastive approach, the chapter includes a descriptive analysis

of collocations and idiomatic expressions within AGORA, providing a more nuanced view of lexical choices and complementing the findings of the contrastive analysis.

- Chapter 4 explores language acquisition through a detailed error analysis of the AGORA essays, identifying recurring errors, exploring their potential sources (with particular attention paid to L1 interference), and offering pedagogical insights.
- Chapter 5 considers the role of artificial intelligence language models, particularly ChatGPT, in shaping L2 writing, comparing original student essays from AGORA with AI-generated revisions.
- Chapter 6 concludes the thesis by summarising the main findings, outlining pedagogical implications, highlighting contributions to the field, and addressing limitations with suggestions for future research.

The **novelty** of this research lies in the creation and in-depth analysis of the AGORA corpus, which is the first digitalised English learner corpus specifically designed for Romanian pre-university settings. This enables a nuanced, data-driven analysis of linguistic features, grammatical categories, lexical diversity, cohesion, and interlanguage influence (L1 Romanian vs. L1 Hungarian), all of which had not been previously analysed. Furthermore, this was done while simultaneously accounting for AI usage on school essay writing, which had a large impact to the coherence of language styles throughout highschoolers' essays.

The **obtained results** reveal several key findings:

- Romanian and Hungarian EFL learners display distinct linguistic profiles.
- Both learner groups significantly underuse the definite article compared to native speakers.
- In both learner groups, there was an overuse of additive and summarising cohesive devices compared to other cohesive devices.
- Romanian and Hungarian L1 students rely heavily on formulaic structures.
- Romanian and Hungarian L1 students tend to use stative verbs more frequently, while underusing other verbs with multiple semantic values.
- Common grammar and lexical errors were identified, as well as areas for improvement in terms of sentence structure, lexical diversity, and style.

- Several types of errors are attributable to L1 interference from Romanian, including instances of double comparatives, incorrect preposition use, article misuse, and word order errors.
- AI-generated revisions of students' essays enhanced quality across multiple aspects, including coherence, structural organization, grammatical accuracy, and lexical sophistication.
- Essays revised by AI show a quantitative convergence in lexical diversity, approaching proficient native-speaker levels.
- AI tools like ChatGPT can be used to help students develop their essays, but there must be a balance so as not to threaten originality and to avoid homogenisation of writing styles, with time and effort still required on the students' part.

This study's **personal contribution** involves the development of new methods that incorporate AI into language learning, as well as providing a novel application of corpus linguistics in a pre-university setting.

In terms of **pedagogical implications**, the thesis offers specific recommendations for Romanian EFL teachers, including the development of contrastive awareness exercises, targeted vocabulary instruction, and strategies for balancing AI assistance with authentic writing skills. The thesis emphasises the importance of addressing recurrent grammatical errors, L1-triggered lexical errors, and the overuse of formulaic expressions. Furthermore, the research suggests the need for more explicit instruction in argumentative writing strategies and techniques for promoting balanced argumentation.

On the whole, this thesis has **impact** in the field of English language education by providing empirical evidence and practical recommendations, enriching pedagogical practices for non-native English learners in Romania. By bridging the gap between corpus linguistics and EFL teaching, this research offers valuable insights for educators, curriculum designers, and researchers aiming to enhance the writing proficiency of Romanian students and prepare them for success in academic and professional settings. The work also contributes to the discourse on ethical and productive use of AI in pedagogical settings.

Rezumat

Creșterea tot mai acerbă de competențe lingvistice în limba engleză într-o lume globalizată subliniază necesitatea unei înțelegeri aprofundate a dificultăților cu care se confruntă elevii care învață engleza ca limbă străină. În timp ce lingvistica de corpus oferă instrumente valoroase pentru analiza utilizării limbii, integrarea sa în predarea limbii engleze ca limbă străină, în special în contexte precum sistemul preuniversitar din România, rămâne limitată. Prezenta teză de doctorat abordează această lacună, explorând potențialul lingvisticii de corpus de a informa și îmbunătăți predarea redactării de eseuri argumentative elevilor de liceu din România care studiază limba engleză. Acest studiu contribuie la literatura de specialitate existentă, oferind dovezi empirice și recomandări pentru abordările pedagogice, cu scopul final de a îmbunătăți abilitățile de scriere ale elevilor vorbitori de engleză non-nativi din România.

Cercetarea este motivată de discrepanța dintre potențialul lingvisticii de corpus și utilizarea sa relativ rară la orele de limba engleză. În ciuda disponibilității diverselor corpusuri, doar o mică parte dintre profesori se implică activ în cercetarea pe corpus. Această discrepanță necesită o integrare mai bună a metodologiilor bazate pe corpus în educația lingvistică, depășind analizele cantitative pentru a include perspectivele oferite de abordările calitative. Concentrându-se pe redactarea de eseuri argumentative, o abilitate crucială evaluată în examenele naționale și internaționale de limbă engleză, această teză abordează direct o nevoie practică din cadrul sistemului educațional românesc.

Scopul acestei teze este de a explora modul în care analiza lingvisticii de corpus poate fi utilizată pentru a informa și îmbunătăți predarea limbii engleze în liceele din România. Acest obiectiv central este atins prin investigarea unui corpus nou compilat și prin realizarea unei analize comparative cu date de la vorbitori nativi, culminând cu recomandări pedagogice practice. Teza își propune să răspundă la următoarele **întrebări de cercetare**:

1. Cum utilizează elevii de liceu din România care învață engleza ca limbă străină categoriile gramaticale, precum adjectivele, substantivele, verbele și pronumele, în manieră diferită față de vorbitorii nativi de engleză?
2. Cum se compară diversitatea lexicală din eseurile argumentative scrise de elevii de liceu din România care învață engleza ca limbă străină cu cea a vorbitorilor nativi de engleză?

3. Care sunt principalele asemănări și diferențe în utilizarea conectorilor între eseurile argumentative scrise de elevii de liceu din România care învață engleza ca limbă străină și cele scrise de vorbitorii nativi de engleză?
4. Care sunt cele mai recurente erori lexicale, gramaticale și de coeziune în eseurile argumentative scrise de elevii de liceu din România care învață engleza ca limbă străină și cum contribuie influența interlingvistică (L1 română vs. L1 maghiară) la aceste erori?
5. Cum diferă stilurile de scriere ale elevilor de liceu din România care învață engleza ca limbă străină, în funcție de L1 (română vs. maghiară), și ce perspective oferă acest lucru pentru îmbunătățirea predării limbii engleze în școlile românești?
6. Cum influențează integrarea inteligenței artificiale organizarea structurală și coerența eseurilor argumentative produse de elevii de liceu din România care învață engleza ca limbă străină?
7. Cum pot fi aplicate perspectivele din provocările lexicale, gramaticale și de coeziune în redactarea de eseuri argumentative de către elevii de liceu din România care învață engleza ca limbă străină pentru a îmbunătăți strategiile de instruire în predarea scrierii argumentative în engleză?

Pentru a răspunde la aceste întrebări de cercetare, teza folosește o **metodologie** hibridă care integrează tehnici de analiză de corpus cantitativă și calitativă. În centrul acestei abordări se află corpusul AGORA (ArGumentative & Opinion Romanian learner essAys), un corpus compilat de autorul tezei de doctorat, format din 1.083 de eseuri argumentative și de opinie în limba engleză, scrise de elevi de liceu (clasele a IX-a – a XII-a) din România, între 2021 și 2023. Eseurile acoperă nivelurile de competență B1, B2 și C1, iar metadatele includ informații demografice, detalii academice, atribute textuale și date administrative. Datele au fost colectate cu sprijinul a 104 profesori de engleză din 74 de școli la nivel național, asigurând o acoperire de 81% a județelor României. Corpusul AGORA a fost comparat cu un eșantion din Learner Corpus of Native English Essays (LOCNESS), folosit în calitate de corpus de referință. Au fost utilizate instrumente precum AntConc și LanksBox X pentru generarea de liste de cuvinte, liste de frecvență și analize de cuvinte cheie. Analiza calitativă suplimentară a implicat instrumentul KWIC (Key Words in Context), urmat de analiza erorilor utilizând LanguageTool API pentru a automatiza detectarea erorilor, care a fost apoi rafinată manual. În cele din urmă, mai multe eseuri au fost rescrise folosind ChatGPT și comparate cu textele originale.

Structura tezei reflectă o progresie logică prin procesul de cercetare:

- Capitolul 1 oferă cadrul teoretic pentru înțelegerea predării limbii engleze ca limbă străină în România, concentrându-se pe dezvoltarea curriculumului, predarea redactării de eseuri și evaluările standardizate.
- Capitolul 2 detaliază dezvoltarea și analiza corpusului AGORA, prezentând criteriile de proiectare, reprezentativitatea, considerațiile etice și pașii de procesare.
- Capitolul 3 prezintă o analiză contrastivă interlingvistică utilizând corpusul AGORA, comparând eseurile redactate de vorbitori nativi de română și de maghiară și folosind corpusul LOCNESS ca referință pentru a analiza stilurile de scriere și trăsăturile lexico-gramaticale și discursive. Pe lângă această abordare contrastivă, capitolul include și o analiză descriptivă a colocațiilor și expresiilor/locuțiunilor idiomatice din AGORA, oferind o perspectivă mai nuanțată asupra alegerilor lexicale ale elevilor și completând concluziile analizei contrastive.
- Capitolul 4 explorează achiziția limbii printr-o analiză detaliată a erorilor din eseurile AGORA, identificând greșelile recurente, examinând sursele lor posibile (cu un accent deosebit pe interferența L1) și oferind perspective pedagogice.
- Capitolul 5 ia în considerare rolul modelelor lingvistice de inteligență artificială, în special ChatGPT, în modelarea scrierii L2, comparând eseurile originale ale elevilor din AGORA cu revizuirile generate de inteligența artificială.
- Capitolul 6 concluzionează teza prin rezumarea principalelor observații, prezentarea implicațiilor pedagogice, evidențierea contribuțiilor la domeniu și abordarea limitărilor cu sugestii pentru cercetări viitoare.

Noutatea acestei cercetări constă în crearea și analiza aprofundată a corpusului AGORA, care este primul corpus digitalizat obținut din și destinat în special învățământului preuniversitar din România. Acest lucru permite o analiză nuanțată, bazată pe date, a caracteristicilor lingvistice, a categoriilor gramaticale, a diversității lexicale, a coeziunii și a influenței interlingvistice (L1 română vs. L1 maghiară), toate acestea nefiind analizate în studii anterioare. În plus, studiul de față a analizat, de asemenea, impactul utilizării inteligenței artificiale asupra scrierii de eseuri școlare și asupra stilurilor lingvistice din eseurile liceenilor din România.

Rezultatele obținute dezvăluie mai multe descoperiri cheie:

- Elevii români și maghiari care studiază limba engleză ca L2 au profiluri lingvistice distincte.
- Ambele grupuri de elevi prezintă o puternică tendință de a evita sau omite articolul hotărât în eseurile lor în comparație cu vorbitorii nativi de engleză.
- În cele două grupuri, s-a remarcat utilizarea mult mai frecventă a conectorilor copulativi și de sinteză/rezumare în comparație cu alți conectori.
- Elevii români și maghiari se bazează foarte mult pe structuri de tip șablon.
- Vorbitorii nativi de română și maghiară tind să folosească mai frecvent verbe statice, în timp ce utilizează mai rar verbe dinamice sau cu valori semantice multiple.
- Au fost identificate erori gramaticale și lexicale frecvente, precum și curențe în ceea ce privește structura propozițiilor, diversitatea lexicală și elementele stilistice.
- Mai multe tipuri de erori pot fi atribuite transferului lingvistic din limba maternă, inclusiv utilizarea comparativelor duble (folosirea simultană a doi indicatori ai gradului de comparație), folosirea incorectă a prepozițiilor, erori în utilizarea articolelor și greșeli de ordonare a cuvintelor.
- Revizuirile generate de inteligența artificială asupra eseurilor elevilor au îmbunătățit în mod consecvent calitatea acestora pe mai multe planuri, inclusiv în privința coerenței, a organizării structurale, a preciziei gramaticale și a bogăției lexicale.
- Eseurile revizuite de inteligența artificială demonstrează o convergență cantitativă în diversitatea lexicală, apropiindu-se de nivelurile vorbitorilor nativi avansați.
- Instrumentele AI precum ChatGPT pot fi utilizate pentru a ajuta elevii să-și dezvolte eseurile; cu toate acestea, este esențial să se păstreze un echilibru care să protejeze originalitatea și să evite uniformizarea stilurilor de scriere, menținându-se, în mod necesar, implicarea reală (în timp și efort) a elevilor.

Contribuția personală a acestui studiu constă în dezvoltarea de noi metode care să încorporeze inteligența artificială în predarea-învățarea limbilor străine, precum și în promovarea unei abordări inovatoare de aplicare a lingvisticii de corpus în învățământul preuniversitar.

În ceea ce privește **implicațiile pedagogice**, teza oferă recomandări specifice pentru profesorii români care predau limba engleză ca limbă străină, inclusiv elaborarea de exerciții de sensibilizare a elevilor la diferențele lingvistice prin analize contrastive, de predare orientată spre îmbogățirea vocabularului și strategii pentru armonizarea echilibrată a utilizării inteligenței artificiale cu scrierea autentică. Teza subliniază importanța abordării erorilor gramaticale recurente, a erorilor lexicale care își au originea în L1 și a utilizării excesive de expresii-șablon. Mai mult, cercetarea sugerează necesitatea unei instruirii mai explicite în strategii de scriere argumentativă și tehnici pentru promovarea argumentării echilibrate.

În ansamblu, această teză are un **impact** în domeniul educației lingvistice în limba engleză, oferind dovezi empirice și recomandări practice, îmbogățind practicile pedagogice pentru elevii non-nativi de limba engleză din România. Prin reducerea decalajului dintre lingvistica de corpus și predarea limbii engleze ca limbă străină, această cercetare oferă informații valoroase pentru profesori, creatorii de curriculum și cercetătorii care urmăresc să îmbunătățească competența de scriere a elevilor din România și să-i pregătească pentru a avea succes în mediile academice și profesionale. Lucrarea contribuie, de asemenea, la demersul academic de explorare și de înțelegere a utilizării inteligenței artificiale în mediile pedagogice în mod etic și productiv.

References (a selection)

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.
- Ädel, A., & Erman, B. (2012). *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.
- AlAfnan, M. A., Dishari, S., Jovic, M. & Lomidze, K. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60–68. <https://doi.org/10.37965/jait.2023.0184>.
- Altenberg, B. (1991). Amplifier collocations in spoken English. In S. Johansson & A-B. Stenström (Eds.), *English computer corpora: Selected papers and research guide* (pp. 127-147). Mouton de Gruyter.
- Bernardini, S. (2004). Corpora in the classroom. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.12.05ber>.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.
- Biber, D., & Barbieri, F. (2007). Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*, 26, 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>.
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173.
- Briana, J. C. D. (2024). Is ChatGPT-produced text authentic? A contrastive analysis of cohesive markers in human and AI-generated text. *Journal of English and Applied Linguistics*
- Cenoz, J. (2003) The Additive Effect of Bilingualism on Third Language Acquisition: A Review. *International Journal of Bilingualism*, 7, 71-87.

- Chen, Y.-H., & Baker, P. (2010). *Lexical bundles in L1 and L2 academic writing*. *Journal of English for Academic Purposes*, 9(3), 183–197.
- Chitez, M. (2012). *Lexical frequency profile applications on learner corpora: A Romanian learner English explorative analysis*. In P. Rodríguez-Puente, D. Tizón-Couto, B. Tizón-Couto, & I. Pastor-Gómez (Eds.), *New trends and methodologies in applied English language research II: Studies in variation, meaning and learning* (pp. 15–36). Peter Lang.
- Chitez, M. (2014). *Learner corpus profiles: The case of Romanian Learner English*. Bern, Berlin: Peter Lang.
- Chitez, M. (2018). Corpus linguistics meets academic writing: Examples of applications in the Romanian EFL context. In Chitez, M., Doroholschi C. I., Kruse, O., Salski, L. & Tucan, D. (Eds.), *University writing in Central and Eastern Europe: Tradition, transition, and innovation* (pp. 133-149). Springer International Publishing.
- Chitez, M. (2022). *Corpus of Romanian Academic Genres ROGER*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1700>.
- Chitez, M., Bercuci, L., Dincă, A., Rogobete, R., & Csűrös, K. (2021). *Corpus of Romanian Academic Genres (ROGER)*. West University of Timisoara. Available at <https://roger-corpus.org/>.
- Chitez, M., Rogobete, R. & Foitos, A. (2020). Digital Humanities as an Incentive for digitalisation strategies in Eastern European HEIs: a case study of Romania. In A. Curaj, L. Deca and R. Pricopie (Eds.), *European Higher Education Area: Challenges for a New Decade* (pp. 545-564). Cham: Springer.
- Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London, England: Continuum.
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202-224.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Cristea, D., & Forăscu, C. (2006). Linguistics Resources and Technologies for Romanian Language. *Journal of Computer Science of Moldova*, 14(1), 33-73.
- Crossley, S. A., & McNamara, D. S. (2012). *Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication*. *Journal of Research in Reading*, 35(2), 115–135.

- Crossley, S. A., Kyle, K., & McNamara, D. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1-16. <https://doi.org/10.1016/j.jslw.2016.01.003>
- Dagneaux, E., Denness, S. & Granger, S. (1998). Computer-aided error analysis, *System*, 26(2), 163-174.
- Dalton-Puffer, C. (2011). *Content and Language Integrated Learning: From Practice to Principles?*. Annual Review of Applied Linguistics, 31, 182-204.
- Dincă, A., & Chitez, M. (2021). Assessing learners' academic phraseology in the digital age: a corpus-informed approach to ESP texts. *Journal of Teaching English for Specific and Academic Purposes*.
- Ellis, R. (2008). *The Study of Second Language Acquisition (2nd ed.)*. Oxford University Press.
- Evert, S. (2004). *The statistics of word cooccurrences: Word pairs and collocations* (PhD dissertation). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Faigley, L., & Witte, S. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, 32(2), 189–204.
- Flor, M. & Beigman Klebanov, B. (2018). Catching Idiomatic Expressions in EFL essays. In *Proceedings of the Workshop on Figurative Language Processing* (pp. 34-44). New Orleans, Louisiana: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W18-0905>
- Flowerdew, L. (1999). A corpus based-analysis of referential and pragmatic errors in students' writing. Hong Kong University of Science and Technology, China.
- Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL*, 24(2), 152-168.
- Fontenelle, T. (1994). What on the earth are collocations: An assessment of the ways in which certain words cooccur and others do not. *English Today*, 10(4), 42–48.
- Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. New York and London: Routledge Taylor & Francis Group.
- Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse?. *Frontiers in Education*. 8. 10.3389/feduc.2023.1166682.
- Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, 8(4), 1-37 <http://www.tesl-ej.org/ej32/a1.html>.
- Gass, S. M. & Selinker, L. (2008). *Second language acquisition: An introductory course*. (3rd ed.). New York and London: Routledge.

- Gavioli, L. (2000). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Houston, TX: Athelstan/Bologna: CLUEB.
- Gilquin, G. (2015). From design to collection of learner corpora. In Granger S., Gilquin G. & Meunier F. (Eds.) *The Cambridge handbook of learner corpus research* (pp. 9–34). Cambridge: Cambridge University Press.
- Gilquin, G., Granger, S. & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-335.
- Gilquin, G., Papp, S., & Diez-Bedmar, M. B. (2008). *Linking up contrastive and learner corpus research*. Brill Publishers.
- González-Lloret, M. (2024). The future of language learning teaching in a technology-mediated 21st century. *The Modern Language Journal*, 108(2), 541-547.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford University Press.
- Granger, S. (1998). *Learner English on Computer*. Addison Wesley Longman.
- Granger, S. (2002). A bird's eye view of learner corpus research. In Granger, D., Hung, J. & Petch-Tyson, S. (Eds.). *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam and Philadelphia: Benjamins.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Language and Computers*, 52(1), 123-145.
- Granger, S. (2008). Learner corpora in foreign language education. In Van Deusen-Scholl, N. & Hornberger, N. H. (Eds.), *Encyclopedia of language and education* (Vol. 4, pp. 337-351). New York: Springer.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam/Philadelphia: John Benjamins.
- Granger, S., & Tyson, S. (1996). *Connector usage in the English essay writing of native and non-native EFL speakers*. *World Englishes*, 15(1), 17–27.
- Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Harmer, J. (2004). *How to teach writing*. Pearson Education.

- Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hyland, K. (2003). *Second Language Writing*. Cambridge University Press.
- Hyland, K. (2004a). *Genre and Second Language Writing*. University of Michigan Press.
- Hyland, K. (2004b). *Disciplinary Discourses: Social Interactions in Academic Writing*. University of Michigan Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4–21.
- Ionin, T., Ko, H. & Wexler, K. (2004). Article semantics in L2-acquisition: The role of specificity. *Language Acquisition*, 12, 3-69.
- Jarvis, S. (2013). *Defining and measuring lexical diversity*. Amsterdam: John Benjamins Publishing.
- Jelson, A. & Lee, S. W. 2024. An empirical study to understand how students use ChatGPT for writing essays and how it affects their ownership. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants*. Association for Computing Machinery, New York, NY, USA, 26–30. <https://doi.org/10.1145/3690712.3690720>.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund University Working Papers in Linguistics*, 53(2), 61-79.
- Kellogg, R. & Raulerson, B. (2007). Improving the writing skills of college students. *Psychonomic Bulletin and Review*, 14(2), 237-242.
- Kenesei, I., Vago, R. M., & Fenyvesi, A. (1998). *Hungarian: Descriptive grammars* (pp.151-155). Routledge.
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81(2), 345-381 <https://doi.org/10.1353/lan.2005.0071>.
- Kővári, A. (2025). Ethical use of ChatGPT in education—Best practices to combat AI-induced plagiarism. *Frontiers in Education*.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.

- Leech, G. (1997). Teaching and language corpora: a convergence. In Wichmann, A., Fligelstone, S., McEnery, A.M. & Knowles, G. (Eds.), *Teaching and language corpora* (pp. 1-23). Addison Wesley Longman.
- Leki, I. (1992). *Understanding ESL writers: A guide for teachers*. Boynton/Cook Publishers.
- Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (pp. 53-66). Longman.
- Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.
- Manda, I., & Chitez, M. (2022). Writing culture clash: English essay writing challenges in Romanian pre-university settings. In *Conference Proceedings of The Future of Education 2022* (pp.403-409). Filodiritto Editore https://doi.org/10.26352/G630_2384-9509.
- Marzuki, et al. (2023). *The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective*. *Cogent Education*, 10(1), 2236469.
- McCarthy, P., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392.
- McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 364-380). Routledge.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. Routledge.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication*, 27(1), 57-86.
- Mititelu, V. B., Tufiş, D., & Irimia, E. (2018). The reference corpus of the contemporary Romanian language (CoRoLa). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Available at: <http://corola.racai.ro/>.
- Mitrofan, M, Barbu Mititelu, V., & Mitrofan, G. (2019). MoNERo: a Biomedical Gold Standard Corpus for the Romanian Language. In *Proceedings of the BioNLP 2019 workshop*, Florence, Italy (pp. 71-79). Available at: <https://www.aclweb.org/anthology/W19-5008.pdf>.
- Mukherjee, J., & Rohrbach, J.-M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In Kettemann, B. & Marko, G. (Eds.), *Planning, gluing and painting corpora: Inside the applied corpus linguist's workshop* (pp. 205-232). Peter Lang.

- Naber, D. (2003). *A rule-based style and grammar checker* (Master's thesis). Bielefeld University. <https://languagetool.org>.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In Sinclair, J. (Ed.), *How to use corpora in language teaching* (pp. 125-152). John Benjamins. <https://doi.org/10.1075/scl.12.11nes>.
- Nesselhauf, N., & Cornelia T. (2002). Collocations in CALL: An investigation of vocabulary-building software for EFL. *Computer Assisted Language Learning*, 15(3), 251-279.
- O’Keeffe, A., McCarthy, M. J. & Carter, R. A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Osborne, J. (2002). Top-down and bottom-up approaches to corpora in language teaching. In Connor, U. and Upton, T. (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 251–265). Rodopi.
- Pană Dindelegan, G. (Ed.). (2013). *The grammar of Romanian* (pp. 415-417). Oxford University Press.
- Paquot, M. (2013). *Lexical bundles and L1 transfer effects*. *International Journal of Corpus Linguistics*, 18(3), 391–417.
- Paradis, C. (1997). *Degree modifiers of adjectives in spoken British English*. Lund University Press.
- Park, S. (2022). A Corpus-Based Comparison of Syntactic Complexity in Spoken and Written Learner Language. *Canadian Journal of Applied Linguistics*, 25(2), 47–70. <https://doi.org/10.37213/cjal.2022.32477>
- Pavičić Takač, V., et al. (2019). *Frame markers and coherence in L2 argumentative essays*. *Discourse and Interaction*, 12(2), 46–71.
- Pery-Woodley, M.M. (1990). Contrasting discourses: Contrastive analysis and a discourse approach to writing. *Language Teaching*, 23, 143-151.
- Popescu, T. (2013). A Corpus-based Approach to Translation Error Analysis. A Case-study of Romanian EFL Learners. *Procedia - Social and Behavioral Sciences*, 83, 242-247.
- Pungă, L., & Pârlog, H. (2015). ‘He is a criminal in series’: a foray into errors by Romanian Learners of English. *Professional Communication and Translation Studies*, 8, 161-176.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-134

- Römer, U. (2008a). Corpora and language teaching. In Lüdeling, A. & Kytö, M. (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 112-131). Walter de Gruyter.
- Römer, U. (2008b). Origin and history of corpus linguistics: Corpus linguistics vis-a-vis other disciplines. In Lüdeling A. & Kytö M. (Eds.), *Corpus linguistics: An international handbook* (Vol. 1 pp. 112-122). De Gruyter Mouton.
- Schiftner, B. (2014). (Non-)signalling of coherence structures in English learner writing. In H. Gruber & G. Redeker (Ed.), *The Pragmatics of Discourse Coherence: Theories and applications* (pp. 243-266). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.254.09sch>.
- Schmitt, N. & Carter, R. (2004). Formulaic sequences in action. *Language Learning & Language Teaching*, 1-22. <https://doi.org/10.1075/llt.9.02sch>.
- Seidlhofer, B. (2002). Pedagogy and local learner corpora: working with learning-driven data. In Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213-234). John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (1996). Preliminary recommendations on corpus typology. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards). Retrieved from www.ilc.cnr.it/EAGLES96/corpus typ/corpus typ.html (last accessed on July 11, 2023).
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Timmis, I. (2013). Corpora and materials: Towards a working relationship. In Tomlinson, B. (Ed.), *Developing materials for language teaching* (pp. 461-475). Bloomsbury.
- Timmis, I. (2015). *Corpus Linguistics for ELT: Research and Practice*. Routledge.
- Tseng, W., & Warschauer, M. (2023). *AI-writing tools in education: if you can't beat them, join them*. *Journal of China CALL*, 3(2), 150–168. ([AI-writing tools in education: if you can't beat them, join them](#)).
- Tufiş, D., Ion, R. Ceaşu A., & Stănescu, D. (2008). RACAI'S Linguistic Web Services. In *Proceedings of The 6th Language Resources and Evaluation Conference* (pp. 28-30).
- Virtanen, T. (1998). Direct questions in argumentative student writing. In S. Granger (Ed.), *Learner English on Computer* (pp. 94–106). Routledge.
- Zamel, V. (1992). Writing one's way into reading. *TESOL Quarterly*, 26, 463-485. <https://doi.org/10.2307/3587174>.

Webpage references (a selection)

- Anthony, L. (2024). AntConc (Version 4.3.1). Tokyo, Japan: Waseda University.
<https://www.laurenceanthony.net/software/AntConc>.
- Ask AI. (n.d.). Chat. <https://askaichat.app/chat>
- Baiango. (2023). List of English idioms. GitHub. https://github.com/baiango/english_idioms
- Brezina, V. & Platt, W. (2024) #LancsBox X. Lancaster University, <http://lancsbox.lancs.ac.uk>.
- Cambridge English. (2020). Teacher guide for writing B2 First.
https://www.cambridgeenglish.org/pl/Images/603901-cer_6647_v1b_jul20_teacher-guide-for-writing_b2_first_schools_prf2.pdf
- Cambridge English. (2020). Teacher guide for writing C1 Advanced.
https://assets.cambridgeenglish.org/schools/CER_6647_V1d_JUL20_Teacher-Guide-for-Writing-C1_Advanced.pdf
- Cambridge English. (n.d.). B2 First exam format. <https://www.cambridgeenglish.org/exams-and-tests/first/exam-format/>
- Cambridge English. (n.d.). C1 Advanced exam format.
<https://www.cambridgeenglish.org/exams-and-tests/advanced/>
- Cambridge English. (n.d.). Exams and tests. <https://www.cambridgeenglish.org/exams-and-tests/>
- Farlex. (n.d.). Go out with someone. In The Free Dictionary.
<https://idioms.thefreedictionary.com/go+out+with+someone>
- Lancaster University. (n.d.). Corpus linguistics: Method, analysis, interpretation: A practical introduction to the methodology of corpus linguistics for researchers in the social sciences and humanities. Future Learn. <https://www.futurelearn.com/courses/corpus-linguistics>
- O'Donnell, M. (2013). The UAM Corpus Tool. Retrieved from
<http://www.wagsoft.com/CorpusTool/>.
- ROGER. (2017). Retrieved from <http://roger.projects.uvt.ro>.
- ROMTEXT, digital corpus of texts, annotated and dated, of the Romanian language, 16th – 21st centuries. Available at <https://www.romtext.ro/>
- The Idioms. (n.d.). The Idioms Dictionary. <https://www.theidioms.com/>.
- UCREL. (n.d.). CLAWS 7 tagset. Lancaster University. <https://ucrel.lancs.ac.uk/claws7tags.html>.
- ZEROGPT AI DETECTOR www.zerogpt.com.