University of Zadar

International scientific conference

Corpora in Language Learning, Translation and Research

BOOK OF ABSTRACTS

University of Zadar

Zadar

August 27th - 28th, 2025

International scientific conference

Corpora in Language learning, Translation and Research

Book of Abstracts

Publisher

University of Zadar

For the Publisher

Josip Faričić

Edited by

Larisa Grčić and Marija Brkić Bakarić

ISBN

978-953-331-579-9

PROGRAMME COMMITTEE

Marija Brkić Bakarić (University of Rijeka, Croatia), Elisa Corino (University of Turin, Italy), Larisa Grčić (University of Zadar, Croatia), Ivana Lalli Paćelat (University Jurja Dobrile of Pula), Marek Łukasik (Pomeranian University in Słupsk, Poland), Paula de Santiago (University of Valladolid, Spain), Vanja Slavuj (University of Rijeka, Croatia), Špela Vintar (University of Ljubljana, Slovenia)

ORGANIZING COMMITTEE

Larisa Grčić (University of Zadar, Croatia), Marija Brkić Bakarić (University of Rijeka, Croatia)

Program

August 27th, 2025

08:00-8:30	***Registration***
08:30-9:00	**Opening – welcome speech**
09:00-10:15	***Plenary lecture*** **Mojca Pecman** (Université Paris Cité, France) *Terminology, Specialized Corpora, and LLMs: Show Me Your Weaknesses, and I Will Help You Develop Your Strengths*
10:15-10:30	**Coffee break**
10:30-12:00	**Session 1**
10:30-11:00	**Pertti Hietaranta** (University of Helsinki, Finland) *Language and Gender: Why Does Androcentrism in Finnish Disappear so Slowly?*
11:00-11:30	**Siyuan Liu** (University of Bologna, Italy) *Bridging User Needs with Interlanguage Corpus Design and Applications: The Case of MICICL*
11:30-12:00	**Andreja Drašler** and **Monika Kavalir** (University of Ljubljana, Slovenia) *Compilation of a Corpus-based Word List for Teaching English Geographical Vocabulary*
12:00-12:30	**Francisco Javier Fernández-Polo** and **Mario Cal Varela** (University Santiago de Compostela, Spain) *From Corpus to Classroom: Leveraging SUNCODAC to Enhance Interpersonal Communication Skills in Academic Online Forums* (online)
12:30-14:30	**Lunch break**

	Session 2
14:30-15:00	**Goretti Faya Ornia**, **Natalia Barranco Izquierdo** and **Teresa Calderón Quindós** (University of Valladolid, Spain) *The Use of an Aligned Parallel Corpus as a Didactic Tool for Phrasal Verbs: an Approach from Cognitive Linguistics* (online)
15:00-15:30	**Khalid Elasri** (Faculty of Educational Sciences, Mohammed V University, Rabat, Morocco) *Lexicogrammatical Patterns of Emotion Expression in Moroccan Arabic and English: A Corpus-Based Analysis*
15:30-16:00	**Sanja Virovec** (Faculty of Humanities and Social Sciences, Croatia) *Corpus-based Research as a Useful Tool in Contrastive Linguistics*
16:00-16:30	**Coffee break**
16:30-17:00	**María Sampedro Mella** (Catholic University of Louvain, Belgium) *Corpus in L2 Teaching: Enhancing Discourse Skills through Pragmatically Annotated Corpora*
17:00-17:30	**Filip Kalaš** and **Pavol Lipták** (University of Economics in Bratislava, Slovakia) *Phrase-frames in Economic Discourse: A Corpus-based and AI-assisted Comparative Study*
17:30-18:00	**Meilin Chen** and **Linfeng Cai** (Hong Kong Baptist University) *Authorial Stance in Academic Writing by Hong Kong University Students: Findings from a Novice Cross-discipline Corpus and Implications for Academic Writing Pedagogy*

August 28th, 2025

09:00-10:15	***Plenary lecture*** **Simon Krek** (Jožef Štefan Institute, Centre for Language Resources and Technologies, University of Ljubljana, Slovenia) *Squeezing Linguistic Data into a Database Model: the Case of Slovenian*
10:15-10:30	**Coffee break**
	Session 1
10:30-11:00	**Noriko Matsumoto** (Osaka Electro-Communication University, Osaka, Japan) *A Possible Explanation as to Why 'Go to Shop' is Unacceptable* (online)
11:00-11:30	**Oleksandra Palchevska, Petro Hubych** (Lviv State University of Life Safety, Ukraine) and **Irena Snikhovska** (Zhytomyr Medical Institute, Ukraine) *Features of Compiling the English-Polish-Ukrainian Fire and Rescue Terms Dictionary Based on the Specialized Texts Corpora*
11:30-12:00	**Asmar Ibadova** (University of Debrecen, Hungary) *The Role of Mobile Technology in Language Learning: A Study on Azerbaijani University Students*
12:00-12:30	**Daša Farkaš, Vanja Štefanec** and **Marko Tadić** (Faculty of Humanities and Social Sciences, Zagreb, Croatia) *CLARIN's Services as a Translation Technology Access Point*
12:30-14:30	**Lunch break**
	Session 2
14:30-15:00	**Marija Brala Vukanović** (University of Rijeka, Croatia) *Corpora and AI Translation Tools in Maritime English: Ethical, Legal, and Factual Challenges in Translating Specialized Language*
15:00-15:30	**Lidija Cvikić** (Faculty of Teacher Education, University of Zagreb, Croatia) and **Gordana Dobravac** (University of Rijeka, Croatia)

	Enhancing Early Disciplinary Literacy through Corpus Research: The Croatian RAPID Corpus Project
15:30-16:00	**Marina Banjac** (University of Novi Sad, Serbia) and **Jelena Banjac** (HTEC Group, Serbia) *English Borrowings in Serbian IT Discourse: A Corpus-Based and NLP-Assisted Study of Anglicisms and Englishisms (2011–2021)*
16:00-16:30	**Martina Pavić** (Institute for Croatian language, Croatia) *Adjectives in Disease Terms: A Comparison of Scientific and Popular Texts*
16:30-17:00	**Bogdan Vuruna** (Faculty of Philology, University of Belgrade, Serbia) *Serbian for Foreign Students: Variations of Positions of Reflexive Pronoun 'Se' in Past and Future Tense of Reflexive Verbs*
17:00-17:15	Closing ceremony

Marina Banjac (University of Novi Sad, Serbia), **Jelena Banjac** (HTEC Group, Serbia)
<banjac.marinaa@gmail.com>; <msjelenabanjac@gmail.com>

English Borrowings in Serbian IT Discourse: A Corpus-Based and NLP-Assisted Study of Anglicisms and Englishisms (2011–2021)

This study explores the presence and evolution of English-derived terminology in Serbian academic writing within the domain of Information Technology. By examining research paper abstracts from two points in time—2011 and 2021—it provides a diachronic perspective on language contact and lexical borrowing. The analysis is based on the distinction between anglicisms (English-derived words adapted into the Serbian linguistic system) and englishisms (terms retaining their original English orthographic form), as defined by Prćić (2019). The central aim is to assess how English, as the dominant global language in technological discourse, permeates Serbian academic writing and shapes terminological conventions over time.

The dataset includes 387 abstracts (209 from 2011 and 178 from 2021), totaling approximately 20,000 words, drawn from the journal Zbornik radova Fakulteta tehničkih nauka (Proceedings of the Faculty of Technical Sciences), with a focus on Electrical Engineering and Computer Science. The corpus was chosen to reflect developments in domain-specific language within a glocalized context—capturing both global linguistic influence and local adaptation, in line with Sharifian's (2011) framework of linguistic glocalisation.

The study applies a hybrid methodology combining Natural Language Processing (NLP) tools, manual annotation, and large language model (LLM) assistance. Initial annotation was conducted using Stanza, an NLP toolkit supporting POS tagging and lemmatization for both Serbian and English. However, due to the multilingual and domain-specific nature of the corpus, automated annotation was often unreliable. Common errors included incorrect lemma generation for specialized terms, misclassification of englishisms as Serbian words and vice versa, and general lemma inaccuracies. To address this, a Python-based pipeline was created for manual review and correction of all content words by part of speech. This included correcting lemmas, reassigning misclassified tokens, and removing irrelevant data.

Following cleanup, the dataset was processed using OpenAI's GPT API for lexical classification. The model was prompted to assess whether a Serbian word was likely an

anglicism, categorizing each item as “Yes” (70% confidence), “Maybe” (30–70%), or “No” (30%). If classified as an anglicism, the model was further instructed to identify its English source and provide a brief etymology. This approach enhanced interpretability and minimized ambiguity in classification.

A particularly challenging aspect of the analysis involved categorizing borderline cases—terms that resisted clear classification. Many showed partial morphological or orthographic adaptation, blurring the line between Serbian and English. For instance, waveleta and widget-a carry Serbian suffixes but preserve English features like w or dg, which are uncommon in Serbian. Words like monitoringu are even more ambiguous: their seamless integration into Serbian morphology and spelling raises questions about whether their adaptation is systematic or incidental. Other examples revealed multiple strategies of adaptation, such as lejat and lejaut (from layout), pejsmejker/pejsmeker (from pacemaker), or standardi/standarti (from standard). Competing forms like virtualizacija vs. virtuelizacija further suggest inconsistent standardization. Stanza was used to extract content words—nouns, adjectives, adverbs, verbs, and English words—while discarding function words. The total number of content words identified was 7,710 in 2011 and 7,276 in 2021. Nouns were the most frequent category, with 4,419 in 2011 and 4,146 in 2021. OpenAI's classification detected 1,167 noun tokens as anglicisms in 2011 (26.4% of all nouns), and 971 in 2021 (23.4%). In terms of type frequency, there were 922 unique nouns in 2011, 275 of which were anglicisms (29.8%), and 970 in 2021, with 262 anglicisms (27.0%).

In addition to adapted anglicisms, the corpus contained raw English words—unadapted terms often written in English orthography, including brand names and acronyms like TLS, Android, Python, SCADA, and DMS. A total of 645 such words were found in 2011, making up 8.4% of content words, compared to 539 in 2021 (7.4%). The theoretical framework of cultural linguistics (Sharifian, 2011) is particularly useful here. It allows for refracting research results through a cultural lens and highlights the shared international practices in technological discourse. In this view, Serbian academics contribute to and draw from a global "technoscape"—a term popularized by Arjun Appadurai to describe the rapid, trans-boundary flow of technology, knowledge, and practices (1996).

This research also contributes to broader debates on language contact, standardisation, and language policy in technologically driven domains. Particularly contentious are instances where

English words are used in their original orthographic form—such as Bootloader-a or web—which Prćić (2011) considers linguistically unacceptable. As a proponent of linguistic purism, he argues that such forms violate Serbian orthographic norms and reflect a broader process of cultural degradation. Similar concerns are raised in Croatian linguistic discourse. Stojaković and Malčić (2006) emphasize the sociolinguistic risks associated with the absence of standardized computer terminology, advocating for structured standardisation efforts. Likewise, Škifić and Mustapić (2012) analyze the interplay between anglicisms and domestic terms in Croatian, concluding that while conflict is minimal, deviations from orthographic norms often trigger purist reactions. In this context, Milroy's (2001) critique of standardisation as an ideologically motivated process becomes especially relevant. Rather than viewing variation as linguistic decline, this study positions it as a reflection of ongoing negotiation between global terminology and local linguistic identity.

References

Appadurai, Arjun. *Modernity at Large: Cultural Dimensions of Globalization*. University of Minnesota Press, 1996.

Milroy, James. "Language Ideologies and the Consequences of Standardization." *Journal of Sociolinguistics* 5, no. 4 (2001): 530–555. https://doi.org/10.1111/1467-9481.00163

Prćić, Tvrtko. "English as the Nativized Foreign Language and its Impact on Serbian." *English Today* 30, no. 1 (2014): 3–11. https://doi.org/10.1017/S0266078413000515.

Prćić, Tvrtko. Engleski u srpskom (3rd electronic ed.). Filozofski fakultet, Univerzitet u Novom Sadu, 2019. https://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-512-9

Sharifian, Farzad. *Cultural Linguistics: The Relationship between Language and Culture*. 2011.

Škifić, Sanja, and Emilija Mustapić. (2012). "Anglicisms and the Croatian Computer Terminology through the Prism of Language Conflict and Language Ideology." *Jezikoslovlje* 13, no. 3 (2012): 809–839. Retrieved from https://hrcak.srce.hr/clanak/140139.

Stojaković, Biljana, and Goran Malčić. "Standardizacija hrvatskog računalnog nazivlja." In *Computers in Education (CE): Conference Proceedings*, 262-265. Accessed February 8, 2025. https://www.croris.hr/crosbi/publikacija/prilog-skup/517541.

Marija Brala Vukanović (University of Rijeka, Croatia)
<marija.brala@ffri.uniri.hr>

Corpora and AI Translation Tools in Maritime English: Ethical, Legal, and Factual Challenges in Translating Specialized Language

This paper explores the use of AI-powered translation tools and language corpora in the domain of Maritime English, with a particular focus on English-Croatian and Croatian-English translation. Drawing from real-world maritime documents and domain-specific corpora, the study critically examines how AI tools (such as Google Translate, ChatGPT, and Glosbe) cope with the linguistic, technical, and terminological complexity inherent in maritime language. In doing so, it highlights key ethical, legal, and factual challenges associated with deploying these tools in professional translation contexts.

The study adopts a mixed-methods approach, combining qualitative and quantitative methodologies. The research is structured around a survey distributed to students of Maritime Studies, English Studies, and IT Studies at Croatian universities. The survey investigates their experiences with AI translation tools, their perceived reliability of such tools in specialized translation tasks, and their awareness of the ethical, legal, and factual limitations these tools entail. Students are also asked to comment on potential pedagogical strategies for integrating a critical awareness of AI translation tools into classroom and translation practice.

The overarching aim of this research is twofold: (a) to gauge the extent to which students from different disciplinary backgrounds recognize the limitations of AI-based translation tools in the context of maritime language; and (b) to identify effective strategies for raising awareness and fostering critical, responsible usage of these tools in educational and professional settings. Central to the inquiry is the pedagogical question of how teachers—and by extension, future professionals—can be equipped to meaningfully engage with AI tools, balancing their practical utility with a nuanced understanding of their risks.

The paper contributes to ongoing debates about the role of artificial intelligence in language education and professional translation, particularly in domains where precision, compliance with legal standards, and terminological consistency are critical. It argues that while AI tools offer unprecedented support for language processing, they cannot be used uncritically in specialized fields such as Maritime English without significant ethical and legal considerations.

Issues such as data privacy, machine bias, and mistranslation of legal or technical terminology are discussed in depth.

Ultimately, the study offers a set of practical recommendations for incorporating AI translation tools into both language education and translation workflows in ways that are ethically sound, legally compliant, and pedagogically meaningful. These findings aim to inform educators, policymakers, and translation professionals about the importance of developing critical digital literacy and domain-specific translation competence among students and practitioners alike.

References

Borucinsky, Mirjana, Jana Kegalj, and Sandra Tominac Coslovich. *Maritime English Corpus*. Rijeka: Pomorski fakultet (Faculty of Martime studies), University of Rijeka, 2024.

Crompton, Helen, Adam Edmett, and Neenaz Ichaporia. *Artificial Intelligence and English Language Teaching: A Systematic Literature Review*. London: British Council, 2023.

Edmett, Alan, Neenaz Ichaporia, Helen Crompton, and Ross Crichton. *Artificial Intelligence and English Language Teaching: Preparing for the Future*. London: British Council, 2023. https://doi.org/10.57884/78EA-3C69

Farrokhnia, Mohammad, Sima K. Banihashem, Omid Noroozi, and Allan Wals. "A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research." *Innovations in Education and Teaching International* 61, no. 3 (2023): 460–474. https://doi.org/10.1080/14703297.2023.2195846

Gartner, Smiljana and Marjan Krašna. "Artificial Intelligence in Education - Ethical framework." *12th Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro, 2023, 1–7. 10.1109/MECO58584.2023.10155012.

Gašparović, Marijana, Marija Brala Vukanović, and Marija Brkić Bakarić. *Idioms and Machine Translation: Assessing Translation Accuracy in the Age of AI.* Unpublished manuscript.

Kazu, Ibrahim Y. and Murat Kuvvetli. "The Influence of Pronunciation Education via Artificial Intelligence Technology on Vocabulary Acquisition in Learning English." *International Journal of Psychology and Education Studies* 10, no. 2 (2023): 480–493. https://doi.org/10.52380/ijpes.2023.10.2.1044

Musk, Nicholas. “Using Online Translation Tools in Computer-Assisted Collaborative EFL Writing.” *Classroom Discourse* 13, no. 2 (2022): 119–144. https://doi.org/10.1080/19463014.2021.2025119

Naveen, Palanichamy and Petr Trojovský. “Overview and Challenges of Machine Translation for Contextually Appropriate Translations.” *iScience* 27, no. 10 (2024). https://doi.org/10.1016/j.isci.2024.110878.

Okolo, Chinwe Jane, Chinyere Grace Ezeonwumelu, Chioma Ihuoma Barah, and Ugwu Nnenna Jovita. “Language Education in the Age of AI: Opportunities and Challenges.” *Newport International Journal of Research in Education* 4, no. 1 (2024): 39–44. https://doi.org/10.59298/NIJRE/2024/41139448

Rowe, Lindsey W. “Google Translate and Biliterate Composing: Second-Graders' Use of Digital Translation Tools to Support Bilingual Writing.“ *TESOL Quarterly* 56, no. 3 (2022) DOI: https://doi.org/10.1002/tesq.3143

Russell, Stuart, and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach* (3rd ed.). Boston: Pearson Education.

Schmidt, Thomas, and Thomas Strasser. “Artificial Intelligence in Foreign Language Learning and Teaching: A CALL for Intelligent Practice.” *International Journal of English Studies* 33, no. 1 (2022): 165–184.

Vogt, Kerstin A., and Lars Flindt. “Artificial Intelligence and the Future of Language Teacher Education: A Critical Review of the Use of AI Tools in the Foreign Language Classroom.” In *The Future of Teacher Education: Innovations Across Pedagogies, Technologies, and Societies*, edited by P. Hohaus and J.-F. Heeren, 179–199. Leiden: Brill, 2023.

Meilin Chen, Linfeng Cai (Hong Kong Baptist University)
<meilin22@hkbu.edu.hk; <linfeng_cai@hkbu.edu.hk>

Authorial stance in academic writing by Hong Kong university students: Findings from a novice cross-discipline corpus and implications for academic writing pedagogy

As Hyland (2005b: 173) has pointed out, academic writing is a "persuasive endeavour", rather than "an objective, faceless, and impersonal form of discourse". This requires authors to be able to appropriately establish author presences and claim solidarity with readers, strategically express stances and indicate epistemic commitment to claims made, as well as critically evaluate opinions of others via the use of a wide range of metadiscoursal devices such as self-mentions, reader pronouns, hedges, boosters, attitude markers, directives, etc. (Charles 2006, Chang 2012, Lancaster 2016; Xu & Zhang 2019).

Establishing authorial stances in academic writing is also discipline-specific (Hyland, 2004a, 2004b, 2005a). Mastery of very nuanced stance manipulation and author-reader relationship in disciplinary academic writing has been found highly challenging for novice writers, especially for those whose first language is not English (Aull & Lancaster 2014; Bruce, 2016, Hyland, 2016; Jiang, 2015; Sawaki 2014). This may be due to various factors, such as higher anxiety or less self-efficacy in the writing process (Huerta et al. 2017; Martinez et al. 2011), writers' lack of knowledge of metadiscouse devices (Sawaki 2014, Zhang & Zhang, 2021, 2023), and teaching effects (Petrić 2010).

While a plethora of studies have investigated authorial presences in student writing, very few have systematically examined them in novice writing across disciplines. To fill this gap, this study aims to scrutinize authorial stance in a novice academic corpus that consists of academic writing by Hong Kong (HK) undergraduate students from 26 disciplines across five major disciplinary areas, including Humanities, Creative Arts, Science, Business and Management, and Social Sciences. The corpus includes 2,600 written assignments that students submitted for their subject courses.

Using the Authorial Voice Analyzer (AVA) tool developed by Yoon (2018) to tag the corpus, followed by manual checking, we compared engagement devices in the HK novice academic corpus with those in the BAWE corpus (Nesi & Gardner, 2018). Results show that compared

to academic writing by British undergraduate students, Hong Kong university students from most disciplines make less frequent use of hedges, boosters and attitude markers and show problematic use of some stance devices.

Critical attitudes in HK student writing are often expressed unmitigatedly, with abrupt switches between positive and negative attitudes. On the other hand, HK students tend to use self-mentions more frequently in writing in comparison to their British counterparts. This is especially the case the HK business students, using far more reader pronouns in writing. These might indicate students' poor knowledge of academic genres at the tertiary level, showing possible traces of secondary-school writing habits. These observations point to the need for discipline- and genre-specific instructions on constructing authorial presences. Pedagogical implications will be discussed in the presentation.

References

Aull, Laura L., and Zak Lancaster. "Linguistic Markers of Stance in Early and Advanced Academic Writing: A Corpus-based Comparison." *Written Communication* 31, no. 2 (2014): 151–183.

Bruce, Ian. "Constructing Critical Stance in. University Essays in English Literature and Sociology." *English for Specific Purposes* 42 (2016): 13–25.

Chang, Peichin. "Using a Stance Corpus to Learn about Effective Authorial Stance-Taking: A Textlinguistic Approach." ReCALL 24, no. 2 (2012): 209–236.

Charles, Maggie. "The Construction of Stance in Reporting Clauses: A Cross-Disciplinary Study of Theses." *Applied Linguistics* 27, no. 3 (2006). 492–518.

Huerta, Margarita, Patricia Goodson, Mina Beigi, and Dominique Chlup. "Graduate Students as Academic Writers: Writing Anxiety, Self-efficacy and Emotional Intelligence." Higher Education Research and Development 36, no. 4 (2017): 716–729.

Hyland, Ken. *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor: University of Michigan Press, 2004.

Hyland, Ken. "Disciplinary Interactions: Metadiscourse in L2 Postgraduate Writing." *Journal of Second Language Writing* 13, no. 2 (2004): 133–151.

Hyland, Ken. “Representing Readers in Writing: Student and Expert Practices.” *Linguistics and Education* 16, no. 4 (2005): 363–377.

Hyland, Ken. “Stance and Engagement: a Model of Interaction in Academic Discourse.” *Discourse Studies* 7, no. 2 (2005): 173–192.

Hyland, Ken. “Writing with Attitude: Conveying a Stance in Academic Texts.” In *Teaching English Grammar to Speakers of Other Languages* edited by E. Hinkel. New York: Routledge, 2016.

Jiang, Feng. “Nominal Stance Construction in L1 and L2 Students’ Writing.” *Journal of English for Academic Purposes* 20 (2015): 90-102.

Lancaster, Zak. “Expressing Stance in Undergraduate Writing: Discipline-specific and General Qualities.” *Journal of English for Academic Purposes* 23 (2016): 16–30.

Nesi, Hilary, and Sheena Gardner. “The BAWE Corpus and Genre Families Classification of Assessed Student Writing.” *Assessing Writing* 38 (2018): 51-55.

Petrić, Bojana. “Students’ Conceptions of Voice in Academic Writing.” In *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres*, edited by R. Lorez-Sanz, P. Mur-Duenas, and E. Lafuente-Millan, 324–336. Newcastle upon Tyne, England: Cambridge Scholars Publishing, 2010.

Sawaki, Tomoko. “On the Function of Stance-neutral Formulations: Apparent Neutrality as a Powerful Stance Constructing Resource.” *Journal of English for Academic Purposes* 16 (2014): 81–92.

Xu, Linlin, and Lawrence Jun Zhang. “L2 Doctoral Students’ Experiences in Thesis Writing in an English-medium University in New Zealand.” *Journal of English for Academic Purposes* 41 (2019): 1–13.

Yoon, Hyung-Jo. “Textual Voice Elements and Voice Strength in EFL Argumentative Writing.” *Assessing Writing* 32 (2017): 72-84.

Zhang, Lu, and Lawrence Jun Zhang (2021). “Fostering Stance-taking as a Sustainable Goal in Developing EFL Students’ Academic Writing Skills: Exploring the Effects of Explicit Instruction on Academic Writing Skills and Stance Deployment.” *Sustainability* 13, no. 8 (2021): 4270

Zhang, Lu, and Lawrence Jun Zhang. “Development and Validation of the Questionnaire on EFL Students’ Perceptions of Authorial Stance in Academic Writing.” *Applied Linguistics Review*, 14, no. 5 (2023): 1219-1251.

Lidija Cvikić (Faculty of Teacher Education, University of Zagreb, Croatia), **Gordana Dobravac** (University of Rijeka, Croatia)
<lidija.cvikic@ufzg.hr>; <gdobravac@gmail.com>

Enhancing Early Disciplinary Literacy through Corpus Research: The Croatian RAPID Corpus Project

This paper explores the potential of corpus research to support disciplinary literacy, defined as the ability to interpret, produce, and critically engage with texts specific to academic and professional fields (Shanahan & Shanahan, 2008; Moje, 2008). Developing it involves mastering discipline-specific conventions, structures, text types, vocabulary, reasoning practices, and epistemological norms, as each field exhibits distinct linguistic features (Schleppegrell, 2004; Coffin, 2006). Schools play a central role in this process, requiring targeted linguistic support tailored to disciplinary demands. Textbooks, in this context, function as both knowledge sources and key inputs for developing disciplinary literacy.

Research across languages shows that disciplinary language differs from general language at morphosyntactic, lexical, and discourse levels (Hyland, 2008; Schleppegrell, 2004), and that scientific disciplines use language in distinct ways (Coffin, 2006; Bergh Nestlog et al., 2024). However, most studies focus on global languages like English and Spanish, analyzing secondary-level textbooks and individual disciplines. The research project *Early Disciplinary Literacy in Croatian: Features of Textbook Discourse – RAPID* (HRZZ-IP-2024-05-2755) addresses this gap by examining an under-researched language across multiple disciplines at the primary level. Using corpus-based methods, it investigates disciplinary textbooks in Croatian for students aged 6–12.

This paper has two main objectives. First, it discusses the process of the RAPID corpus design and compilation, focusing on methodological issues encountered. Second, it presents several examples to illustrate how the corpus can be used to deepen our understanding of the specific characteristics of disciplinary language. The RAPID corpus comprises textbooks aimed at developing early disciplinary literacies in mathematics, science, and general literacy for ISCED Level 1 students (ages 6–12). It includes materials for courses: Science and Society (grades 1–4), Science (grades 5–6), Mathematics (grades 1–6), and History (grades 5–6). The corpus

consists of 62 officially approved textbooks from three publishers, whose editions are used by about 80% of Croatian primary school students.The selection criteria ensured that the corpus is representative, though not balanced (Lemnitzer & Zinsmeister, 2010), since some publishers provide multiple textbooks for a specific grade and subject. The RAPID corpus was compiled and is being analyzed using the Sketch Engine tool (Kilgarriff et al., 2014), chosen for its compatibility with the Croatian language. The corpus is tagged with metadata for subject area (discipline), grade, publishing house, and author, allowing for the extraction of sub-corpora and comparative analyses across grades and disciplines. Upon compilation, the corpus contains 2,873,486 tokens, of which 1,718,125 are word tokens. It can be categorized as a small (McEnery et al., 2006), specialized (Bowker & Pearson, 2002), and representative corpus (Leech, 1991), ensuring the reliability of its analytical outcomes. Several methodological challenges encountered during compilation will be addressed in the paper.

To meet the second objective, preliminary analyses of corpus data will be presented. Through targeted corpus queries, the paper demonstrates how the RAPID corpus can be used to explore the specific features of disciplinary language. The paper concludes with a discussion on the implications of the RAPID corpus for educational practice and future research. The argument is made that corpus-based approaches can empower both teachers and learners by increasing awareness of how knowledge is constructed and communicated in different fields. By bridging the gap between linguistic research and educational practice, the RAPID corpus offers a valuable model for integrating corpus linguistics into literacy development in other languages and educational systems.

References

Bergh Nestlog, Ewa, Kristina Danielsson, and Fredrik Jeppsson. “Disciplinary Content and Text Structures Communicated in the Classroom – Pathways in Science Lessons.” *Linguistics and Education* 84 (2024): 101343. https://doi.org/10.1016/j.linged.2024.101343

Bowker, Lynne, and Jennifer Pearson. *Working with Specialized Language: a Practical Guide to Using Corpora*. London/New York: Routledge, 2002.

Coffin, Caroline. “The Voices of History: Theorizing the Interpersonal Semantics of Historical Discourses.” *Text & Talk* 22, no. 4 (2002): 503-528. https://doi.org/10.1515/text.2002.020

Hyland, Ken. “As Can Be Seen: Lexical Bundles and Disciplinary Variation.” *English for Specific Purposes* 27, no. 1 (2008): 4–21.

Kilgarriff, Adam, Vit Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vit Suchomel. “The Sketch Engine: Ten Years On.” *Lexicography* 1 (2014): 7-36.

Lemnitzer, Lothar, and Heike Zinsmeister. *Korpuslinguistik. Eine Enfuerung*. Narr Verlag, 2010.

Leech, Geoffrey. “The State of the Art in Corpus Linguistics.” In *English Corpus Linguistics: Studies in Honor of Jan Svartvik*, edited by A. Karin and B. Altenberg. Longman, 1991.

McEnery, Anthony, Richard Xiao, Yukio Tono. *Corpus-based Language Studies. An Advanced Resource Book.* London/New York: Routledge, 2006.

Moje, Elizabeth Birr. “Doing and Teaching Disciplinary Literacy with Adolescent Learners: A Social and Cultural Enterprise.” *Harvard Educational Review* 85, no. 2 (2015): 254-278.

Schleppegrell, Mary J. *The Language of Schooling*. Taylor and Francis, 20024. Kindle Edition.

Shanahan, Timothy, and Cynthia Shanahan. (2008). “Teaching Disciplinary Literacy to Adolescents: Rethinking Content-area Literacy.” *Harvard Educational Review* 78, no. 1 (2008): 40-59

Andreja Drašler, Monika Kavalir

(University of Ljubljana, Slovenia)

<andreja.drasler@ff.uni-lj.si>; <monika.kavalir@ff.uni-lj.si>

Compilation of a Corpus-based Word List for Teaching English Geographical Vocabulary

For highly specialised English for Specific Academic Purposes (ESAP) courses that focus mainly on the development of reading skills and technical vocabulary learning, there are often no textbooks available. As a consequence, teachers must often design their own materials, which has traditionally been done in an intuitive way. To make this process more systematic, an empirically grounded corpus-informed approach is presented in this study where a purpose-built corpus was designed as a foundation for assembling a corpus-based word list for pedagogical purposes in teaching English for university students of geography, serving as the basis for the design of customised teaching materials.

Identifying technical vocabulary is a key factor in developing teaching materials (Kwary 2011). To determine what technical vocabulary should be taught, a specialised corpus of texts for the course should be first created (Kennedy and Bolitho 1991; Hüttner et al. 2009; Coxhead 2013). According to Toriida (2016), the teaching context and students' needs should be considered for the design of the corpus for teaching purposes. Four needs analyses conducted since 2020 have consistently shown that, in their ESAP courses, students of geography at the University of Ljubljana most want to improve their geographical vocabulary. The ESAP course takes place in the second year of geography studies. A review of the first-year Geography BA curriculum revealed that students develop foundational geographical knowledge during their initial year of study. Geography professors confirmed this finding, emphasizing that this foundational knowledge serves as a basis for more advanced coursework in the second and third years.

A specialised corpus of geographical texts comprising of approx. 1.4 million words was built in order to find out what specialised vocabulary is relevant for teaching an ESAP course. In line with the needs analysis, the review of the Geography BA curriculum and consultations with geography professors, the corpus was designed to include texts that introduce basic geographical topics, meaning the criterion for the selection of texts was the articles and textbooks that were relevant to the first-year subjects in the university programme of geography

studies. The corpus was balanced according to the criterion of the number of lectures. On the basis of the collected texts, which were suggested by the geography professors who teach the first year courses, a proportion of 2,000 words per contact hour of lectures was used for textbooks and a proportion of 500 words for articles. In other words, the more lectures a certain course comprised of, the more related text was included in the corpus.

To assemble the specialised word list for pedagogical purposes, a combined approach of lexical profiling was used. First, the online corpus tool Sketch Engine was employed for extracting keywords. Then the proportion of technical words in the compiled keyword lists was determined, using technical dictionaries in the field of geography and consulting the geography professors when necessary. A list of 259 geographical single-words and 249 multi-word items was compiled according to the following criteria: (1) Sketch Engine keyness score, (2) minimum document frequency, (3) technicality (whether a keyword or a multi-word term is present in a technical dictionary or approved by a discipline-specific expert). If a keyword or a term was found in one of the technical dictionaries, it was considered a technical term and included in the word list. For keywords and terms that were not listed in technical dictionaries, professors of geography were asked to give their expert opinions on whether a particular term was technical or not.

The grand total of geographical terms in the word list is 507, i.e. 259 keywords and 248 multi-word items. The keyword analysis has shown that the vast majority (94.24%) of keywords and multi-word items extracted by Sketch Engine are technical and have been thus included in the list. The fact that only 31 out of 538 keywords and multi-word items, representing merely 5.76%, were neither found in technical dictionaries nor considered technical items by specialist experts underscores the efficacy of Sketch Engine's terminology extraction function.

References

Coxhead, Averil. “Vocabulary and ESP” In *The Handbook of English for Specific Purposes*, edited by Brian Paltridge and Sue Starfield. Wiley-Blackwell, 2013.

Hüttner, Julia, Ute Smit, and Barbara Mehlmauer-Larcher. “ESP Teacher Education at the Interface of Theory and Practice: Introducing a Model of Mediated Corpus-Based Genre Analysis.” *System* 37, no. 1 (2009): 99−109.

Kennedy, Chris, and Rod Bolitho. *English for Specific Purposes*. MacMillan Press Ltd, 1991.

Kwary, Deny Arnos. "A Hybrid Method for Determining Technical Vocabulary." *System* 39, no. 2 (2011): 175–185.

Toriida, Marie-Claude. "Steps for Creating a Specialized Corpus and Developing and Annotated Frequency-Based Vocabulary List." *TESL Canada Journal/Revue TESL du Canada* 34, no 11 (2016): 87–105.

Khalid Elasri

(Faculty of Educational Sciences, Mohammed V University, Rabat, Morocco)

<elasrik@gmail.com>

Lexicogrammatical Patterns of Emotion Expression in Moroccan Arabic and English: A Corpus-Based Analysis

This study investigates the lexicogrammatical encoding of emotions in Arabic and English using a corpus-based approach. While emotions are universal, their linguistic representation varies across languages due to structural and cultural differences (Wierzbicka 1992, 1999; Kövecses 2003). This research explores how emotions are expressed through lexicogrammatical patterns, particularly focusing on the interaction between lexis and grammar in the use of verbs, adjectives, and argument structures.

The study addresses three key research questions: (1) What lexicogrammatical patterns are used to express emotions in Arabic and English? (2) How do these patterns differ cross-linguistically in terms of morpho-syntactic preference and lexical selection? (3) What role does culture play in shaping these lexicogrammatical choices? Data were drawn from the British National Corpus (BNC) and the ArabiCorpus. The data from both languages was analyzed using frequency analysis, collocation extraction, and syntactic pattern identification.

The findings show that English predominantly encodes emotions through adjectival constructions (e.g., He is sad, She feels worried), using copular verbs, stative passives (to be worried), change-of-state verbs (to get), and experiencer-object structures (It made me sad reflecting an individualistic and static conceptualization of emotions, often portraying them as involuntary states beyond the experiencer's control. In contrast, Arabic predominantly uses dynamic verb-based patterns (e.g., aḥzana-nī al-khabar 'The news saddened me'; ghaḍiba al-rajul 'The man became angry'), suggesting that Arabic conceptualizes emotions as externally caused, dynamic events rather than as static inner states. Additionally, Arabic constructions frequently highlight external causation and volitional control (e.g., al-ḥuzn yaqtulunī 'Sadness is killing me'), making the force of emotions explicit and externalized, in contrast to English, where this force remains more implicit.

By systematically analyzing the lexicogrammatical encoding of emotions, this study contributes to corpus-based research on cross-linguistic emotion representation and highlights the role of

culture in shaping grammatical choices. These findings have implications for linguistic typology, corpus linguistics, and cross-cultural pragmatics.

References

Dewaele, Jean-Marc. Emotions in Multiple Languages. Palgrave Macmillan, 2010.

Elasri, Khalid. "The Linguistic Representation of Emotions in English and Moroccan Arabic." In *Cultural linguistics applied: Trends, directions and implications*, edited by A. Peters and N. Mundt, 105–126, 2021. Peter Lang. https://doi.org/10.3726/b17907.

Keltner, Dacher, Keith Oatley, and Jennifer M. Jenkins. *Understanding Emotions.* 2nd ed. Blackwell, 2006.

Pavlenko, Aneta. "Bilingualism and Emotions." *Multilingua* 21, no. 1 (2002): 45–78. https://doi.org/10.1515/mult.2002.004.

Pavlenko, Aneta. Emotions and Multilingualism. Cambridge University Press, 2005.

Pavlenko, Aneta. "Structural and Conceptual Equivalence in the Acquisition and Use of Emotion Words in a Second Language." *The Mental Lexicon* 3, no. 1 (2008): 91–120. https://doi.org/10.1075/ml.3.1.07pav.

Pavlenko, Aneta. "Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning." In *The bilingual mental lexicon: Interdisciplinary approaches*, edited by A. Pavlenko, 125–160. Multilingual Matters, 2009.

Pavlenko, Aneta, and Viktoria Driagina. "Russian Emotion Vocabulary in American Learners' Narratives. *The Modern Language Journal* 91, no. 2 (2007): 213–234. https://doi.org/10.1111/j.1540-4781.2007.00541.x.

Russell, James A. "Culture and the Categorization of Emotions." *Psychological Bulletin* 110, no. 3 (1991): 426–450. https://doi.org/10.1037/0033-2909.110.3.426.

Sharifian, Farzad. *Cultural Linguistics: Cultural Conceptualisations and Language.* John Benjamins, 2017.

Wierzbicka, Anna. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-specific Configurations*. Oxford University Press, 1992.

Wierzbicka, Anna. *Emotions across languages and cultures*: Diversity and universals. Cambridge University Press, 1999.

Daša Farkaš, Vanja Štefanec, Marko Tadić
(Faculty of Humanities and Social Sciences, Zagreb, Croatia)
<dfarkas@ffzg.hr>; <marko.tadic@ffzg.hr>; <vstefane@ffzg.hr>

CLARIN's services as a translation technology access point

Considering the results of new research in the field of translation, translation technology improves and facilitates the translation process itself on several levels. Translation technology significantly reduces the time required to translate large volumes of text, and thereby increases efficiency. Translation technology also ensures consistent use of terminology and phrasing, which is vital for maintaining consistency and accuracy across multiple documents or projects. It also increases accessibility by making information and research results accessible to a wider audience by breaking down language barriers. The use of translation technology ensures real-time communication (e.g. by using social media interactions, international collaboration, etc.) and cost-effectiveness because automating the translation process can lower costs associated with translation (Frontini et al. 2024).

CLARIN (Common Language Resources and Technology Infrastructure) was born out of a consensus on the continuous importance of language even in the digital communication age. CLARIN as a research infrastructure for language resources and technology has made great strides in creating and maintaining an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences (Krauwer & Maegaard 2022:3). As CLARIN is a distributed infrastructure whose tools and resources are made accessible through certified repositories hosted by national consortia, findability is facilitated by the so-called Virtual Language Observatory (VLO; Goosen & Eckart 2014; Lenardič & Fišer 2022).

Virtual Language Observatory (VLO) is playing a central role in CLARIN's infrastructure as a means of discovering language (Windhouwer & Goosen 2022:210), which automatically harvests the metadata from the repositories and thereby provides a catalogue of all the available tools and resources.

In the context of language resources, scholars and students of translation studies as well as practical translators may be assumed to be active data providers of language resources, as well as prolific users of the CLARIN services (Lušicky & Wissik 2017:1). Since the framework of research has been transformed into e-research, CLARIN services can provide sustainable access for researchers in the translation studies to digital language data and tools. CLARIN's VLO (https://vlo.clarin.eu) is certainly the first step in this e-research as an access point to language data. As a catalogue of data and services, it offers to active users facetted search, links to landing pages, download options, details on licences, details on technical features, overview of tools that match the data and instructions for citing language resources.

Translators and researchers in the field also have at their disposal CLARIN Resource Families (https://www.clarin.eu/resource-families) that provide a user-friendly overview per data type of the available language resources in the CLARIN infrastructure. Many types of corpora (e.g. parallel corpora, corpora of academic texts, corpora of disordered speech, historical corpora, L2 learner corpora, legal corpora, literary corpora, manually annotated corpora and others) can be found through Resource Families, as well as different lexical resources (conceptual resources, dictionaries, glossaries, language models, lexica, wordlists) useful to the translation studies community. Another service offered by CLARIN is CLARIN's tool inventory called Language Service Switchboard that users helps to find tools that can process a broad range of data (https://switchboard.clarin.eu/tools).

CLARIN's network of Knowledge Centres is also available to researchers in the field. K-centres share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure. Since the focus of CLARIN is on language resources, K-centres serve researchers and educators from any discipline where language plays one of its many roles, including translation. There are many CLARIN K-centres with expertise in translation studies, translation corpora, machine translation, translation memories (e.g. TRTC – K-Centre for Terminology Resources and Translation Corpora; LLMs4SSH – K-Centre for usage of LLMs in Social Sciences and Humanities, CLARIN-SMS – K-Centre for Swedish in a Multilingual Setting; K-Dutch – K-Centre for Dutch; PORTULAN – K-Centre for the Science and Technology of the Portuguese Language; CROATINA – K-Centre for Croatian language etc.).

Most K-centres offer a helpdesk, documentation and learning resources.

The Croatian participation in CLARIN – HR-CLARIN (clarin.hr) – recently established a repository (repository.clarin.hr), opening up access to the storing and sharing of language resources for Croatian scholars, and the first resources have already been published. Users can store their language resources in the repository, with each language resource receiving a unique persistent identifier (PID) and clear reference that can be used in publications. Croatian language resources stored in the HR-CLARIN repository are directly included in the European federation of CLARIN ERIC repositories and are visible and, depending on the chosen license, accessible to all authenticated researchers and/or the wider interested public. In the translation context we present here the resources that are particularly interesting: RomCro v.2.0 – Parallel corpus of seven Romance languages and Croatian, large language models for Croatian language and Croatian language data collection. Our centre has made a notable contribution to the digital humanities and social sciences community through the development of HR-CLARIN repository and we invite all researchers to join as users and to include their resources in the repository.

References

Baker, Mona, and Gabriela Saldanha, eds. *Routledge Encyclopedia of Translation Studies*. 3rd ed. London: Routledge, 2020.

Frontini, Francesca, Iulianna van der Lek, Vesna Lušicky, Janiça Hackenbuchner, Ralph Krüger, and Vincent Vandeghinste. "CLARIN Café: Teaching & Training Translation Technologies in SSH Disciplines." Web publication, 2024. https://doi.org/10.5281/zenodo.12703632

Goosen, Twan, and Thomas Eckart. "Virtual Language Observatory 3.0: What's New." In *Proceedings of the CLARIN annual conference 2014*, 23-25 October 2014, Soesterberg, The Netherlands. 2014.

Krauwer, Steven, and Bente Maegaard. "CLARIN – How It Started." In *CLARIN – The Infrastructure for Language Resources,* edited by Fišer, Darja and Andreas Witt, 3–24. Berlin/Boston: De Groyer, 2022.

Lenardič, Jakob, and Darja Fišer. "The CLARIN Resource and Tool Families." In *CLARIN – The Infrastructure for Language Resources,* edited by Fišer, Darja and Andreas Witt, 343–372. Berlin/Boston: De Groyer, 2022.

Lušicky, Vesna and Tanja Wissik. “Discovering Resources in the VLO: A Pilot Study with Students of Translation Studies.” In *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016,* edited by L. Borin, 63–75. CLARIN Common Language Resources and Technology Infrastructure. Linköping University Electronic Press, 2017. http://www.ep.liu.se/ecp/136/005/ecp17136005.pdf

Goretti Faya Ornia, Natalia Barranco Izquierdo, Teresa Calderón Quindós

(Universidad de Valladolid, Spain)

<mariagoretti.faya@uva.es>; <natalia.barranco@uva.es>; < teresa.calderon@uva.es>

The Use of an Aligned Parallel Corpus as a Didactic Tool for Phrasal Verbs: an Approach from Cognitive Linguistics

Didactic purposes of parallel corpora are common (Aijmer 2008; Faya-Ornia et al. 2022 and 2023), but no study has been found where they are used to teach phrasal verbs, and mainly from a cognitive approach. Here we present a teaching initiative carried out at the University of Valladolid, using an aligned parallel corpus (consisting of TV subtitles) and based on cognitive linguistics to teach English phrasal verbs.

First, in small groups, participants are asked to look for a series of particles and try to make cognitive connections, taking into account their context. Following the guidelines of Rudzka-Ostyn (2003), students are asked to reflect on issues such as the separability of the phrasal verbs, the different types of movement and trajectories, the spatial or metaphorical meaning of the particle. Then, their observations are discussed in a focus group where the teacher takes the opportunity to ask questions to make the students reflect on the results obtained and to ask them to read out the sentences they have worked on covering the phrasal verb so that they can fill in the gap. This intervention is also combined with a role-play activity. Therefore, once the meaning is clear, the students are asked to create their own sentences using that particular phrasal verb and they have to role-play a conversation similar to the one they found in the corpus (they can refer to the episode and watch it).

In the first activity (i.e. identifying the particles), students work mainly on analysing the meanings and nuances that particles give to verbs and their linguistic characteristics. This process allows students to see that particles are the meaningful part of phrasal verbs and it allows them to create cognitive patterns and connections that can help them to remember and put them into practice. The second activity (i.e. discussion in a focus group) is challenging as it requires students to reflect further on their own and other students' ideas. Finally, the third

activity (i.e. conducting a conversation in a role play) focuses on making the new vocabulary active and works mainly on students' oral skills.

Although the process is still at an early stage, the results so far have been very positive and encouraging, with students emphasising that they find these activities highly motivating and useful in improving their communicative competence. It is our intention to implement these activities in other English courses in order to obtain further results.

References

Aijmer, Karin. Corpora and Language Teaching. John Benjamins Publishing Company, 2008.

Faya-Ornia, Goretti, Natalia Barranco-Izquierdo, Teresa Calderón-Quindós, and Carmen Quijada-Diez. "Subtitle Creation and Subtitle Alignment as Didactic Resources for Foreign Language Teaching." *Culture and Education [Cultura y Educación]*, 34, no. 1 (2022): 180-187. https://doi.org/10.1080/11356405.2021.2007617.

Faya-Ornia, Goretti, Natalia Barranco-Izquierdo, Carmen Quijada-Diez, and Teresa Calderón-Quindós. "Combining Subtitle Creation and Subtitle Alignment in Foreign Language Teaching: a Means to Foster Translation Competence and Communication Skills." In *Instrumentalizing Foreign Language Pedagogy in Translator and Interpreter Training: Methods, Goals and Perspectives*, edited by Olaf Immanel Seel, Silvia Roiss and Petra Zimmermann-González, 99–111. John Benjamins. 2023. https://doi.org/10.1075/btl.161.

Rudzka-Ostyn, Brygida. *Word Power: Phrasal Verbs and Compounds: a Cognitive Approach*. Mouton de Gruyter, 2003.

Francisco Javier Fernández-Polo, Mario Cal-Varela

(University Santiago de Compostela, Spain)

<xabier.fernandez@usc.es>; <mario.cal@usc.es>

From Corpus to Classroom: Leveraging SUNCODAC to Enhance Interpersonal Communication Skills in Academic Online Forums

This presentation explores the application of learner corpora to foreign language teaching and learning (Brezina & Flowerdew, 2018; Götz & Mukherjee, 2019), specifically demonstrating how SUNCODAC (Santiago University Corpus of Online Discussions in Academic Contexts) can be utilized to develop participation skills in online forum discussions.

Integrated into most virtual learning environments, discussion forums play a crucial role in collaborative learning in higher education (Garrison & Cleveland-Innes, 2005), promoting reflective participation, encouraging shy students to engage, and providing lecturers with valuable insights into students' work (Garrison et al., 2001; Hew & Cheung, 2012). These collaborative online exchanges represent significant learning environments where students must deploy sophisticated persuasion strategies and manage interpersonal relationships in digital spaces, skills that are increasingly essential for their future academic and professional endeavors.

Our presentation is structured in two parts. First, we will describe the setting of the forums that constitute SUNCODAC: a blended learning translation course where students collaborate in the translation of texts by making suggestions to improve draft translations by classmates. These forums feature students from different lingua-cultural backgrounds who use English as a lingua franca to participate in the discussions. The corpus represents authentic interaction between learners working together on meaningful tasks, providing a rich source of data on naturally occurring academic discourse (Flowerdew, 2015). We will outline the corpus's holdings and search interface capabilities, briefly demonstrating the types of searches that can be performed.

The second part explores the applicability of the corpus by showing how it can be used by students to learn how to frame their contributions politely, to preserve interpersonal relationships and secure effective interaction. This will develop in two phases. Initially, we will share our previous research on how students ‘pre-empt’ the negative impact of criticism and suggestions by giving praise and minimizing the importance of their comments. Our findings

reveal significant patterns in the linguistic realization of these pragmatic functions, including the use of hedging devices, positive evaluative language, and personal stance expressions (Hyland, 2005).

In the second phase, we will present a series of practical activities where students explore the corpus to identify and critically assess different possibilities for giving praise and mitigating criticism. Specifically, we will focus on a recurrent rhetorical element in students' posts—the pre-proposal move—located in the opening sections of feedback contributions. This critical rhetorical element serves to prepare the ground for the criticism and suggestions that constitute the core of the feedback, helping to establish interpersonal rapport before presenting potentially face-threatening content.

Our pedagogical approach is scaffolded, moving systematically from macro to micro levels of analysis. Students begin examining the broader rhetorical purpose of pre-proposal moves within the communicative context of peer feedback, then progress to analyzing the structural components that typically constitute this move, and finally focus on specific linguistic features and lexicogrammatical choices used to realize these functions. This structured progression ensures students develop a comprehensive understanding of both why these communicative strategies are important and how they are effectively implemented in authentic discourse.

Through guided corpus exploration activities, students will work inductively, searching for patterns in the SUNCODAC data, comparing more and less successful examples of pre-proposal moves, in order to develop their own repertoire of strategies for framing constructive criticism. These hands-on activities include concordance searches for praise markers, analysis of hedging devices preceding critical comments, identification of personal stance expressions, and collaborative assessment of the interpersonal effectiveness of different pre-proposal formulations.

We conclude by reflecting on the value of authentic materials like the SUNCODAC corpus for studying and developing effective communicative skills in academic settings, highlighting how corpus-based approaches can bridge the gap between theoretical linguistic knowledge and practical communication needs in higher education contexts. By developing these digital communication skills through corpus-informed instruction, students acquire transferable competencies that extend far beyond the immediate academic context into their professional futures.

References

Brezina, Vaclav, and Lynne Flowerdew, eds. *Learner Corpus Research: New Perspectives and Applications*. Bloomsbury Academic, 2018.

Flowerdew, Lynne. "Data-driven Learning and Language Learning Theories: Whither the Twain Shall Meet." In *Multiple Affordances of Language Corpora for Data-driven Learning*, edited by A. Leńko-Szymańska and A. Boulton, 15–36. John Benjamins Publishing Company, 2015.

Garrison, D. Randy. *E-learning in the 21st Century: A Community of Inquiry Framework for Research and Practice* (3rd ed.). Routledge, 2017.

Garrison, D. Randy, Terry Anderson, and Walter Archer. "Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education." *American Journal of Distance Education* 15, no. 1 (2001): 7–23.

Garrison, D. Randy, and Martha Cleveland-Innes. Facilitating Cognitive Presence in Online Learning: Interaction Is Not Enough. *American Journal of Distance Education* 19, no. 3 (2005): 133–148.

Götz, Sandra, and Joybrato Mukherjee, eds. *Learner Corpora and Language Teaching*. John Benjamins Publishing Company, 2019.

Hew, Khe Foon, and Wing Sum Cheung. *Student Participation in Online Discussions: Challenges, Solutions, and Future Research*. Springer Science & Business Media, 2012.

Hyland, Ken. *Metadiscourse: Exploring Interaction in Writing*. Continuum, 2005.

Pertii Hietaranta
(University of Helsinki, Finland)
<pertti.hietaranta@professori.fi>

Language and Gender: Why Does Androcentrism in Finnish Disappear so Slowly?

This paper examines the situation of (in)equality between men and women in references to persons, with special reference to translation-induced language change, as this situation is reflected in a large Finnish-language corpus of just under 600 million words.

The hypothesis here is that androcentrism, i.e. the 'perception of people as male and male as people' (Engelberg 2002: 128) might no longer be as strong in Finnish as it used to be because Finnish has been under notable linguistic and cultural influence from English via translation, in particular.

While the data seem to indicate that some changes have indeed taken place, the overall picture has not changed over the last decades.

The paper argues that such reluctance to language change is explainable especially by reference to three cognitive and/or psychological properties of human beings: (1) a change may be perceived as a threat (cf. e.g. Mortensen 2024); (2) human cognition is not an isolated function but is connected especially with emotions (Baars and Gage 2010: 431); and (3) resorting to existing types of reference is often intuitive (Remmers et al. 2024).

References

Baars, Bernard J. and Nicole M. Gage. *Cognition, Brain, and Consciousness. Introduction to Cognitive Neuroscience.* 2nd ed. Academic Press, 2010.

Engelberg, Mila. "The communication of gender in Finnish". In *Gender Across Languages,* edited by Marlis Hellinger and Hadumod Bußman, 106–132. Benjamins, 2002.

Mortensen, Janus. Beyond Threat or Opportunity: English and Language Ideological Tensions in the Nordic Countries. In *English in the Nordic Countries: Connections, Tensions, and*

Everyday Realities, edited by Elizabeth Peterson and Kristy Beers Fägersten, 104–124. Taylor & Francis, 2024.

Remmers, Carina, Sascha Topolinski, Christine Knaevelsrud, Thea Zander-Schellenberg, Sebastian Unger, Albert Anoschin, and Johannes Zimmermann. “Go With Your Gut! The Beneficial Mood Effects of Intuitive Decisions.” *Emotion*, 24, no. 7 (2024):1652–1662.

Asmar Ibadova
(University of Debrecen, Hungary)
<aibadova10527@ada.edu.az>

The Role of Mobile Technology in Language Learning: A Study on Azerbaijani University Students

With the rapid advancement of technology, Mobile-Assisted Language Learning (MALL) has emerged as a powerful tool for enhancing language learning experiences. The widespread use of smartphones and mobile applications has provided learners with flexible and accessible platforms to improve their language skills beyond traditional classroom settings. MALL allows learners to practice listening, speaking, vocabulary, and pronunciation skills through interactive and gamified applications (Kukulska-Hulme & Shield, 2008). Through an analysis of their usage patterns, favorite apps, and opinions of mobile learning resources, this study investigates the function of mobile technology in language learning among university students in Azerbaijan.

In the field of language education, mobile learning plays a crucial role in Mobile-Assisted Language Learning (MALL), a subcategory of m-learning that focuses on enhancing language acquisition through mobile technology (Kukulska-Hulme & Shield, 2008). Language learners benefit from mobile apps that provide interactive vocabulary exercises, pronunciation practice, real-time feedback, and AI-powered language tutors. Due to its adaptive learning mechanisms, research suggests that m-learning can improve language retention, engagement, and motivation (Stockwell, 2013). However, despite its advantages, mobile learning also presents challenges. These include digital distractions, internet dependency, lack of structured learning pathways, and varying levels of technological literacy among students (Zou et al., 2018). Additionally, while many mobile learning applications offer free content, premium features often require paid subscriptions, which may limit access for some students.

An online survey of 103 undergraduate students from a variety of fields, including literature, engineering, translation, architecture, and English language instruction, served as the basis for the study. According to research, most students (more than 90%) regularly use mobile devices to learn languages, with smartphones being the most popular platform. YouTube, Duolingo, BBC Learning English, Instagram, and other well-known apps demonstrate the increasing

demand for interactive, multimedia-rich learning environments (Rosell-Aguilar, 2018; Crompton, 2013). According to students, mobile learning improves speaking and listening the most, followed by vocabulary growth and reading comprehension. Accessibility, interactive features, and customized learning experiences are major drivers of mobile app use, but issues such as a lack of incentive, limited free content, and commercials make it difficult to engage consistently (Soyoof et al. 2023).

The study finds gaps in institutional support for incorporating mobile learning into formal education, despite the clear benefits of MALL. Although mobile apps are actively used by students for self-study, university courses rarely include them in a structured manner. According to the study, Azerbaijani higher education institutions could improve language learning results by integrating mobile-assisted learning techniques into their curricula, offering advice on how to use apps practically, and tackling issues with accessibility and motivation. Overall, by offering empirical insights on MALL's acceptance in Azerbaijan—a context that has received little attention—this study adds to the expanding corpus of knowledge on the subject. The findings suggest that while MALL is widely used and considered effective, its implementation in formal education settings should address content limitations and engagement strategies. Teachers, legislators, and app developers interested in enhancing language learning through mobile technology may find these findings valuable.

References

Crompton, Helen. “A Historical Overview of Mobile Learning: Towards Learner-centered Education.” In *Handbook of Mobile Learning*, edited by Zane L. Berge and Lin Muilenburg, 3–14. Routledge, 2013.

Kukulska-Hulme, Agnes, and Lesley Shield. “An Overview of Mobile-assisted Language Learning: From Content Delivery to Supported Collaboration and Interaction.” *ReCALL* 20, no. 3 (2008): 271–289. https://doi.org/10.1017/S0958344008000335

Rosell-Aguilar, Fernando. “Autonomous Language Learning through a Mobile Application: a User Evaluation of the Busuu App.” *Computer Assisted Language Learning* 31 (2018): 854–881. 10.1080/09588221.2018.1456465

Ali Soyoof, Barry Lee Reynolds, Boris Vazquez-Calvo, and Katherine McLay. Informal Digital Learning of English (IDLE): A Scoping Review of What Has Been Done and a Look Towards What Is To Come. *Computer Assisted Language Learning, 36(4), 608-640.* https://doi.org/10.1080/09588221.2021.1936562

Stockwell, G. (2013). Technology and Motivation in English-language Teaching and Learning. In *Contemporary Computer-assisted Language Learning*, edited by M. Thomas, H. Reinders, & M. Warschauer, 156–175. Bloomsbury, 2023.

Zhou, L., Xue, S., & Li, R. "Extending the Technology Acceptance Model to Explore Students' Intention to Use an Online Education Platform at a University in China." *Sage Open* 12, no. 1 (2022): 21582440221085259.

Filip Kalaš, Pavol Lipták

(University of Economics in Bratislava, Slovakia)

<filip.kalas@euba.sk>; <pavol.liptak@euba.sk>

Phrase-frames in Economic Discourse: A Corpus-based and AI-assisted Comparative Study

In recent years, phraseology has become an essential lens for examining specialised academic discourse, particularly through the concept of phrase-frames (p-frames). A phrase-frame is a recurrent discontinuous multiword sequence containing one or more variable slots, enabling flexible yet formulaic constructions across registers and genres (Ren, 2022). For example, the frame *the * of the study* produces variants such as *the aim of the study* or *the focus of the study*.

Corpus-based studies have shown that p-frames serve as fundamental building blocks for fluency, native-like proficiency, and discourse community membership (Ren, 2022; Wray, 2006). In academic writing, p-frames simultaneously capture structural patterns and facilitate discourse organization, providing greater flexibility compared to fixed lexical bundles (Gray & Biber, 2013; Lu, Yoon, & Kisselev, 2018). As Golparvar and Barabadi (2020) emphasise, p-frames function as primary carriers of meaning within academic genres such as research articles, and their strategic use is essential for meeting disciplinary expectations.

Following the perspective of "language as phraseology" (Hunston, 2002), researchers highlight the need to identify p-frames not only based on frequency but also considering their functional and genre-specific value (Lu et al., 2018). This becomes particularly important in fields like economics, where writers rely on highly conventionalised phraseological patterns to convey abstract relationships.

Although many studies have focused on corpus-driven identification of p-frames (Cunningham, 2017; Casal & Kessler, 2020), relatively few have explored the competitiveness and accuracy of emerging AI-assisted extraction methods, such as those offered by ChatGPT. Traditional corpus-based extraction, conducted through tools like AntConc, depends on linguistic and frequency thresholds. In contrast, AI models simulate domain-specific phraseological knowledge, although often without direct empirical validation (Lu et al., 2018).

Given these developments, the present study investigates phrase-frame extraction in economic discourse using two competing approaches: traditional corpus-based extraction via AntConc and AI-assisted generation via ChatGPT-4o. The study aims to assess the structural accuracy, domain specificity, and pedagogical potential of phrase-frames produced by each method.

Hypothesis: *Corpus-based extraction will yield phrase-frames that are more specialised, structurally consistent, and contextually aligned with economic discourse, while AI-assisted extraction will produce phrase-frames that are more creative but less domain-specific. Consequently, corpus-extracted frames will be rated as more pedagogically useful for academic writing in economics.*

RQ_1: What types of phrase-frames are extracted from economic discourse using corpus-based methods compared to AI-assisted (ChatGPT-4o) methods?

RQ_2: How do corpus-extracted and AI-extracted phrase-frames differ in terms of structural consistency and variability?

RQ_3: To what extent are corpus-extracted and AI-extracted phrase-frames perceived as pedagogically useful for academic writing in economics?

The study uses an approximately 1,000,000-token corpus compiled from authentic economic discourse. The corpus includes economic research monographs. The material covers subdomains such as macroeconomics, international trade, finance, and development policy. Non-textual elements such as metadata, tables, and references have been removed to ensure textual consistency.Phrase-frame candidates are extracted with AntConc. The extraction targets four- to six-word sequences with one variable slot. Frames are selected based on frequency, dispersion, and internal cohesion. Typical examples include *the role of * in* and **a significant impact on* * *. Each extracted frame is classified according to its structural type: function-word, content-word, or verb-based frames.

ChatGPT-4o generates additional phrase-frames. A set of structured prompts requests the model to produce typical five-word phrase-frames from economic writing. Outputs are collected across multiple iterations to ensure variation and consistency. All AI-generated phrase-frames are classified using the same criteria as for corpus-based frames.

The extracted phrase-frames are evaluated based on three main criteria. First, the study assesses the structural diversity by analysing the syntactic variation across the frames. Second, the

domain relevance is determined by comparing the frames to authentic economic discourse examples within the corpus. Third, the frequency of occurrence is verified by searching for AI-generated frames within the corpus.

References

Casal, J. Elliott, and Matt Kessler. "Form and Rhetorical Function of Phrase-frames in Promotional Writing: A Corpus- and Genre-based Analysis." *System 95* (2020): 102370. https://doi.org/10.1016/j.system.2020.102370.

Cunningham, Kelly J. "A Phraseological Frames in Mathematics Research Articles." *Journal of English for Academic Purposes* 25 (2017): 71-83. https://doi.org/10.1016/J.JEAP.2016.11.005.

Golparvar, Seyyed Ehsan, and Elyas Barabadi. "Key Phrase Frames in the Discussion Section of Research Articles of Higher Education." *Lingua,* 236 (2020): 102804. https://doi.org/10.1016/j.lingua.2020.102804.

Gray, Bethany, and Douglas Biber. Lexical Frames in Academic Prose and Conversation. *International Journal of Corpus Linguistics* 18, no. 1 (2013): 109–136. https://doi.org/10.1075/ijcl.18.1.08gra.

Hunston, Susan and Gill Francis. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. John Benjamins, 2000.

Lu, Xiaofei, Jungwan Yoon, and Olesya Kisselev. "A Phrase-frame List for Social Science Research Article Introductions." *Journal of English for Academic Purposes* 36 (2018): 77–89. https://doi.org/10.1016/j.jeap.2018.09.004.

Ren, Junqiang. "A Comparative Study of the Phrase Frames Used in the Essays of Native and Nonnative English Students." *Lingua* 265 (2022): 102970. https://doi.org/10.1016/j.lingua.2022.103376.

Wray, A. "Formulaic Language." In *Encyclopedia of Language and Linguistics,* 2nd ed., edited by Keith Brown, 590–593. Elsevier, 2006.

Siyuan Liu
(University of Bologna, Italy)
<liusiyuan0000@gmail.com>

Bridging User Needs with Interlanguage Corpus Design and Applications: The Case of MICICL

This paper introduces MICICL (Multimodal Interlanguage Corpus for Italian Chinese Learners), a novel interlanguage corpus designed for Italian-speaking learners of Chinese. As a multimodal corpus, MICICL aims to meet the specific needs of Chinese language teachers and learners by offering user-oriented features and practical applications. This abstract outlines MICICL's design philosophy, development progress, and its potential to enhance teaching and learning practices.

1. User-Centered Multimodal Data Structure

To accurately represent learners' reading and writing abilities, MICICL employs a multimodal structure, which integrates three distinct types of data:

- Images: Scanned copies of learners' handwritten texts.

- Audio: Recordings of learners reading their handwritten texts aloud.

- Text: Transcriptions of both the image and audio data.

This structure enables a comprehensive and multidimensional representation of learner performance. By including both written and spoken modalities, MICICL provides an enriched dataset for error analysis and pedagogical interventions.

2. Innovative Transcription Rules

The transcription approach in MICICL is tailored to maintain the authenticity of the original data while addressing the specific needs of researchers and educators.

- Image Transcriptions: Handwritten texts are transcribed using Chinese characters, reflecting learners' orthographic accuracy.

- Audio Transcriptions: Spoken data is transcribed using Pinyin, the phonetic system for Mandarin, offering insights into learners' pronunciation and tonal accuracy.

This dual transcription system ensures that each data type is represented in a manner that best captures its unique linguistic characteristics, providing researchers with a clear view of both written and spoken interlanguage performance.

3. Multi-Level Annotation Framework

MICICL employs a two-layer annotation system to enhance the usability of the corpus for analysis and teaching:

- Error Annotation: Identifies error locations and categorizes error types, allowing detailed tracking of learners' mistakes.

- Correction Annotation: Provides native-speaker corrections for the identified errors, offering an ideal target form for comparison.

This layered annotation system ensures that users can access both the raw data (including errors) and corrected versions, enabling a variety of analytical and instructional applications. The annotation guidelines are based on widely-used standards, such as those from the HSK Dynamic Composition Corpus, but have been adapted to reflect the specific features of MICICL's data.

4. Longitudinal and Cross-Sectional Data Collection

MICICL's data collection spans three academic semesters, with approximately one data collection session per month. It currently includes 151 image files and 153 audio files from 38 learners of varying proficiency levels. Among these, 151 image files are paired with corresponding audio recordings.

Participation among learners varies, with some contributing data from a single session and others participating in up to ten sessions. This design enables both:

- Cross-sectional Analysis: Comparisons between learners of different proficiency levels.

- Longitudinal Tracking: Observation of individual learners' progress over time.

This combination of cross-sectional and longitudinal data makes MICICL a rich resource for investigating interlanguage development and variation.

5. Applications for Teachers and Learners

MICICL's primary aim is to maximize its utility for Chinese language educators and learners. Beyond serving as a research tool, the corpus is actively used to design learning materials and pedagogical interventions. By analyzing learner errors and identifying patterns, MICICL helps

develop targeted resources that address common challenges faced by learners. For instance, instructional materials can incorporate authentic learner errors and corrections, making them more relevant and practical for classroom use.

Furthermore, MICICL's multimodal design allows for innovative teaching applications, such as training learners to self-correct their errors by comparing their handwritten and spoken data with native-speaker corrections. This holistic approach ensures that MICICL not only provides a corpus for academic research but also delivers actionable insights for educators and learners.

MICICL bridges the gap between user needs and interlanguage corpus design by offering a multimodal structure, detailed annotation, and longitudinal data. It provides a versatile tool for error analysis, learner assessment, and material design. As the corpus continues to develop, future work will focus on expanding its scope, refining its usability, and gathering user feedback to further enhance its relevance and impact.

References

Atkins, Sue, Jeremy Clear, and Nicholas Ostler. Corpus Design Criteria. *Literary and Linguistic Computing* 7, no. 1 (1992): 1–16.

Callies, Marcus. "Current Perspectives on Learner Corpus Research." *Arbeiten aus Anglistik und Amerikanistik* 48, no. 1 (2023): 37–51.

Callies, Marcus. Challenges in the Compilation, Annotation, and Analysis of Learner Corpus Data. In *Studies in Corpus Linguistics*, vol. 118, edited by M. Kaunisto & M. Schilk, 55–67. John Benjamins Publishing Company, 2024.

Corder, Stephen Pit. "Error Analysis, Interlanguage, and Second Language Acquisition." *Language Teaching* 8, no. 4 (1975): 201–218.

Zhang, Bao Lin, and Cui Xi Liang. "Quanqiu Hanyu Zhongjieyu Yuliaoku Jianshe he Yanjiu de Sheji Linian [Design Concepts of "the Construction and Research of the Inter-language Corpus of Chinese from Global Learners"]." *Language Teaching and Linguistic Studies* 5 (2013): 27–34.

Granger, Sylviane. “The Contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching: A Critical Evaluation.” In *Corpora and Language Teaching, edited by* A. O’Keeffe & M. McCarthy, 13–32. Cambridge University Press, 2009.

Huang, Wei. “Duomotai Hanyu Zhongjieyu Yuliaoku Jianshe Chuyi [A Discussion on the Building of Multimodal Chinese Interlanguage Corpora.” *Journal of International Chinese Teaching* 3 (2015): 60–66.

Iurato, Alessia. “Learner Corpus Research Meets Chinese as a Second Language Acquisition: Achievements and Challenges.” *Annali di Ca’ Foscari. Serie Orientale* 58 (2022a): 1–34.

Iurato, Alessia. Investigating Chinese Learner Corpus Research and Learner Corpora. Language Learning 26, no. 1, (2022b): 531–562. https://doi.org/10.13135/1825-263X/7017

Meunier, Fanny. Introduction to Learner Corpus Research. In *The Routledge Handbook of Second Language Acquisition and Corpora*, edited by N. Tracy-Ventura and M. Paquot, 23–36. Routledge, 2021.

Ren, H. B. “Guanyu Zhongjieyu Yuliaoku Jinashe de Jidian Sikao——Yi HSK Dongtai Zuowen Yuliaoku Weili [Towards to the Construction of the Inter-language Corpus of Chinese——Using the Dynamic Corpus of Writing Papers from HSK as an Example].” *Language Teaching and Linguistics studies*, no. 6 (2010): 8–15.

Sun, Dekun K. “Zhongjieyu Lilun yu Hanyu Xide Yanjiu [Interlanguage theory and Chinese acquisition research].” *Applied Linguistics*, no. 4 (1993): 82–92. https://doi.org/10.16499/j.cnki.1003-5397.1993.04.011.

Noriko Matsumoto
(Osaka Electro-Communication University, Osaka, Japan)
<matsumoto@osakac.ac.jp>

A Possible Explanation as to Why 'Go to Shop' is Unacceptable

This talk proposes a possible explanation as to why '*go to shop*' is unacceptable and why '*go shopping*'. is acceptable, because no current grammar textbooks for advanced students and their teachers have treated such a fundamental distinction. The explanation provided here should be supported by three things—corpus findings from the ICLE (*International corpus of learner English*) as a learner corpus and the *Collins Wordbanks Online* as a native speaker corpus, and descriptions of English grammar proposed by linguists working in the field of theoretical linguistics. Based on these three things, this paper empirically demonstrates why the distinction between '*go to shop*' and '*go shopping*' is essential for both learners and teachers. The distinction offers a descriptively adequate and intuitively acceptable account of how both '*go to shop*' and '*go shopping*' function.

In this talk, it is not enough to only inform the learners and/or teachers that a particular element belongs to a given category and not to another. A pedagogical grammar is a description of a language aimed at foreign language learners and/or teachers. Its purpose is to promote insight into the foreign language, thereby facilitating the foreign language. In accordance with the purpose of pedagogical grammar, it is clear that explanations constitute a powerful promoter of insight. As pointed out by Plunkett and Marchman (1993), learning by insight is more effective than mere rote learning. If the learner and/or teacher can see no reason why a foreign language should be as it is, the learner and/or teacher perceive its rules as arbitrary. Promoting insight means reducing the perceived arbitrariness of the foreign language. This talk argues that the reliability and validity of this idea are confirmed by the corpus-based approach to the distinction between '*go to shop*' and '*go shopping*'.

Both native-speaker and learner corpora are vital for English language teaching. Our corpus methodology depends on both learner and native speaker corpora to identify specific features of '*go to shop*' and '*go shopping*'. Comparing the corpus findings from learner and native speaker corpora enables researchers to identify not only particular grammatical forms that

learners have learned, but also the ones that learners have not learned. It is well known that comparing data from learner and native-speaker corpora enables researchers to identify overuse, underuse, and misuse in the English of learners (e.g., Barlow 2005, Gilquin, Granger, & Paquot 2007). Since the frequency data rests upon the presupposition of correctness, this key feature poses one significant problem that previous studies have seldom considered. The problem is that learner corpora never include linguistic phenomena that learners have not learned. This problem that the key feature of learner corpus research poses arises when researchers try to identify particular grammatical forms that learners have not learned. To solve the problem, this talk discusses the distinction between '*go to shop*' and '*go shopping*'.

In general, learner corpus research presupposes that learners have already learned particular grammatical forms, and that learner corpora include linguistic phenomena that learners have learned. If researchers try to identify particular grammatical forms that learners have not learned, learner corpora do not function as an effective means for identifying grammatical forms displayed by learners. This is because the learner's overuse, underuse, and misuse are almost nonexistent in this situation where correctness which is closely related to overuse, underuse, and misuse does not always function properly. In such a case, if a learner corpus is contrasted with a native-speaker corpus, data from a learner corpus reflects an area or areas in which learner language differs from native-speaker language in terms of frequency of distribution rather than correctness. Through presenting such corpus findings, we can establish the key area(s) where we should focus, when we provide grammatical input for learners and/or teachers. This means that a direct link between corpus findings and pedagogical relevance is established. The explanation provided here is a pilot scheme that can be extended to other grammatical areas. This paper thus claims that providing a possible explanation as to why '*go to shop*' is unacceptable and why '*go shopping*' is acceptable clearly makes a positive contribution to grammatical instruction.

References

Barlow, Michael. "Computer-based Analyses of Learner Language." In *Analyzing Learner Language*, edited by R. Ellis and G. Barkhuizen, 335–357). Oxford: Oxford University Press, 2005.

Gilquin, Gaëtanelle, Sylviane Granger, and Magali Paquot. “Leaner Corpora: The Missing Link in EAP Pedagogy.” *Journal of English for Academic Purposes* 6, no. 4 (2007): 319-335.

Plunkett, Kim, and Virginia Marchman. From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition* 48, no. 1 (1993): 21–69.

Martina Pavić
(Institute for Croatian language, Croatia)
<mpavic@ihjj.hr>

Adjectives in Disease Terms: A Comparison of Scientific and Popular Texts

Medicine is a field that sparks interest from diverse speaker groups, many of whom are neither experts nor semi-experts. Their engagement with the domain stems from the practical need for a basic understanding of concepts related to diseases, procedures, diagnoses, etc.

Multi-word terms referring to diseases often include adjectives that specify the nature of the disease, its severity, or duration, contributing to the consistency, precision, and classification of terms into semantic groups. Given that language use in specialized fields is influenced by contextual factors, M. A. K. Halliday's (1978) systemic functional linguistics framework highlights how context determines language choice through social roles, power dynamics, subject matter, and communication mode. Halliday (1964) defines register as a functional variation of language adapted to particular communicative situations.

As Sager (1997) notes, the primary function of terms is to facilitate knowledge transfer within a specialized field. Consequently, terminological variation is an inherent feature of medical language, as different levels of expertise among participants require linguistic adaptation (Daille et. al. 1996). In medicine, variation is further shaped by the limited or inconsistent application of standardized medical terminology, leading to communication challenges among professionals and between professionals and patients (Bowker and Hawkins 2006). One key factor influencing variation is the level of specialization. While the core meaning of a concept remains unchanged, different aspects may be emphasized depending on the communicative context – such as the underlying cause of a disease, its symptoms, or available treatment options. As the level of specialization decreases, the medical language increasingly mirrors the general language. Evidence of this shift can be found in varied expressions, redundancy, ambiguity, and a greater reliance on synonyms and paraphrasing to convey concepts (Cabré Castellví 1998, in Freixa 2006). Within this framework, the description of terminological variants considers semantic, pragmatic, and rhetorical aspects, highlighting their importance as part of the lexical richness of specialized language (Grčić Simeunović 2021).

In this study, the similarities and differences in the use of adjectives in terms referring to acute, chronic, infectious, and rare diseases will be analyzed across two different communication registers. The selection of acute, chronic, infectious, and rare diseases enables an exploration of contrasting medical concepts, from time-limited to long-term conditions, from common to uncommon cases, and from individual to public health concerns.

This analysis provides insight into linguistic strategies for shaping medical terminology and adapting specialized language for non-expert audiences. The study is based on two corpora previously compiled using *Sketch Engine* tools (Pavić 2022): a scientific corpus of research papers (5,318,395 words) and a corpus of texts from medical portals intended for the general public (5,022,639 words). The types of adjectives used in disease terms – such as relational adjectives (*infectious disease*), descriptive adjectives (*severe anemia*), or metaphorical adjectives (*silent killer*) – and their functions in shaping medical communication, whether by providing precision, evoking urgency, or making information more accessible to a general audience will be discussed.

Scientific texts are expected to use standardized terminology predominantly, in which relational adjectives are primarily used to describe specific characteristics of diseases accurately. These adjectives often indicate etiology (e.g., *bacterial meningitis*), duration (e.g., *chronic illness*), location (e.g., *cardiovascular disease*), or pathological features (e.g., *malignant tumor*) of the disease. This use of adjectives ensures clarity and precision in professional medical communication.

Popular texts may favor more emotive and simplified expressions accessible to the general audience (e.g., *severe, dangerous, well-known, deadly, mysterious*). They also emphasize severity and urgency, while scientific texts prioritize precise medical descriptions. The primary objective of this study is to identify patterns in adjective usage that exist between scientific and popular discourses, and how such differences influence the public perception of diseases. Ultimately, this research seeks to enhance our understanding of the role language plays in shaping public discourse on health.

References

Bowker, Lynne, and Shane Hawkins. “Variation in the Organization of Medical Terms: Exploring Some Motivations for Term Choice.” *Terminology* 12, no. 1 (2006): 79–110.

Freixa, Judit. “Causes of Denominative Variation. A Typology Proposal.” *Terminology* 12, no. 1 (2006): 51–77.

Daille, Béatrice, Benoît Habert, Christian Jacquemin, and Jean Royauté. “Empirical Observation of Term Variations and Principles for their Description.” *Terminology* 3, no. 2 (1996): 197–257.

Grčić Simeunović, Larisa. *Terminološki opis u službi stručnoga prevođenja. Dinamično modeliranje specijaliziranoga znanja*. Zadar – Zagreb: Sveučilište u Zadru – Institut za hrvatski jezik i jezikoslovlje, 2021.

Halliday, Michael Alexander Kirkwood. *The Linguistic Sciences and Language Teaching*. London: Longmans, 1964.

Halliday, Michael Alexander Kirkwood. *Language as Social Semiotic: the Social Interpretation of Language and Meaning*. London: Edward Arnold, 1978.

Pavić, Martina. *Uloga pridjeva u hrvatskome medicinskom nazivlju*. . PhD diss. Filozofski fakultet u Zagrebu, Zagreb, 2022.

Sager, Juan C. “Term Formation.” In *Handbook of Terminology Management*, vol. 1, edited by Wright, S. E.; Budin, G. 25–41. Amsterdam – Philadelphia: John Benjamins Publishing Company, 1997.

Oleksandra Palchevska (Lviv State University of Life Safety, Ukraine), **Petro Hubych** (Lviv State University of Life Safety, Ukraine), **Irena Snikhovska** (Zhytomyr Medical Institute of Zhytomyr Regional Council, Ukraine)
<palch56@ukr.net>; <irena29@gmail.com>

Features of Compiling the English-Polish-Ukrainian Fire and Rescue Terms Dictionary Based on the Specialized Texts Corpora

This project presents the research study that will result in compiling English-Polish-Ukrainian Fire and Rescue Terms Dictionary Based on the Specialized Texts Corpora. The trilingual dictionary will provide a list of terms and abbreviations used in Fire and Rescue sphere. The corpus-based approach also aims at identifying the best way to develop English and Polish to Ukrainian translators' competencies in the Fire and Rescue field.

Today, there are a large number of specialised software tools for automatic terminology extraction to work with texts of various sizes and topics. A significant contribution to the development of corpus-based terminology has been made by recent works, such as the study by Rigouts Terryn et al. (2021), which proposed methods of hybrid term extraction based on corpus data and deep learning models. In the field of corpus analysis of terms, the work of Malamatidou (2018) also stands out, proposing new approaches to corpus analysis in the field of translation studies. A considerable contribution was also made by Havrylenko (2019), who introduced the concept of 'terminology for translation' for adaptive terminology in specialised domains. At the practical level, tools for corpus analysis continue to improve. Versions of systems such as Sketch Engine make the corpus approach even more accessible for terminology research (Kilgarriff et al. 2014).

English and Polish to Ukrainian terminology processing in the Fire and Rescue field is one of the biggest challenges in this process. Dew to the development of safety sphere (new types of firefighting facilities and equipment, fireproof clothing, fire extinguishers and so on) English and Polish language Fire and Rescue action terminology is developing at a rapid pace, which leads to the entry of a large number of completely new terminological units. Translation of such language units is a considerable problem not only for Ukrainian Fire and Rescue professionals, users of the English language for the safety sphere but also for professional translators as well. The implementation of the project involves the following stages: building English, Polish and

Ukrainian text corpora on mine action, using language tools and concordances for distinguishing terms and compiling a multilingual dictionary.

At the first stage the task we have set was to build 3 corpora comprising the special texts on fire rescue thematic. A special corpus is usually designed from the texts already stored in a general corpus. Obviously, text samples included in our special corpora belong to the Fire and Rescue field subject, with emphasis on special aspects and properties of the terminological units that we want to extract and explore. We have implemented the Sketch Engine corpus manager to form 3 corpora of special texts (separate one-language corpus for each language) dealing with the main regulations and particularities of the firefighting and rescue sphere. For this study, the commercial Sketch Engine software (2025) was chosen because it gives the possibility to bult a corpus of special texts in English, Polish and Ukrainian compare frequencies with the other corpora of the languages mentioned and thus identify unique terminological vocabulary. Sketch Engine gives several possibilities for corpus building – we may build a parallel texts corpus, build a corpus from your own files and a corpus from the web. We stopped on building corpus from the web because the parallel texts on the fire and rescue subject are not available. On the second stage of managing the corpora we added our own files. The key words for building the English language corpus were fire security, firefighting, rescue, extinguish, Polish – gaśnica, gaszenie pożaru, sprzęt pożarniczy, halligan and Ukrainian – пожежна безпека, гасіння, пожежний гідрант, вогнерезистентність.

The main terminology describing the fire safety field was extracted using several corpus tools: frequency words lists construction, key words and terms identification, detecting collocations and word combinations with the Word Sketch tool, synonyms and similar words identification using the Thesaurus tool, multy-word expressions extraction with the help of N-gram and SQL tools.

Results showed that a large group of firefighting terms are metaphors derived from the names of human body parts, including terms generated by resemblance to the head, neck, torso, upper and lower limbs, and other organs. Thus, the lexeme головка, which is derived from the body part head, can be found in the following terms: з‘єднувальні головки (пожежні), головка ізолятора, сполучна головка, головка рукава, головка типу Богданова, «ікла» з‘єднувальних головок, гвинтова головка, рукавна (пожежна) з‘єднувальна головка; (пожежна) головка – заглушка. The component head in emergency terminology is used in the

following nominative units-metaphors: hermetic bulkheads – waterproof bulkheads headband – a helmet suspension that can be adjusted to fit the head) band – a belt, headfire - forest wildfire (covers the entire forest from the soil layer to the treetops; headgear - helmet; headphone – headset. The following metaphors have been observed in Polish: głowica (derived from the word głowa – head) – in firefighting terminology means the head "the working part (end) of a device", e.g: głowica otworowa – exhaust head, głowica sonaru – sonar head, głowica turbo standardowa – turbo standard head, głowica wkrętu – screw head, głowica stojaka – riser head

References

Havrylenko, Valeriia. "Comparison of Automatic Systems of Term' Extraction." *Advanced Linguistics* 4 (2019): 29–33.

Kilgarriff, Adam, Vit Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vit Suchomel. "The Sketch Engine: Ten Years On." *Lexicography* 1 (2014): 7-36.

Malamatidou, Sofia. *Corpus Triangulation. Combining Data and Methods in Corpus Based Translation Studies*. London: Routledge, 2018.

Rigouts Terryn, Ayla, Véronique Hoste, and Els Lefever. HAMLET: Hybrid Adaptable Machine Learning Approach to Extract Terminology. *Terminology* 27, no. 2 (2021): 290-329.

Sketch Engine. https://www.sketchengine.eu/

María Sampedro Mella
(Catholic University of Louvain, Belgium)
<maria.sampedro@uclouvain.be>

Corpus in L2 Teaching: Enhancing Discourse Skills through Pragmatically Annotated Corpora

Since last decades, corpora have been extensively used in research. However, there are still not many research proposals based on corpus data, and L2 teachers are not highly motivated to use corpora in their courses. Various reasons have been identified, such as logistical difficulties, teachers' skepticism regarding the use of corpora in language courses, and the time-consuming nature of preparing both the materials and the learners to use these tools (Granger & Gilquin, 2010, Asención-Delaney, et al. 2022). Furthermore, most corpus resources have been developed for lemmatization and morphosyntactic labeling of texts, enabling various types of morphosyntactic searches, as well as the retrieval of specific information corresponding to the queries made. However, only a few corpora include pragmatic annotation systems, which allow for the automatic extraction of discursive and pragmatic features, such as speech acts, which are more closely related to didactic approaches based on learning communicative functions.

This paper aims to promote the teaching and learning of pragmatic skills through the use of corpora (Aston, 2001; Sinclair, 2004) within the framework of the so-called 'Data-driven Learning' (DDL) (Leech, 1997). This approach facilitates foreign language acquisition through guided learning by using corpus data.

In particular, I will focus on L2 Spanish. Although in recent years there has been a growing number of L2 Spanish didactic proposals based on DDL, most of them still focus on grammar or vocabulary rather than pragmatics (e.g., Abad Castelló, 2019; Marcos Miguel, 2020; Abad Castelló & Álvarez Baz, 2021; Ballestero de Celis & Sampedro Mella, 2021, among others). In this presentation, firstly, three pragmatically annotated Spanish corpora are briefly presented: the ESLORA corpus (Corpus for the Study of Oral Spanish) (Barcala et al., 2018), the COREMAH corpus (Spanish Multimodal Corpus of Speech Acts) (Vacas Matos, 2020), and the CORDIAL corpus (Linguistically Annotated Didactic Oral Corpus) (Nicolás Martínez, 2012). I will give examples of the didactic potential of these corpora. In particular, I will show how linguistic material can be easily retrieved for L2 courses, both for class preparation and for

the design of language learning activities. Additionally, I will compare a selection of corpus examples of communicative functions (such as greetings, requests, and offers) with equivalent examples from L2 Spanish textbooks to highlight why teachers, learners, and textbook developers should use corpora.

Finally, the advantages of using pragmatically annotated corpora in the context of L2 Spanish are summarized.

References

Abad Castelló, Magdalena. "Uso de corpus lingüísticos por y para profesores de español como lengua extranjera." *redELE: Revista Electrónica de Didáctica ELE* 31 (2019): 148–167. https://doi.org/10.4438/1571-4667-redele.

Abad Castelló, Magdalena, and Antonio Álvarez Baz. "Aprendizaje basado en datos en español como lengua extranjera: Ampliando el ámbito." *Revista Internacional de Lenguas Extranjeras* 16 (2022): 1–20.

Asención-Delaney, Yuly, Joseph Collentine, Jersus Jose Colmenares, and Alfredo Urzúa. "Training Teachers to Use Corpus Tools in the Spanish Language Classroom." *Journal of Spanish Language Teaching* 9, no. 2 (2022): 134–147.

Aston, Guy. *Learning with Corpora*. Bolonia: CLUEB, 2001.

Ballestero de Celis, Carmen, and María Sampedro Mella, M. "La didáctica del español como lengua extranjera: un acercamiento desde la lingüística de corpus." In *Aportes del CAES a la enseñanza del español como lengua extranjera: estudios aplicados de gramática y pragmática*, edited by C. Ballestero de Celis & M. Sampedro Mella, 9–20. Universidade de Santiago de Compostela, 2021.

Barcala, Mario, Eva Domínguez, Alba Fernández, Raquel Rivas, María Paula Santalla, Victoria Vázquez, and Rebeca Villapol. "El corpus ESLORA de español oral: diseño, desarrollo y explotación." *CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Lingüísticos* 5, no. 2 (2018): 217–237.

Leech, Geoffrey. “Teaching and Language Corpora: A Convergence.” In *Teaching and Language Corpora*, edited by A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles, 1–23. Routledge: 1998.

Gilquin, Gaëtanelle, and Sylvie Granger. “How Can Data-driven Learning be Used in Language Teaching?” In *The Routledge Handbook of Corpus Linguistics,* edited by A. O’Keeffe y M. McCarthy, 359–370. Routledge, 2010.

Marcos Miguel, Nausica. “Aplicaciones para la didáctica de ELE del corpus del español.” *E-SEDLL* 3 (2020): 32–44.

Nicolás Martínez, Carlota. “C-Or-Dial (Corpus Oral Didáctico Anotado Lingüísticamente).” Cima Press, 2012.

Vacas Matos, Marta. “El aprendizaje de la pragmática a través de un corpus multimodal de nativos y aprendientes de español.” *Revista Nebrija de lingüística aplicada a la enseñanza de lenguas* 14, no. 29 (2020): 17–32.

Sinclair, John. *Corpus and Text — Basic Principles, in Developing Linguistic Corpora: A Guide to Good Practice Corpus*, edited by M. Wynne, 5–24. Oxbow Books, 2004.

C-OR-DIAL: Linguistically Annotated Didactic Oral Corpus.

COREMAH: Spanish Multimodal Corpus of Speech Acts.

ESLORA: Corpus for the Study of Oral Spanish.

Sanja Virovec
(Department of Turkology, Faculty of Humanities and Social Sciences)
<svirovec@m.ffzg.hr>

Corpus-based Research as a Useful Tool in Contrastive Linguistics

This paper researches the importance of corpus (or, in other terms, corpora) in contrastive linguistics. The languages examined are Croatian and Türkiye Turkish, two genetically and structurally different languages. The linguistic category in the focus of this research is *factitive* or, as some scholar call it, *factitive-causative*.[1] The factitive is a voice category that expresses the meaning of "making someone do something", "forcing someone to do something", "stimulating someone to do something". While Turkish grammar gives much attention to this voice category, Croatian grammars do not even mention it.[2] There is no term or dedicated chapter in Croatian grammar books that addresses the factitive. That is, on the one hand, logical because Turkish, as an agglutinative language, has specialized suffixes for expressing the factitive: -DIr-, -Ir-, -Ar-, -It- and -t-. As we can conclude from the presence of multiple suffixes, factitive is a quite productive verb category in Turkish. Nevertheless, Croatian language also can and does express the factitive voice. However, there are only a few linguistic articles that discuss it, and they do so in a quite rudimentary way.[3] If we were to compare the two languages, Turkish and Croatian, only by comparing their grammar books, we might come to the conclusion that Turkish has suffixes to form factitive verbs whereas Croatian does not have a way to express factitive meaning, the meaning of "making someone do something". That conclusion is quite misleading and incorrect. Like many languages, Croatian also possesses the meaning of "making someone do something" and a like within its "mental grammar". It expresses it in a few ways. However, if we cannot rely on grammar books to supply us this information, how can we conduct research on this matter? The best way, in our opinion, is to

[1] Compare Marković, Ivan (2013) Uvod u jezičnu morfologiju. Zagreb: Disput and Čaušević, Ekrem (1996) Gramatika suvremenoga turskog jezika. Zagreb: Sveučilišna naklada.
[2] Compare Babić et al. (2005) Hrvatska gramatika. Zagreb: Školska knjiga, Maretić, Tomo (1963) *Gramatika hrvatskoga ili srpskoga književnog jezika*. Zagreb: Matica hrvatska, Polančec, Jurica (2015) „O definiciji kategorije glagolskog stanja". *Filologija,* 65: 113–130 and Žagar-Szentesi, Orsolya (2011) „Funkcionalne varijante konstrukcije dati (se) + infinitiv u hrvatskom jeziku – u okviru gramatikalizacije". *Suvremena lingvistika,* 37: 295–318.
[3] Compare Virovec (2019) Usporedba faktitiva-kauzativa u turskome i hrvatskome (unpublished master thesis).

conduct a corpora-based research. By corpora, we do not mean computer-supported corpora[4], although such an upgrade would most definitely be useful for providing statistical and numerical data. For example, it would make easier to show which ways of expressing factitive are more common, which are less common etc. Given that in Turkish factitive is expressed in a morphological way and in Croatian in a syntactical one, the best way of making a corpus is parallel reading of the original and the translated text. Here by "original" we mean the Turkish text because Turkish is the language with widely used factitive voice, therefore it must be the source-language of the research. By "translated" text we mean the Croatian text - it is a sort of "control-corpus" - it is a corpus from which we will determine whether Croatian, and if so, in which way(s) it expresses the category we are researching. During the process of reading the texts, a table with two columns should be created: one listing all the Turkish factitive examples one by one and another providing their Croatian equivalents. It should be pointed out that the person conducting the research and reading the texts should be a linguistic expert in both languages. Besides that, the texts that form the corpus should be written by master writers (for Turkish texts) and translated by qualified translators (for Croatian texts). It would also be excellent if the original texts are of different authors, in that way the conclusion based on the corpus would be more reliable. If all the texts are of the same author, scholars could argue that the corpus is just a result of a style of one author, not so much a reflection of the linguistic reality. After such a research, we become able to list the ways in which Croatian expresses factitive: 1) by using transitive verbs, 2) by using the verb učiniti "to do" + instrumental case, 3) through semantical means, for example umrijeti - ubiti ("to die" – "to kill"), 4) by using infinitive of the verbs with the meaning of "to persuade" and a like + co-sentence and 5) by using the verb dati "to give, to make" + infinitive of a verb that carries the main meaning. To conclude, corpus-based research is, in our opinion, quite helpful and useful in contrastive linguistics. Ultimately, it can even lead to revolutionary findings and contributions of big importance for the grammars of the researched languages.

[4] See Zalizniak et al. (2012) and Mc Enery and Hardie (2012).

Bibliography:

Barić, E., M. Lončarić, D. Malić, S. Pavešić, M. Peti, V. Zečević, M. Znika. *Hrvatska gramatika*. Zagreb: Školska knjiga, 2005.

Čaušević, Ekrem. *Gramatika suvremenoga turskog jezika*. Zagreb: Sveučilišna naklada, 1996.

Mc Enery, Tony, and Andrew Hardie. *Corpus Linguistic*. Cambridge: Cambridge University Press, 2012.

Maretić, Tomo. *Gramatika hrvatskoga ili srpskoga književnog jezika*. Zagreb: Matica hrvatska, 1963.

Marković, Ivan. Uvod u jezičnu morfologiju. Zagreb: Disput, 2013.

Polančec, Jurica. "O definiciji kategorije glagolskog stanja". *Filologija* 65 (2015): 113–130.

Virovec, Sanja. *Usporedba faktitiva-kauzativa u turskome i hrvatskome*. Unpublished master thesis, 2019.

Zalizniak, Anna A., Maria Bulakh, Dmitrij Ganenkov, Ilya Gruntov, Timur Maisak, And Maxim Russo. "The Catalogue of Semantic Shifts as a Database for Lexical Typology". *Linguistics* 50, no. 3 (2012): 633–669.

Žagar-Szentesi, Orsolya. "Funkcionalne varijante konstrukcije dati (se) + infinitiv u hrvatskom jeziku – u okviru gramatikalizacije". *Suvremena lingvistika* 37 (2011): 295–318.

Bogdan Vuruna
(Faculty of Philology, University of Belgrade, Serbia)
<bogdan93vuruna@yahoo.com>

Serbian for Foreign Students: Variations of Positions of Reflexive Pronoun 'Se' in Past and Future Tense of Reflexive Verbs

This study examines the challenges foreign students are facing when acquiring the correct position of the reflexive pronoun "se" in past and future tense of reflexive verbs in Serbian. The analysed corpus consists of essays from thirty students from diverse language and cultural backgrounds, predominantly from African and Asian countries, all of them being proficient in English, French, or Portuguese, who have been learning Serbian for seven months. This study is based on the framework of contrastive analysis (Kurteš, 2006; Odlin, 2003), with particular attention to language transfer and cross-linguistic interference. The observed errors are analysed as evidence of structural influence from learners' first and second languages, as well as developmental features of their interlanguage (Odlin, 2003; Moulton, 2015).

The analysis revealed three main categories of errors: misplacement of the pronoun "se" using reflexive verbs in sentences, insertion where it is not needed, and omission where it is essential - even though instructions given included verbs were in infinitive form, together with the pronoun "se". Our hypothesis is that the acquisition of reflexive constructions in Serbian is affected by several interrelated factors: the typological distance between Serbian and learners' L1/L2 (Miličević, 2008a; Rivero Sheppard, 2001), the absence or different functioning of reflexivity in their known languages (Miličević, Butt & King, 2009), and individual learning strategies such as self-study and metalinguistic awareness (Gilquin, 2024; Ilic, 2021).

The findings underscore the necessity of applying targeted teaching strategies in the acquisition of Serbian reflexive verb syntax, particularly in past and future tense where, besides the pronoun "se", students need to pay attention to the position of auxiliary verbs and pronouns (Miličević, 2008b; O'Connor, 2002; Milićević, 2009). The research aims to contribute to the understanding of second-language acquisition among students from multilingual backgrounds, as well as to provide insights into patterns of the cross-linguistic interference and the cognitive processes involved in second-language acquisition.

References:

Gilquin, Gaëtanelle. "From Second Language Acquisition Research to Foreign Language Teaching through the Prism of Corpora." *Ampersand* 13 (2024): 100204.

Ilic, Nina. The Acquisition of se-verbs in Serbian as L1 (Usvajanje glagola sa klitikom se u srpskom jeziku kao maternjem). PhD diss., University of Novi Sad (Serbia), 2021.

Kurteš, Svetlana. *Reflexivity and Middleness in English and Serbian (a Contrastive Study)*. PhD diss., University of Novi Sad (Serbia), 2006.

Milićević, Jasmina. "Linear Placement of Clitics and Syntactic Dependencies (Serbian Second-Position Clitics)". In *Dependency in Linguistic Description.* John Benjamins Publishing Company, 2009.

Miličević, Maja. "On the Productivity of Reflexive and Reciprocal se in Serbian." In *FASSBL6*, (2008): 85–91.

Miličević, M. (2008). "Reflexive and Reciprocal Forms in L2 Serbian. In *Bosnisch, Kroatisch, Serbisch. B. K. S. als Fremdsprachen an den Universitäten der Welt. Die Welt der Slaven*, edited by B. Golubović and J. Raecke, vol. 31, 143–154. München: Verlag Otto Sagner.

Miličević, Maja. "On the Status of Clitic Reflexives and Reciprocals in Italian and Serbian." In *Proceedings of LFG09*, 441–458. 2009.

Moulton, Erin Elizabeth. *Rethinking Reflexivity: SJA/SE-verbs in Russian and Bosnian/Croatian/Serbian*. PhD diss., University of Kansas, 2015.

O'Connor, Rob. *The Placement of Enclitics in Bosnian, Croatian and Serbian*. Manuscript, University of Manchester, 2002.

Odlin, Terence. "Cross-linguistic Influence." In *The Handbook of Second Language Acquisition*, edited by Catherine J. Doughty, and Michael H. Long, 436-486. 2003.

Rivero, María-Luisa, and Milena Sheppard. "On Impersonal se/się in Slavic." *Current Issues in Formal Slavic Linguistics,* 137–147. Peter Lang, 2002.