

University of Zadar

International conference

*Corpora in Language Learning, Translation and
Research (CLLTR)*

BOOK OF ABSTRACTS

University of Zadar

Zadar

August 23rd – 24th, 2023

International conference
Corpora in Language learning, Translation and Research
Book of Abstracts

Publisher

University of Zadar

For the publisher

Josip Faričić, Rector

Edited by

Larisa Grčić

ISBN

978-953-331-464-8

PROGRAMME COMMITTEE

Maja Bratanić (Croatia), Paula de Santiago (Spain), Adriana Mezeg (Slovenia), Maja Lončar (Croatia), Ana Ostroški Anić (Croatia), Mojca Pecman (France), Ivanka Rajh (Croatia), Špela Vintar (Slovenia)

ORGANIZING COMMITTEE

Larisa Grčić (President), Vanda Mikšić, Maja Bahnik (University of Zadar, Croatia)

Programme

August 23, 2023 Corpora as resources in specialized and literary translation

08:00-08:30	<i>Registration</i>
08:30-09:00	Opening – welcome speech
09:00-10:30	<i>Plenary lecture: Marija Brkić Bakarić</i> (Faculty of Informatics, University of Rijeka)
10:30-12:00	Session 1
10:30-11:00	Bojana Mikelenić, Antoni Oliver (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia) <i>A multilingual literary parallel corpus and its applications in machine translation</i>
11:00-11:30	Jelena Parizoska (University of Zagreb, Croatia) <i>Teaching idiomatic variation in EFL: How corpus data can enhance students' metaphoric competence</i>
11:30-12:00	Marta Petrak, Ivana Franić (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia)

	<i>Interface of derivational morphology and translation: prefixation in parallel corpora</i>
12:00-12:30	Discussion
12:30-14:30	Lunch break
	Session 2
14:30-15:00	José Luis Poblete (Universidad de Santiago, Chile), Kevin Frank Gerigk (Lancaster University, UK) <i>Corpus Methods for Translation Studies – An exploration of the perceptions of Chilean undergraduate translation students</i>
15:00-15:30	Antonio Hermán-Carvajal (University of Granada, Spain), <i>Emotional characterization of English-language mental health texts for youth: a corpus study</i>
15:30-16:00	Anita Lemić Stagličić (Polytechnic “Nikola Tesla” Gospić, Croatia) <i>Mistranslations and misinterpreting in the legal context</i>
16:00-16:30	Kaja Mandić (University of Mostar, Bosnia and Hercegovina) <i>A nursing corpus and the Academic Collocation list</i>
16:30-17:00	Frane Malenica (Department of Anglistics, University of Zadar, Croatia) <i>Picking up the scraps – creating a specialized corpus using web-scraping tools</i>
17:00-17:30	Goranka Antunović, Jelena Čović, Patrik Posedi (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia) <i>Sitting and pondering the meaning and potential Croatian renderings of Swedish pseudocoordinated constructions with posture verbs</i>
17:30-18:00	Discussion

August, 24, 2023 Data-driven learning (DDL) in foreign language education

09:00-10:30	Plenary lecture: Nikola Ljubešić (Institut Josef Štefan, Ljubljana, Slovenia) ONLINE
10:30-12:00	Session 1
10:30-11:00	Yusuf Öztürk, Samed Yasin Öztürk (Mus Alparslan University, Turkey) <i>(er) .. when we at school: A Corpus-Based Analysis of Errors and Speech Fluency</i>
11:00-11:30	Emir Feridun Caliskan, Yusuf Öztürk (Mus Alparslan University, Turkey) <i>Employing a Corpus-based Approach in First Language</i>
11:30-12:00	Samed Yasin Öztürk, Yusuf Öztürk (Mus Alparslan University, Turkey) <i>In the Aftermath of an Unprecedented Catastrophe: Examining Changing Themes in a Corpus of News Articles</i>
12:00-12:30	Discussion
12:30-14:30	Lunch break
	Session 2
14:30-15:00	Jana Kegalj, Mirjana Borucinsky (Maritime faculty, University of Rijeka, Croatia) <i>Data-Driven Learning for Writing Skills Development</i>
15:00-15:30	Pedro Humánez-Berral (University of Cantabria. Santander (Spain), <i>Unveiling Pronunciation and Motivation: An Examination using an Oral Learner Corpus</i>

15:30-16:00	<p style="text-align: center;">Zorica Đurović (Faculty of Maritime Studies Kotor, Montenegro)</p> <p style="text-align: center;"><i>Some corpus linguistics methods for vocabulary research and teaching</i></p>
16:00-16:30	<p style="text-align: center;">Eriola Qafzezi, (Fan S. Noli University, Korca, Albania)</p> <p style="text-align: center;"><i>Inside Out: Investigating the relationship between culture and language as reflected on uses of expressions with parts of body in Albanian language</i></p>
17:00-17:30	Discussion
17:30-18:00	Closing ceremony

PRESENTATIONS IN SESSIONS

SESSION I: **Corpora as resources in specialized and literary translation**

Bojana Mikelenić (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia)

Antoni Oliver (Universitat Oberta de Catalunya, Spain)

A multilingual literary parallel corpus and its application in machine translation

We start by presenting the building of the parallel multilingual and multidirectional corpus RomCro, which contains six languages: five Romance languages (Spanish, French, Italian, Portuguese, and Romanian) and Croatian. The corpus is composed of literary texts from the 20th and 21st centuries that have been selected manually and aligned with their translations to the rest of the languages. The language distribution of the original texts is as follows: 37.2% Spanish, 19.8% French, 16% Italian, 13.9% Romanian, 7% Portuguese, 6.2% Croatian. The alignment was performed automatically using Hunalign (Varga et al., 2005) via an open source program LF-Aligner and post-edited manually. In this way, the possible errors are minimised and the resource is more reliable. The lemmatisation and POS tagging were performed automatically through Sketch Engine (Kilgarriff et al., 2004), using the taggers available for each language. RomCro counts with 142,470 segments and 15.9 million words and the original order of sentences is scrambled. The corpus is available on platforms Sketch Engine and ELRC (European Language Resource Coordination), the latter being an untagged version in TMX and TSV formats under the CC-BY-NC-4.0 license. The fact that the corpus is composed of original texts in each language and their translation to the remaining languages makes it a valuable resource for translation studies and practice (cf. Baker, 1993; Laviosa, 1998; Johansson y Oksefjell, 1998).

This corpus can be very useful for several tasks related to literary translation. When looking at the available parallel corpora in the Opus Corpus Collection (Tiedemann, 2012), we can observe that there are almost no literary corpora in the collection, and when there is one available, it is very small. In the article, we test the use of RomCro for training neural machine

translation (NMT) systems tailored to literature between Spanish and the rest of languages in the corpus. As the number of segments is not enough for training, we extend the corpus by automatically selecting relevant segments in a large general domain corpus for those language pairs. This task will be performed using a Python script which performs the following steps: (1) language model calculation from the source part of the RomCro corpus for a given language; (2) calculation of the perplexity of all the source segments from the large parallel corpus; (3) selection of a given number of parallel segments from the large corpus sorted by the calculated perplexity, that is, selection of the parallel segments with smaller perplexity. We will train and evaluate NMT systems trained using Marian with a transformer configuration. The training corpus will be a combination of a part of RomCro and the selected parallel segments from the large corpus; the validation and the evaluation corpus will be formed by the remaining segments from RomCro. The following systems will be trained: Spanish-French, Spanish-Italian, Spanish-Portuguese, Spanish-Romanian and Spanish-Croatian. The CCMatrix corpus (Schwenk et al., 2021) will be used for the automatic selection of suitable parallel segments for all language pairs except Spanish-Croatian. For this language pair the larger available parallel corpus is MultiCC Aligned (El-Kishky et al., 2020). In table 1 the size in the number of segments is presented. As we can see, the size for the Spanish-Croatian language pair is not enough, and we will use transfer learning techniques for training the NMT systems. The trained systems will be evaluated using the following automatic metrics: BLEU (Papineni et al., 2001), TER (Snover et al., 2006) and COMET.

References

- Baker, M., 1993. Corpus Linguistic and Translation Studies: Implications and Applications. In *Text and Technology: In Honour of John Sinclair*. 233-250, John Benjamins Publishing Company, Philadelphia / Amsterdam.
- El-Kishky, A., Chaudhary, V., Guzmán, F., Koehn, P. 2020. CCAIined: A Massive Collection of Cross-Lingual Web-Document Pairs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 5960-5969.
- Laviosa, S. 1998. The Corpus-based Approach: A New Paradigm in Translation Studies. *Meta* 43(4). 474-479.
- Johansson, S., Oksefjell, S. (eds.). 1998. *Corpora and Cross-linguistic research: Theory, Method, and Case Studies*. Rodopi, Amsterdam / Atlanta.

Kilgarriff, A., Rychlý, P., Smrž, P., Tugwell, D. 2004. *The Sketch Engine*. *En Proc Eleventh EURALEX International Congress*. 105-116. Lorient, France.

Papineni, K., Roukos, S., Ward, T., Zhu, W. J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311-318.

Schwenk, H., Wenzek, G., Edunov, S., Grave, É., Joulin, A., Fan, A. 2021. CCMatrix: Mining Billions of High-Quality Parallel Sentences on the Web. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 6490-6500.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. 223-231.

Tiedemann, J. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* 2214-2218.

Varga, D., Halácsy, P., Kornai, A., Nagy, V., Nemeth, L. & Tron, V. 2005. Parallel corpora for medium density languages. In: *Proceedings of the RANLP 2005*. 590-596 (Borovets, Bulgaria).

Jelena Parizoska (University of Zagreb, Croatia)

Teaching idiomatic variation in EFL: How corpus data can enhance students' metaphoric competence

In the Common European Framework of Reference for Languages a good command of idiomatic expressions is an integral part of the lexical repertoire at C levels. In addition, it has been argued that the ability to understand and produce metaphor, which is termed metaphoric competence, is essential in L2 learning (Littlemore and Low 2006). That is why a number of EFL textbooks feature tasks with figurative expressions.

When it comes to idiomatic variation, teaching materials for EFL typically list idioms in only one form. This is in line with the traditional view of idioms as fixed expressions. For instance, it has been claimed that you cannot replace one word with another (*kick the bucket* → *kick the pail), you cannot add a word (*the long and the short of it* → *the very long and the short of it)

and you cannot change an idiom's grammatical structure (*face the music* → *the music was faced).

However, corpus-based studies of idioms show that variation is widespread (Moon 1998, 2015; Langlotz 2006; Omazić 2015). In fact, a number of English idiomatic expressions occur in two or more conventionalized forms, e.g. *call the tune/shots*, *like a (hot) knife through butter*, *let the cat out of the bag / the cat is out of the bag*.

It has also been shown that verb phrase (VP) idioms are particularly prone to changes and vary their lexico-grammatical structure systematically to construe the same event from different perspectives (Author 2020), e.g. *come to light* and *bring something to light*, *give someone the cold shoulder* and *get the cold shoulder*, *bury your head in the sand* and *take your head out of the sand*. Furthermore, idioms are frequently subject to creative use, especially in journalism and advertising (Naciscione 2010; Jaki 2014), e.g. *Old Hatreds Die Hard* (a newspaper headline), *Comfort is in the eye of the beholder* (an advert for contact lenses).

Cognitive linguistic studies show that such ad hoc manipulations are governed by similar restrictions as conventionalized variations (Author 2022). Given that variation is widespread and that creative use of idioms displays some regularity, it is important to raise learners' awareness of idiomatic variation by including it in the EFL classroom, particularly at advanced levels.

The aim of this paper is to show how linguistic findings about idiomatic variation can be applied in teaching EFL. Based on a study in SKELL (Sketch Engine for Language Learning, 1 billion words), it will be shown how teachers can create activities for students to explore idiomatic variation in a corpus themselves. We conducted a study of 100 VP idioms whose components and grammatical structure vary frequently (Moon 1998). Data from the SKELL corpus shows that the most common types of conventionalized variations are lexical substitution (*back/paint/force someone into a corner*), addition (*bitten by the travel/political/showbiz bug*), and lexico-grammatical changes (*come under the spotlight* and *put someone/something under the spotlight*).

Corpus data also shows that creative use follows specific patterns. For example, in the expression *beauty is in the eye of the beholder*, the idiomatic component *beauty* is regularly replaced by other words that denote concepts about which people have different opinions (e.g. cleanliness, quality, success). Based on the findings on the types of changes that idioms typically undergo, teachers can create activities which involve specific corpus searches

(substitution, addition etc.). The ultimate goal of such activities is to show students that idiomatic variation is systematic, even in cases when speakers play with idioms. In addition to letting students explore real language in use, teaching idiomatic variation will enhance their metaphoric competence.

References

Author. 2020.

Author. 2022.

Jaki, Sylvia. 2014. *Phraseological Substitutions in Newspaper Headlines: "More than Meats the Eye"*. Amsterdam, Philadelphia: John Benjamins.

Langlotz, Andreas. 2006. *Idiomatic Creativity*. Amsterdam, Philadelphia: John Benjamins.

Littlemore, Jeannette, and Graham Low. 2006. *Figurative Thinking and Foreign Language Learning*. Basingstoke: Palgrave Macmillan.

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.

Moon, Rosamund. 2015. 'Multi-word items'. In *The Oxford Handbook of the Word*, edited by John R. Taylor, 120–140. Oxford: Oxford University Press.

Naciscione, Anita. 2010. *Stylistic Use of Phraseological Units in Discourse*. Amsterdam, Philadelphia: John Benjamins.

Omazić, Marija. 2015. *Phraseology through the looking glass*. Osijek: Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences.

Marta Petrak (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia)

Ivana Franić (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia)

Interface of derivational morphology and translation: prefixation in parallel corpora

The goal of this paper is to analyse how morphologically complex words are translated, i.e. which strategies translators use when rendering derived words formed through prefixation. The question is addressed by comparing languages from three different branches of the Indo-European family: Romance (French), Germanic (English) and Slavic (Croatian). All of these

languages use prefixation as a rather productive word-formation mechanism (cf. Babić 2002; Apothéloz 2002; Bauer 2013).

Even though the general guiding principle in translation is that translators should be more concerned with meaning than with structure (Vinay and Darbelnet 1958: 21), linguistic microstructure also has a role in translation (Menzel 2018: 182). More precisely, translators should pay attention to “word-internal structures in complex words and the role of morphemes as the smallest individually meaningful units” (*ibid.*). This seems of particular interest in morphologically rich languages, such as Croatian or German, but in no way does it exclude more analytical languages such as English and French, which still retain some rather productive derivational mechanisms. This question makes part of a larger phenomenon of morphological awareness, which is the “reader’s conscious awareness of the morphemic structure of lexemes and their ability to reflect on and manipulate that structure” (Carlisle 1995: 194). Morphological awareness is not only important for the awareness of inflectional and word-formation processes, but also for textual and discourse competence, in particular the recognition of links between words that share related morphemes (*ibid.*). Owing to corpus linguistics, we know today that there are large differences in the use of derivational morphology across registers, genres and domains (Lefer 2012: 1325). When related specifically to translation, some earlier studies have shown that the so-called prefix-by-prefix translation (Lefer 2015) is the most common mechanism used, for instance between French and Croatian texts (Petrač 2021).

In our analysis, we shall start from French words prefixed with *sur-* ‘above; over’, and compare them with their translation equivalents in English and Croatian. In order to do so, we shall use EU DGT-UD parallel corpora for the three mentioned languages, which contain legal texts of the European Union from various specialized domains (such as motor vehicles, finance, medicine, etc.). We shall take into consideration *sur-* words (nouns, verbs and adjectives) with a minimal frequency of 10 occurrences.

According to preliminary data, words prefixed with *sur-* are rendered in English and Croatian texts either by corresponding prefixes (e.g. *super-* as in *supercharger* for *dispositif de suralimentation* or *over-*, as in *overestimation* for *surestimation* in English, and *pre-*, as in *pregrijavanje* for *surchauffe* or *nad-* as in *nadfrekvencija* for *surfréquence* for Croatian), descriptively (e.g. with the adjective *additional*, as in *additional collateral* for *surantissement* or *previsok* as in *previsoka procjena* for *surestimation*), or with completely different linguistic material (e.g. *survêtement* ‘lit. above-garment’ is translated into English as *tracksuit* and *trenirka* into Croatian).

It remains to be seen: 1) which is the most common strategy adopted in English and Croatian texts for translating sur- prefixed words, and 2) whether there are specific tendencies related to the translation of various parts of speech formed with that prefix.

References

Apothéloz, Denis. 2002. *La construction du lexique français*. Paris: Ophrys.

Babić, Stjepan. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. Zagreb: Nakladni zavod Globus.

Bauer, Laurie; Lieber, Rochelle; Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford & New York: Oxford University Press.

Carlisle, Joanne F. 1995. Morphological Awareness and Early Reading Achievement. In: Laurie Beth Feldman (ed.), *Morphological Aspects of Language Processing*. Hillsdale, NJ: Erlbaum, 189- 209.

Lefer, Marie-Aude. 2012. La préfixation française à travers les genres et les domaines : étude de corpus. *SHS Web of Conferences* 1, 3e Congrès Mondial de Linguistique Française, 1325-1349.

Lefer, Marie-Aude. 2015. Super-creative and over-bureaucratic: A cross-genre corpus-based study on the use and translation of evaluative prefixation in TED talks and EU parliamentary debates. *Across Languages and Cultures* 16 (2), 187-208.

Menzel, Katrin. 2018. Morphological Awareness and Translation Studies. In: Thomas Tinnefeld (ed.), *Challenges of Modern Foreign Language Teaching: Reflections and Analyses*, Saarbrücken: Htw Saar, 181-206.

Petrak, Marta. 2021. Strategije prevođenja motiviranih riječi: primjer prefiksacije. In: Mikić Čolić, Ana; Maja Glušac (ed.). *Jezik u digitalnom okruženju - Knjiga sažetaka*. Osijek: Filozofski fakultet Osijek, 106-107.

Corpora

EU DGT-UD French: https://www.clarin.si/noske/all.cgi/first_form?corpname=dgtud_fr

EU DGT-UD Croatian: https://www.clarin.si/noske/all.cgi/first_form?corpname=dgtud_hr

EU DGT-UD English : https://www.clarin.si/noske/all.cgi/first_form?corpname=dgtud_en

José Luis Poblete (Universidad de Santiago / Pontificia Universidad Católica de Valparaíso)

Kevin Gerigk (Lancaster University)

Corpus Methods for Translation Studies – An exploration of the perceptions of Chilean undergraduate translation students

This study investigates Chilean undergraduate translation students' perceptions regarding the use of corpus linguistics for the development of translation and L2 English skills. In translation studies, the focus lies on the development of students' intercultural-translational and linguistic competences (Contreras, 2016; Singer et al., forthcoming). In Chile, the development of language skills receives priority, since the majority of students achieve A2 Level (Council of Europe, 2020) of English at high school (Singer et al., 2021). This study reports on a teaching intervention, using a Content-and-Language-Integrated-Learning (CLIL) approach, conducted with a cohort of undergraduate students (n=8), enrolled in the elective module, at a university in Chile in collaboration with a UK university. Using a CLIL approach catered to two goals: 1) teaching new content to students, whilst 2) simultaneously creating opportunities for them to use the L2 in more complex ways (Wiesemes, 2009). To capture students' perceptions, we devised three research questions, investigating the development and inter-linking between 1) learner autonomy, 2) translation and language skills, and 3) corpus-based research skills. Initially, we captured the expectations, language learning and translation routines of the cohort, using an open-answer questionnaire (n=11). After the intervention, we devised a post-intervention Likert-Scale questionnaire (n=14) supported by an open-answer questionnaire (n=11) for data triangulation. This allowed us to assess the students' perceived progress and effectiveness of the intervention in terms of their translation and language skills. Our results indicate overall positive perceptions of our students in terms of language and translation skills development. Furthermore, we highlight aspects of the teaching intervention that were perceived particularly successful as well as those that require more attention in future. We conclude by explicating practical implications for translation courses at undergraduate level in the Chilean context.

References

Singer Contreras, N. D. (2016). A proposal for language teaching in translator training programmes using data-driven learning in a task-based approach. *International Journal of English Language and Translation Studies* 4(2). 155–167.

Singer, N., Poblete, J. L., & Velozo, C. (2021). Design, implementation and evaluation of a data-driven learning didactic unit based on an online series corpus. *Literatura y Linguística* 43. 391–420.

Singer, N., Velozo, C., & Poblete, J. L. (forthcoming). *Developing translation sub-competences by implementing Content and Language Integrated Learning (CLIL)*. John Benjamins Publishing Company.

Wiesemes, R. (2009). Developing Theories of Practices in CLIL: CLIL as Post-method Pedagogies? In Y. R. de Zarobe & R. M. Jimenez Catalan (Eds.), *Content and language integrated learning. Evidence from research in Europe*. 41–59. Multilingual Matters.

Antonio Hermán-Carvajal (University of Granada, Spain)

Emotional characterization of English-language mental health texts for youth: a corpus study

The mental health situation has worsened worldwide in recent years, especially after the outbreak of the COVID-19 pandemic. According to the World Health Organization (2020), back in 2020, almost 1 billion people were living with a mental disorder, 3 million people died every year from the harmful use of alcohol and one person died every 40 seconds by suicide. However, within this alarming frame, a “positive” effect of this deterioration in mental health can be found—the taboo surrounding this human dimension is weakening and mental health issues are being discussed more and more openly. Many young people, who are among the groups who suffer the most from mental health deterioration, do not have the financial means to afford psychotherapy and instead seek information online to better understand their situation. In this regard, it would be interesting to determine whether the inclusion of emotions (lemmas such as “joy,” “sadness” or “disappointment”) or emotional elements (such as colors or shapes that trigger specific emotions) in mental health texts designed for young people improves their understanding of these texts. As a preliminary step to this analysis, it is essential to comprehensively compile and analyze mental health texts designed specifically for young people, paying attention to the use of emotions, which improve learning, as well as the processing and comprehension of texts (Citron, 2012; Tyng et al. 2017, inter alia). In this paper, a corpus of English texts has been compiled and analyzed in terms of hedonic valence, arousal (Warriner et al., 2013) and according to the classification of emotion lemmas proposed by Láinez Ramos-Bossini and Tercedor Sánchez (2021). As this classification refers to emotion

lemmas in Spanish, this paper provides a translation of said lemmas into English by using the Lexemos project terminological database (<https://varimed.ugr.es/lexemos/>). This database provides insightful bilingual (Spanish-English) information on concepts and variants within the “emotion” domain (López-Rodríguez & Tercedor-Sánchez, 2022). Once the texts were compiled, SketchEngine (Kilgarriff et al., 2014) was used to extract the most representative words of the corpus. This extraction allowed us to determine the emotional load of the corpus. Certain linguistic patterns of emotion expression that occurred in these texts on mental health aimed specifically at young people were also extracted. Once they were identified, we analyzed how they differed from the patterns used in texts aimed at the general population. The results of this study revealed the need to adapt the texts on mental health aimed at young people not only from a terminological point of view, but also from an emotional point of view.

References

- Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and language*, 122(3), 211-226.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.
- Láinez Ramos-Bossini, A. J. y Tercedor Sánchez, M. (2021). Lenguaje y emoción: estudio basado en corpus de foros de debate sobre salud mental. In S. A. Flores Borjabad y R. Pérez Cabaña (eds.), *Nuevos retos y perspectivas de la investigación en Literatura, Lingüística y Traducción*, 1712-1737. Dykinson.
- López-Rodríguez, C. I., & Tercedor Sánchez, M. (2022). Caracterización léxicosemántica de emociones para la comunicación interlingüística inglésespañol. *Pragmalinguística* 30. 261-282.
- Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in psychology* 1454.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods* 45. 1191-1207.
- World Health Organization (2020). *World Mental Health Day: an opportunity to kick-start a massive scale-up in investment in mental health* [News release].

Anita Lemić Stagličić (Veleučilište "Nikola Tesla", Gospić, Croatia)

Mistranslations and misinterpreting in the legal context

Language and its grammatical codes are necessary for an understanding of the outside world, and to interact with members of the community. However, what if the community is not national, or do not share the same grammatical codes? The closest example of this type of community to us is the European Union, Economic Community whose main goal is to enable all its citizens' equality and freedom of movement within its borders. Equality of citizens cannot be realized without equal participation in community life, and equal participation presupposes knowledge of the law and regulations of society. When learning a foreign language, the speaker primarily observed rules and standards, and adopts the same. The correct use of grammar codes allows us to communicate with other community members, and eliminate possible noise during the communication process. If the recipient is unable to decode the message, there is a need for a translator. This paper deals with the problems in translating legal documents, and administrative style as very specific functional style, which characteristics prevent deviation from the norm. If legal text leaves space to style selection, the validity of the provided information is questionable, and there is a misunderstanding in communication because participants are not aware of how their codes differ. The aim of this paper is not to offer a solution, but to enable understanding of the core problems, how it could be a valid approach.

References

Bajčić, M.; Stepanić, M. 2010. (Ne)dosljednost pri prevođenju pojmova iz prava tržišnog natjecanja Europske unije. *Zbornik Pravnog fakulteta u Zagrebu*, Vol. 60 No. 3-4. 747-772.

Gačić, M. 2011. *Anatomija prevođenja zakonskoga teksta*. Zagreb : Institut za hrvatski jezik i jezikoslovlje – Hrvatska sveučilišna naklada

Husinec, S. 2011. Metafora u pravnom diskursu. u: Karabalić, Vladimir; Varga, Aleksa Melita; Pon, Leonard (ur.), *Diskurs i dijalog: teorije, metode i primjene* Zagreb: Hrvatsko društvo za primijenjenu lingvistiku; Osijek: Filozofski fakultet u Osijeku. 69–84.

Doranić, J. 2007. O prevođenju pravnih tekstova. *Prevoditelj* 29, 86. 17-19.

Knežević, B. 2010. Passive and passive-like constructions in the translation of a Croatian legislative text into English. *Jezikoslovlje* 11. 1. 25-51.

Novak, J. (ur.). 2006. *Priručnik za prevođenje pravnih propisa Republike Hrvatske na engleski jezik*. Zagreb : Ministarstvo vanjskih poslova i europskih integracija.

Pavlović, S. 2004. Prevođenje pravnih tekstova u kontekstu približavanja Europskoj uniji. *Prevoditelj* 29. 80/81. 24-26.

Kaja Mandić (University of Mostar)

A nursing corpus and the Academic Collocation list

Knowledge of frequent collocations is very important in the scientific discourse and writers should become familiar with collocations valued in a particular scientific discipline. Collocations enable English language learners to put words commonly used by native speakers to appropriate use in appropriate contexts. The study is based on the comparison of collocations from a corpus of articles written by native speakers of English and collocations from a corpus of English articles written by native speakers of the Croatian, Bosnian and Slovenian language with the Academic Collocation List (ACL). The objective of the study is to extract the most frequent two-word collocations from the two nursing corpora and compare them with the ACL. The study will produce two corpora that are classified as specialized corpora. Both, native and non-native corpora will include only English scientific research articles from the field of nursing science. The articles for the corpus of native speakers of English were chosen on the basis of the NAHRS Selected List of Nursing Journal 2016, and articles for the corpus of non-native speakers of English were chosen as the only available in this field of research and in countries of non-native speakers of English who are the subjects of the study. The focus is on noun-noun and noun-adjective collocations. There is a necessity to generate field specific academic collocations in order to strengthen learners' academic reading and writing proficiency.

References

Ackermann, K. & Chen, Y. H. 2013. Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes* 12(4), 235-247.

Bal, B. 2010. *Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars*. PhD Thesis. Georgia State University, Georgia, USA.

- Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Byrd, P.; Coxhead, A. 2010. On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *Papers in TESOL* 5, 31-64.
- Chen, Y.; Baker, P. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* (14)2. 30-49.
- Conrad, S. 1996. Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education* 8. 299-326.
- Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2). 213-238.
- Delač, D. 2009. *TermeX v1.0*. University of Zagreb, Faculty of Electrical Engineering and Computing.
- Delač, D., et. al. 2009. TermeX: A Tool for Collocation Extraction. In: Glebukh, A. (Ed.), *Computational Linguistics and Intelligent Text Processing*. 149-157.
- Demir, C. 2017. Lexical Collocations in English: A Comparative Study of Native and Non-native Scholar of English. *Journal of Language and Linguistic Studies* 13(1). 75-78.
- Donesch-Jezo. E. 2011. Comparison of generic organization of research papers in English and Polish. *Journalism and Mass Communication*. 185-200.
- Duan, M.; Qin, X. 2012. Collocation in English Teaching and Learning. *Theory and Practice in Language Studies* 2(9). 1890-1894.
- Durrant, P. L. 2008. High frequency collocations and second language learning. Ph. D. Thesis, University of Nottingham.
- Durrant, P.; Schmitt, N. 2009. To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47, 157-177.
- Fan, M. 2009. An exploratory study of collocational use by ESL students – A task based approach. *System* 37. 110-123.
- Farooqui, A., S. 2016. *A corpus-based study of academic-collocation use and patterns in postgraduate Computer Science students*. Ph. D. thesis, Department of Language and Linguistics, University of Essex.
- Gilquin, G., Granger, S., Paquot, M. 2007. Learner Corpora: The Missing Link in EAP Pedagogy. *English for Academic Purposes*, 1-26.

- Gledhill, C. 2000. Collocations in Science Writing. *Language in Performance Series* 22. 7-20.
- Gray, B., Egbert, J., Biber, D. 2017. *Exploring Methods for Evaluating Corpus Representativeness*. The 9th International Corpus Linguistics Conference. University of Birmingham.
- Hill, J. 1999. Collocational competence. *English Teaching Professional* 11. 3-6.
- Howarth, P. A. 1996. Phraseology in English Academic Writing. *Lexicographica Series Maior*: 75.
- Hsu, W. 2013. Bridging the vocabulary gap for EFL medical undergraduates: the establishment of a medical word list. *Language Teaching Research* 17(4). 454-484.
- Huang, P. 2017. Collocations in Computer Science Research Articles: A Corpus-based Study. Birmingham Conference Archive.
- Hyland, K. 2008. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*. 18(1).
- Hyland, K., Tse, P. 2007. Is there an academic vocabulary? *TESOL Quarterly* 41(2). 235-253.
- Jones, M., Durrant, P. 2010. What can a corpus tell us about vocabulary teaching materials? In O'Keeffe, A., McCarthy, M. (Ed.) *The Routledge Handbook of Corpus Linguistics*. 387-400.
- Lindquist, H. 2009. *Corpus Linguistics and the Description of English*. Edinburgh University Press.
- McEnry, T., Xiao, R., Tono, Y. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. Routledge.
- McIntosh, M. 1961. Patterns and ranges. *Language* 37 (3). 325-337.
- Men, H. 2015. *Vocabulary Increase and Collocation Learning: A Corpus-Based Cross-Sectional Study of Chinese EFL Learners*. Ph. D. thesis. The Faculty of Arts, Design and Media. University of Birmingham.
- Nation, P. 1990. *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P. 2001. *Learning a vocabulary in another language*. Cambridge: Cambridge University Press.
- Navarro Gil, G., Martinez Caro, E. 2019. Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language & Literature* 12(1). 65-90.

- Nesselhauf, N. 2005. Collocations in a Learner Corpus. John Benjamins Publishing Company.
- Pan, F., Reppen, R., Biber, D. 2016. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes* 21, 60-71.
- Paquot, M. & Granger, S. 2012. Formulaic Language in Learner Corpora. *Review of Applied Linguistics* 32. 130–149.
- Sherwill-Navarro, P., Kennedy, J. C., Allen, M. 2014. Developing an evidence-based list of journals for nursing. *Journal of the Medical Library Association*. 105-109.
- Shin, D., Nation, P. 2008. Beyond single words: the most frequent collocations in spoken English. *ELT Journal* 62(4). 339-348.
- Sinclair, J. 2001. Preface to small corpus studies and ELT. In Ghadessy M., Henry, A. and Roseberry R. (Eds.), *Small corpus studies and ELT* 7-15. Amsterdam: John Benjamins.
- Wood, A. 2001. International scientific English: The language of research scientists around the world. In Flowerdew J., Peacock M. (Eds.), *Research perspectives on English for academic purposes* 71-83. Cambridge: Cambridge University Press.
- Wu, W-S. 2003. A quantitative study of university students' collocational competence in an ESP context. *ESP Malaysia*. 9. 95-119.
- Xia, G. 2017. A cross-disciplinary corpus-based study on English and Chinese native speakers' use of first-person pronouns in academic English writing. *Text&Talk* 38(1). 93-113.
- Xue, G., Nation, P. 1984. A University Word List. *Language Learning and Communication* 3, 2. 215-229.
- Yang, M.-N. 2015. A nursing academic word list. *English for Specific Purposes* 37, 27-38.

Frane Malenica (Department of English Studies, University of Zadar, Croatia)

***Picking up the scraps – creating a specialized
corpus using web-scraping tools***

The methods for creating corpora from websites have been in use for almost two decades (Baroni & Ueyama 2006, Baroni et al. 2009) and numerous tools for extracting textual data and

metadata from websites have been developed since either as standalone programs, browser extensions, or as packages and libraries in programming languages such as Python and R (cf. Bradley and James 2019, Diouf et al. 2019, Kumar and Roy 2023). The widespread availability of these tools has allowed scholars to create custom corpora on a wide array of very specific topics, such as song lyrics (Kreyer & Mukherjee 2009, Werner 2012, Motschenbacher 2016), poetry (Toivanen et al. 2012), comics (Dunst et al. 2017, Unser-Schutz 2011) and video games (Heritage 2020).

Video games and video game reviews in particular have become a valuable source of linguistic information and the focus of linguistic inquiry in recent decades as research has revealed a potentially interesting correlation between video games and L2 acquisition (Sylvén & Sundqvist 2012, Chen and Yang 2013, Santos 2017, Kasemap 2017, Zhonggen 2018, Vásquez & Ovalle 2019). Similarly, analysis of video game reviews as sources of data has revealed a positive correlation between user-generated reviews and playing time (Guzsvinecz 2022), as well as different aspects of video games and emotions experienced by the players (Arik 2022, Britto & Pacifico 2020, Anees et al. 2020, Guzsvinecz 2022). Previous research in this domain conducted by Cho et al. (2020) has also shown that the NLP methods can be used to extract and identify the main themes of video games.

In this paper, I will present the results of research conducted on a corpus of video game reviews collected from the GameSpot website (www.gamespot.com/) through use of the rvest package (Wickham 2021) for webscraping in R and analysed using the a combination of traditional corpus linguistic (CL) methods, such as frequency lists, n-grams, and keywords and the Natural Language Processing (NLP) methods, such as sentiment analysis and word embeddings, available in the quanteda package (Benoit et al. 2018).

The main aims of this paper are to: i) identify words and phrases typical for different genre of video game reviews; ii) test the applicability of web-scraping and NLP methods for linguistic research; and iii) create a word embedding model from a small corpus and test its usefulness.

References

- Anees, A. F., Shaikh, A., Shaikh, A., & Shaikh, S. (n.d.). *Survey Paper on Sentiment Analysis: Techniques and Challenges*. 5.
- Arik, K. (2022). Social Media Content Review of MMORPG Games: Reddit Comment Scraping and Sentiment Analysis. *Journal of Emerging Computer Technologies* 2(1) 13–21.

- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3), 209–226. <https://doi.org/10.1007/s10579-009-9081-4>
- Baroni, M., & Ueyama, M. (n.d.). Building general- and special-purpose corpora by Web crawling. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software* 3(30). 774. <https://doi.org/10.21105/joss.00774>
- Bradley, A., & James, R. J. E. (2019). Web Scraping Using R. *Advances in Methods and Practices in Psychological Science* 2(3). 264–270. <https://doi.org/10.1177/2515245919859535>
- Britto, L. F. S., & Pacifico, L. D. S. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *SBC Proceedings of SBGames*. <https://dx.plos.org/10.1371/journal.pone.0232525>
- Chen, H.-J. H., & Yang, T.-Y. C. (2013). The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive Learning Environments* 21(2). 129–141. <https://doi.org/10.1080/10494820.2012.705851>
- Cho, H., Bossaller, J. S., Adkins, D., & Lee, J. H. (2020). Human versus machine: Analyzing video game user reviews for plot and narrative. *Proceedings of the Association for Information Science and Technology* 57(1). <https://doi.org/10.1002/ptra2.235>
- Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. *IEEE International Conference on Big Data (Big Data)*. 6040–6042. <https://doi.org/10.1109/BigData47090.2019.9005594>
- Dunst, A., Hartel, R., & Laubrock, J. (2017). The Graphic Narrative Corpus (GNC): Design, Annotation, and Analysis for the Digital Humanities. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 15–20. <https://doi.org/10.1109/ICDAR.2017.286>
- Guzsvinecz, T. (2022). The correlation between positive reviews, playtime, design and game mechanics in souls-like role-playing video games. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-12308-1>
- Heritage, F. (2020). Applying corpus linguistics to videogame data: Exploring the representation of gender in videogames at a lexical level. *Game Studies* 20(3).

- Kasemap, K. (2017). The Fundamentals of Game-Based Learning. In T. Kidd & L. R. Morris, Jr. (Eds.), *Handbook of Research on Instructional Systems and Educational Technology*. 174–185. IGI Global. <https://doi.org/10.4018/978-1-5225-2399-4>
- Kreyer, R., & Mukherjee, J. (2007). The Style of Pop Song Lyrics: A Corpus-linguistic Pilot Study. *Anglia - Zeitschrift Für Englische Philologie*. 125(1). <https://doi.org/10.1515/ANGL.2007.31>
- Kumar, S., & Roy, U. B. (2023). A technique of data collection. In *Statistical Modeling in Machine Learning*. 23–36. Elsevier. <https://doi.org/10.1016/B978-0-323-91776-6.00011-7>
- Motschenbacher, H. (2016). A corpus linguistic study of the situatedness of English pop song lyrics. *Corpora* 11(1). 1–28. <https://doi.org/10.3366/cor.2016.0083>
- Santos, A. (2017). Instructional Strategies for Game-Based Learning. In T. Kidd & L. R. Morris, Jr. (Eds.), *Handbook of Research on Instructional Systems and Educational Technology*. 164–173. IGI Global. <https://doi.org/10.4018/978-1-5225-2399-4>
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302–321. <https://doi.org/10.1017/S095834401200016X>
- Toivanen, J. M., Toivonen, H., Valitutti, A., & Gross, O. Corpus-Based Generation of Content and Form in Poetry.
- Unser-Schutz, G. (2011). Developing a text-based corpus of the language of Japanese comics (manga). In J. Newman, H. Baayen, & S. Rice (Eds.), *Corpus-based Studies in Language Use, Language Learning, and Language Documentation*. 213–238. BRILL. <https://doi.org/10.1163/9789401206884>
- Vásquez, G. C., & Ovalle, J. C. (2019). The Influence of Video Games on Vocabulary Acquisition in a Group of Students from the BA in English Teaching. *Gist Education and Learning Research Journal*, 19, 172–192.
- Werner, V. (2012). Love is all around: A corpus-based study of pop lyrics. *Corpora* 7(1), 19–50. <https://doi.org/10.3366/cor.2012.0016>
- Wickham, H. (2021). *Harvest: Easily Harvest (Scrape) Web Pages*. R Package Version 1.0.2.

Goranka Antunović (Faculty of Humanities and Social Sciences, University of Zagreb, Croatia)

Jelena Čović

Patrik Posedi

Sitting and pondering the meaning and potential Croatian renderings of Swedish pseudocoordinated constructions with posture verbs

Pseudocoordination of verbs in Swedish is a linguistic phenomenon that is commonly covered in grammars (e.g. SAG 1999, Josefsson 2009) but is nevertheless felt to require further research (Lødrup 2019). Traditionally, pseudocoordination has been described as a type of verb coordination in which the first verb (V1), typically belonging to one of a small number of verbal sets such as posture verbs, does not carry much lexical meaning but rather contributes aspectual (progressive) meaning to the whole phrase (Thorell 1973, Josefsson 2009). The construction's syntactic features set it apart from standard coordination of two verbs (Hultman 2003). Researchers have been revisiting the topic, discussing the construction's syntactic behaviour, questioning the claim of desemantisation and offering new interpretations of its semantics (Lødrup 2019). The complex character of this fairly frequent construction poses a challenge for Swedish language learners, who first need to realize that the meaning of V1 is not what they expect it to be and later on have to cope with not being certain of its exact meaning. Translators experience the same challenge when trying to render the meaning of the verbal construction in the target language (TL), struggling not only to interpret it but also to find an adequate way to render it with TL resources. Consequently, pseudocoordination becomes a highly relevant topic for both Swedish language teachers and translator trainers, who need to draw the students' attention to the construction but lack a solid basis for clear guidelines.

Exposure to authentic sentences containing pseudocoordination might prove beneficial for language learners in an attempt to develop an understanding of the construction similar to that of native speakers'. Likewise, an awareness of translation solutions commonly employed by professional translators may relieve translation students' insecurity regarding the construction, its meaning as well as potential renderings. In a university setting it is often one and the same group of students that may find both kinds of information valuable and electronic corpora, as growing evidence from research as well as teaching practice indicates, may be a valuable resource when attempting to provide it (Vintar 2007, Laviosa 2008, Yao 2019).

Since the primary interest of this study relates to translator education, it aims to establish how professional translators deal with verbal pseudocoordination when translating from Swedish into Croatian. It focuses on pseudocoordinated constructions including the posture verbs *ligga* “lie”, *stå* “stand”, *sitta* “sit” and the verb of motion *gå* “go”. In order to gain insight into translators’ interpretation of such constructions and their preferred translation solutions, a parallel Swedish–Croatian corpus has been constructed. It consists of a dozen novels in Swedish and their translations into Croatian done by ten translators. When selecting the texts an effort has been made to avoid overrepresentation of a single translator and his/her personal preferences – a rather challenging requirement in the Swedish-Croatian translation context. The language pair has also made it necessary that a “DIY corpus” (Mikhailov and Cooper 2016) be compiled since no suitable corpora are readily available. Due to the evasive character of the construction that has prompted this study, the analysis of translation solutions identified in the corpus is not expected to point to a specific norm guiding the translation of pseudocoordination with the four verbs. More likely, it shall provide an array of adequate translation solutions that may serve as an inspiration in future translation tasks, while potentially also shedding light on the implied meaning of the construction. It is also deemed probable that in Croatian translations verbs of posture will occur more often than the aspectual interpretation of the construction would require, for which interference is a likely but by no means only potential explanation. The list of translation solution types produced by the study, coupled with authentic examples, will be used in class as the basis for a directed discussion regarding the identified solution types and their adequacy and acceptability in their respective contexts. Potential benefits and limitations of the exercise will be discussed in the presentation.

References

- Hultman, T. 2003. *Svenska Akademiens språklära*. Stockholm: Svenska Akademien.
- Josefsson, G. 2009. *Svensk universitetsgrammatik för nybörjare*. Lund: Studentlitteratur.
- Laviosa, Sara. 2008. “Discovery and Justification Procedures in the Corpus-Based Translation Classroom”, *Translation Challenges: From Training to Profession*. Hammamet.
- Lødrup, H. 2019. “Pseudocoordination with posture verbs in Mainland Scandinavian: A grammaticalized progressive construction?”. *Nordic Journal of Linguistics* 42: 87-110.
- Mikhailov, M. & Cooper, R. 2016. *Corpus Linguistics for Translation and Contrastive Studies*. London and New York: Routledge.
- Teleman, U. & Hellberg, S. & Andersson, E. 1999. *Svenska Akademiens Grammatik*. Norstedts.

Thorell, O. 1973. *Svensk grammatik*. Stockholm: Esselte Studium

Vintar, Š. 2007. "Corpora in Translator Training and Practice: A Slovene Perspective". in: Anderman, G. & Rogers, M. *Incorporating Corpora. The Linguist and the Translator. Multilingual Matters*.

Yao, G. 2019. "Vocabulary learning through data-driven learning in the context of Spanish as a foreign language". *Research in Corpus Linguistics*. 7: 18-46.

SESSION II: Data-driven learning (DDL) in foreign language education

Yusuf Öztürk (Mus Alparslan University, Turkey)

Samed Yasin Öztürk (Mus Alparslan University, Turkey)

(er) .. when we at school: A Corpus-Based Analysis of Errors and Speech Fluency

Accuracy and fluency are two prominent aspects of language use for L2 learners. In the context of speaking, accuracy is producing error-free speech, while fluency is usually referred to as utterance fluency measured by temporal aspects of speech such as speech rate and pauses (Segalowitz, 2010). On the one hand, examining errors in L2 production can reveal how dynamic learner language can be (Selinker, 1972) and benefit researchers, curriculum designers and practitioners in many ways including new insights on learner interlanguage and more focused language teaching materials. On the other hand, considering that fluency and accuracy "go hand in hand" (VanPatten, 2015; p.15) in language performance, investigating various aspects of speech fluency in learner language as well as how fluency and accuracy develop over time can unfold the relationship between these two aspects of language acquisition. In this respect, the present study sets out to examine Turkish EFL learners' grammatical and lexical errors and speech fluency in a developmental corpus of learner spoken English. The corpus contains informal interviews with 40 Turkish university students conducted over a six-month period (i.e. at the beginning and end of the semester). At the time of the first interview, the students had just started an EFL prep program and did not have any experience in

conversational skills. After the interviews were transcribed, the corpus data contained nearly 60,000 words. The interviews have been examined for utterance fluency through speech and articulation rates, mean pause duration and mean length of runs. Following the fluency analysis that has just been completed, the students' errors are being tagged in the corpus by following the error-tagging manual proposed by Genç-Yöntem and Eveyik-Aydın (2022) and that includes the main categories of syntactic, morphological, morpho-syntactic and lexical errors. Descriptive statistics regarding the types and frequencies of errors in the corpus will be retrieved by means of WordSmith Tools 6 (Scott, 2011). Correlation analysis and inferential statistics will then be performed to see what aspects of speech fluency can be related to grammatical and lexical accuracy, and the nature of this relationship in the same learner group over a six-month period. The findings can be expected to provide important insights to language educators and applied linguists regarding the instructional emphasis on accuracy and fluency. The results will be discussed with reference to the contemporary approaches to foreign language teaching.

References

Genç-Yöntem, E., & Eveyik-Aydın, E. (2022). The compilation of a developmental spoken English corpus of Turkish EFL learners. *Research in Corpus Linguistics*, 10(1), 45-62.

Scott, M. (2011). *WordSmith tools version 6*. Liverpool: Lexical Analysis Software.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232.

Van Patten, B. (2015). *Communication and skill*. Routledge.

Samed Yasin Öztürk (Mus Alparslan University, Turkey)

Yusuf Öztürk (Mus Alparslan University, Turkey)

*In the Aftermath of an Unprecedented Catastrophe: Examining Changing
Themes in a Corpus of News Articles*

Media can influence what people perceive as the most important issues. Media outlets have the power to prioritize certain stories and topics over others, which can shape public opinion on what is considered the most pressing issue of the day (McCombs, 2014). Moreover, media coverage of catastrophic events can shape how people perceive the risks associated with these events, and can also influence their subsequent behavior in response (Slovic, 2010). In this regard, the present study aims to investigate how the Turkish media's approach and focus on the recent earthquake that occurred in the eastern part of Turkey on February 6, 2023, has changed over time.

By creating a corpus of news articles from various Turkish media outlets, this study will conduct a qualitative analysis of key words, collocations, and three- to five-word lexical bundles to identify the shifts in the media's coverage of the earthquake. The earthquake, which had a magnitude of 7.8 on the Richter scale, caused significant damage to the region and resulted in the loss of many lives. The initial response from the Turkish media was to provide factual reporting of the event, with a focus on the number of casualties and the extent of the damage.

Using a corpus analysis, the study will examine the news articles published by three biggest Turkish media outlets within the four weeks following the earthquake. The corpus will include articles from both print and online sources, allowing for a comprehensive analysis of the media's response towards earthquake and amount of its coverage over time.

Key words and collocations will be identified using the software package WordSmith Tools 6 (Scott, 2011), which will help to identify the most frequently used terms in the corpus. Three to five-word lexical bundles, which are frequently used multi-word expressions, will be automatically identified and analyzed.

The KeyWord function of WordSmith will be employed to make statistical comparisons of the theme shift on a day and week basis. This approach will allow for a nuanced understanding of the media's changing narrative around the earthquake. The study provides important insights into the role of the media in shaping public perception of natural disasters. By identifying the shifts in the media's narrative and the amount of coverage of the earthquake over time, this study highlights the power of the media to influence public opinion and shape public policy. The findings also have implications for disaster response and preparedness measures, as they highlight the need for effective communication strategies and the importance of addressing the emotional and psychological impacts of natural disasters.

It is thought that the study will offer a comprehensive analysis of the Turkish media's coverage of the February 6, 2023, earthquake in the eastern part of Turkey. Through the creation of a corpus and the qualitative analysis of key words, collocations, and lexical bundles, the present study identifies important shifts in the media's narrative and how much the catastrophic earthquake is covered over time. The findings have implications for disaster response and preparedness measures, as well as for the role of the media in shaping public opinion and policy.

References

McCombs, M. (2014). *Setting the agenda: The mass media and public opinion*. John Wiley & Sons.

Scott, M. (2011). *WordSmith tools version 6*. Liverpool: Lexical Analysis Software.

Slovic, P. (2010). *The feeling of risk: New perspectives on risk perception*. Earthscan.

Emir Feridun Çalışkan (Mus Alparslan University, Turkey)

Yusuf Öztürk (Mus Alparslan University, Turkey)

Employing a Corpus-based Approach in First Language Teaching:

Vocabulary Profiles of Elementary School Turkish Language Coursebooks

In the last two decades, corpus-based approaches are widely used in developing materials and coursebooks for teaching foreign languages, particularly the teaching of English as a foreign language (Burton, 2012). This is mostly because corpus data can simply provide researchers and educators frequency information, and frequent words are usually the ones that learners are more likely to encounter and that can be processed better (Schmitt, 2010). Nevertheless, it is argued that discrepancies may exist how a language is described and modeled in a coursebook and how native speakers actually use that language in discourse (Cheng & Warren, 2007; Cullen & Kuo, 2007). Furthermore, particularly in the context of first language education, the intuitions and experience of language experts, rather than corpus-based data, seem to be more influential on the materials and coursebooks. In this respect, this study aims to examine Turkish language coursebooks used in elementary schools in Turkey in terms of vocabulary profiles. The research corpus was compiled by collecting the reading texts included in the first, second, third and fourth-grade elementary school coursebooks. In total, the corpus contained 176 reading texts

and 36,782 words. Word lists were generated by using WordSmith Tools 6 (Scott, 2011), and the types and tokens of words were retrieved for each grade. The word lists and the relative frequencies of words were then compared to the list of most frequent 5,000 words in Turkish, which was compiled by Aksan et al. (2017) based on the 50-million word Turkish National Corpus. The results of this analysis would reveal to what extent the most frequent 5,000 words in Turkish (i.e. split into lists of most frequent 1k, 2k, 3k, 4k and 5k words) are used in second to fourth-grade elementary school Turkish language coursebooks. Data analysis is still in progress, and the findings will be discussed with reference to the literature on vocabulary research and language acquisition. Suggestions will be offered to curriculum designers and coursebook writers in first language education regarding the use of corpus-based approaches to materials and coursebook design.

References

- Aksan, Y., Aksan, M., Mersinli, Ü. & Demirhan, U. U. (2017). *A frequency dictionary of Turkish: Core vocabulary for learners*. Routledge.
- Burton, G. (2012). Corpora and coursebooks: destined to be strangers forever?. *Corpora* 7(1), 91- 108.
- Cheng, W. and Warren, M. (2007) ‘Checking Understandings: Comparing Textbooks and a Corpus of Spoken English in Hong Kong’. *Language Awareness* 16(3): 190–207.
- Cullen, R. and Kuo, I.-C. (2007) ‘Spoken Grammar and ELT Course Materials: A Missing Link?’ *TESOL Quarterly* 41(2): 361–86.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Scott, M. (2011). *WordSmith tools version 6*. Liverpool: Lexical Analysis Software.

Jana Kegalj (Faculty of Maritime Studies, University of Rijeka, Croatia)

Mirjana Borucinsky (Faculty of Maritime Studies, University of Rijeka, Croatia)

Data-Driven Learning for Writing Skills Development

The rapid progress and increasing availability of language technologies has greatly influenced various areas of language production and language teaching. Language technologies comprise various software and tools that include natural language processing, lexical computing and

speech technologies. They can be broadly divided into language resources, language tools and commercial products. The most prominent among the language resources are corpora, which enable the processing of large amounts of linguistic data.

This study shows how corpora, corpus tools, i.e. text collections and software packages, and corpus methods, i.e. techniques for analysing corpus data (Römer 2008), can be used to develop students' writing skills, while enabling them to improve their digital competencies, which is in line with current trends in education. The hypothesis is that corpus-derived materials for teaching writing skills will engage students during their learning process and enable them to conduct their own research in the future.

Johns (1991) was one of the first to suggest that corpora can be used in the classroom as an effective way to engage students and make them active participants in the learning process, i.e. to allow them to discover language and language patterns. This was the basis for the Data-Driven Learning (DDL) method, which uses large amounts of data as input for students to observe, analyse, interpret, explore, compare, hypothesize and draw conclusions from. Since then, corpora have been introduced in teacher education, translator training, teaching literature and assessment (Flowerdew 2012). The use of corpora in teaching has been attested by various studies (cf. Kennedy and Miceli 2001, 2010; Cheng et al. 2003; Chambers and O'Sullivan 2004; Gaskell and Cobb 2004; Lee and Swales 2006; Boulton, 2009; Chujo et al 2012; Vyatkina 2016, 2020; Boulton and Cobb 2017). This type of teaching stimulates student motivation, develops critical thinking and lifelong learning skills. The role of the teacher changes significantly as the teacher becomes more of a guide, a mentor, an advisor. The approach of using corpora as a source of information is problem-based as the students are confronted with a problem that they need to investigate. It is also a form of self directed learning as students can take initiative, identify their problems and sources, and then apply the strategies they have acquired to solve the problem. Corpora provide large amounts of authentic data in a particular form of output. They can be used directly by teachers and learners in class or indirectly in the production of textbooks and teaching materials (Römer 2008). In class, the students can work indirectly with corpus results in the form of filtered and printed concordances, or they can work directly with corpora, which has become possible owing to the availability of software programs (e.g. AntConc, Skell, etc).

The most commonly mentioned corpus tools are concordances which are lines of immediate context around (as defined by <http://ihjj.hr/mreznik/page/pojmovnik/6/>. a searched keyword). Johns (1991) suggested them as a way for students to infer meanings and study functions of

words and thus become aware of some typical combinations. Other data include various frequency lists and lists of collocations that can provide insight into word meaning, typical grammatical patterns, stylistic preferences, conceptual framework and semantic network. In this study, the corpus data was used as input for guided writing tasks and as the basis for productive tasks. The study also took into account that students need to be trained in corpus research before they can engage in autonomous corpus-informed learning. The study also pointed out some possible drawbacks, pitfalls and challenges of such an approach, mainly related to the compilation of the corpus, the selection of appropriate material, the time required for mastering the skills and for querying the corpus, and the training of learners in the use of corpora.

References

- Boulton, A. (2009). Corpora for all? Learning styles and data-driven learning. In M. Mahlberg, V. González-Díaz & C. Smith (eds), *Proceedings of 5th Corpus Linguistics Conference*.
- Boulton, A., Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2): 348-393. DOI 10.1111/lang.12224.
- Chambers, A., O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL* 16.1, 158–172.
- Cheng, W., Warren, M., Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System* 31. 173–186.
- Chujo, K., Laurence, A., Oghigian, K., Uchibori, A. (2012) Paper-based, computer-based, and combined data-driven learning using a web-based concordancer. *Language Education in Asia*, 3 (2). 132–145. https://doi.org/10.5746/LEiA/12/V3/I2/A02/Chujo_An...
- Flowerdew, L. (2012). Corpora in the Classroom: An Applied Linguistic Perspective. In: *Corpus Applications in Applied Linguistics*, Ken Hyland, Chau Meng Huat and Michael Handford (eds.). Continuum International Publishing Group: London and New York. 208–225.
- Gaskell, D., Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System* 32/3: 301–19.
- Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal* 63(4).
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns and P. King (eds.), *Classroom concordancing: ELR Journal*, 4. Birmingham: Centre for English Language Studies, University of Birmingham. 1–16.

Kennedy, C., Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5.3, 77–90.

Kennedy, C., Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology* 14.1, 28–44.

Lee, D., Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes* 25.1, 56–75.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling and M. Kytö (eds.), *Corpus linguistics: An international handbook*. Berlin: Mouton de Gruyter. 112–130.

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System* 29/1: 91–102.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, 20(3), 159–179.

Vyatkina, N. (2020a). Corpus-informed pedagogy in a language course: Design, implementation, and evaluation. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching*. Hershey, PA: IGI Global. 306–335.

Pedro Humánez-Berral (University of Cantabria, Santander, Spain)

Unveiling Pronunciation and Motivation: An Examination using an Oral Learner Corpus

Unveiling Pronunciation and Motivation: An Examination using an Oral Learner Corpus
Content and Language Integrated Learning (CLIL) is a pedagogical approach that combines the learning of a foreign language, typically English, with academic subjects in schools (Coyle et al., 2010). Extensive research has demonstrated the positive impact of CLIL on various linguistic aspects, including receptive vocabulary, oral comprehension, and fluency (Dalton-Puffer, 2011; Pérez-Cañado, 2018). Moreover, it has been shown to influence non-linguistic factors such as motivation and anxiety, as increased exposure to the language may negatively

affect anxiety and positively affect motivation levels (Lasagabaster & Doiz, 2017). However, limited attention has been given to the influence of CLIL on pronunciation and affective factors. Therefore, this study aims to investigate this matter. Additionally, CLIL has been observed to reduce gender differences in foreign language learning, as girls traditionally outperformed boys in traditional approaches, whereas in CLIL, this difference is less pronounced (Heras & Lasagabaster, 2014). However, the extent to which this gender gap-narrowing effect applies to pronunciation and motivation remains to be established. The main objective of this study is to analyse the effects of CLIL and non-CLIL approaches on English pronunciation and motivation levels, employing an oral learner corpus. The study involves third and fourth-grade students from both bilingual and non-bilingual primary schools in the Spanish region of Cantabria. To collect linguistic data for the learner corpus, participants were asked to individually narrate and describe a wordless story involving a dog. The recorded narratives are analysed based on three dimensions: degree of foreign accent (DFA), foreign accent intelligibility (FAIN), and foreign accent irritation (FAIR). As the assessment of these dimensions depend on judgement, two independent listeners are in charge of assessing each recording under the same circumstances (vid. Gallardo-del-Puerto et al., 2009) Additionally, a contextual questionnaire is administered to gather background information, which contributes to the learners' linguistic profiles. This questionnaire also explores other variables such as gender, age, years of English study, and engagement in informal language learning activities. Finally, a motivation questionnaire is employed to assess the motivation levels of each student in relation to various constructs, including ideal L2 self, ought-to L2 self, and L2 learning experience. Overall, this comprehensive study aims to illuminate the relationship between CLIL, pronunciation proficiency, and motivation levels, shedding light on the less explored aspects of this dual-focus approach. In addition to investigating the current effect of CLIL and gender variables on motivation towards English pronunciation, this study seeks to gather linguistic data from the same cohort of students over a period of two years. By employing a pseudolongitudinal design, the study aims to examine the evolution of students' pronunciation skills within both CLIL and non-CLIL contexts. This expanded learner corpus will provide more results and conclusions on the success of CLIL as an approach.

References

Coyle, D., Hood, P., Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: From Practice to Principles? *Annual Review of Applied Linguistics* 31, 182-204. 10.1017/S0267190511000092

Gallardo del Puerto, F., Gómez Lacabex, E., García Lecumberri, M. L. (2009). Testing the effectiveness of CLIL in foreign language contexts: The assessment of English pronunciation. In Y. Ruiz de Zarobe, R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe*. 63–80. Bristol: Multilingual Matters

Heras, A., & Lasagabaster, D. (2014). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research* 19(1), 70–88. 10.1177/1362168814541736

Lasagabaster, D., Doiz, A. (2017). A Longitudinal Study on the Impact of CLIL on Affective Factors. *Applied Linguistics*, amv059. 10.1093/applin/amv059

Pérez Cañado, María Luisa. (2018). CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes. *Porta Linguarum* 29. 51-70. 10.30827/Digibug.54022

Zorica Đurović (Faculty of Maritime Studies Kotor, Montenegro)

Some corpus linguistics methods for vocabulary research and teaching

The paper presents an overview of the research and methods applied within the author's doctoral dissertation. Its aim was to explore the opportunities to enhance a specific branch of ESP – English for Marine Engineering Purposes, by applying certain contemporary corpus linguistics methods. In attempting to do this, we were led by the latest findings made by recognized authors in the field of corpus linguistics and word list creation methodologies, but also by some of the specific and official requirements of the International Maritime Organization (IMO) in terms of the English language skills of seafarers, as well as the practical needs of the learners. To begin with, an adequate Marine Engineering professional corpus was selected and formatted, reaching a final total of 1,769,821 words from ship's instruction books and manuals, as the most technical and necessary genre used by this professional discourse community. We examined the lexical profile of the corpus by comparing the coverage with the available General English and engineering word lists. For this purpose, we used the Laurence Anthony's software AntWordProfiler (version 1.4.0w).

In addition to the unambiguous lexical demand of the corpus in terms of both the abundance and specificity of its technical vocabulary, the obtained results clearly pointed to the need to build a specialized marine engineering word list, with the intention of achieving a desired (known) vocabulary coverage threshold of 95%. Following on from this, the other part of research was dedicated to creating a list of the most frequently occurring vocabulary from ship instruction books and manuals using the same software. The list was being built as an upgrade to the 3,000 most frequent English words (from the British National Corpus and Corpus of Contemporary American English – BNC/COCA), together with the lists of the most frequent abbreviations, proper names and marginal words. Finally, we came up with the list of 337 word families with 73 transparent compounds reaching the desired threshold for adequate reading comprehension.

In addition to building and evaluating the obtained frequency vocabulary list, we used an additional AntConc programme (Anthony, 2012) which provided us with additional key vocabulary from our professional corpus. This is done by comparing a corpus to a referent general English corpus, which, in our case, was The Freiburg-LOB Corpus of British English. In this way, we obtained an additional list of 124 key word families with 43 key transparent compounds. By combining the two methods, we eventually produced a glossary of marine technical manuals comprising of 577 headwords in total. Although the list obtained in this way still contains fewer than 1,000 words, it is both much more comprehensive and more specific, given that it also includes the most frequent key words of the genre, i.e. those appearing much more frequently than in the general English corpora. Bearing in mind that our final aim was the enhancement of marine engineering vocabulary teaching, we wanted to test the application of the corpus methods and the produced list in the classroom. Assuming that the technical vocabulary would be acquired more efficiently through a greater number of individual activities on the part of the students, we conducted a pilot-research with Marine Engineering students of the Faculty of Maritime Studies Kotor within the English Language II subject. As instructed and consulted by the teacher, the experimental group worked on home assignments which anticipated working with the word list and the selected instruction books, i.e. through the Data-driven Learning (DDL) method. The results achieved on the Vocabulary Size Tests point to a slight advantage among the experimental group. Taking into account the problematic generalization of the results due to the small group size of the participants, as well as the limitations in the comparability of the achieved results, the statistical analysis and presentation of the results was achieved through a semi-structured interview conducted with the experimental group representatives. In this way, we were able to get feedback from the same

students, recording their impressions and perception of the applied methods, something that is of inestimable value for future activities and improvements in ESP teaching. Apart from the specific scientific contribution provided by the production of the specialized word list and its specific pedagogical implications, the positive reactions and suggestions from our students have been especially encouraging and energizing in terms of further research related to innovative teaching methods. What is of special importance here is that the presented and applied methodology can be used for any ESP area, and thus we hope to provide a modest contribution to all our hard working and enthusiastic colleagues facing the challenges of contemporary teaching methods and the ever more demanding needs and requirements of language learners.

References

- Anthony, L. 2012. *AntConc* (Version 3.3.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>.
- Bernardini, S. 2002. Corpora in the classroom. In J. McH. Sinclair (ed.) *How to use corpora in language teaching*, 15-36. Amsterdam: John Benjamins.
- Tribble, C. 2012. Teaching and language corpora: Quo Vadis? *10th Teaching and Language Corpora Conference (TALC)*. Warsaw, 11th-14th July 2012.

Eriola Qafzezi (Fan S. Noli University Korça, Albania)

Inside Out: Investigating the relationship between culture and language as reflected on uses of expressions with parts of body in Albanian language

The aim of this paper is to observe the reciprocal relation between language and culture, based on a corpus driven study. Research that is based on corpora, either in foreign language education, or in specialized translation is an area of study yet to be explored in Albanian language. This is due to the fact that corpora of Albanian language have only recently been developed and, thus, corpus based linguistic research and/or corpus driven research present challenging and innovative pursuits. Apart from the usefulness of corpora search and analysis in the field of grammar, corpora can also help teachers and researchers with vocabulary.

Linguists have always been interested in discovering and analyzing how words combine in collocations and how such collocations have unitary meanings and specialized functions. It is specifically due to collocation and the neoFirthian approach to word meaning that we conclude that meaning does not reside with the word in isolation, but with the other words that are combined with it. Furthermore, collocations result from repeated combinations used by the speakers of any language. In support of this thesis, we have focused on parts of body and searched for expressions that use parts of body in Albanian language.

Our study adopts a qualitative approach, based on a corpus driven analysis. We concentrate on expressions that use parts of body in Albanian language in order to observe how language is reflected in culture and vice versa. We have extracted examples from Albanian National Corpora (ANC). The corpus of Albanian language did not exist until the end of 2011, when the Corpus was developed as a result of efforts of the creative community of linguists from Saint Petersburg (Institute for Linguistic Studies of the Russian Academy of Sciences) and Moscow (School of Linguistics at HSE). The current version of ANC uses the morphological analyzer and the tsakorpus platform and provides reference data for both professional linguists and anyone interested in Albanian language and history, Albanian lexicon and grammar, as well as language changes which happened in Albanian in the previous centuries. Our aim is to show that ANC can also be used for discourse studies, cultural studies, sociolinguistic studies, among many other purposes.

In the first part of the paper we bring some information about ANC and methodology of study. In the second part of the paper we bring illustrative examples of the most typical expressions that we have found that contain parts of the body. We bring such examples within a tentative taxonomy that will be supported with examples, based on all expressions we have observed in the corpus of our study. Conclusions of our research support the view that culture is manifested in linguistic choices and in language use. The paper will be concluded with some suggestions of how to extend the study in the future, possibly with quantitative analyses about examples found and commented upon.

References

- Cushing, F. H. (1892). Manual Concepts: A Study of the Influence of Hand-Usage on Culture-Growth. *American Anthropologist* 5/4. 289-318.
- Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. In *Literary and Linguistic Computing* 25/ 4. 447-464.

Granger, S., Meunier, F. (eds) (2008). *Phraseology: An Interdisciplinary Perspective*. John Benjamins Publishing Co., Amsterdam, the Netherlands.

Goddard, C. (2008). Contrastive semantics and cultural psychology: English heart vs. Malay hati. In Sharifian, F. et al. (eds) *Culture, Body, and Language. Conceptualizations of Internal Body Organs across Cultures and Languages*. Mouton de Gruyter, Berlin, Germany.

Kövecses, Z. (2004). *Metaphor and Emotion. Language, Culture, and Body in Human Feeling*. Cambridge University Press. Cambridge, United Kingdom.

Lopes, A. (2013). Exploring Culture-related Content in the COCA with Task-based Activities in the EFL Classroom. In L. Bradley & S. Thouwsny (Eds.), *20 Years of EUROCALL: Learning from the Past, Looking to the Future*. Proceedings of the 2013 EUROCALL Conference, Évora, Portugal. 169-174. Dublin/Voillans: Research-publishing.net.

Maria Morozova, Alexander Rusakov, Timofey Arkhangelskiy. Albanian National Corpus.

Morozova, M. & Rusakov, A. (2013). Korpusi elektronik i shqipes: përpunimi, përmbajtja dhe përdorimi. In *The XXXII International Seminar For Albanian Language, Literature And Culture*. Prishtinë, Fakulteti i Filologjisë.

Morozova, M. & Rusakov, A. (2015). Albanian National Corpus: Composition, Text Processing and Corpus-oriented Grammar Development. In *Albanische Forschungen. Sprache und Kultur der Albaner. Zeitliche und räumliche Dimensionen Akten der 5*. 270-306. Germany, Hubert & Co., Göttingen.

Morrow, P. R. (2009). Hand and Heart: A Study of the Uses and Phraseology Associated with Two Common Nouns. In *The Nagoya Gakuin Daigaku RonshU; Journal of Nagoya Gakuin University; Language and Culture*. 20/2. 11-20, <http://doi.org/10.15012/00000535>.

O’Keeffe, A., McCarthy, M., Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. New York, Cambridge University Press.

Internet Sources

http://albanian.web-corpora.net/albanian_corpus/search, accessed February to August, 2023