



**Doctoral School of Linguistics**  
**Pazmany Peter Catholic University**  
2 Bertalan Lajos utca, Budapest, H-1111 Hungary

## **Complex examination in Corpus Linguistics**

### **Topics**

- 1 Elaborate on the steps in compiling and annotating a corpus for linguistic research, with a focus on the challenges and considerations.
- 2 Discuss how corpora can be categorised, including issues such as sampling and representativeness.
- 3 Describe the descriptive characteristics in corpus searches, including issues of frequency, dispersion, co-occurrence, concordancing.
- 4 Discuss corpus-related research issues related to grammar, lexis, registers and genres.
- 5 Discuss how corpus-based approaches can enhance discourse analysis, providing examples of their application in studying specific discourse features.
- 6 Discuss how corpus-based approaches can enhance discourse analysis from a pragmatic perspective, providing examples of their application in studying specific discourse features.
- 7 Discuss how corpus-based approaches can help to understand the key differences between human- and machine-generated translation.
- 8 Discuss the most important issues in learner corpus research.
- 9 Elaborate on the use of corpora in language teaching.
- 10 Elaborate on the most important issues related to corpora and lexicography.

### **Compulsory readings**

- Paquot, M. & Gries, S. (eds.). (2020). *A Practical Handbook of Corpus Linguistics*. Springer, DOI: 10.1007/978-3-030-46216-1. **OR**  
O’Keeffe, A. & McCarthy, M.J. (eds.) (2022). *The Routledge Handbook of Corpus Linguistics*. Routledge.

### **Recommended readings**

- Bago, K. & Karlic, V. (2023). DirKorp: *A Croatian Corpus of Directive Speech Acts (v3.0)*  
DOI: <https://doi.org/10.4312/slo2.0.2023.1.189-217>.  
Biber, D. (2019). Text-linguistic approaches to register variation, *Register Studies*, Vol. 1, No. 1, p. 42–75, DOI: 10.1075/rs.18007.bib

- Brunni, S., Lehto, L. M., Jantunen, J. H. & Airaksinen, V. (2015). How to annotate morphologically rich learner language. Principles, problems and solutions, *BeLLS 2015* Vol. 6.
- Clancy, B. (2015). "Hurry up baby son all the boys is finished their breakfast": Examining the use of vocatives as pragmatic markers in Irish Traveller and settled family discourse, in C. P. Amador-Moreno, K. McCafferty & E. Vaughan (eds.), *Pragmatic Markers in Irish English*. John Benjamins, 229–247.
- Conrad, S. (2023). Register in corpus linguistics: the role and legacy of Douglas Biber, *Corpus Linguistics and Linguistic Theory*, Vol. 19, No. 1, p. 7–23.
- Davies, M. (2023). English-Corpora.org: Introduction, <https://www.youtube.com/watch?v=yJSyhYLLfnc>
- Donaj, G., Antloga, S. (2023). *ParaDiom – A Parallel Corpus of Idiomatic Texts*, DOI: [https://doi.org/10.1007/978-3-031-40498-6\\_7](https://doi.org/10.1007/978-3-031-40498-6_7)
- Ferrari, L. A. (2012) Building a corpus for comparative analysis of language attrition, Heliana Mello, Massimo Pettorino, Tommaso Raso (eds.), *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*. ISBN 978-88-6655-351-9
- Jaworska, S. & Ryan, K. (2018). Gender and the language of pain in chronic and terminal illness: A corpus- T based discourse analysis of patients' narratives, *Social Science & Medicine* 215, 107–114.
- Lange, C. & Leuckert S. (2019). *Corpus Linguistics for World Englishes*. Routledge.
- Larsson, T., Plonsky L. & Hancock, G. (2022). On learner characteristics and why we should model them as latent variables, *International Journal of Learner Corpus Research*, Volume 8, Issue 2, 237–260.
- Liu, M. & Kobayashi, I. (2022). Construction and Validation of a Japanese Honorific Corpus Based on Systemic Functional Linguistics, *Proceedings of the 1<sup>st</sup> Workshop on Dataset Creation for Lower-Resourced Languages*, 19–26.
- O'Keeffe, A. & Mark, G. (2017). The English Grammar Profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, Volume 22, Issue 4, 457–489.
- Podesva, J. & Sharma, D. (eds.) (2013). *Research Methods in Linguistics*. Cambridge University Press.
- Shawaqfeh, A. T. & Khasawneh, M. A. S. (2023). Incorporating Corpus Linguistics Tools in the Training and Professional Development of Lecturers in Translation Studies. *Studies in Media and Communication*, Vol. 11, No. 7; December 2023.
- Stoykova, V. (2014). Teaching Corpus Linguistics, *Procedia – Social and Behavioral Sciences*, 143, 437–441.
- Viana, V., & O'Boyle, A. (2021). *Corpus Linguistics for English for Academic Purposes*. Routledge.
- Wen, X., Xu, L., Ye, S., Sun, Z., Huang, P. & Qian, X. (2023). Personality differences between children and adults over the past two centuries: Evidence from corpus linguistics, *Journal of Research in Personality*, 102, 104336.