

Exploring the affordances of CEFR-based learner corpora in Data-driven learning Kris Heylen¹, Ilan
Luciana Forti (University for Foreigners of Perugia, Italy)

Abstract

Data-driven learning (DDL) practices are known for being predominantly based on L1 corpora. According to the review by Boulton & Vyatkina (2021), in fact, only 4% of the total number of DDL studies published over 30 years (n = 489) rely on learner corpora instead. Yet, claims on the potentially beneficial applications of learner corpora in DDL have been made in many learner corpus research (LCR) papers, deeming learner corpora as rich repertoires of linguistic input for the second language learner to explore.

This presentation will discuss not only the benefits of using learner corpora in DDL, but also present the specific pedagogical advantages of learner corpora based on CEFR-categorised texts. Special attention will be devoted to some of the pedagogical applications of the CELI corpus (Spina et al., 2022), a corpus of learner Italian texts produced under language certification exam conditions, evenly divided into proficiency levels B1, B2, C1 and C2.

References

- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning and Technology*, 25(3), 66–89.
- Spina, S., Fioravanti, I., Forti, L., Santucci, V., Scerra, A., & Zanda, F. (2022). Il Corpus CELI: Una nuova risorsa per studiare l'acquisizione dell'italiano L2. *Italiano LinguaDue*, 14(1), 116–138.