

In business with the Apprentice Multiple Business GEnRes (AMBER) corpus

Business texts represent a wide variety of genres, which tend to be identified on the basis of communicative purpose(s) and situation(s) (Bhatia 1993, Koester 2010). A broad distinction (Nelson 2000) can be made between genres used to do business and communicate to get work done within the framework of companies'/organisations' activities (e.g. business meetings, social media posts) and genres not issued by companies/organisations that are used to talk or write about business (e.g. news articles about the world of business, business studies lectures). An examination of the Learner Corpora around the World webpage (<https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>) reveals that learner corpora which specifically target business communication by L2 learners are very much few and far between and tend to include only one or two genres (e.g. Connor et al. 2002). In addition, existing corpora are rarely readily accessible to the research community or provide no or very limited information about the learners (e.g. Allan 2018).

To fill this gap in learner corpus research, this paper sets out to introduce the recently launched Apprentice Multiple Business GEnRes (AMBER) corpus and to further develop the AMBER network of partners. The paper aims (1) to explain the rationale behind this new multi-L1 apprentice computerised corpus collection project (e.g. providing a strong empirical basis to investigate genre awareness and knowledge among novice users of business genres) and (2) to outline the main corpus design criteria with a specific focus on the 'apprentices' targeted (i.e. novices at professional business communication genres, typically students; predominantly but not limited to L2 learners) and on the business genres included (e.g. press releases, LinkedIn summaries, cover letters, tourist brochures). Information about the core metadata recorded (following Paquot et al.'s 2024 Core Metadata Schema) is also provided (e.g. degree scheme, genre training, language(s) of instruction, L1(s), writing anxiety; detailed task instructions). The paper is rounded off by discussing the main takeaways from the AMBER pilot data collection.

References

- Allan, R. (2018). 'Recycling the data: Building and using a learner business English writing corpus'. In *Call Your Data. Proceedings of the XIXth CALL Conference, Antwerp*. Available from <https://miun.diva-portal.org/smash/get/diva2:1250868/FULLTEXT01.pdf>
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Connor, U., Precht, K. & T. Upton (2002). 'Business English in a learner corpus'. In Granger, S., Hung, J. & S. Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 175-194.
- Koester, A. (2010). *Workplace discourse*. London and New York: Continuum.
- Nelson, M. (2000). *A corpus-based study of the lexis of business English and business English teaching materials*. Unpublished PhD thesis. University of Manchester.
- Paquot, M., König, A., Stemle, E. W., & J.-C. Frey (2024). 'The Core Metadata Schema for Learner Corpora: Collaborative efforts to advance data discoverability, metadata quality and study comparability in L2 research.' *International Journal of Learner Corpus Research* 10(2), <https://www.jbe-platform.com/content/journals/10.1075/ijlcr.24010.paq>.

