

Introducing AMBER The Apprentice Multiple Business GEnRes corpus

Sylvie De Cock and Jennifer Thewissen

UCLouvain and Universiteit Antwerpen

sylvie.decock@uclouvain.be; Jennifer.Thewissen@uantwerpen.be

Why do we need the new AMBER corpus?

1. **Few existing learner business corpora** and generally **limited to one or two genres** (e.g. Indiana Business Letter Corpus - IBLT, Connor et al. 2002; Italian Learner Corpus of Business English - ItalBEC, Romagnuolo 2015)
2. **Limited accessibility, visibility and metadata** (e.g. local corpus used in Allan 2018; Learner Corpus of English for Business Communication, <http://langbank.engl.polyu.edu.hk/corpus-search.asp>; EF-Cambridge Open Language Database – EFCAMDAT, Xia et al. 2024)
3. **Becoming rather dated** (e.g. IBLT, Connor et al. 2002)

The AMBER corpus = a strong empirical basis to study

- **genre awareness and knowledge** (Bremner 2018) among learners / novice users of business genres
- the extent to which learners' / novice users' **productions vary according to business genre**
- learners' / novices' **writing process** when drafting professional genres, using Inputlog (Leijten & Van Waes 2013)
- learner / novice business communication **profiles** by extracting typical features of how learners / novices **create meaning** in various professional **contexts** (e.g. pragmatic competence including persuading, strategic vagueness, negotiating), functional adequacy, and more local linguistic features such as accuracy or complexity)

Introducing the new AMBER corpus project



A variety of business-related degree schemes, e.g. Business Administration, Business Communication, HR, International Business, Management, Marketing & Advertising, Tourism, ...

Main design criteria

- **Specialised and multi-genre:** professional business communication genres
- **Mode:** written (to start with) and oral genres
- **Multilingual:** English; Dutch, French, German, Spanish, ...
- **'Apprentice':**
 - novices at professional business communication genres, typically students
 - both foreign language learners (intermediate to advanced proficiency levels) and native speakers
- **Multi-L1:** a variety of mother-tongue backgrounds

- genres identified on the basis of communicative purpose(s) and situation(s) (Bhatia 1993, Koester 2010)
- genres used to do business and communicate to get work done within the framework of companies'/organisations' activities (Nelson 2000)

Genre examples:

- press/new releases
- cover letters
- LinkedIn summaries
- sales email messages
- sales newsletters
- social media posts
- mission statements
- minutes of meetings
- business proposals
- business reports
- tourist brochures
- business presentations
- business meetings

Rich metadata documentation

Based on the Core Metadata Schema for Learner Corpora (Paquot et al. 2024)

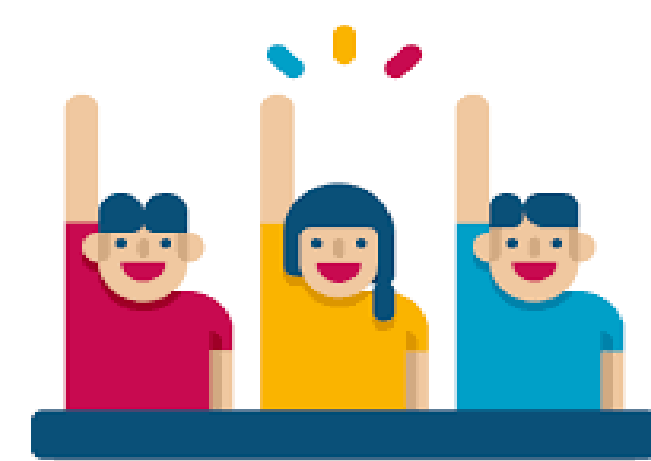
Learner / Novice, e.g.

- Degree scheme and level (BA, MA)
- Genre training / familiarity
- Native language(s)
- Target language (TL) education and use
- Language(s) of instruction
- Knowledge of other languages
- Self-assessed proficiency in the TL
- Frequency and exposure to TL
- TL motivation levels; TL writing anxiety
- ID: age, gender, country

Task and context, e.g.

- Detailed instructions (incl. intertextuality, task materials)
- Topic(s)
- Duration and planning
- Use of reference tools and AI
- Exam, in-class activity, home assignment, ...

Interested in joining this open-access corpus project?



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.
- Bremner, S. (2018). *Workplace Writing. Beyond the Text*. London and New York: Routledge.
- 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.
- Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze Writing Processes. *Written Communication* 30(3): 358-392.
- 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).
- Romagnuolo, A. (2015). Teaching English Professional Writing in an E-learning Environment: An Italian Case Study. *International Journal of Language and Linguistics* 3(6): 383-393.
- Xia, D., Sulzer, M. & Pae, H. (2024). Phrase-frames in business emails: a contrast between learners of business English and working professionals. *Text & Talk* 44(5): 693-714.