

## Formulaicity of English used in the pharmaceutical sector: A corpus-driven study of the use and functions of lexical bundles and phrase frames<sup>1</sup>

Lukasz Grabowski (Opole University, Poland)

This study, conducted from a register-perspective (Biber & Conrad 2009), aims to provide a corpus-driven description of the use and function of contiguous and non-contiguous multi-word units used across four pharmaceutical registers: patient information leaflets, summaries of product characteristics, clinical trial protocols and chapters from academic textbooks on pharmacology. This research extends an earlier preliminary study limited to two pharmaceutical text varieties only (Grabowski 2013).

The research material is a purpose-designed corpus of English pharmaceutical texts with approximately 2.2 million words. The study – largely based on the methodology proposed by Biber (2006, 2009), Goźdz-Roszkowski (2011) and Roemer (2009, 2010), with some modifications – was conducted in three stages, namely the functional analysis of 4-word lexical bundles (top-50 by frequency; short 4LBs) and phrase frames (top 20 by frequency; short 4PFs) based on these bundles. Finally, the most frequent structural patterns (after Liu 2012) based on the most frequent 4LBs and 4PFs were identified in order to further increase the pedagogical value of the results.

Due to large swaths of descriptive data obtained from the analyses, the results are only summarized in this presentation. All in all, the study revealed some patterns of intra-disciplinary register variation between different text varieties used within the pharmaceutical sector. More specifically, the use and functions of 4LBs and 4PFs were found to be text type-specific (with the clinical trial protocols being the most repetitive and patient information leaflets the most formulaic); considerable differences with respect to the choice of referential 4LBs and 4PFs show that register variation is not only function-related, but also content-related. These differences arise from the context of use and communicative functions of the text varieties under scrutiny. Finally, it was revealed that the problem of measuring pattern variability of 4PFs requires further research in the future.

### References

- Biber, D. 2009. A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.
- Biber, D. 2006. *University Language. A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins.
- Biber, D. & S. Conrad 2009. *Register, Genre and Style*. Cambridge: CUP.
- Goźdz-Roszkowski, S. 2011. *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Frankfurt am Main: Peter Lang Verlag.
- Grabowski, Ł. 2013. Register Variation Across English Pharmaceutical Texts: A Corpus-Driven Study of Keywords, Lexical Bundles and Phrase Frames in Patient Information Leaflets and Summaries of Product Characteristics. *Procedia – Social and Behavioral Sciences*, 95(C), 391-401.
- Liu, D. 2012. The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31, 25-35.
- Roemer, U. 2009. English in Academia: Does Nateness Matter?. *Anglistik: International Journal of English Studies*, 20 (2), 89-100
- Roemer, U. 2010. Establishing the phraseological profile of a text type. The construction of meaning in academic book reviews. *English Text Construction*, 3 (1), 95-119.

---

<sup>1</sup> This research project was funded by Narodowe Centrum Nauki (*National Science Centre*): Sonata 5 programme (Nr rej. 2013/09/D/HS2/00543).