A corpus of contemporary Czech poetry

Michal Škrabal (Institute of the Czech National Corpus, Charles University, Czech Republic), Karel Piorecký (Institute of Czech Literature of the Czech Academy of Sciences, Czech Republic), Petr Plecháč (Institute of Czech Literature of the Czech Academy of Sciences, Czech Republic), Robert Kolár (Institute of Czech Literature of the Czech Academy of Sciences, Czech Republic), Petr Kuběnský (Institute of Czech Literature and Library Studies, Masaryk University, Czech Republic) and Pavel Procházka (Institute of the Czech National Corpus, Charles University, Czech Republic)

The aim of this joint project of two academic institutions in the Czech Republic – Institute of Czech Literature and Institute of the Czech National Corpus – is to build a corpus of *contemporary* poetry (cf. Plecháč & Kolár, 2015) that would, with the methods of corpus linguistics and also within the methodological intentions of *digital literary studies* (cf. e.g. Jockers, 2013; Moretti, 2013; Piorecký, 2015, 2016), investigate Czech language poetry texts published in their first editions since 1990 and going up to the present time (we have set the year 2015 as a boundary line). The goal is, above all, to create a broad empirical foundation on whose basis it shall be possible to identify and verify the specifics of contemporary poetic language and to strengthen the literary-historical knowledge of Czech poetry of the recent past and of the present, thus in the period when it was significantly influenced by a change in the media paradigm given by the arrival of digital media.

In the corpus there are to be found texts compromising, thanks to their elite artistic level, the notional centre of the literary system (its poetic subsystem): especially books published by respected publishers, which follow the standard editorial process and whose edition profile is built on the rigorously chosen principle. A second segment of the corpus are texts situated outside of the demarcated centre of the literary system – primarily because of the not matter-of-course literary competence of their creators. This will above all concern a wide field of texts published on open internet platforms, at which editorial selection and the processing of texts before they are presented to the public is eliminated (so called literary servers, blogs, etc., but also some self-published titles). The representativeness of the corpus should not be secured by the canonization method (by the choice of what is ostensibly the best) but, to the contrary, by a wide base of sources collecting a relatively exhaustive typological and qualitative spectrum of contemporary poetry (from representative publishers to amateur web pages).

Besides the creation of a database for the investigation of literary language, the project would also focus on research questions which it is possible with the collection of the data to solve. Also of significance is the interdisciplinary nature of the entire project: the quantity of themes is mutual both for linguists (e.g., the common lexical characteristics of poetic language, especially in contrast to non-poetic language, and the specifics of lexical semantics; also potentially the problem of the metaphor, poetisms, etc.) and for literary studies (e.g., versification

characteristics – cf. Plecháč, 2016, thematic assemblages and genre structures of poetry production in individual types of media).

References

Jockers, M. L. (2013). *Macroanalysis. Digital methods and literary history*. Urbana: University of Illinois Press.

Moretti, F. (2013). *Distant reading*. London/New York: Verso.

Piorecký, K. (2015). Literatura a digitální kultura: Potřebujeme digital literary studies? *Česká literatura*, 63(6), 935-949.

Piorecký, K. (2016). *Česká literatura a nová média*. Praha: Academia.

Plecháč, P. (2016). Czech Verse Processing System KVĚTA: Phonetic and Metrical Components. *Glottotheory*, 7(2), 159-174.

Plecháč, P., & Kolár, R. (2015). The Corpus of Czech Verse. *Studia Metrica et Poetica*, 2(1), 107-118.