Multilingual practices in Late Modern written English

Arja Nurmi, Päivi Pahta, Jukka Tyrkkö, Anna Petäjäniemi

University of Tampere, Finland

Research on multilingual practices during historical periods of English – and in written language – is still lacking a credible overview of the frequency and type of switching practices evident on the basis of systematic corpus-based study. While single genres and topic domains have received some attention, most of the focus of previous research has been on small datasets. With the current availability of large masses of electronic text from historical periods of English, work on this scale is finally possible. This paper aims to start correcting this oversight, focusing here on the Late Modern period.

With the 333-text, 34-million-word Corpus of Late Modern English 3.0 as our point of departure, we have identified foreign-language passages in running text using a method that makes use of a range of complementary automatic and semi-automatic techniques. We have enhanced the corpus with sociolinguistic background information (gender, social status, age, education, etc.) regarding the authors. We have also expanded the basic text typological data assigned by the corpus compilers with variables such as the suggested audience of each text (for example, specialist vs. lay readers), and possible triggering elements for multilingual passages, such as co-textual references to foreign locations (for example, novels taking place in France vs. those confined to England).

The analysis of this stratificational data on authors and texts will allow us to present a fully evidence-based overview of 1) the frequency of foreign-language passages in written English in 1710–1920, 2) the variety of languages used in these texts in addition to English, 3) the connections of multilingual practices and the social variables describing the authors of each text and 4) the further qualities evident in texts where multilingual resources are drawn upon. The patterns in the data will be identified through multivariate analysis.

Keywords: corpus-based sociolinguistics, Late Modern English, written language.