Quantitative Text Typology

Peter Grzybek\textsuperscript{1} and Ernst Stadlober\textsuperscript{2}

\textsuperscript{1} University Graz, Department of Slavic Studies, A-8010 Graz, Austria
\textsuperscript{2} Graz University of Technology, Department of Statistics, A-8010 Graz, Austria

Abstract. This presentation concentrates on the question in how far quantitative approaches may contribute to an empirically based typology of texts. For a long time, attempts to study ‘language’, not texts, have relied on large corpora, believing that an increase in the data base results in a better approximation of the assumed representative norms. Even in quantitative stylistics, research has rather concentrated on particular functional styles, (e.g., administrative, journalistic, or private) than on individual texts. It has turned out, however, that any combination of two or more texts, or parts of texts, results in a heterogeneous mixture. As a result, different text organizing mechanisms intermingle; the results of corresponding analyses are thus based on abstract linguistic units, rather than on concrete texts, and consequently imply the danger of being artificial.

More recently, quantitative linguistics has begun to shift the attention from corpus analysis to text analysis. Yet, whenever ‘language’ has been the objective, research has been accompanied by the assumption that particular genres might display prototypical characteristics. Letters have been postulated to be such prototypical texts, representing a combination of oral and written elements. By way of an example, this presentation concentrates on word length, only (cf.: http://www-gewi.uni-graz.at/quanta). Submitting the above-mentioned argument to statistical testing, approximately 200 letters of different kinds (e.g., private and open letters, readers’ letters to the editors, letters from an epistolary novel) are analyzed in a balanced study. Applying multivariate methods, such as cluster and discrimination analyses, it shall be demonstrated that even between these closely related text types, there are differences, which allow for specific classifications.

References


Keywords

LANGUAGE, TEXT TYPOLOGY, QUANTITATIVE TEXT ANALYSIS, WORD LENGTH