Authorship Attribution with Delta Measures

Authorship attribution is a highly active research topic in the field of Digital Humanities. Traditional approaches using a wide variety of stylometric criteria have recently been replaced by the much simpler, but highly successful Delta measure (Burrows 2002), which is based on standardized frequency counts of the most frequent words (MFW) in a text collection. Despite many theoretical (e.g. Argamon 2008) and empirical (e.g. Rybicki & Eder 2011; Jannidis et al. 2015) studies on Delta and its variants, there is still no convincing explanation why and how such measures are able to identify the stylistic “fingerprint” of an author with great accuracy.

In this talk, I present the results of a series of experiments on authorship attribution with Delta measures carried out within the BMBF-funded Digital Humanities centre Kallimachos. In particular, I will focus (i) on the influence of parameters such as the number of MFW, text length and distance measure, (ii) on the insight that Delta exploits patterns of over- and underuse rather than the precise magnitude of such deviations, (iii) on questions of statistical significance and (iv) on recent attempts to make sense of the high-dimensional feature vectors underlying Delta.

References

