

# Combining quantitative and qualitative methods in digital analyses of literary style

*J. Berenike Herrmann*

German Department, Göttingen University (Germany)

bherrma1@gwdg.de

This talk reports on a methodological agenda for digital analyses of literary style. Combining quantitative measures from stylometry and qualitative sample studies, I aim to fill the gap between large-scale methods of “distant reading” (e.g., Moretti 2000) and the examination of more intuitively visible patterns of language usage. Putting at the center of my study the prose of Franz Kafka embedded in a corpus of Modern German-written Literature (ca. 1880 - 1930, ca. 5 million words) I propose the following:

**(1) Use quantitative measures to test literary hypotheses:**

Specifically, I examine the question whether the claimed ‘solitariness’ of Franz Kafka’s writing style (e.g., Oschmann 2010) is observable also in quantitative, formal terms (such as distance measures, vocabulary richness, keyword analysis, but also word and sentence length, cf. Stamatatos 2009; see also Craig et al. 2014). For example, a first application of distance measures confirms the idea that Kafka’s writing is not very similar to his contemporaries (Herrmann 2013a).

**(2) Use quantitative results as heuristics for ensuing qualitative style analysis:**

The findings obtained in step 1 are not only suited to probe well-delineated hypotheses, but often also generate new kinds of data, allowing insights about what specific textual features may be indicative of a particular style. For example, a keyword analysis of Kafka’s prose, combined with a manual post-hoc linguistic analysis indicates that Kafka appears to use more modal particles than other authors across genres (Herrmann 2013b). What is more, close inspection of features suggested that these are used in Kafka’s text in a counter-intuitive way (Herrmann under revision).

**(3) Feed qualitative analysis back to run finer-grained quantitative studies:**

Findings obtained in step 1 and 2 may be fed back into quantitative style analyses, using measures that directly test new hypotheses about particular linguistic features. For example, I am currently building a corpus of Modern German literature that will be annotated for part-of-speech, in order to assess lexico-semantic variation across authors, text passages (including fictional characters), and time.

Step 1 is about the empirical examination of widely held scholarly findings, on formal and systematic terms and on a sizable data basis. Step 2 allows making sense of the quantitative

macro results on a micro-level, looking at stylistic phenomena that are closer to readers' experiences of style, thereby providing means of validation for the quantitative measures. Step 3 then returns to the power of quantitative measures and hypothesis testing, however, with adjusted variables and stylistic measures. My talk will run through these three steps, reporting on my ongoing project on the digital analysis of Kafka's style, arguing that a combination of quantitative and qualitative methods in digital literary scholarship can be highly productive and valid, as well as thought-provoking.

## References

- Craig, H., Eder, M., Jannidis, F., Kestemont, M., Rybicki, J., & Schöch, C. (2014). 'Validating computational stylistics in literary interpretation', *Paper presented at the Digital Humanities Conference 2014, Lausanne*. <http://dharchive.org/paper/DH2014/Paper-857.xml>
- Herrmann, J. B. (2013a). 'Kafka among the authors. Stylometric analyses', *Paper presented at the Expert Workshop "Stylometry@Kraków"*, Krakow, Poland.
- Herrmann, J. B. (2013b). 'Computing Kafka - How keyness and collocation analysis help explain paradoxical style', *Paper presented at the International Herrenhausen Conference "(Digital) Humanities Revisited – Challenges and Opportunities in the Digital Age"*, Hanover, Germany.
- Herrmann, J. B. (under revision). 'Läuse im Pelz der Sprache? Zu Funktionen von Modalpartikeln in narrativen (De-)Motivierungsstrategien bei Franz Kafka'. In M. Horváth & K. Mellmann (Eds.), *Die biologisch-kognitiven Grundlagen narrativer Motivierung*. Berlin: De Gruyter.
- Moretti, F. (2000). 'Conjectures on world literature', *New Left Review*, vol. 1, pp. 54-68.
- Oschmann, D. (2010). 'Kafka als Erzähler', In M. Engel & B. Auerochs (Eds.), *Kafka-Handbuch* (pp. 438-449). Stuttgart.
- Stamatatos, E. (2009). 'A survey of modern authorship attribution methods', *Journal of the American Society for Information Science and Technology*, vol. 60(3), pp. 538-556.