

# Stylometry for medieval chronicles on the example of the “Kaiserchronik”

*Łukasz Gagala*

Universität Göttingen

lukaszgagala@wp.pl

A mammoth part of recent research on computational stylometry and non-traditional authorship attribution concerns modern languages and their modes of usage (e.g. corpora for PAN-contests involving blogs, essays or newspaper articles). Research conditions notably alter when one takes into account pre-modern texts and their specificity. It is not only a matter of research infrastructure for extinct languages/dialects or earlier stages of modern languages (sparse data, uniqueness of a given text in its category etc.), but also application of methodological concepts of stylometry for manuscripts.

Authorship as a pivotal term of philology has undergone quantification by means of stylometry. A contested climax of these scientific proceedings turns a so-called Stylome Hypothesis claiming existence of a linguistic DNA concealed in an author’s every text. According to that – to say it in the manner of the DH slang – each author should be seen as a textual machine with its unique distribution of language elements. This idea, being nowadays very intensively tried and tested in research on texts of the print age, must be considerably tailored for pre-modern manuscripts.

Those restraints have been orderly formulated by Herman Brinkman in the form of ‘Ten Commandments’ for stylometric authorship attribution of medieval texts. Pre-modern manners of text (re-)usage and transmission hinged upon scribes and their ‘editorial’ decisions (ranging from a transcription convention up to slicing, mingling and reworking the original), that in many cases should be defined as co-authorship, judging the issue by our contemporary standards. To solve the issue of scribe-dependency, one should either concern only (close-to-) prototype manuscripts or search for stable text elements not influenced (in the terms of statistics) by scribes, like rhyme words in rhymed texts. If a scribe potentially might have modified statistical features of a text-to-transcribe, then considering the Stylome Hypothesis he/she should have left his/her linguistic fingerprint in a transcribed/compiled text and this must be an object of statistical analysis as well.

In view of prolificacy of the Middle High and Middle Low German literature potential advances in stylometric analysis of author-scribe relations will give insights in the medieval process(es) of writing longer works with potential multiple authorship. One of my case studies (regarded as a kind of pilot research) is the 12th century *Kaiserchronik* (‘The Book of Emperors’/‘Imperial Chronicle’), the first versed chronicle written in (Middle High) German, thought of as composed by many authors using plenty of sources. One of the arguments for multiple authorship and multiple sources of the *Kaiserchronik* is a short 200 verses long fragment of *Annolied* (‘Song of Anno’), a 100 years older apology of the Cologne archbishop Anno, that is to find at the beginning of the chronicle. Analysing rhyme words, most frequent words, function words and part-of-speech tags I try to examine the authorial structure of the whole Vorau manuscript of the *Kaiserchronik* (fully annotated in *Reference Corpus Middle High German (1050-1350)*) and compare it with other copies/fragments at disposal.

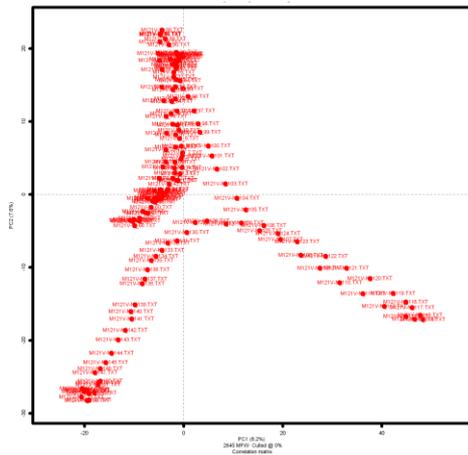


Fig. 1 A PCA plot for rhyme words (more than 17.000 entries) of the main text of *Kaiserchronik*. Each point represents a text slice of 1000 rhyme words. The smoothed make-up of clustering is caused by the windowing procedure (here a step of 100 rhyme words). [Performed with *Stylo R* package].

The preliminary results of unsupervised analysis with different alternative stylometric features (PoS-tags, MFW, rhyme words) indicate a strong clustering of the *Kaiserchronik* into many fragments (inclusively the aforementioned *Annolied*-fragment). The evident coherence of this phenomena over a set of different features supports the compilation theory. Furthermore, statistical differences between the stylistic clusters have been evaluated in comparison to other narrative texts written in the same time and dialect region. In this way we reassure that the received results should be treated as authorial signals and not as stylometric variability within an oeuvre. Those empirical findings encourage further research on different chronicle versions digitalised by the Cambridge-based *Kaiserchronik Project*.

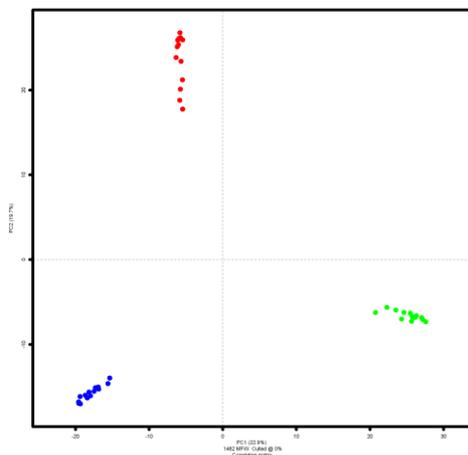


Fig. 2 The ‘control’ group of three narrative rhymed texts from manuscripts written in the same time period and dialect region as the *Vorau* manuscript: Wolfram von Eschenbach’s *Parzival* (green), Hartmann von Aue’s *Iwein* (red), Gottfried von Straßburg’s *Tristan* (blue). The PCA plot reveals three distinct clusters for three authors; 15 samples of length 1000 rhyme words in each cluster. [Performed with *Stylo R* package]

As a result I expect identification of not only authorial signals but also ‘scribal’ ones in the *Kaiserchronik*. The expertise gained through the stylometric research on the *Kaiserchronik* should serve us also in the analysis of the *Sächsische Weltchronik* (‘Saxon World Chronicle’), the first (Middle Low) German chronicle written in prose, being a very important work of medieval vernacular historiography.

This empirical research is paired with theoretical reflexion how to accommodate the Stylome Hypothesis (which as such is a matter of hot dispute) to the context of the medieval text culture and its feeling of authorship. In the near future, when having digitalised more than one manuscript of a particular work, one will be able to treat the scribe-factor as an element of statistical analysis, not a threat or data noise jeopardising reliability of results.

### Selected bibliography

- Adamska A., Kraś P. (2014), *Kultura pisma w średniowieczu. Znane problemy - nowe metody*, Lublin: Wydawnictwo KUL.
- Andersen E. (1998) *Autor und Autorschaft im Mittelalter*, Berlin: Walter de Gruyter.
- Dickhut-Bielsky J. (2015) *Auf der Suche nach der Wahrheit in "Annolied" und "Kaiserchronik": Poetisch-historiographische Wahrheitssuche in frühmittelhochdeutschen Geschichtsdichtungen*, Stuttgart: Hirzel S. Verlag
- van Dalen-Oskam K.H. (2012) „The secret life of scribes. Exploring fifteen manuscripts of Jacob van Maerlant's *Scolastica* (1271)“, in: *LLC: The journal of digital scholarship in the Humanities*, 27, Oxford University Press, pp. 355-372
- Jannidis F., Lauer G., Martínez M. and Winko S. (1999) *Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen: Niemeyer.
- Kestemont M., Moens S. and Deploige J. (2015), „Collaborative Authorship in the Twelfth Century. A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux“, in: *The Journal of Digital Scholarship in the Humanities*, Oxford University Press, 30(2), pp. 199-224.
- Kestemont M. (2013) *Het gewicht van de auteur. Stylometrische auteursherkenning in Middel nederlandse literatuur*, Ghent.
- Love H. (2002) *Attributing authorship. An introduction*, Cambridge: Cambridge University Press
- Matthews A. (2012) *The Kaiserchronik: A Medieval Narrative*, Oxford: Oxford University Press
- Mierke G. (2014) *Riskante Ordnungen: Von der Kaiserchronik zu Jans von Wien*, Berlin: Walter de Gruyter
- Mostert M. and Adamska A. (2014a) *Writing and the Administration of Medieval Towns: Medieval Urban Literacy I*, Turnhout: Brepols
- Mostert M. and Adamska A. (2014b), *Uses of the Written Word in Medieval Towns: Medieval Urban Literacy II*, Turnhout: Brepols
- Müller S. (1999), *Vom Annolied zur Kaiserchronik: zu Text- und Forschungsgeschichte einer verlorenen deutschen Reimchronik*, Heidelberg: Winter Verlag
- Piotrowski M. (2012) *Natural Language Processing for Historical Texts*, Morgan & Claypool Publishers.
- Stolz M. and Mettauer A. (2006) *Buchkultur im Mittelalter: Schrift – Bild – Kommunikation*, Berlin: Walter de Gruyter
- Wolf J. (1997) *Die Sächsische Weltchronik im Spiegel ihrer Handschriften: Überlieferung, Textentwicklung, Rezeption*, Paderborn: W. Fink