

Robust Rhymes or Stable Stresses? Finding the Authorial Features Most Resilient to Variation

Valentina Modolo^{1,2}, Théo Moins¹, Marie Bizais-Lillig^{1,3}, Cecile Vermaas¹, Jean-Baptiste Camps¹,
Mathias Garnier¹

LT

¹ École nationale des chartes - PSL

² Scuola Normale Superiore di Pisa

³ Université de Strasbourg

Authorship attribution in medieval French literature poses specific methodological challenges due to manuscript transmission, orthographic variation, and the instability of textual corpora. While computational stylometry has shown that such texts can be analysed using supervised methods such as Support Vector Machines, the reliability of stylometric analyses depends critically on the robustness of the features employed. In this perspective, Kestemont, Dalemans and Sandra [1] suggested that metrical constraints, rather than surface lexical forms, may preserve authorial signals more effectively. Building on this insight, recent work by Camps et al.[2] has shifted the focus toward prosodic features, using stress patterns as quantifiable stylistic markers. Approaches based on prosodic annotation, such as the Metronome framework [3], allow rhythmic structures of medieval verse to be modelled independently of lexical variation. The present study situates itself within this line of research and examines the behaviour of stress-based features under conditions of textual variation in medieval octosyllabic verse.

The corpus consists of texts written by Chrétien de Troyes (ca. 1130-1190), Robert Wace (ca. 1100-1175) and Gerbert de Montreuil (born ca. 1200). For each author two texts are selected and for each text two witnesses (see appendix). Therefore it is possible to compare relative distances between different witnesses, texts and authors. The samples taken from each witness overlap in halfth their length with the other witness from the same text, to enable us to see the influence of scribal intervention.

To process the corpus we transcribe three hundred verses of each witness and apply lemmatization and POS-tagging to the selection. Finally stress patterns are created using Anochre [4] and the distances between the stress patterns of each witness are computed (see appendix for examples).

Corpus processing on overlapping segments reveals that lexical features and rhymes are highly sensitive to textual variation and therefore poorly suited for capturing authorial signals in medieval verse. By contrast, stressed-based features display a remarkable stability across witnesses and texts, confirming their resilience to scribal intervention and supporting their relevance as a foundation for authorship-oriented analyses. On the other hand, the experiments based on non-overlapping segments suggest a less clear-cut authorial fingerprint.

References

- [1] Kestemont, M., Daelemans, W., and Sandra, D. (2012). Robust Rhymes? The Stability of Authorial Style in Medieval Narratives*. *Journal of Quantitative Linguistics*, 19(1), 54-76. <https://doi.org/10.1080/09296174.2012.638796>
- [2] Camps, J.-B. et al. (2025a). Style in Eight Syllables: Metric Annotation and Stylometry of Chrétien de Troyes and Contemporaries. *Digital Humanities Benelux 2025 Conference*, 1-12.
- [3] Nagy, B. et al. (2025). Metronome: Tracing Variation in Poetic Meters via Local Sequence

Appendix

List of corpus

Table 1: Composition of the corpus (all texts are romances in octosyllabic verse).

Author	Title	Date (Work)	Manuscript(s)	Date (MS)
Chrétien de Troyes	<i>Erec et Enide</i>	ca. 1170	fr. 01376, BnF; fr. 24403, BnF	1288; mil. XIII c.
Chrétien de Troyes	<i>Cligès</i>	ca. 1176	fr. 1374, BnF; fr. 01450, BnF	mil. XIII c.; 1st q. XII c.
Robert Wace	<i>Rou</i>	ca. 1170	BnF, fr. 794 (C)	c. 1235
Robert Wace	<i>Brut</i>	ca. 1155	Ottob. Lat. 1869 (Vatican Library); BL Add. 45103 (British Library)	XII c.; XIII c.
Gerbert de Montreuil	<i>Violette</i>	ca. 1227–1229	fr. 1553, BnF; fr. 1374, BnF	1284; mil. XIII c.
Gerbert de Montreuil	<i>Continuation</i>	start XIII c.	fr. 12576, BnF; nouv. acq. fr. 6614, BnF	beg. XIII c.; XIII c. (frag.)

Numerical analysis

The two plots below are preliminary results to support the arguments. See https://github.com/Bizais-Lillig/robust_rhythm for the code and more detailed examples.

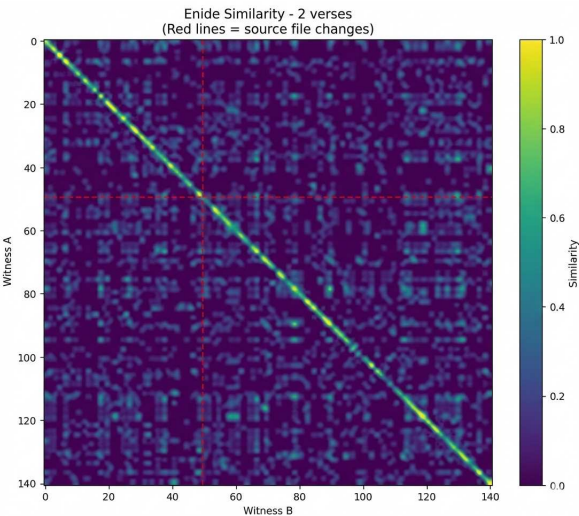


Figure 1: Similarity between witnesses for *Erec et Enide*. A high intensity on the diagonal means proximity between the two witnesses.