The Translators' Touch
A Computational Stylometric Inquiry into Medieval Greek-Latin Translations

Pieter Beullens (Research Foundation Flanders, FWO – KU Leuven)
Wouter Haverals (University of Antwerp)
Ben Nagy (Polish Academy of Sciences)

The present study focuses on the examination of translators of Greek philosophical and technical texts from the 12th and 13th centuries. The Latin translations they produced, particularly those of Aristotle's works, exerted a profound influence on the intellectual advancements during the later Middle Ages and the Early Modern period. Despite the widespread impact of their work, many of these translators remained modest and their works have been transmitted anonymously. Nevertheless, the identification of these individuals would provide valuable insight into the context in which the process of cultural transfer took place.

It is important to recognize that conventional methods that have demonstrated effectiveness in the identification of translators (such as analysing patterns of preferred words) cannot be straightforwardly applied to translations from the Late-Medieval period, given the unique transmission history and inherent qualities of these works. Medieval translators of Greek philosophical and technical texts adhered to a rigorous methodology, which imposed the requirement for a precise correspondence of each Greek word with a suitable Latin equivalent. Consequently, the translations exhibit a uniformity in style that is very distinct from original Latin texts on related topics written during the same period.

From the first half of the 20th century onward, scholars have recognized the value of function words to identify the individual stylistic 'fingerprint' in the written works of different translators (e.g. Minio-Paluello 1947a and 1947b; Bossier 1989, see also Figure 1). Through a meticulous comparison of the renderings of function words and other semantic and grammatical elements in the Latin versions with their Greek counterparts, scholars have been able to determine key features of the methods employed by the foremost figures in this translation movement (e.g. Beullens 2020).

γάρ = enim (46Ph, 37M, 63m) — quippe (12Ph, 14M, 8m) — namque (4Ph, 9M, 11m) — siquidem (5Ph, 6M, 13m) nam (2Ph, 6M, 12m) — om. (1Ph, 2M, 3m) — etenim (1Ph, 2M, 2m) — nimirum (5M, 11m) — quia (2Ph, 2m) — quoniam (1Ph, 2m) — vero (1Ph, 1m) — autem (4m) — vere (1Ph) — quidem (1m).

δέ = vero (31Ph, 45M, 111m) — autem (37Ph, 45M, 44m) — om. (39Ph, 25M, 49m) — -que (4Ph, 22M, 60m) — sane (10Ph, 13M, 24m) — quoque (6Ph, 7M, 21m) — profecto (2Ph, 8M, 14m) — utique (2Ph, 4M, 12m) — nempe (2Ph, 4M, 11m) — certe (3Ph, 7M, 5m) — atque (1Ph, 1M, 5m) — et (1Ph, 1M, 1m) — quidem (1Ph, 18m) — itaque (1Ph, 5m) — enim (1Ph, 4m) — at (1M, 3m) — quippe (2Ph, 1M) — equidem (1M, 2m) — sed (4m) — etiam (3Ph) — revera (3M) — ergo (2m) — nimirum (2M) — atqui (1m) — immo (1m) — necnon (1M) — ceterum (1m) — igitur (1m).

δή = om. (21Ph, 30M, 2m) — vero (3Ph, 6M, 2m) — utique (5Ph, 5M, 1m) — autem (3Ph, 8M) — quidem (2Ph, 2M) — sane (1Ph, 3M) — certe (1Ph, 2M) — profecto (3M) — nempe (2M) — quoque (2M) — etiam (1Ph) — equidem (1M) — revera (1M).

	sed (+et) adhuc	sed et	sed etiam	sed tamen	quinimmo	insuper	sed	at vero
De celo, rec. 1	36	13	2	3	1	1	0	0
Al. In Met.	2	6	1	1	1	5	0	0
De gen. an., rec. 1 I	5	1	0	1	0	0	0	0
II	4	0	0	0	0	0	0	0
III	2	0	0	0	0	0	0	0
IV	1	0	0	0	0	0	0	0
V	—	—	—	—	—	—	—	—
De mo. an.	2	0	0	0	0	0	0	0
Polit. imp.	2	1	0	0	0	0	0	6
De part. an.	0	0	0	0	0	0	0	3
Metaph. XI	—	—	—	—	—	—	—	—
XIII-XIV	0	0	0	0	0	0	0	21
De pro. an.	0	0	0	0	0	0	0	2
Al. In De sensu	0	0	0	0	0	0	1	15
De hist. an.	0	0	0	0	0	0	0	1
Meteor.	0	0	0	0	1	0	0	8
Rhetor.	—	—	—	—	—	—	—	—
Al. De fato	0	0	0	0	0	0	0	14
Ptol. Iudic.	—	—	—	—	—	—	—	—
Polit. II.11-VIII	0	0	0	0	0	0	0	24
Simpl. In Cat.	0	0	0	0	3	0	2	7
Them. De an.	0	0	0	0	0	0	1	5
Am. Peri herm.	0	0	0	0	3	0	3	8
Poet.	—	—	—	—	—	—	—	—
Procl. Tria opusc.	0	0	0	0	0	0	0	1
In Parm.	0	0	1	0	0	0	1	53

Figure 1 – **Left**: Lorenzo Minio-Paluello, "Henri Aristippe, Guillaume de Moerbeke et les traductions latines médiévales des «Météorologiques» et du «De Generatione et Corruptione» d'Aristote", *Revue Philosophique de Louvain*, 45 (1947), 206-235, list taken from 215. **Right**: Fernand Bossier, "Méthode de traduction et problèmes de chronologie", in *Guillaume de Moerbeke. Recueil d'études à l'occasion du 700ᵉ anniversaire de sa mort (1286)*, ed. Jozef Brams, Willy Vanhamel (Leuven: University Press, 1989), 257-294, table taken from 273.

In the current project, we aim to evaluate the established attributions through the application of computational stylometric techniques (i.e., the quantitative study of writing style), ultimately clustering anonymously transmitted translations with other Latin versions by known translators. Our initial methodology involves the establishment of a reference corpus of texts, comprised of works whose translators are either explicitly identified in the manuscripts preserving them or for whom the attribution is widely accepted. The selected texts were translated by six prominent scholars from the 12th century (Burgundio of Pisa, James of Venice, Henricus Aristippus, and an unnamed translator) and the 13th century (William of Moerbeke, and Bartholomew of Messina).[1] The treatises encompass a wide range of topics, including philosophy, medicine, astrology, and theology, allowing us to assess the potential impact of genre characteristics on the translations. Moreover, the corpus includes instances of Latin versions of the same Greek texts made by different translators, offering an opportunity to precisely evaluate the salient features that display variation, even when utilizing the same source text. Such a comparison allows for a clearer differentiation between the translator's personal style and the characteristics inherent to the source text. This differentiation can also be achieved by including the translators' prefaces in our analyses, thereby distinguishing the style of the translator and their unique qualities as an autonomous author.

In our study, we aim to assess the effectiveness of two stylometric methods for the analysis of translation style. The first method employs a pre-determined list of function words that have been considered to be indicative of Greek-to-Latin translation style in previous research. An initial examination of this approach, utilizing the k-nearest neighbours (kNN) algorithm for classification, has already yielded intriguing results (Figure 2). In the current, final stages of our research, this algorithm has demonstrated remarkable success in differentiating individual translators from each other, with a few interesting exceptions. Additionally, the Random Forest Classification technique will be employed, where the most distinctive features for each author will be identified through the automatic comparison of multiple decision trees. Particularly useful is Random Forest Classification's ability to identify the most important features for classification. By considering the frequency of feature selection across numerous randomly generated decision trees, it is possible to determine which features are most impactful in determining class labels (Haverals & Kestemont, forthcoming). This information can be used to gain insight into the characteristics that define the translation style of individual authors.

Ultimately, our study will contribute to the fields of translation studies and stylometry, with a particular emphasis on similarities in the style of late-medieval translators. The techniques employed will be compared with each other, making this research intriguing as it provides us with a better understanding of stylometry based on careful, premeditated feature selection on the one hand versus automated feature selection using Random Forests on the other.

---

[1] For the composition of our corpus, we primarily utilize the critical editions of the translated works if they have been published. Many of them are available through digital repositories such as *Brepolis*. It must be noted that there is of course a risk that any errors made by scribes and/or non-contemporary editorial interventions can have an impact on the transmission of the text. Nevertheless, we are convinced that given the amount of texts, this impact is negligible and furthermore, the high-frequency function words remain largely unaffected (see also Kestemont 2014).
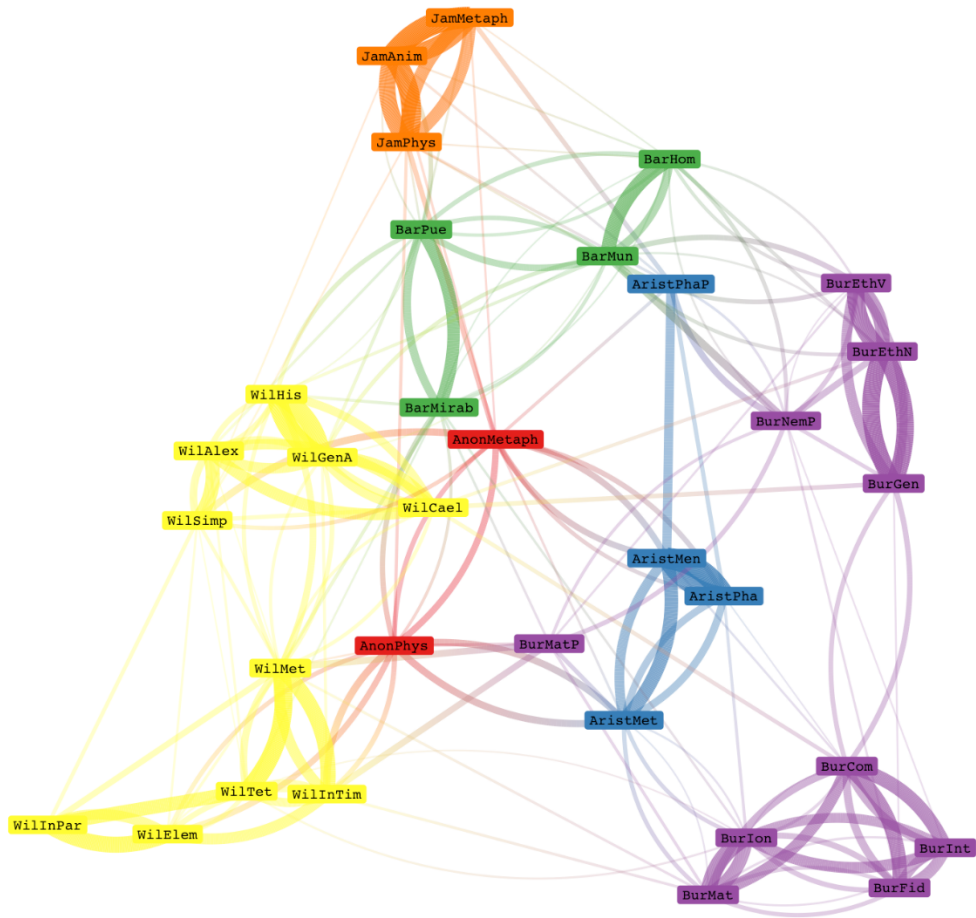
Figure 2. Network built based on 100 k-nearest neighbours votes, showing the stylistic relations between various late-medieval translators (using a random subset of 25 features, distance: cosine).

Bibliography

Pieter Beullens, *A Methodological Approach to Anonymously Transmitted Medieval Translations of Philosophical and Scientific Texts. The Case of Bartholomew of Messina*. PhD dissertation (KU Leuven, 2020).

Fernand Bossier, "Méthode de traduction et problèmes de chronologie", in *Guillaume de Moerbeke. Recueil d'études à l'occasion du 700ᵉ anniversaire de sa mort (1286)*, ed. Jozef Brams, Willy Vanhamel (Leuven: University Press), 1989, 257-294.

Jeroen Deploige, Jeroen de Gussem, "Medieval Authorship and Canonicity in the Digital Age – an Introduction", in *Interfaces*, 8 (2021), 113-124. DOI: https://doi.org/10.54103/interfaces-08-07

Wouter Haverals, Mike Kestemont, "From exemplar to copy: The scribal appropriation of a Hadewijch manuscript computationally explored", in *Journal of Data Mining and Digital Humanities* (forthcoming). DOI: https://doi.org/10.48550/ARXIV.2210.14061

Mike Kestemont, "Function Words in Authorship Attribution. From Black Magic to Theory?", in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature* (2014), 59-66. URL: https://aclanthology.org/W14-0908

Lorenzo Minio-Paluello, "Guglielmo di Moerbeke, traduttore della *Poetica* di Aristotele (1278)", in *Rivista di Filosofia Neo-Scolastica*, 39 (1947a), 1-17. Reprinted in Id., *Opuscula. The Latin Aristotle* (Amsterdam: Hakkert, 1972), 40-56.

Lorenzo Minio-Paluello, "Henri Aristippe, Guillaume de Moerbeke et les traductions latines médiévales des «Météorologiques» et du «De Generatione et Corruptione» d'Aristote", in *Revue Philosophique de Louvain*, 45 (1947b), 206-235. Reprinted in Id., *Opuscula. The Latin Aristotle* (Amsterdam: Hakkert, 1972), 57-86.