

Delta und Merkmalsselektion: Welche Wörter unterscheiden arabisch-lateinische Übersetzer?

Delta hat sich als stilometrisches Verfahren zur Autorschaftsattribuierung bereits für verschiedene Sprachen und Textkorpora bewährt. Unser Vortrag wendet die Methode auf arabisch-lateinische Übersetzungen wissenschaftlicher Texte aus dem 12. Jahrhundert an und widmet sich dabei insbesondere dem Problem, dass die unterschiedliche inhaltliche Ausrichtung der Texte den individuellen Stil der Übersetzer überlagert.

In unseren Experimenten zeigen wir, dass eine Verwendung der Häufigkeiten der häufigsten Wörter auf unserem Textkorpus in erster Linie eine Gruppierung der Texte nach den Disziplinen erzeugt, aus denen die Texte stammen. Durch eine Merkmalsselektion, in unserem Fall die rekursive Merkmalseliminierung, können jedoch Wörter gefunden werden, deren Häufigkeitsverteilung nur für die Übersetzer der Texte oder nur für die Disziplinen der Texte spezifisch ist. Insbesondere zeigt sich, dass der Ausschluss der für eine der beiden Kategorien entscheidenden Wörter eine Verbesserung der Klassifizierung hinsichtlich der jeweils anderen Kategorie zur Folge hat. Das Verfahren erweist sich so als wirksames Instrument zur Bestimmung von übersetzer- und disziplintypischen Merkmalen und bietet zudem die philologische Chance, anstatt einer undurchsichtigen statistischen Maschinerie überschaubare Wortlisten als nachvollziehbaren Kriterien bei der Klassifizierung von Texten einzusetzen.

Andreas Büttner

JMU Würzburg

Institut für Philosophie

Thomas Proisl

FAU Erlangen-Nürnberg

Professur für Korpuslinguistik