

Regionale Sprachmerkmale und ihre Kartierung: Korpusgestützte Ansätze in der forensischen Autorschaftsanalyse

Forensisches Autorschaftsprofiling befasst sich mit der Analyse sprachlicher Merkmale, um auf soziodemografische Eigenschaften unbekannter Autor*innen wie etwa Alter, Geschlecht oder den Einfluss einer Erstsprache rückzuschließen (Nini, 2018). Die regionale Herkunft von Schreibenden wurde bislang hingegen nur begrenzt berücksichtigt, obwohl Einzelfälle wie die Identifikation des Regionalismus *devil strip* in einem US-Erpresserbrief (Shuy, 2001) das Potenzial dieses Ansatzes verdeutlichen. Die vorliegende Studie knüpft an solche Arbeiten an und erweitert sie durch korpuslinguistische Methoden zur großskaligen Analyse regionalsprachlicher Variation.

Im Vortrag wird ein korpusgestützter Zugang zum geolinguistischen Autorschaftsprofiling vorgestellt, der sich von der herkömmlichen, auf Wörterbücher oder Dialektatlanten gestützten Analyse regionaler Sprachmerkmale löst. Basis ist ein Korpus von 21 Millionen geolokalisierten Social-Media-Beiträgen von der Plattform Jodel (Hovy & Purschke, 2018). Zunächst erfolgt eine Untersuchung der regionalen Verteilung sprachlicher Merkmale im Korpus, wobei auch Herausforderungen wie Datensparsamkeit und urbane Verzerrung thematisiert werden. Durch Anwendung geostatistischer Verfahren, insbesondere des *Ordinary Kriging* (Wackernagel, 2003), wird eine flächendeckende Kartierung regionaler Varianten ermöglicht. Darüber hinaus wird ein Algorithmus zur Vorhersage von Dialektregionen diskutiert.

Die Ergebnisse unterstreichen den Mehrwert korpuslinguistischer Verfahren für das forensische Autorschaftsprofiling. Die Nutzung großer, geolokalisierter Korpora verbessert nicht nur die Genauigkeit regionaler Zuordnungen, sondern schafft auch eine empirische Grundlage für qualitative Analysen. Der hier vorgestellte Ansatz leistet somit einen Beitrag zur Entwicklung automatisierter Verfahren und zur methodischen Erweiterung der forensischen Linguistik.

Hovy, D., & Purschke, C. (2018). Capturing Regional Variation with Distributed Place Representations and Geographic Retrofitting. *Proceedings of the 2018 Conference on EMNLP*, 4383–4394. <https://doi.org/10.18653/v1/D18-1469>

Nini, A. (2018). Developing forensic authorship profiling. *Language and Law / Linguagem e Direito*, 5(2), 38–58.

Shuy, R. W. (2001). DARE's role in linguistic profiling. *DARE Newsletter*, 4(3), 1–5.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer.