

# Interaktive Analyse historischen Grammatikwandels: Konstruktionsgrammatik trifft auf *machine learning*

Dr. Marie-Luis Merten & Prof. Dr. Doris Tophinke (Universität Paderborn)

Im Paderborner Digital Humanities-Projekt InterGramm arbeiten Linguisten, Computerlinguisten und Informatiker zusammen, um den Sprachausbau des Mittelniederdeutschen (Tophinke 2012, Merten i. Dr.) zu erforschen (Seemann et al. 2017). Korpusbasiert werden das Entstehen komplexer Konstruktionen, die wachsende Präferenz für verdichtende *construal*-Techniken sowie der Ausbau (lese-)textorganisierender Konstruktionen – also Prozesse des Konstruktionswandels und der Konstruktionalisierung (Traugott/Trousdale 2013) – genauer in den Blick genommen. Unter Berücksichtigung von Ambiguität(en), kategorialer Gradienz, diachroner Gradualität und Fällen von Annotatorenunsicherheit folgt die Annotation der Daten – ein ca. 1,2 Millionen Tokens umfassendes Korpus, das größtenteils aus juristischen Texten besteht – einem *human-in-the-loop*-Ansatz: Auf Basis einer ersten partiellen, tlw. unsicheren Annotation auf POS- sowie CC<sup>1</sup>-Ebene, die der linguistische Experte vornimmt, induziert das lernende System grammatische Muster, die wiederum durch den Linguisten korrigiert werden. Die Phasen *Experten-Korrektur* und *maschinelles Lernen & Induzieren* folgen iterativ solange aufeinander, bis ein zufriedenstellendes maschinelles Analyseergebnis erzielt wird. Insgesamt gesehen ermöglicht der Vortrag nicht nur einen Einblick in die disziplinenübergreifende Erforschung der mnd. Grammatik, sondern schärft v. a. den (sprachwissenschaftlichen) Blick für die korpuslinguistische Beschäftigung mit konstruktional-dynamischen Wandelphänomenen.

## Quellen

Merten, Marie-Luis (i. Dr.): Literater Sprachausbau kognitiv-funktional. Funktionswort-Konstruktionen in der historischen Rechtsschriftlichkeit. Berlin: de Gruyter (Reihe Germanistische Linguistik 311).

Seemann, Nina / Marie-Luis Merten / Michaela Geierhos / Doris Tophinke / Eyke Hüllermeier (2017): Annotation Challenges for Reconstructing the Structural Elaboration of Middle Low German. In: Proceedings of Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, and Literature. Vancouver, BC, August 4, 2017, 40–45.

Tophinke, Doris (2012): Syntaktischer Ausbau im Mittelniederdeutschen. Theoretisch-methodische Überlegungen und kursorische Analysen. In: Niederdeutsches Wort 52, 19–46.

Traugott, Elizabeth C. / Graeme Trousdale (2013): Constructionalization and constructional change. Oxford: Oxford University Press.

---

<sup>1</sup> Dabei handelt es sich um die Ebene *complex constructions* (keine atomaren Konstruktionen), für die ein eigens entwickeltes konstruktional-Tagset im Einsatz ist.