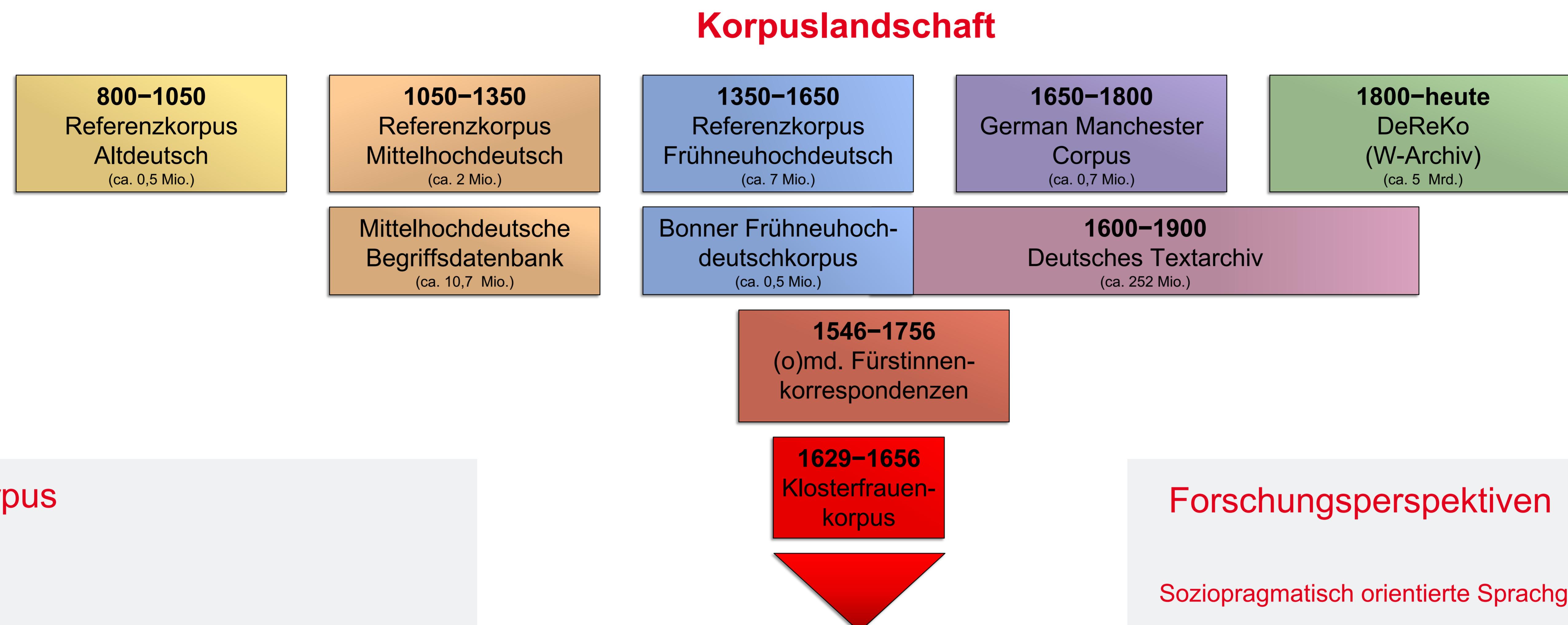


Sprachgebrauch süddeutscher Klosterfrauen des 17. Jahrhunderts

Linguistische Aufbereitung und Annotation eines historischen Spezialkorpus



Zum Korpus

Motivation

In historischen Referenzkorpora des Deutschen sind Texte, die von Frauen verfasst wurden, nach wie vor stark unterrepräsentiert. Gerade aus kultur-, aber auch aus sprachhistorischer Perspektive kann sich jedoch der Vergleich von Texten, die von Schreiberinnen stammen, mit solchen, die von Männern verfasst wurden, als aufschlussreich erweisen, wie etwa die Befunde zeigen, die auf Grundlage des Fürstinnenkorrespondenz-Korpus von Lühr (2016, 2018) erzielt wurden. Mit Nolting (2010) liegt eine handschriftennahe Edition von insgesamt acht chronikalischen sowie Brieftexten vor, die von vier katholischen Klosterfrauen in der ersten Hälfte des 17. Jahrhunderts verfasst wurden. Diese Edition bildet den idealen Ausgangspunkt für eine korpuslinguistische Aufbereitung der Texte, die sich methodisch und inhaltlich an bereits existierenden historischen Referenz- und Spezialkorpora orientiert.

Aufbereitung

Die edierten Texte wurden zunächst mit CAB (Jurish 2012), einer Sammlung von Tools, die fürs Deutsche Textarchiv speziell zur Analyse historischer deutschsprachiger Texte entwickelt wurde, automatisch normalisiert und auf dieser Grundlage lemmatisiert und POS-getaggt. Für die weitere halbautomatische sowie manuelle Annotation wird GATE (Cunningham et al. 2002) genutzt. Mit Hilfe von PEPPER (Zipser & Romary 2010) werden die GATE-XML-Dateien schließlich für das browserbasierte Such- und Visualisierungstool ANNIS (Krause & Zeldes 2016) aufbereitet.

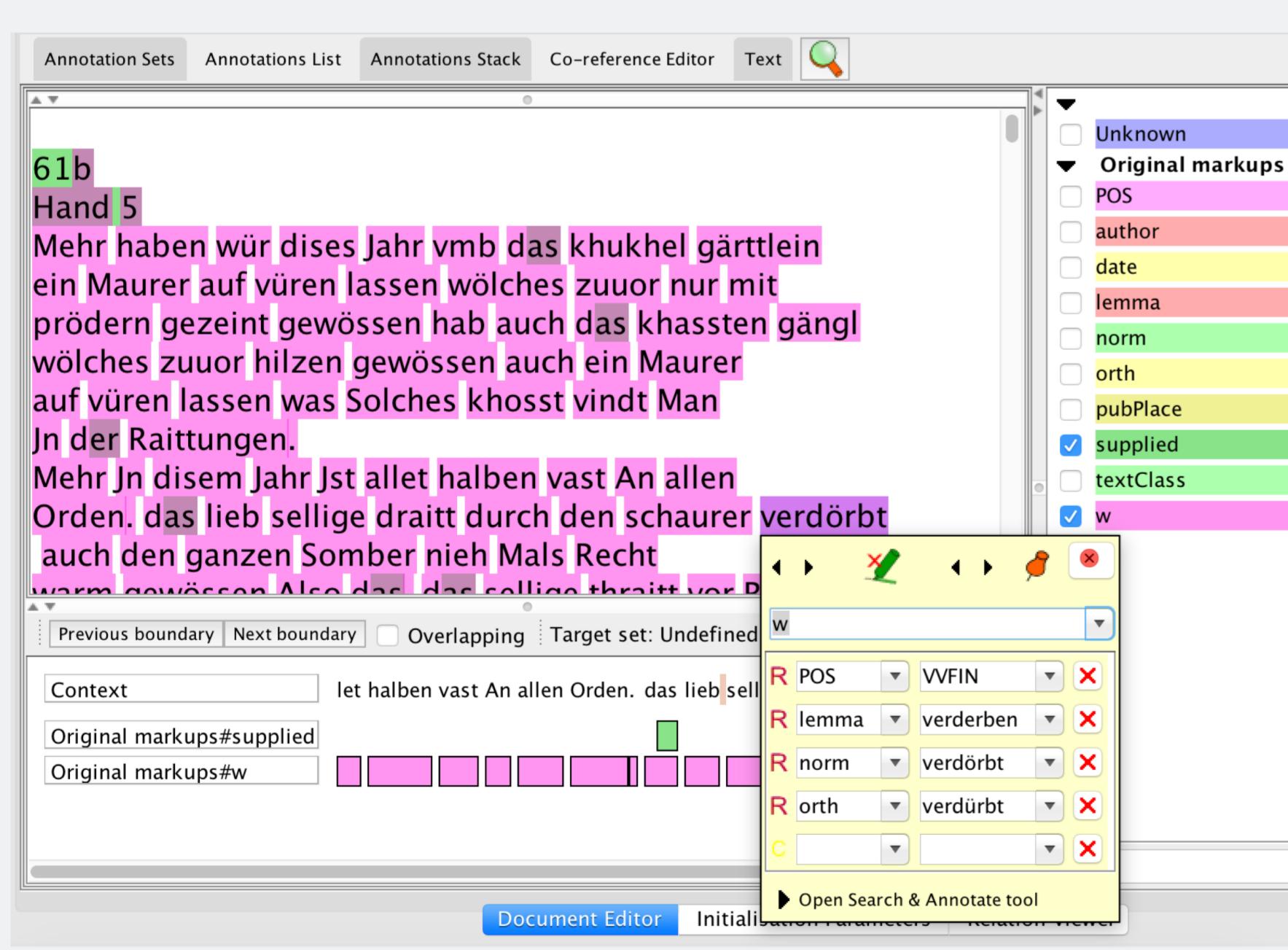


Abb. 1: Annotation in GATE.

Korpusaufbau

Das Korpus besteht aus chronikalischen Texten (Klosterhistoriographie) und Brieftexten, wobei Erstere jedoch deutlich überwiegen. Auch unterscheiden sich die einzelnen chronikalischen Texte deutlich hinsichtlich ihrer Größe. Dieser Aufbau ist der Edition geschuldet, die explizit "als exemplarische Zusammenstellung konzipiert ist" (Nolting 2010: 51) und keinen Anspruch auf Repräsentativität erhebt.

Bei den Verfasserinnen handelt es sich um "geübte Schreiberinnen, jedoch keine gelehrten 'Schreibprofis'" (Nolting 2010:261)

Übersicht Korpustexte

Teilkorpus	Chroniken	Briefe	ges.
Tokens	197.654	1.552	199.206
Schreiberin			
Ernst	6.510	319	6.829
Junius	53.961	---	53.961
Staiger	101.213	1.113	102.326
Frauenchiemsee	34.538	---	34.538

Tab. 1: Überblick über die Korpustexte.

Annotationsebenen

Die Texte liegen in diplomatischer Transkription (**dpl**) und in zum Nhd. hin normalisierter (**norm**) Form vor. Sie wurden automatisch lemmatisiert (**lemma**) und POS-getaggt (**POS**), weiterhin wurden editorische Ergänzungen auf der Ebene **supplied** als "added_editors" markiert (automatisch mittels des Markup in der Edition). In Kürze werden zudem die gelöschten Textstellen ausgezeichnet (**del**). Weitere syntaktische und inhaltliche Annotationen sind geplant.

Sprachgeographische Zuordnung

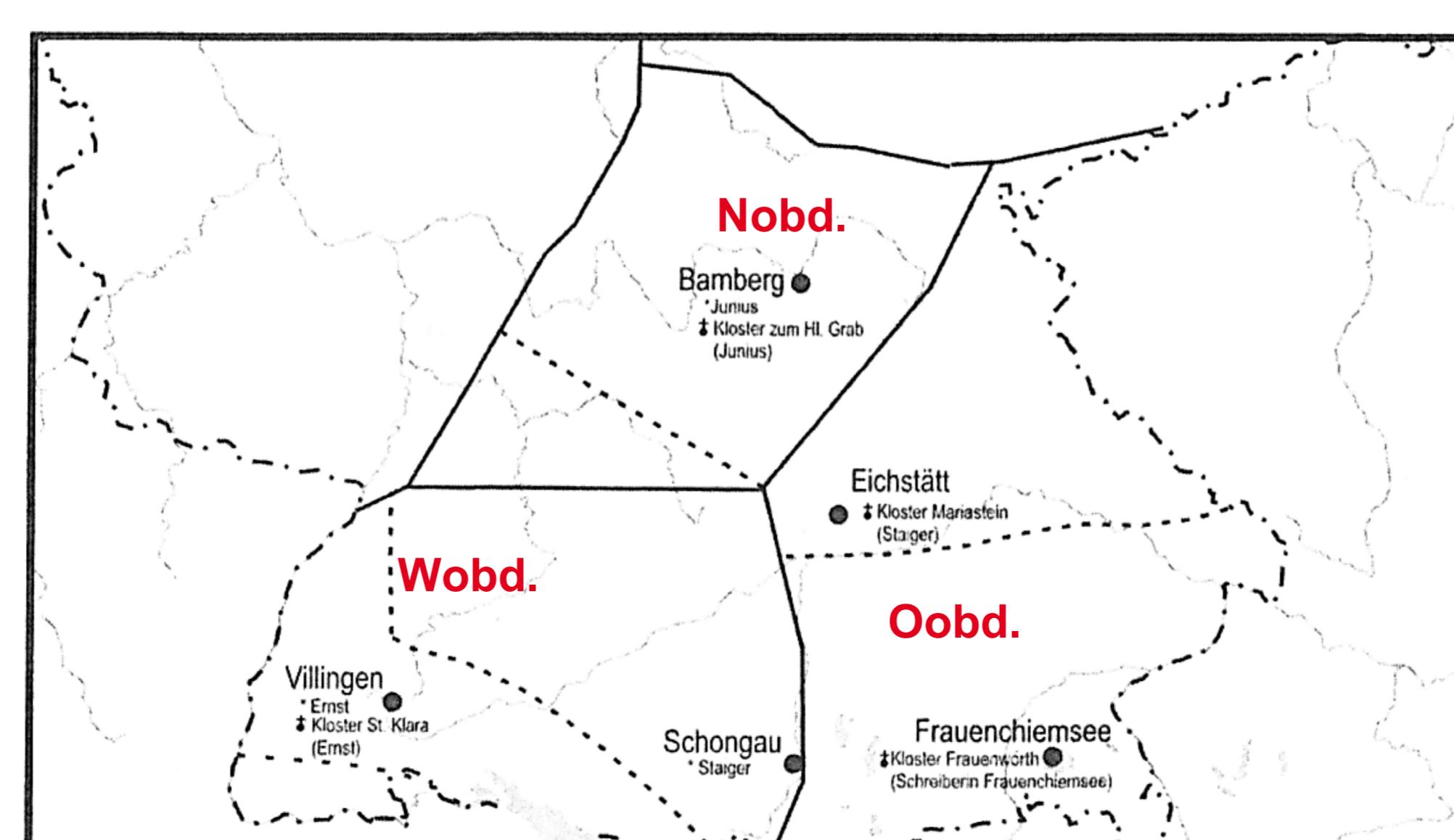


Abb. 3: Sprachgeographische Verteilung. Graphik aus Nolting (2010:55)

Forschungsperspektiven

Soziopragmatisch orientierte Sprachgeschichte

Bereits Nolting (2010) konnte zeigen, dass sich die Klosterfrauentexte hervorragend zur Untersuchung sprachlicher Phänomene aus soziopragmatischer Perspektive eignen. Ziel einer soziopragmatisch orientierten Sprachgeschichte ist die Erschließung der sprachhistorischen Wirklichkeit in ihrer ganzen Breite und Vielfalt unter Einbezug situativ-kommunikativer Variation. Zentrale Aspekte dabei sind beispielsweise die sprachliche Flexibilität der Schreiberinnen sowie Regionalität und Entregionalisierung auf unterschiedlichen Analyseebenen:

- lautlich-graphisch;
- morphologisch;
- syntaktisch-stilistisch;
- lexikalisch-stilistisch.

Desiderate

- Manuelle Annotation syntaktischer Phänomene sowie pragmatischer Marker (z.B. Höflichkeitsfloskeln nach Vorbild des Fürstinnenkorrespondenzkorpus)
- Korpuslinguistische Erschließung weiterer von Frauen verfasster Texte: Erweiterung des Korpus nach Zeit und Raum sowie um weitere Textsorten.

Dank

Wir danken dem Waxmann-Verlag für die freundliche Genehmigung, die Editionstexte zu verwenden und als Korpus öffentlich zugänglich zu machen. Für hilfreiche Kommentare und technische Unterstützung gilt außerdem Fabian Barteld (Hamburg/Bochum) unser besonderer Dank.



Mehr Informationen:

tinyurl.com/klosterfrauen

Literatur

- Cunningham, Hamish et al. (2002): GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of ACL*.
- Jurish, Bryan (2012): *Finite-state Canonicalization Techniques for Historical German*. PhD Thesis, University of Potsdam, urn:nbn:de:kobv:517-opus55789.
- Krause, Thomas & Amir Zeldes (2016): ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1), 118–139.
- Lühr, Rosemarie (2016): Satzkomplexität in fürstlichen Korrespondenzen der frühen Neuzeit. In Sarah Kwekkeboom & Sandra Waldenberger (eds.), *PerspektivWechsel*, Bd. 1, 455–478. Berlin.
- Lühr, Rosemarie et al. (eds.) (2018): *Genderspezifika in thüringischen Fürstinnenkorrespondenzen der Frühen Neuzeit*. Hamburg.
- Nolting, Uta (2009): Sprachgebrauch süddeutscher Klosterfrauen des 17. Jahrhunderts. Münster u.a.
- Zipser, Florian & Laurent Romary (2010): A model oriented approach to the mapping of annotation formats using standards. Workshop on Language Resource and Language Technology Standards, LREC 2010. La Valette, Malta. <https://hal.inria.fr/inria-00527799>.

Abb. 2: Benutzeroberfläche in ANNIS.