

58. Jahrestagung des Leibniz-Instituts für Deutsche Sprache vom 15. bis 17. März 2022  
(als Online-Konferenz)

## KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

### METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

#### **Das Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650) (ReN) als Impuls für die historische Grammatikforschung**

Sarah Ihden & Ingrid Schröder, Universität Hamburg

Das Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650) (ReN) erschließt eine Schriftkultur, die durch ihre prominenten Funktionen als internationale Verkehrssprache der Hanse im Nord- und Ostseeraum, als Rechts- und Verwaltungssprache in den Städten, als Fachsprache v.a. der Medizin, aber auch als Sprache religiöser Alltagspraxis im Mittelalter und in der frühen Neuzeit geprägt wird.

Das nach raum-zeitlichen Kriterien sowie Textgruppen („Feldern der Schriftlichkeit“) gegliederte Korpus umfasst in der Version 1.1 vom 06.01.21 (ReN-Team 2021) insgesamt 2,3 Millionen Token, von denen ca. 1,5 Millionen Token grammatisch annotiert sind, und ist über das Such- und Visualisierungstool ANNIS (Krause/Zeldes 2016) im Internet frei zugänglich. Es bietet für jeden Text Metadaten zum Entstehungskontext, zu weiteren quellenrelevanten Aspekten (z.B. Aufbewahrungsort des Originals, Editionen und weiterführende Forschungsliteratur) sowie zur Bearbeitung (z.B. Umfang des aufgenommenen Textes/Ausschnitts, Name der Bearbeiter:innen, Hinweise zum Vorgehen der Transkription und Annotation). Das ReN liefert auf der Textebene eine originalgetreue sowie eine hinsichtlich der grammatischen Worteinheiten normalisierte Transkription und enthält Annotationen zum Lemma, zur Wortart (PoS), zur Flexionsmorphologie sowie zu Satzeinheiten. Das für die grammatische Annotation genutzte Historische Niederdeutsch-Tagset HiNTS (Barteld et al. 2018) orientiert sich grundlegend am Historischen Tagset HiTS (Dipper et al. 2013), das in den anderen Referenzkorpora zur deutschen Sprachgeschichte im Verbund Deutsch Diachron Digital (Altddeutsch, Mittelhochdeutsch, Frühneuhochdeutsch und Deutsche Inschriften) verwendet wird. Die ähnliche Anlage der Referenzkorpora hinsichtlich der Struktur und der Annotationen stellt die Basis für sprachstufenübergreifende korpusbasierte Analysen zum älteren Deutschen dar.

Zu ReN-spezifischen Besonderheiten zählen drei nach Grad der Diplomtizität unterschiedene Textansichten in ANNIS und eine von ANNIS unabhängige Leseversion aller Texte im PDF-Format sowie die verschiedenen Darstellungsformen des Lemmas von der eindeutigen Form wie im Wörterbuch (MNWB 1956 ff.) bis zu einer um sämtliche Sonderzeichen reduzierten Form zum einfachen Einstieg in die lemmabasierte Suche. Ein besonderes Merkmal der flexionsmorphologischen Annotation stellt der Einsatz von Portmanteau-Tags (z.B. Dat-Akk für Dativ oder Akkusativ) dar, welche die genaue Abbildung formaler Ambiguitäten erlauben. Auf diese Weise sollen Interpretationen der Annotator:innen, welche die Analysen vorwegnehmen und so zu Zirkelschlüssen führen, vermieden werden.

Mit seiner hinsichtlich formaler Ambiguitäten umsichtigen Annotation stellt das ReN eine geeignete Datengrundlage für Analysen im Rahmen der Erarbeitung einer neuen mittelniederdeutschen Grammatik (DFG: SCHR 999/9-1) bereit. Gegenwärtig wird auf der Basis des

ReN die Flexion der Appellativa analysiert. Erste Auswertungen zeigen, dass die Struktur und die Annotationen im Korpus variationssensitive Untersuchungen ermöglichen, die neue Einblicke in die Substantivflexion im Mittelniederdeutschen, beispielsweise bezüglich diachroner Entwicklungen, liefern (vgl. Ihden/Schröder 2021).

Im Bereich der Lexembildung bietet das ReN gemeinsam mit den anderen historischen Referenzkorpora eine geeignete Basis für das geplante Projekt Wortfamilien diachron (WoDia) – Eine Forschungsumgebung zur historischen Wortbildung des Deutschen, das den alt- und mittelhochdeutschen, altsächsischen und mittelniederdeutschen Wortschatz anhand der Wortfamilienstruktur abbilden und miteinander verknüpfen soll. Über Strukturformeln werden die Lexembildungseinheiten sowie ihre hierarchischen Verbindungen dargestellt. In WoDia sind außerdem die Lemmata der Referenzkorpora mit denen der Zeitstufenwörterbücher verlinkt, sodass für Analysen die unterschiedlichen Vorteile dieser Ressourcen sinnvoll kombiniert werden können. Die Korpora belegen das Vorkommen bestimmter Lexembildungseinheiten und -arten in verschiedenen Schreibsprachlandschaften, Zeiträumen, Textsorten und Texten.

In der Posterpräsentation soll das ReN mit seinen Spezifika erläutert und die Anwendungsmöglichkeiten der historischen Korpora, insbes. des ReN, sollen anhand der zwei genannten korpusbasierten Projekte demonstriert werden. Zur Veranschaulichung sollen exemplarische Suchen im Korpus in ANNIS am Rechner präsentiert werden.

## **Literatur:**

Barteld, Fabian/Ihden, Sarah/Dreessen, Katharina/Schröder, Ingrid (2018): HiNTS: A Tagset for Middle Low German. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 3940–3945. Online unter: <http://www.lrec-conf.org/proceedings/lrec2018/pdf/870.pdf> (Stand: 15.11.2021).

Barteld, Fabian/Dreessen, Katharina/Ihden, Sarah/Schröder, Ingrid (2017): Das Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650) – Korpusdesign, Korpuserstellung und Korpusnutzung. In: Becker, Anja/Hausmann, Albrecht (Hrsg.): Mittelniederdeutsche Literatur. Mitteilungen des deutschen Germanistenverbandes 64/3, 226–241.

Dipper, Stefanie/Donhauser, Karin/Klein, Thomas/Linde, Sonja/Müller, Stefan/Wegera, Klaus-Peter (2013): HiTS: ein Tagset für historische Sprachstufen des Deutschen. In: Journal of Language Technology and Computational Linguistics 28/1, 85–137. Online unter: [https://www.linguistics.ruhr-uni-bochum.de/~dipper/pub/jlcl13\\_webVersion.pdf](https://www.linguistics.ruhr-uni-bochum.de/~dipper/pub/jlcl13_webVersion.pdf) (Stand: 15.11.2021).

Ihden, Sarah/Schröder, Ingrid (2021): Mittelniederdeutsche Grammatik: Konzeption und erste Analysen. In: Niederdeutsches Jahrbuch 144, 79–104.

Krause, Thomas/Zeldes, Amir (2016): ANNIS3: A new architecture for generic corpus query and visualization. In: Digital Scholarship in the Humanities 31, 118–139. Online unter: <http://dsh.oxfordjournals.org/content/31/1/118> (Stand: 15.11.2021).

MNWB (1956 ff.) = Lasch, Agathe/Borchling, Conrad/Cordes, Gerhard/Möhn, Dieter/Schröder, Ingrid (Hrsg.) (1956 ff.): Mittelniederdeutsches Handwörterbuch. Bd. 1 ff. Neumünster.

Korpus:

ReN-Team 2021: Reference Corpus Middle Low German/Low Rhenish (1200–1650); Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650). Version 1.1 vom 06.01.2021. <http://doi.org/10.25592/uhhfdm.9195> (Stand: 15.11.2021).