

Varianz, Ambiguität, Unsicherheit. Methodische Schlaglichter zur mittelniederdeutschen Grammatikographie

Ihden, Sarah

sarah.ihden@uni-hamburg.de
Universität Hamburg, Deutschland

Eine besondere Herausforderung in der Grammatikographie historischer Sprachstufen des Deutschen stellt der Umgang mit Varianz, Ambiguitäten und Unsicherheiten dar. Hinzu kommt die Gefahr, dass durch die den Analysen für die Grammatikschreibung zugrunde gelegten Daten, insbesondere grammatische Annotationen in Korpora, die Ergebnisse dieser Analysen gewissermaßen vorgeprägt sind. Diese Besonderheiten sind auch bei der geplanten Bearbeitung der Flexionsmorphologie als Teil einer neuen wissenschaftlichen mittelniederdeutschen Grammatik zu berücksichtigen. Im Vortrag sollen die Methoden und Grundsätze dieser neuen Grammatik vorgestellt werden, wobei ein Fokus auf der Variationssensitivität und dem Korpusbezug liegt. Zudem soll beschrieben und anhand erster Analysen veranschaulicht werden, wie in der Erforschung der mittelniederdeutschen Flexionsmorphologie dem potentiellen Risiko einer zirkulären Darstellung begegnet wird und wie auf der Basis von Daten, die möglichst oberflächenbezogen annotiert und in denen Ambiguitäten ausgezeichnet sind, flexionsmorphologische Variation ermittelt und vor dem Hintergrund potentieller außer- und innersprachlicher Parameter beschrieben werden kann.

Eine umfassende wissenschaftliche Grammatik des Mittelniederdeutschen, die modernen Ansprüchen genügt, stellt ein dringendes Forschungsdesiderat dar. Die gegenwärtig sowohl für die Forschung als auch die akademische Lehre herangezogenen Grammatiken von Colliander (1912), Lasch (1914, / ²1974 / Nachdruck 2011), Lübben (1882) und Sarauw (1921–1924) sind methodisch veraltet. Ihrer Entstehungszeit entsprechend folgen sie einem junggrammatischen Paradigma und liefern Darstellungen der mittelniederdeutschen Grammatik mit einem deutlichen Fokus auf der Laut- und Formenlehre. Andere Sprachebenen wie Satz und Text bleiben weitestgehend unberücksichtigt. Auch die unflektierbaren Wortarten werden, wenn überhaupt, nur am Rand betrachtet wie bei Lübben (1882: 120–132) und Sarauw (1924: 229–234).

In einer neuen wissenschaftlichen Grammatik des Mittelniederdeutschen sollen diese Lücken geschlossen

und dabei moderne Methoden der Grammatikschreibung herangezogen werden. Eine wesentliche methodische Anforderung liegt im Korpusbezug. Eine umfassende Beschreibung der tatsächlichen grammatischen Gegebenheiten im Mittelniederdeutschen benötigt ein umfangreiches empirisches Fundament. Hierfür werden die Daten des strukturierten und balancierten Referenzkorpus Mittelniederdeutsch / Niederrheinisch (1200–1650) (kurz: ReN) genutzt, das seit 2013 mit Unterstützung der DFG an den Universitäten Hamburg und Münster entstanden ist und dessen finale Korpusversion im September 2019 online veröffentlicht wird. Das ReN ist nach verschiedenen Zeit- und Sprachräumen sowie Feldern der Schriftlichkeit strukturiert (vgl. Barteld et al. 2017: 227f., Peters / Nagel 2014: 167–169). Zudem können die Texte des Korpus differenziert nach umfangreichen Metadaten, unter anderem zur Kommunikationssituation, zur äußeren Form oder zum Genre (z.B. Prosa vs. Vers), betrachtet werden.

Das nach verschiedenen Parametern strukturierte Korpus als Basis der Analysen ermöglicht die Umsetzung zweier weiterer methodischer Prinzipien der neuen Grammatik: Die Variationssensitivität und die diasystematische Differenziertheit, die in neueren grammatischen Darstellungen zunehmend als wesentliche Parameter erkannt worden sind. So spielt sprachliche Variation unter anderem in neueren grammatischen Studien zur Gegenwartssprache eine Rolle, bspw. in der Variantengrammatik des Standarddeutschen (Dürscheidt / Elspaß / Ziegler 2018) und in der Korpus-Grammatik des IDS (<http://www1.ids-mannheim.de/gra/projekte/korpusgrammatik.html?L=0>). Wie die Abbildung von zeitlich und räumlich sowie durch die Überlieferungsform bedingter Variation auf der Basis eines umfangreichen Korpus für eine historische Grammatik erfolgen kann, lässt sich anhand der Neuarbeitung der Mittelhochdeutschen Grammatik beobachten (vgl. Herbers 2014), von der die Bände zur Wortbildung (Klein u. a. 2009) sowie zur Flexionsmorphologie (Klein u. a. 2018) publiziert sind. Auch im Konzept der geplanten Neuarbeitung der Mittelniederdeutschen Grammatik stehen Korpusbezug und Variation im Mittelpunkt.

Die geplante Gesamtgrammatik des Mittelniederdeutschen soll mit Bezug auf die zu berücksichtigenden Sprachebenen (Graphemik, Phonologie, Lexemklassifizierung, Lexembildung und Flexion, Syntax und Text inklusive Pragmatik) in mehreren Schritten bearbeitet werden. In einem ersten Zugriff steht die Flexionsmorphologie im Fokus. Gerade für flexionsmorphologische Analysen bildet das ReN mit seinen Annotationen zu Wortart (PoS), Flexionsmorphologie und Lemma die ideale Basis. Von den insgesamt ca. 2,3 Mio. Token der finalen Korpusversion ReN 1.0 (<http://hdl.handle.net/11022/0000-0007-D829-8>) sind knapp 1,4 Mio. Token grammatisch annotiert. Damit ist eine unabdingbare Voraussetzung für eine korpusbezogene variationssensitive Grammatik erfüllt. Die Annotation der Wortarten und der Flexionsmorphologie ist mit dem im

ReN entwickelten HiNTS (Historisches-Niederdeutsch-Tagset) erfolgt, das auf dem HiTS (Historisches Tagset; vgl. Dipper et al. 2013) und dem STTS (Stuttgart-Tübingen-Tagset; vgl. Schiller et al. 1999) basiert. Zusätzlich wurde auf der Basis einer digitalen Lemmaliste für das Mittelniederdeutsche eine Lemmatisierung vorgenommen (vgl. Kleymann et al. 2015).

Eine zentrale Anforderung an die neue mittelniederdeutsche Grammatik, die bereits in den grammatischen Annotationen des ReN berücksichtigt wurde, stellt die Vermeidung von Vorgriffen und bestimmten Interpretationen, welche die Ergebnisse entscheidend beeinflussen, dar. Gerade in einer korpusbasierten Grammatikschreibung ist die Auseinandersetzung mit der Frage, inwieweit die genutzten Daten und deren Annotation das Ergebnis der Analyse vorprägen, von besonderer Bedeutung. So basiert bspw. die Festsetzung der PoS-Tags im Tagset auf Vorannahmen zur Differenzierung zwischen bestimmten Wortarten in der jeweiligen Sprache. Auch im Bereich der mittelniederdeutschen Flexionsmorphologie können durch Interpretationen der Annotatorinnen und Annotatoren, z.B. aufgrund vorhandener grammatischer Darstellungen oder durch Übertragungen aus dem Gegenwartsdeutschen, die Ergebnisse vorweggenommen werden. Gleichzeitig jedoch kann keine grammatische Annotation absolut frei von bestimmten Annahmen über die Grammatik der jeweiligen Sprache erfolgen. Wie gezeigt werden soll, wird im ReN daher eine Balance zwischen notwendigen und vermeidbaren Vorannahmen sowie zwischen Oberflächenbezogenheit und Einbeziehung des sprachlichen Kontextes angestrebt. So gilt im Bereich der Valenz von Verben beispielweise die Regel, dass für das Subjekt im Satz der Nominativ annotiert werden kann (auch wenn der Beleg rein formal z.B. nach Nominativ oder Akkusativ flektiert ist), sofern keine eindeutig davon abweichende Form vorliegt. Eine Unterscheidung zwischen Genitiv-, Dativ- und Akkusativobjekt hingegen kann lediglich auf der Basis einer eindeutigen Flexionsform erfolgen. Da bislang kein Valenzwörterbuch des Mittelniederdeutschen existiert, ist bei ambigen Formen eine Auflösung nicht möglich, ohne dadurch wiederum die Ergebnisse zur Valenz mittelniederdeutscher Verben vorwegzunehmen.

Im Vortrag soll erläutert werden, wie der beschriebenen Gefahr einer zirkulären Darstellung begegnet werden kann, um eine oberflächenbezogene grammatische Analyse zu ermöglichen. Im ReN wurde mit dem HiNTS die Auszeichnung von Ambiguitäten auf flexionsmorphologischer Ebene mithilfe von Portmanteau-Tags (Leech et al. 1994) vorgenommen (vgl. Barteld et al. 2014).¹ So erhält bspw. eine Form wie *mî* (Personalpronomen), die sowohl Dativ als auch Akkusativ repräsentieren kann und bei der aufgrund des vorliegenden Kontextes keine Disambiguierung möglich ist, das Tag „Dat-Akk“. Der Vorteil eines solchen Tags im Vergleich z.B. zum Asterisk, der für ambige Formen im STTS genutzt wird, besteht darin, dass die konkrete Überschneidung

– hier zwischen Dativ und Akkusativ – abgebildet wird und weitere Alternativen ausgeschlossen werden. Statt die bestehenden Ambiguitäten auf der Basis bestimmter Vorannahmen aufzulösen und auf diese Weise mit den Annotationen dieses Vorwissen zu bestätigen, werden mehrdeutige Formen als solche gekennzeichnet.

Bei einem solchen Vorgehen ist kritisch zu hinterfragen, ob durch die Vergabe unterspezifizierter Tags die Annotationsentscheidung unnötigerweise auf einen späteren Zeitpunkt verlagert wird. Sollten nämlich bestimmte zunächst ausgewiesene Ambiguitäten nachträglich aufgelöst werden, erfordert dies eine erneute Sichtung der Belege, was insgesamt einen Mehraufwand bedeutet. Diese spätere Analyse der Sprachdaten mit potentieller Neubewertung der grammatischen Annotation ist jedoch unabdingbar, um die oben erwähnte Zirkularität zu vermeiden. Kann bspw. auf Basis der ReN-Daten die Valenz eines Verbs im Mittelniederdeutschen (z.B. *rôpen*) mit Hilfe der Fälle nicht-ambiger Kasusannotation bei Objekten dieses Verbs (z.B. *he rep dat kint* Neut.Akk.S) eindeutig bestimmt werden, wäre bei Objekten mit ambiger Form (z.B. *he rep den sone* Masc.Dat-Akk.Sg) eine Auflösung der Ambiguität zugunsten des ermittelten Objektkasus möglich (z.B. *he rep den sone* [Masc.Akk.Sg]). Während der Annotation im Rahmen der Korpuserstellung sind diese Informationen noch nicht vorhanden, sodass eine Disambiguierung ohne Vorannahmen nicht möglich ist. Neben diese Fälle, in denen die Auszeichnung der Ambiguität im ersten Schritt notwendig ist, um die Ergebnisse nicht vorzuprägen, eventuell aber in einem zweiten Schritt anhand neuer korpusbasierter Erkenntnisse eine nachträgliche Disambiguierung erfolgt, treten solche Fälle, in denen auch zu einem späteren Zeitpunkt keinerlei Auflösung der Ambiguität möglich ist. Dies betrifft unter anderem die Ebene des Genus, wo formal bestehende Ambiguitäten (z.B. *dit is en spegel* Masc-Neut.Nom.Sg) ausschließlich durch den sprachlichen Kontext (z.B. durch einen eindeutig nach einem Genus markierten Determinierer) aufgelöst werden können, was bereits in der Annotation im ReN berücksichtigt wird. Diese Beispiele belegen, dass durch die Vergabe von Portmanteau-Tags auf flexionsmorphologischer Ebene Entscheidungen zugunsten eines Wertes vermieden werden, die entweder gar nicht oder zum Zeitpunkt der Korpuserstellung noch nicht getroffen werden können, und dass die Auszeichnung von Ambiguitäten für eine möglichst vorurteilsfreie Annotation der mittelniederdeutschen Sprachdaten zwingend notwendig ist.

Um die Anwendbarkeit dieses Verfahrens zu evaluieren und potentielle Vor- und Nachteile der Auszeichnung von Ambiguitäten durch Portmanteau-Tags im ReN gegenüber der Annotation mithilfe des Asteriks wie im STTS auszumachen, wurden Inter-Annotator-Agreement-Experimente durchgeführt (vgl. Barteld et al. 2018: 3943). Diese ergaben unter anderem, dass der Wert der Übereinstimmung zwischen den Annotierenden auf flexionsmorphologischer Ebene unter Verwendung des

HiNTS ähnlich hoch ausfällt, wie wenn statt der Portmanteau-Tags solche mit Asterisk gesetzt würden. Die vergleichsweise höhere Zahl an potentiellen Tags im HiNTS führt somit nicht zu einer geringeren Qualität der Annotation. Dabei bieten jedoch die Portmanteau-Tags den entscheidenden Vorteil, eine bestehende Ambiguität in konkreter Form abzubilden.

Für eine variationssensitive Analyse der mittelniederdeutschen Flexionsmorphologie können solche Auszeichnungen von Ambiguitäten auf unterschiedlichen Ebenen genutzt werden, bspw. um den Gebrauch von Substantiven in verschiedenen Genera zu untersuchen. Hierfür können in einem ersten Schritt all diejenigen Lemmata ermittelt werden, bei denen eine Genusambiguität annotiert ist. Anschließend kann für spezifische Lemmata geprüft werden, wo sie in einer eindeutigen Genusform vorkommen. Disambiguierung wird hier durch den sprachlichen Kontext erreicht, z.B. durch einen eindeutig nach einem Genus flektierten Determinierer. In Analysen exemplarisch betrachteter Substantive wie „lîf“ und „dêl“ wird eine auf der Ebene des Genus zum Teil sehr unterschiedlich stark ausgeprägte Variation sichtbar. Während bei „dêl“ der Anteil der als Maskulinum und der als Neutrum annotierten Belege annähernd ähnlich hoch ausfällt, dominieren bei „lîf“ deutlich die als Neutrum annotierten Belege. Zudem zeigen sich vereinzelt je nach Text und Sprachraum unterschiedliche Verteilungen.

Ein Beispiel für die erwähnte Ausbalancierung von Vorwissen einerseits und reiner Oberflächenbezogenheit andererseits findet sich bei der Rektion von Präpositionen. Ausgehend von den Angaben im Mittelniederdeutschen Handwörterbuch von Lasch et al. (1956ff.) wurde bei der Annotation eine Eingrenzung auf bestimmte Kasus vorgenommen, z.B. bei *in* auf den Dativ und den Akkusativ. Gleichzeitig aber wurde eine Einbeziehung des semantischen Kontextes, z.B. bei *in* die Annotation von Dativ bei lokaler und Akkusativ bei direktonaler Semantik, vermieden und stattdessen die Ambiguität mithilfe von Portmanteau-Tags abgebildet. Bei Belegen, die eine von den Angaben im Wörterbuch abweichende eindeutige Form aufweisen, wurde eine Annotation zugunsten der tatsächlich vorliegenden Form vorgenommen. Auf diese Weise kann z.B. an bestimmten Stellen der sich im Niederdeutschen vollziehende Kasussynekretismus der Substantive, bei denen Dativ und Akkusativ zu einem obliquen Kasus auf der Basis der Akkusativform zusammenfallen, beobachtet und untersucht werden. Erste Analysen, die im Vortrag vorgestellt werden sollen, liefern im ReN mehrere Belege, in denen auf die Präpositionen *nâ* und *tô*, die im Mittelniederdeutschen überwiegend den Dativ regieren, Nominalphrasen im Akkusativ als Teil der Präpositionalphrase folgen (*nâ*: 22 Belege, *tô*: 14 Belege). Hierbei zeigt sich eine deutliche Konzentration auf Texte des 16. und 17. Jahrhunderts. Dies stützt die Hypothese, dass im späteren Mittelniederdeutschen der Kasuszusammenfall einsetzt. Zudem zeigt sich ein starkes

Gewicht der Belege auf einer Quelle, der Seekarte von 1577, das durch den Inhalt des Textes, der zahlreiche Richtungsangaben enthält, zu erklären ist. Bei *nâ* fällt außerdem auf, dass es mit Akkusativ vereinzelt auch in früheren Texten aus dem niederrheinischen Sprachraum vorkommt, was auf eine gewisse diatopische Variation hindeutet.

Wie anhand der Ergebnisse erster Analysen gezeigt wird, kann dank der Oberflächenbasiertheit und der Auszeichnung von Ambiguitäten im ReN Variation erfasst und vor dem Hintergrund potentieller außer- und innersprachlicher Parameter beschrieben werden. Auf diese Weise leistet die geplante korpuslinguistisch basierte variationssensitive Grammatik einen entscheidenden Beitrag für die moderne mittelniederdeutsche Grammatikographie.

Fußnoten

1. Die Herausforderung ambiger bzw. unterspezifizierter Einheiten für linguistische Annotationen wurde zunächst vor allem für die semantische und syntaktische Ebene diskutiert (s. z.B. Bunt 2007, Kountz et al. 2007). Auch für die Auszeichnung von grammatischer Ambiguität in historischen Sprachdaten wurden bereits Vorschläge gemacht (s. z.B. Pauly et al. 2012, Dipper et al. 2013), die sich jedoch auf die Annotation von Wortarten und syntaktischen Strukturen konzentrieren und flexionsmorphologische Mehrdeutigkeiten nicht berühren.

Bibliographie

Barteld Fabian / Ihden, Sarah / Schröder, Ingrid / Zinsmeister, Heike (2014): “Annotating descriptively incomplete language phenomena”, in: *Proceedings of LAW VIII – The 8th Linguistic Annotation Workshop* 99–104 <http://www.aclweb.org/anthology/W14-4915> [letzter Zugriff 09. Dezember 2019].

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

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 (LREC 2018)* 3940–3945 <http://www.lrec-conf.org/proceedings/lrec2018/summaries/870.html> [letzter Zugriff 09. Dezember 2019].

Bunt, Harry (2007): “Semantic underspecification. Which technique for what purpose?” in: Bunt, Harry / Muskens, Reinhard (eds.): *Computing Meaning* Bd. 3 (=

Studies in Linguistics and Philosophy 83). Dordrecht: Springer 55–85.

Colliander, Elof (1912): *Mittelniederdeutsches Elementarbuch* (Photokopie der Druckfahnen des nicht zur Veröffentlichung gelangten Werkes, das mit § 365 abbricht). Heidelberg: Winter.

Dipper, Stefanie / Donhauser, Karin / Klein, Thomas / Linde, Sonja / Müller, Stefan / Wegera, Klaus-Peter (2013): „HiTS: ein Tagset für historische Sprachstufen des Deutschen“, in: *Journal for Language Technology and Computational Linguistics*, Special Issue, 28(1): 85–137 <https://jcl.org/content/2-allissues/9-Heft1-2013/5Dipper.pdf> [letzter Zugriff 09. Dezember 2019].

Dürscheid, Christa / Elspaß, Stephan / Ziegler, Arne (2018): *Variantengrammatik des Standarddeutschen*. Ein Online-Nachschlagewerk <http://mediawiki.ids-mannheim.de/VarGra/index.php/Start> [letzter Zugriff 09. Dezember 2019]

Herbers, Birgit (2014): „Prozessualität und Variabilität in der Grammatikographie des Mittelhochdeutschen“ in: Ágel, Vilmos / Gardt, Andreas (eds.): *Paradigmen der aktuellen Sprachgeschichtsforschung* (= Jahrbuch für germanistische Sprachgeschichte 5). Berlin, Boston: de Gruyter 135–149.

IDS (o.J.): Korpusgrammatik – grammatische Variation im standardsprachlichen und standardnahen Deutsch: <http://www1.ids-mannheim.de/gra/projekte/korpusgrammatik.html?L=0> [letzter Zugriff 09. Dezember 2019].

Klein, Thomas / Solms, Hans-Joachim / Wegera, Klaus-Peter (2009): *Mittelhochdeutsche Grammatik*. Teil III: Wortbildung. Tübingen: Niemeyer.

Klein, Thomas / Solms, Hans-Joachim / Wegera, Klaus-Peter (2018): *Mittelhochdeutsche Grammatik*. Teil II: Flexionsmorphologie. Berlin, Boston: de Gruyter.

Kleymann, Verena / Nagel, Norbert / Peters, Robert (2015): „Die digitale Lemmaliste für das Mittelniederdeutsche im DFG-Projekt ‚Referenzkorpus Mittelniederdeutsch / Niederrheinisch (1200–1650)‘“, in: *Korrespondenzblatt des Vereins für niederdeutsche Sprachforschung* 122/2: 95–100.

Kountz, Manuel / Heid, Ulrich / Eckart, Kerstin (2007): „A LAF/GrAF based Encoding Scheme for underspecified Representations of syntactic Annotations“, in: *Proceedings of the International Conference on Language Resources and Evaluation 2008* 2262–2269 http://www.lrec-conf.org/proceedings/lrec2008/pdf/569_paper.pdf [letzter Zugriff 09. Dezember 2019].

Lasch, Agathe (1914 / ²1974 / Nachdruck 2011): *Mittelniederdeutsche Grammatik* (= Sammlung kurzer Grammatiken germanischer Dialekte. A. Hauptreihe 9). Tübingen: Niemeyer.

Lasch, Agathe / Borchling, Conrad (1956ff.): *Mittelniederdeutsches Handwörterbuch*. Fortgef. von Gerhard Cordes und Dieter Möhn. Hg. von Ingrid Schröder. Neumünster / Kiel / Hamburg: Wachholtz.

Leech, Geoffrey / Garside, Roger / Bryant, Michael (1994): „CLAWS4: The tagging of the British National Corpus“, in: *Proceedings of the 15th International Conference on Computational Linguistics (COLING)* 1: 622–628.

Lübben, August (1882): *Mittelniederdeutsche Grammatik*. Nebst Chrestomathie und Glossar. Leipzig: Weigel.

Pauly, Dennis / Senyuk, Ulyana / Demske, Ulrike (2012): „Strukturelle Mehrdeutigkeit in frühneuhochdeutschen Texten“, in: *JLCL* 27(2): 65–82 <https://jcl.org/content/2-allissues/10-Heft2-2012/5Pauly.pdf> [letzter Zugriff 09. Dezember 2019].

Peters, Robert / Nagel, Norbert (2014): „Das digitale ‚Referenzkorpus Mittelniederdeutsch / Niederrheinisch (ReN)‘“ in: Ágel, Vilmos / Gardt, Andreas (eds.): *Paradigmen der aktuellen Sprachgeschichtsforschung* (= Jahrbuch für germanistische Sprachgeschichte 5). Berlin, Boston: de Gruyter 165–175.

ReN-Team (2019): Referenzkorpus Mittelniederdeutsch/ Niederrheinisch (1200–1650). Archiviert im Hamburger Zentrum für Sprachkorpora. Version 1.0. Publikationsdatum 14.08.2019. <http://hdl.handle.net/11022/0000-0007-D829-8> [letzter Zugriff 09. Dezember 2019].

Sarauw, Christian (1921): *Niederdeutsche Forschungen*. Bd. 1: Vergleichende Lautlehre der niederdeutschen Mundarten im Stammlande (= Historisk-filologische Meddelelser 5.1). Kopenhagen: Høst.

Sarauw, Christian (1924): *Niederdeutsche Forschungen*. Bd. 2: Die Flexionen der mittelniederdeutschen Sprache (= Historisk-filologische Meddelelser 10.1). Kopenhagen: Høst.

Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine (1999): „Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)“ <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> [letzter Zugriff 09. Dezember 2019].