

Automatisierter Abgleich des Lautstandes althochdeutscher Wortformen

Abstract

Um Texte einer Sprache automatisiert auf ihren möglichen Entstehungszeitraum und ihre dialektale Zugehörigkeit hin zu untersuchen, werden für jedes erwartete Graphem und jede Flexionsendung zunächst Entsprechungsregeln zwischen einer idealisierten Sprachform und den Sprachformen in Grammatiken beschriebener Zeit-Dialekt-Räume erfasst. Anschließend werden mithilfe eines Computerprogramms unter Anwendung dieser Regeln die belegten Wortformen mit ihren Entsprechungen in der idealisierten Sprachform abgeglichen und für jeden Text die Übereinstimmungsgrade mit den einzelnen Zeit-Dialekt-Räumen angegeben. Exemplarisch wird dieser Abgleich für eine althochdeutsche Wortform beschrieben und das Ergebnis der Analyse des zugehörigen Gesamttextes dargestellt.

1 Untersuchungsthema

Seit jeher verändern sich Sprachen im Laufe der zeitlichen Entwicklung. Sobald ihre Sprechergemeinschaften in verschiedene Gruppen zerfallen, die nicht mehr dauerhaft miteinander in Kontakt stehen, entwickeln sie zudem verschiedene Varietäten. Solange die Normierung einer Sprache nicht erfolgt ist, bleibt die textliche Überlieferung daher sprachlich uneinheitlich.

Auch innerhalb eines Textes können Schwankungen auftreten, etwa wenn Sprecher verschiedener Dialekte am selben Text arbeiten oder einen bestehenden Text korrigieren (vgl. etwa BRAUNE/REIFFENSTEIN 2004, § 3 und Anm. 1). Ein einzelner Autor kann ebenfalls verschiedenen dialektalen Einflüssen unterworfen sein oder die im Laufe seines Lebens erfolgte sprachliche Veränderung in seinen Niederschriften wiedergeben. Da vor der Erfindung des Buchdrucks Texte allein durch Abschrift vervielfältigt wurden, kam es schließlich auch seitens der Kopisten – bewusst oder unbewusst – zu sprachlichen Anpassungen bei dialektalen Formen bzw. infolge der zeitlichen Entwicklung.

Sind zu einem Teil der textlichen Überlieferung einer Sprache keine genaueren zeitlichen und örtlichen Angaben bekannt, erscheint es denkbar, diese automatisiert auf ihre Übereinstimmung mit verschiedenen Zeit-Dialekt-Räumen – also Zeitabschnitten mit Bezug auf die verschiedenen örtlichen Varietäten – zu untersuchen. Diese Untersuchung wird im Folgenden beschrieben. Voraussetzung dafür ist, dass Angaben zu den üblichen Entsprechungen der verschiedenen Phonem-Graphem-Entsprechungen (Lautverschriftungen, vgl. MITTMANN 2015b, 248) und der Flexionsendungen in den einzelnen Zeit-Dialekt-Räumen vorliegen.

2 Untersuchungsobjekt und Datengrundlage

Als Untersuchungssprache wird das Althochdeutsche gewählt, das ab etwa 750 überliefert ist und um 1050 ins Mittelhochdeutsche übergeht (vgl. PAUL ET AL. 2007, § E 7). Das althochdeutsche Textkorpus umfasst etwa 560.000 Wortformen und erscheint damit für die Untersuchung hinreichend groß. Ein Teil des Korpus sind umfangreiche Texte mit einheitlicher Sprachform und bekannter Überlieferungsgeschichte (etwa die Werke Notkers des Deutschen mit ca. 320.000 Wortformen)¹ Daneben umfasst es aber auch zahlreiche mittelgroße und kleinere Texte, deren Herkunft oft nicht genau bekannt ist, sodass sich die Angaben in den Grammatiken zu zeitlichen und dialektalen Unterschieden beim Laut- und Formenstand nicht auf sie stützen können. Ein Zirkelschluss ist für diese Texte also ausgeschlossen, und die erstgenannten können zur Überprüfung der Untersuchungsmethode dienen.

Das Althochdeutsche weist eine deutlich erkennbare zeitliche Entwicklung auf und lässt sich in fünf gut bezeugte Dialektgebiete (Alemannisch, Bairisch, Ostfränkisch, Rheinfränkisch und Mittelfränkisch, vgl. BRAUNE/REIFFENSTEIN 2004, §§ 4-6) gliedern. Zur Abgrenzung werden zudem drei westgermanische Nachbarsprachen (Altsächsisch, Altniederfränkisch und Langobardisch, vgl. BRAUNE/REIFFENSTEIN 2004, § 2 u. Anm. 1) bei der Untersuchung hinzugenommen.²

Die für die Untersuchung verwendeten Daten sind auf Grundlage des DFG-geförderten Projektes *Referenzkorpus Altdeutsch* erstellt worden. Das entstandene Korpus umfasst alle althochdeutschen und altsächsischen Texte und weist eine umfangreiche morphologische Annotation auf (vgl. www.deutschdiachrondigital.de, LINDE/MITTMANN 2013 sowie MITTMANN 2013).

3 Vorbereitung der Texte

Beleg	Kestirnis
Lemma	Gistirni
Übersetzung	Gestirn
Wortart Lemma	NA
Wortart Beleg	NA
Flexion Lemma	ja_Neut
Flexion Beleg 1	ja_Neut
Flexion Beleg 2	Sg_Gen

Abbildung 1: Korpusauszug aus der *St. Galler Schularbeit* (vereinfacht)

Ausgehend von den im Korpus angegebenen Lemmata und morphologischen Angaben werden mithilfe eines Computerprogramms zunächst idealisierte Wortformen gebildet, die den für die Untersuchung der zeitlichen und dialektalen Zuordnung benötigten Abgleich mit den belegten Wortformen ermöglichen (vgl. MITTMANN 2015a, 68-74). Dazu werden je nach morphologischen Werten Endungen angefügt oder ersetzt; bei Umlaut oder starken Verben auch Vokalersetzungen im Inlaut durchgeführt. Sowohl das für die Lemmatisierung der althochdeutschen Texte verwendete Wörterbuch (SPLETT 1993) als auch die zur Ermittlung der Flexionsformen verwendete Referenzgrammatik (BRAUNE/REIFFENSTEIN 2004) orientieren sich am ostfränkischen Dialekt um 830, sodass auch die idealisierten Wortformen den größten Teil der im ältesten Althochdeutschen noch bewahrten lautlichen Unterschiede wiedergeben (vgl. MITTMANN 2015b, 249).

Abbildung 1 zeigt eine Wortform aus dem *Referenzkorpus* mitsamt einigen Annotationszeilen. Um die idealisierte Wortform zu erzeugen, wird die Ersetzungsregel für die Angaben „N³, ja,

{*Masc/Neut*}, *Sg, Gen*“ (Genitiv Singular eines maskulinen oder neutralen *ja*-stämmigen Substantivs) – also die grau hinterlegten Informationen – auf das Lemma angewendet. Das auslautende *-i* wird somit durch *-ies* ersetzt, sodass eine idealisierte Wortform *gistirnies* entsteht, die dann mit dem belegten *kestirnis* verglichen werden kann.⁴

4 Vorbereitung der Untersuchung

Vor der Durchführung der Untersuchung müssen noch Entsprechungsregeln zwischen den idealisierten und den belegten Wortformen mit Bezug auf die jeweiligen Zeit-Dialekt-Räume erstellt werden. Dafür wird zunächst in Anlehnung an die Angaben bei BRAUNE/REIFFENSTEIN (2004) die Zeit von 750 bis 1150 in acht 50-Jahres-Abschnitte unterteilt, damit auch sprachliche Entwicklungen zum Mittelhochdeutschen (ca. 1050–1350) hin erfasst sind. Auf diese Weise ergeben sich in Kombination mit den zusammen acht Dialekten und Nachbarsprachen 64 Zeit-Dialekt-Räume, etwa „*Alemannisch, 900–950*“.

Anschließend werden auf Grundlage von Referenzgrammatiken für die althochdeutschen (BRAUNE/REIFFENSTEIN 2004)⁵ und frühmittelhochdeutschen Dialekte (PAUL ET AL. 2007) sowie die Nachbarsprachen – GALLÉE/TIEFENBACH (1993) fürs Altsächsische, BREMMER/QUAK (1992) fürs Altniederfränkische/Altniederländische und BRUCKNER (1895) fürs Langobardische – Entsprechungsregeln erstellt. Die einzelnen Phonem-Graph-Entsprechungen werden dabei nach ihrer Stellung im Wort unterschieden, um der jeweils unterschiedlichen Entwicklung Rechnung zu tragen: Konsonanten nach Anlaut, Inlaut und Auslaut; Vokale (und Diphthonge) nach Präfix, Tonsilbe, Mittelsilbe und Endsilbe.⁶ Darüber hinaus ist in zahlreichen Fällen auch die Umgebung der jeweiligen Phoneme entscheidend, wie Abbildung 2 zeigt:

g	> {g k c} / #_i	für „Alemannisch oder Bairisch, 750–1050“
g	> {g h j Ø} / #_i	für „Altsächsisch“ (ohne zeitliche Eingrenzung)
i	> e / _rC (Tonsilbe)	für „Altniederfränkisch oder Altsächsisch“

Abbildung 2: Beispiele für Entsprechungsregeln für spezifische Zeit-Dialekt-Räume

Neben den Phonem-Graph-Entsprechungen werden für die flektierenden Wortarten auch Flexionsendungen mit gleicher Funktion, aber über den unterschiedlichen Lautstand hinaus auch dialektal unterschiedlicher morphologischer Bildweise, betrachtet. So lässt sich etwa beim Nominativ Plural der *a*-stämmigen maskulinen Substantive die altalemannische Endung *-ā* auf gemeingermanisches rekonstruiertes **-ōz* und schließlich auf urindogermanisches **-oes* zurückführen, die altsächsische Endung *-os* dagegen auf gemeingermanisches **-osiz* und schließlich auf urindogermanisches **-óeses* (vgl. KROGH 1996, 295 i. V. m. BAMMESBERGER 1990, 43 f.).⁷ Bei der althochdeutschen Endungsvariante *-a* wiederum liegt wohl eine Übernahme der gleichlautenden Akkusativendung vor (vgl. KROGH ebd.).

Auf diese Weise ergeben sich 707 Entsprechungsregeln: 203 für Flexionsendungen (plus 21 Regeln für Stammallomorphie) sowie 483 für Phonem-Graph-Entsprechungen (166 für Konsonanten und 317 für Vokale).

5 Durchführung der Untersuchung

Für jeden Text wird eine Wort-für-Wort-Prüfung durchgeführt, indem die Entsprechungsregeln auf dessen einzelne Phonem-Graph-Entsprechungen und Flexionsendungen angewendet werden. Dabei

wird berücksichtigt, dass das Korpus nur wortweise aligniert ist und oft keine 1:1-Zuordnungen der Phonem-Graph-Entsprechungen zueinander vorliegen. Im Althochdeutschen gilt, von Ausnahmen abgesehen, Anfangsbetonung, und aufgrund der schon in ältester Zeit (vgl. BRAUNE/REIFFENSTEIN 2004, § 54) beginnenden Vokalreduktion in den unbetonten Silben – bis hin zum Schwund – sowie des Auftretens epenthetischer Vokale und Konsonanten (vgl. ebd., § 69) ist vor allem in den Mittelsilben mit mangelnden Entsprechungen zu rechnen. Um möglichst viele automatisierte Lautzuordnungen zu ermöglichen, erfolgt die Prüfung daher von beiden Wortenden aus zur Mitte hin. Da Komposita und Präfixe im Korpus nicht markiert sind, wird jedoch auch in Mittel- und Endsilben geprüft, ob ein betonter Vokal vorliegt. Zudem wird die Möglichkeit von Doppel- und Einfachschreibung stets mitberücksichtigt.

Sofern das Wort flektiert, wird zunächst die Endung ermittelt und mit den in den einzelnen Zeit-Dialekt-Räumen bezeugten Formen abgeglichen. Unterschiede zwischen verschiedenen Endungsformen mit dem gleichen zugrundeliegenden Lautstand werden an dieser Stelle nicht berücksichtigt. Unabhängig davon erfolgt dann der Abgleich eines eventuellen auslautenden (sowie eventueller diesem vorangehenden inlautender) Konsonanten, danach in mehrsilbigen Wörtern des Endsilbenvokals und des diesem vorangehenden inlautenden Konsonanten. Ist die Endung hiervon noch nicht vollständig abgedeckt, werden noch weitere Mittelsilbenvokale und inlautende Konsonanten abgeglichen. Anschließend wird der Abgleich am Wortanfang mit einem eventuellen anlautenden (sowie eventuellen diesem folgenden inlautenden) Konsonanten und dem Tonsilbenvokal fortgesetzt, gefolgt von Mittelsilbenvokalen und inlautenden Konsonanten. Falls der Lautstand den Beginn mit einem unbetonten Präfix zulassen könnte, wird diese Möglichkeit für den ersten Vokal mit einbezogen.

Findet sich keine passende Entsprechungsregel, bricht das Programm die Untersuchung der Wortform vom Ende zur Mitte hin bzw. vom Anfang zur Mitte hin ab. Erfolgt der Abbruch unmittelbar im Auslaut oder Anlaut, wird die Prüfung also ausschließlich in die jeweils andere Richtung durchgeführt. Passen sowohl die Endung als auch Auslaut und Anlaut nicht, wird die Wortform vollständig übersprungen, da in diesem Fall von einer Fehlzuzuweisung auszugehen ist, die bei einem Korpus dieser Größe nicht ausgeschlossen werden kann.

Um den Übereinstimmungsgrad eines Textes mit den einzelnen Zeit-Dialekt-Räumen zu ermitteln, wird dieser mit jeder Regelanwendung für alle 64 Zeit-Dialekt-Räume – jeweils beginnend bei 0 – erhöht oder gesenkt. Da eine nur in wenigen Zeit-Dialekt-Räumen verbreitete Lautform für die Zuordnung eines Textes als deutlich signifikanter gelten kann als eine, die in fast allen Zeit-Dialekt-Räumen vorkommt, wird jedes Mal die Gesamtzahl der Zeit-Dialekt-Räume (64) durch die Zahl der (nicht) zutreffenden Zeit-Dialekt-Räume geteilt und 1 davon abgezogen: So ergibt sich etwa eine Veränderung von ± 0 , wenn alle Zeit-Dialekt-Räume zutreffen ($64/64 - 1$), und eine Veränderung von $+ 4,3$ für die in Abbildung 2 zuoberst genannte Regel, die auf zwölf Zeit-Dialekt-Räume zutrifft ($64/12 - 1$). Der ermittelte absolute Übereinstimmungsgrad jedes einzelnen Zeit-Dialekt-Raums wird schließlich prozentual auf den Übereinstimmungsgrad eines fiktiven („idealen“) Zeit-Dialekt-Raums bezogen, der dem jeweiligen Text exakt entspricht, sodass sich ein relativer Übereinstimmungsgrad ergibt.

Automatisierter Abgleich des Lautstandes althochdeutscher Wortformen

Graphem(e)	Idealisierte Wortform	Belegte Wortform	Angewandte Regel	für „Alemannisch, 900–950“			
				Zutreffend?	Gleiche Fälle	Übereinstimmungsgrad	Kumulierte Summe
Endung	g ist ir nies	kestir n is	es → is {Sg, Gen}	-	41	- 0,56	- 0,56
C final	g ist ir nies	kestir n ig	s = s	+	64	± 0,00	- 0,56
V Endsilbe	g ist ir nies	kestir n is	e > i	-	49	- 0,31	- 1,78
prä vokalisches <i>j</i>	g ist ir n ies	kestir n is	j > ∅ / C_V	+	64	± 0,00	- 0,87
C vor V	g ist ir nies	kestir n is	n = n / C_	+	64	± 0,00	- 0,87
C initial	g ist ir nies	k estir n is	g > k / _i	+	12	+ 4,33	+ 3,46
Präfix-V	g ist ir nies	k estir n is	i > e / g_	-	33	- 0,94	+ 2,52
C medial	g ist ir nies	ke s tir n is	s = s	+	64	± 0,00	+ 2,52
C medial	g ist ir nies	ke s tir n is	t = t / s_	+	64	± 0,00	+ 2,52
V Tonsilbe	g ist ir nies	ke s tir n is	i = i / _rC	+	48	+ 0,33	+ 2,85
C medial	g ist ir nies	ke s tir n is	r = r / _C	+	64	± 0,00	+ 2,85

Abbildung 3: Ermittlung des absoluten Übereinstimmungsgrades einer Wortform mit einem Zeit-Dialekt-Raum

6 Beispielfall (Einzelwort)

Abbildung 3 zeigt das Vorgehen des Computerprogramms für die bereits in Abbildung 1 dargestellte Wortform *kestirnis* aus der *St. Galler Schularbeit*. Die rechten vier Tabellenspalten enthalten dabei die Veränderungen der absoluten Übereinstimmungsgrade für den Zeit-Dialekt-Raum „*Alemannisch, 900–950*“. Die Spalte „Zutreffend?“ gibt an, ob nach den Angaben bei BRAUNE/REIFFENSTEIN (2004) die Entsprechungsregel, die das Programm durch Abgleich von idealisierter und belegter Wortform als anzuwenden erkennt, auf diesen Zeit-Dialekt-Raum zutrifft oder nicht. Die Spalte „Gleiche Fälle“ nennt die Anzahl an Zeit-Dialekt-Räumen, die sich ebenso verhalten wie „*Alemannisch, 900–950*“, sodass anhand dieses Wertes die Veränderung des Übereinstimmungsgrades ermittelt werden kann.

Bestünde der Text nur aus diesem einen Wort, würde abschließend das Ergebnis von + 2,63 in Bezug zu dem Ergebnis für den „idealen“ Zeit-Dialekt-Raum gesetzt, das für dieses Wort + 10,77 (= 100 %) beträgt. Der relative Übereinstimmungsgrad des Wortes mit dem Zeit-Dialekt-Raum „*Alemannisch, 900–950*“ beläuft sich nach dieser Methode also auf etwa 26,5 %.

7 Auswertung eines vollständigen Textes

Nach Abgleich aller 104 Wörter der *St. Galler Schularbeit* (Text bei STEINMEYER 1916, 121) mit den einzelnen Zeit-Dialekt-Räumen zeigt Abbildung 4 die Auswertung für den Gesamttext:⁸

	vor 800	800 – 850	850 – 900	900 – 950	950 – 1000	1000 – 1050	1050 – 1100	nach 1100
Langobardisch	-65	-65	-65	-65	-65 ⁹	-65	-65	-65
Alemannisch	-26	-12	+17	+22	+53	+78	+47	+47
Bairisch	-25	-21	+14	+14	+16	+41	+9	+9
Ostfränkisch	-51	-24	+10	+10	+12	+39	+10	+10
Rheinfränkisch	-52	-28	+6	+6	+8	+35	+6	+8
Mittelfränkisch	-60	-55	-19	+3	+3	+29	±0	+2
Niederfränkisch	-58	-47	-45	-46	-46	-45	-35	-13
Altsächsisch	-35	-35	-35	-35	-35	-35	-35	-35

Abbildung 4: Relative Übereinstimmungsgrade der *St. Galler Schularbeit* mit den Zeit-Dialekt-Räumen (gerundet, %)

Die mit Abstand höchste Übereinstimmung, ca. 78 %, wird also für „*Alemannisch, 1000–1050*“ ermittelt, gestützt durch die nächstbesten Werte in den benachbarten Zeitabschnitten sowie Dialekten des Althochdeutschen. Die Annahme SONDEREGGERS (1980, Sp. 1049), die *St. Galler Schularbeit* stamme aus der ersten Hälfte des 11. Jahrhunderts, wird somit bestätigt.

8 Ausblick

Erst nach Analyse sämtlicher althochdeutscher Texte wird eine Aussage dazu möglich sein, inwiefern die beschriebene Untersuchungsmethode tatsächlich dazu geeignet ist, neue Erkenntnisse zu Entstehungszeiten und Schreibdialekten zu liefern. Dafür muss zunächst aber die Analyse der umfangreicheren Texte zeigen, ob die Angaben aus den Grammatiken in Summe einen plausiblen

Regelapparat bilden: Das ist der Fall, wenn der Forschungsstand zur zeitlich-dialektalen Zuordnung dieser Texte einerseits den Zeit-Dialekt-Räumen mit den bei Analyse jeweils höchsten Übereinstimmungsgraden andererseits entspricht. Dann ließe sich die Untersuchung auf die kleineren Texte mit weniger bekannter Überlieferungsgeschichte übertragen und sich so neue Erkenntnisse zu ihrem zeitlich-dialektalen Sprachstand gewinnen. Dies habe ich im Rahmen meiner Dissertation unter-
nommen.

¹ Vgl. etwa BRAUNE/REIFFENSTEIN (2004, XII) zu den für die Erstellung der Grammatik verwendeten „hauptquellen“ und den „nur soweit [...] als nötig“ zugezogenen Quellen.

² Zwar sind aufgrund der Überlieferungsform des Langobardischen die meisten Flexionsendungen dort nicht überliefert (vgl. BRUCKNER 1895, § 98), diese Fälle werden in der Untersuchung jedoch übersprungen und nicht als unzutreffende Übereinstimmung gewertet.

³ Der zweite Teil der Angabe „NA“ steht für „Appellativum“ und ist für die Flexion irrelevant.

⁴ Bei den vokalischen Substantivklassen wird stets die in BRAUNE/REIFFENSTEIN (2004) aufgeführte Form mit erhaltenem /j/ (dort als Ⱬ geschrieben) verwendet, um den Prozess des Schwindens von /j/ nach Konsonant und den damit verbundenen Zusammenfall von Flexionsklassen (vgl. etwa ebd., § 210) deutlicher verfolgen zu können.

⁵ Da für Altalemannische bis heute eine eigene Dialektgrammatik fehlt (vgl. BRAUNE/REIFFENSTEIN 2004, XII f.), kann als Referenzgrammatik für das Althochdeutsche nur eine Grammatik dienen, die alle Varietäten der Sprachstufe beschreibt.

⁶ Ein Ausschnitt aus der Zuordnungstabelle für die konsonantischen Anlaute findet sich bei MITTMANN (2015b, 253).

⁷ Das Doppel-Makron zeigt jeweils Überlänge des Vokals an.

⁸ Nicht negative Werte sind von 0 ausgehend mit sich in 10-Prozentpunkt-Intervallen verdunkelnden Graustufen hinterlegt.

⁹ Die letzte Änderung des langobardischen Lautstandes ist bei BRUCKNER (1895) für „[i]m Laufe des 9. Jhs.“ (§ 48) angegeben und darf somit für um 900 abgeschlossen angenommen werden. Die Angaben zum Langobardischen nach 950 erscheinen daher nur aus systematischen Gründen.

Literatur

- BAMMESBERGER, A. (1990). Die Morphologie des urgermanischen Nomens (= Untersuchungen zur vergleichenden Grammatik der germanischen Sprachen, Band 2). Heidelberg: Winter.
- BRAUNE, W. / REIFFENSTEIN, I.-(2004). Althochdeutsche Grammatik. Band 1: Laut- und Formenlehre (= Sammlung kurzer Grammatiken germanischer Dialekte, A. Hauptreihe Nr. 5). 15. Auflage, bearbeitet von I. Reiffenstein. Tübingen: Niemeyer.
- BREMMER, R. H. JR. / QUAK, A. (Hg.) (1992). Zur Phonologie und Morphologie des Altniederländischen (= North-Western European Language Evolution (NOWELE), Supplement Volume 6). Odense: Odense University Press.
- BRUCKNER, W. (1895). Die Sprache der Langobarden (= Quellen und Forschungen zur Sprach- und Culturgeschichte der germanischen Völker, Bd. LXXV). Straßburg: Trübner.
- GALLÉE, J. H. / TIEFENBACH, H. (1993). Altsächsische Grammatik (= Sammlung kurzer Grammatiken germanischer Dialekte, A. Hauptreihe Nr. 6). 3. Auflage mit Berichtigungen und Literaturnachträgen von H. Tiefenbach. Tübingen: Niemeyer.

- KROGH, S. (1996). Die Stellung des Altsächsischen im Rahmen der germanischen Sprachen (= Studien zum Althochdeutschen, Band 29). Göttingen: Vandenhoeck & Ruprecht.
- LINDE, S. / MITTMANN, R. (2013). „Old German Reference Corpus. Digitizing the knowledge of the 19th century. Automated pre-annotation using digitized historical glossaries.“ In: Bennett, P. et al. (Hg.) (2013). *New Methods in Historical Corpora* (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache / *Corpus Linguistics and Interdisciplinary Perspectives on Language – CLIP*, Band 3). Tübingen: Narr, 235-246.
- MITTMANN, R. (2013). „Old German and Old Lithuanian: the Creation of Two Deeply-Annotated Historical Text Corpora.“ In: Коллектив авторов (Ответственные редакторы: Захаров, В. и др.) [Autorenkollektiv (Verantwortliche Redakteure: Zakharov, V. et al.)] (2013). Труды международной научной конференции «Корпусная лингвистика – 2013» / *Proceedings of the international conference «Corpus linguistics – 2013»*. Санкт-Петербург: Санкт-Петербургский государственный университет, Филологический факультет / St. Petersburg State University, Philological Faculty, 103-III.
- MITTMANN, R. (2015a). „Automated quality control for the morphological annotation of the Old High German text corpus. Checking the manually adapted data using standardized inflectional forms.“ In: Gippert, J. / Gehrke, R. (Hg.) (2015). *Historical Corpora. Challenges and Perspectives. Proceedings of the conference Historical Corpora 2012* (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache / *Corpus Linguistics and Interdisciplinary Perspectives on Language – CLIP*, Band 5). Tübingen: Narr, 65-76.
- MITTMANN, R. (2015b). „Automatisierte Zeit- und Dialektzuordnung althochdeutscher Texte.“ In: Oettinger, N. et al. (Hg.) (2015). *Münchener Studien zur Sprachwissenschaft – MSS*, Heft 69/2 – 2015. Dettelbach: Röhl, 245-256.
- PAUL, H. ET AL. (2007). *Mittelhochdeutsche Grammatik* (= Sammlung kurzer Grammatiken germanischer Dialekte, A. Hauptreihe Nr. 2). 25. Auflage, neu bearbeitet von Th. Klein et al. Mit einer Syntax von I. Schöbler, neubearbeitet und erweitert von H.-P. Prell. Tübingen: Niemeyer.
- SONDEREGGER, S. (1980). Art. „‘St. Galler Schularbeit’.“ In: Ruh, K. et al. (1978–2008). *Die deutsche Literatur des Mittelalters: Verfasserlexikon. Begründet von Wolfgang Stammerl, fortgeführt von Karl Langosch. 2., völlig neu bearbeitete Auflage unter Mitarbeit zahlreicher Fachgelehrter. Band II*. Berlin; New York: de Gruyter, Sp. 1049-1051.
- SPLETT, J. (1993). *Althochdeutsches Wörterbuch. Analyse der Wortfamilienstrukturen des Althochdeutschen, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes. 3 Bände*. Berlin: de Gruyter.
- STEINMEYER, E. V. (1916). *Die kleineren althochdeutschen Sprachdenkmäler*. Berlin: Weidmann.