

Roland Mittmann und Ralf Plate

Das ‚Referenzkorpus Altdeutsch‘ als Lesekorpus. Grammatisch annotierte und mit Wörterbüchern verknüpfte Texte für Lehre und Selbststudium

<https://doi.org/10.1515/mial-2019-0012>

Abstract: Working with Old High German and Old Saxon texts has become rare in German language and literature studies. Nowadays, the brief introductions to historical linguistics and philology do not suffice to enable students to explore the oldest German texts in significant depth. A new tool could at least help move closer to this dept of textual exploration: an on-line publication of all Old High German and Old Saxon texts with a morphological word-by-word annotation.

The data has been collected by the research project ‘Old German Reference Corpus’ (‘Referenzkorpus Altdeutsch’ / ReA) and can already be used for complex database queries. Within the framework of the ‘eHumanities Centre for Historical Lexicography’ (eHumanities-Zentrum für Historische Lexikographie / ZHistLex), it has been converted into a clearly arranged format, focusing on the information needed for comprehending the texts: the ‘Lesekorpus Altdeutsch’ (LeA).

This tool can of course not replace the work with text editions, translations and commentaries, but will help pave the way from the first insight into the language to a profound understanding. Additionally, a supplemental both-way linking between every word token and the corresponding entry in the on-line dictionaries of historical German is being prepared.

Keywords: Old High German, Old Saxon, Corpus Linguistics, Lexicography, Teaching

Kontakt: Dr. Roland Mittmann, Johann Wolfgang Goethe-Universität,
Institut für Empirische Sprachwissenschaft, Postfach 11 19 32, Fach 171,
60054 Frankfurt am Main, E-Mail: mittmann@em.uni-frankfurt.de
Dr. Ralf Plate, Universität Trier, Mittelhochdeutsches Wörterbuch, 54286 Trier,
E-Mail: plate@uni-trier.de

I

Die Arbeit mit althochdeutschen (und altsächsischen) Texten ist aus den germanistischen Studiengängen noch nicht ganz verschwunden, aber stark vom Wegfall bedroht, sei es in der literaturgeschichtlichen Mediävistik, sei es in der historischen Linguistik. Wo das Althochdeutsche überhaupt berücksichtigt wird, begnügen sich die Einführungen in die deutsche Sprachgeschichte und historische Grammatik mit wenigen ganz kurzen Textauszügen; die literaturgeschichtlichen Anthologien bieten zwar meist eine Übersetzung, verzichten aber auf eingehende sprachliche Kommentierung.¹ Auf diese Weise kann die selbstständige Sicherung des sprachlichen Verständnisses mit der Hilfe von Grammatik, Wörterbuch und (wo vorhanden) wissenschaftlichen Textkommentaren nicht eingeübt werden, die Schwelle zum Einstieg in die sprachliche Textarbeit, z.B. mit dem ‚Althochdeutschen Lesebuch‘ (das auch ein sehr zweckmäßig gearbeitetes Wörterbuch mit Hinweisen auf die Grammatik enthält) und dem Abriss der althochdeutschen Grammatik,² bleibt auch nach einer guten Einführung sehr hoch. Vorgestellt wird hier ein einfaches Hilfsmittel, das geeignet sein könnte, diese Schwelle wenigstens leichter überschreitbar zu machen; erste Erfahrungen damit in Einführungskursen zum Althochdeutschen und Altsächsischen am Institut für Empirische Sprachwissenschaft der Goethe-Universität Frankfurt am Main sind ermutigend.

Es handelt sich um die Bereitstellung der gesamten althochdeutschen und altsächsischen Literatur nach den maßgeblichen wissenschaftlichen Editionen zusammen mit der sprachlichen (morphologisch-syntaktischen) Wort-für-Wort-Annotation, die für das ‚Referenzkorpus Altdeutsch‘ (ReA) erarbeitet worden ist. Die Texte selbst sind ohne Annotation seit 1997 sukzessive in die TITUS-Daten-

¹ Vgl. etwa: Rolf Bergmann, Claudine Moulin u. Nikolaus Ruge, *Alt- und Mittelhochdeutsch. Arbeitsbuch zur Grammatik der älteren deutschen Sprachstufen und zur deutschen Sprachgeschichte*, 9., korr. Aufl. Göttingen 2016; Hans-Ulrich Schmid, *Einführung in die deutsche Sprachgeschichte*, 3., aktual. und überarb. Aufl. Stuttgart 2017; Karin Donhauser, Annette Fischer u. Lars Mecklenburg, *Moutons interaktive Einführung in die historische Linguistik des Deutschen* (auf CD). Berlin, New York 2007; Horst Dieter Schlosser (Hg.), *Althochdeutsche Literatur. Mit Proben aus dem Altniederdeutschen*, Neuausgabe. Frankfurt a.M. 1980, S. 116–120; Walter Haug u. Benedikt Konrad Vollmann (Hgg.), *Frühe deutsche Literatur und lateinische Literatur in Deutschland 800–1150* (Bibliothek des Mittelalters 1). Frankfurt a.M. 1991; Horst-Dieter Schlosser (Hg.), *Althochdeutsche Literatur. Eine Textauswahl mit Übertragungen*. Berlin 1998; Stephan Müller (Hg.), *Althochdeutsche Literatur. Eine kommentierte Anthologie, Althochdeutsch/Neuhochdeutsch* (RUB 18491). Stuttgart 2007.

² Wilhelm Braune (Hg.), *Althochdeutsches Lesebuch*. 17. Aufl., bearb. v. Ernst A. Ebbinghaus. Tübingen 1994; Ders., *Abriss der Althochdeutschen Grammatik mit Berücksichtigung des Altsächsischen*. 15., verb. Aufl., bearb. v. Ernst A. Ebbinghaus. Tübingen 1989.

bank der Universität Frankfurt eingestellt worden (titus.uni-frankfurt.de), zusammen mit der ReA-Annotation sind sie seit 2010 in der Datenbank ANNIS für professionelle Recherchen mit einer anspruchsvollen Abfragesprache zugänglich (<http://www.deutschdiachrondigital.de>). Für die textbezogene Arbeit werden sie nun über das TITUS-Portal zusätzlich mit einer vereinfachten Darstellung der Annotation in herkömmlicher Grammatik-Terminologie als Lesekorpus Altdeutsch (LeA) zur Verfügung gestellt (titus.uni-frankfurt.de/lea); hinzukommen soll schrittweise die Verknüpfung mit den Textwörterbüchern, die für die Annotation im ReA-Projekt benutzt worden sind, und mit dem großen ‚Althochdeutschen Wörterbuch‘ der Leipziger Akademie der Wissenschaften (<http://awb.saw-leipzig.de>), das seit kurzem im Trierer Wörterbuchnetz online ist.

Die Idee zu diesem Hilfsmittel steht im Zusammenhang mit den Arbeiten im eHumanities-Zentrum für Historische Lexikographie (www.zhistlex.de), genauer mit dem Teilprojekt ‚Bidirektionale Verknüpfung von Wörterbüchern und Quellentexten‘. In ihm werden exemplarisch die am Projekt beteiligten großen Belegwörterbücher, voran das eben erwähnte ‚Althochdeutsche Wörterbuch‘ (AWB) und das ‚Mittelhochdeutsche Wörterbuch‘ (MWB), über Webservices so mit ihren Quellentexten verknüpft, dass die im Wörterbuch zitierten Textstellen in ihrem Textzusammenhang aufgerufen werden können, und dass umgekehrt von einer Wortform im Text oder von einer Textstelle (z. B. einem Vers oder einer Textzeile) die betreffenden Wörterbuchartikel aufrufbar sind. Für das Mittelhochdeutsche konnte das über das Online-Angebot des Wörterbuchs selbst (www.mhwdw-online.de) in seiner jüngsten Aktualisierung vom August 2018 bereits realisiert werden, weil das MWB zu einem großen Teil aus digitalisierten und lemmatisierten Texten heraus erarbeitet wird. Im Falle des Althochdeutschen handelt es sich um die nachträgliche Verknüpfung der retrodigitalisierten Online-Version des in traditioneller Weise erarbeiteten AWB mit dem Textkorpus, das dem ReA zugrunde liegt. Möglich wird dies zum einen dadurch, dass ReA und AWB in aller Regel dieselben Textausgaben benutzen, zum anderen durch die Einrichtung von Webservices für beide Ressourcen, die im Zuge des ZHistLex-Projekts entwickelt werden.

Im Zusammenhang dieser Arbeiten entstand der Wunsch, nicht nur die ReA-Texte selbst für den ZHistLex-Webservice einzurichten, sondern auch ihre Annotation und die für die Annotation benutzten Textwörterbücher einzubeziehen. Die Einbeziehung der Textwörterbücher kann erst in einem späteren Projektstadium realisiert werden. Vorläufig wird aber das LeA mit seiner Annotation schon auf statischen HTML-Seiten zur Verfügung gestellt in einer Fassung, die für die Nutzung im Unterricht und das Selbststudium gedacht ist. Im Folgenden sollen zunächst das ReA und die Erarbeitung seiner Annotation kurz genauer vorgestellt werden, anschließend dann die LeA-Darstellung der Texte mit Annotation an einem kurzen Textausschnitt erläutert werden.

II

Das ReA ist eine umfassend annotierte Textdatenbank nahezu der gesamten althochdeutschen und altsächsischen (altdeutschen) Textüberlieferung einschließlich der lateinischen Textwörter und Textgrundlagen. Es wurde in den Jahren 2008–2014 unter Leitung von Karin DONHAUSER, Jost GIPPERT und Rosemarie LÜHR an den Universitäten Berlin (HU), Frankfurt am Main und Jena erarbeitet als Teil der Initiative ‚Deutsch.Diachron.Digital‘ (DDD), die sich zum Ziel gesetzt hat, für alle Sprachstufen des Deutschen vergleichbare Referenzkorpora zu erstellen.³ Das Korpus hat einen Umfang von ca. 700.000 Textwörtern.

Die digitale Textgrundlage des ReA wurde im Rahmen des TITUS-Projekts (Thesaurus Indogermanischer Text- und Sprachmaterialien) an der Universität Frankfurt am Main erstellt.⁴ Es handelt sich um die elektronische Aufarbeitung der Editionstexte der maßgeblichen Referenzausgaben. Seit 2018 sind sie auch vollständig seitenweise mit den digitalen Faksimiles der zugrundeliegenden Handschriften verlinkt. Um die morphologische Annotation und die Lemmatisierung der Wortformen des Korpus vorzunehmen, wurden die Textwörterbücher, die zu allen althochdeutschen und altsächsischen Texten vorliegen, digitalisiert und in ein XML-Format überführt.⁵ Aus ihnen wurden teilautomatisiert die für die Annotation benötigten Informationen entnommen; ebenfalls teilautomatisiert wurden ihre morphologischen Angaben dem Tagset des Korpus entsprechend umgeformt.⁶ Lemmata und Bedeutungsangaben wurden einheitlich an die Wör-

3 Karin Donhauser, Referenzkorpus Altdeutsch. Das Konzept, die Realisierung und die neuen Möglichkeiten. In: Jost Gippert u. Ralf Gehrke (Hgg.), *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 5). Tübingen 2015, S. 35–49, hier S. 35.

4 Jost Gippert, The TITUS Project. 25 Years of Corpus Building in Ancient Languages. In: Ingeborg Hafemann (Hg.), *Perspektiven einer corpusbasierten historischen Linguistik und Philologie*. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12.–13. Dezember 2011. Berlin 2012, S. 169–192. http://edoc.bbaw.de/frontdoor.php?source_opus=2439 (Zugriff: 12.12.2018).

5 Roland Mittmann, Digitalisierung historischer Glossare zur automatisierten Vorannotation von Textkorpora am Beispiel des Altdeutschen. In: Armin Hoenen u. Thomas Jügel (Hgg.), *Altüberlieferte Sprachen als Gegenstand der Texttechnologie / Ancient Languages as the Object of Text Technology* (Journal for Language Technology and Computational Linguistics – JLCI 27–2/2012). Berlin 2012, S. 39–52. http://www.jlcl.org/2012_Heft2/3Mittmann.pdf (Zugriff: 12.12.2018).

6 Sonja Linde u. Roland Mittmann, Old German Reference Corpus. Digitizing the knowledge of the 19th century. Automated pre-annotation using digitized historical glossaries. In: Paul Bennett u. a. (Hgg.): *New Methods in Historical Corpora* (Korpuslinguistik und interdisziplinäre Perspektiven

terbücher von SPLETT⁷ für das Althochdeutsche bzw. SEHRT⁸ für das Altsächsische angepasst. Anschließend wurden die aus den Wörterbüchern entnommenen Daten mit den TITUS-Texten verknüpft („vorannotiert“) und ins Format der Software ELAN⁹ umgewandelt,¹⁰ in dem sie manuell überprüft, angepasst und um eine Satzarten-Annotation¹¹ ergänzt wurden. Die ELAN-Dateien wurden schließlich um Metadaten ergänzt und in die Datenbank ANNIS übertragen.¹² Dort sind sie nun online abfragbar; die Texte können darüber hinaus in verschiedenen Formaten vollständig heruntergeladen werden.¹³

Nachfolgend sind die einzelnen ReA-Teilkorpora zusammen mit den der Digitalisierung zugrundeliegenden Editionen und den für die teilautomatisierte Vorannotation verwendeten Textwörterbüchern aufgeführt:

Tabelle 1: Teilkorpora des ‚Referenzkorpus Altdeutsch‘ (ReA)

	Teilkorpus	Entstehungs- zeitraum	Wort- formen	Lat. Text- grundlage	Edition	Textwörter- buch
ahd.	Isidor	etwa 770–810	6.564	5.403	EGGERS ¹⁴	HENCH ¹⁵
	Benediktinerregel	um 800	16.965	14.318	DAAB ¹⁶	HEFFNER ¹⁷

auf Sprache / Corpus Linguistics and Interdisciplinary Perspectives on Language – CLIP 3). Tübingen 2013, S. 235–246, hier S. 238f.

7 Jochen Splett, Althochdeutsches Wörterbuch. Analyse der Wortfamilienstrukturen des Althochdeutschen, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes. 3 Bde. Berlin 1993.

8 Edward H. Sehart, Vollständiges Wörterbuch zum Heliand und zur altsächsischen Genesis. Göttingen 1966.

9 <http://tla.mpi.nl/tools/tla-tools/elan> (Zugriff: 12.12.2018).

10 Linde u. Mittmann (Anm. 6), S. 239.

11 Donhauser (Anm. 3), S. 42.

12 <http://corpus-tools.org/annis> (Zugriff: 12.12.2018) sowie Thomas Krause u. Amir Zeldes, ANNIS3: A new architecture for generic corpus query and visualization. In: Digital Scholarship in the Humanities 31/1 (2016), S. 118–139. <http://doi.org/10.1093/llc/fqu057> (Zugriff: 12.12.2018).

13 <http://hdl.handle.net/11022/0000-0007-C9C7-6> (Zugriff: 12.12.2018).

14 Der althochdeutsche Isidor. Nach der Pariser Handschrift und den Monseer Fragmenten. Hrsg. v. Hans Eggers. Tübingen 1964.

15 Der althochdeutsche Isidor. Facsimile-Ausgabe des Pariser Codex nebst kritischem Texte der Pariser und Monseer Bruchstücke. Mit Einleitung, grammatischer Darstellung und einem ausführlichen Glossar. Mit 22 Tafeln. Hrsg. v. George A. Hench. Straßburg 1893.

16 Die Althochdeutsche Benediktinerregel des Cod. Sang 916. Hrsg. v. Ursula Daab (Altdeutsche Textbibliothek 50). Tübingen 1959.

17 Roe-Merill S. Heffner, A Word-Index to the Texts of Steinmeyer. Die kleineren althochdeutschen Sprachdenkmäler. Madison WI 1961.

	Teilkorpus	Entstehungs- zeitraum	Wort- formen	Lat. Text- grundlage	Edition	Textwörter- buch
ahd. (Forts.)	Monseer Fragmente	um 810	14.538	–	HENCH ¹⁸	
	Murbacher Hymnen	etwa 800–825	3.139	4.319	SIEVERS ¹⁹	
	Tatian	etwa 825–850	55.508	44.941	SIEVERS ²⁰	
	Otfrid	863–871	82.227	–	ERDMANN ²¹	KELLE ²²
	Notker	† 1022	332.016	–	KING/TAX ²³	SEHRT/ LEGNER ²⁴
	Ahd. Physiologus	11. Jh.	1.935	–	STEINMEYER ²⁵	HEFFNER ²⁶
	Kleinere ahd. Denkmäler	etwa 750–1100	32.589	–	(u. a., s. LeA)	
as.	Heliand	um 830	69.770	–	BEHAGHEL/ TAEGER ²⁷	SEHRT ²⁸
	Genesis	um 840	4.042	–		
	Kleinere altsächs. Denkmäler ²⁹	etwa 750–1100	18.493	–		WADSTEIN ³⁰ (u. a., s. LeA)

18 The Monsee Fragments. Newly collated text with notes and a grammatical treatise. A dissertation submitted to the Philosophical Faculty of the Johns Hopkins University for the degree of Doctor of Philosophy. Hrsg. v. George A. Hench. Straßburg 1890.

19 Die Murbacher Hymnen. Mit zwei lithographischen Facsilimes. Hrsg. v. Eduard Sievers. Halle 1874.

20 Tatian. Lateinisch und althochdeutsch mit ausführlichem Glossar. Hrsg. v. Eduard Sievers (Bibliothek der ältesten deutschen Litteratur-Denkmäler V). 2. Aufl. Paderborn 1892.

21 Oskar Erdmann (Hg.), Otfrids Evangelienbuch (Altdeutsche Textbibliothek 49). 6. Aufl., besorgt von Ludwig Wolff. Tübingen 1973.

22 Johann Kelle, Glossar der Sprache Otfrids (Otfrids von Weißenburg Evangelienbuch. Text, Einleitung, Grammatik, Metrik, Glossar 3). Regensburg 1881.

23 Die Werke Notkers des Deutschen. Neue Ausgabe. Hrsg. v. James C. King u. Petrus W. Tax (Altdeutsche Textbibliothek). 18 Bde. Tübingen 1972–2003.

24 Edward H. Sehart u. Wolfram Legner: Notker-Wortschatz. Das gesamte Material zusammengetragen von Edward H. Sehart und Taylor Starck. Halle 1955.

25 Die kleineren althochdeutschen Sprachdenkmäler. Hrsg. v. Elias von Steinmeyer. Berlin 1916.

26 Heffner (Anm. 17).

27 Heliand und Genesis. 9. Aufl., bearb. v. Burkhart Taeger. Hrsg. v. Otto Behaghel (Altdeutsche Textbibliothek 4). Tübingen 1996.

28 Sehart (Anm. 8).

29 Das Freckenhorster Heberegerregister und die Altniederfränkischen Psalmen zählen im ReA zu den Kleineren altsächsischen Denkmälern, werden im LeA aber separat dargestellt.

30 Elis Wadstein, Kleinere altsächsische Sprachdenkmäler. Mit Anmerkungen und Glossar (Niederdeutsche Denkmäler 6). Norden, Leipzig 1899.

Die auf die geschilderte Weise erarbeitete Annotation des ReA enthält für jede laufende Wortform aller Texte Informationen zur Wortart, zur Bestimmung der Flexion und Flexionsklasse, zum entsprechenden Wörterbuch-Stichwort und die Wörterbuch-Bedeutungsangabe(n) zu diesem Lemma; die Wortform-übergreifenden syntaktischen und ggf. metrischen Einheiten sind ebenfalls annotiert mit Informationen zur Satzart und der Auszeichnung der Reimwörter. Für das Textwort *muader* aus der Textstelle Otfrid II,14,7 (Christus am Jakobsbrunnen: *tho gisaz er múader [...] bi einemo brunnen* – „Da setzte er sich müde nieder an einem Brunnen“) sieht die Annotation im ReA wie folgt aus:³¹

Tabelle 2: Annotierte Beispiel-Wortform im ‚Referenzkorpus Altdeutsch‘ (ReA)

ANNIS-Ebenen	ANNIS-Werte
edition	<i>múader</i>
lemma	<i>muodi</i>
translation	<i>müde, ermüdet, erschöpft</i>
lang	<i>goh</i>
posLemma	<i>ADJ</i>
pos	<i>ADJD</i>
inflectionClassLemma	<i>ja,jo</i>
inflectionClass	<i>P</i>
inflection	<i>Pos_Masc_Sg_Nom_st</i>
clause	<i>CF_CS_U_M</i>
rhyme	<i>E</i>

Die Ebene ‚edition‘ zeigt die Wortform des Editionstextes, ‚lemma‘ das Wörterbuchstichwort, ‚translation‘ die Wörterbuch-Bedeutungsangabe(n), ‚lang‘ die Sprachangabe *goh* (*German, Old High*) nach ISO 639-3.³²

³¹ Das Korpus weist darüber hinaus weitere Ebenen mit aus TITUS übernommenen Angaben zur strukturellen Gliederung (z.B. Kapitel und Zeilen für Manuskript und Edition) auf.

³² <http://iso639-3.sil.org> (Zugriff: 12.12.2018).

In den Ebenen ‚posLemma‘ und ‚pos‘ ist die Wortart nach dem Tagset HiTS (Historisches Tagset)³³ aufgeführt.³⁴ Die Unterscheidung der beiden Ebenen erklärt sich dadurch, sich die Wortart einer belegten Wortform von der ihres zugrundeliegenden Lemmas abweichen kann.³⁵ In den Ebenen ‚inflectionClassLemma‘, ‚inflectionClass‘ und ‚inflection‘ sind die morphologischen Angaben zur Wortform annotiert.³⁶ Dabei werden in ‚inflectionClassLemma‘ die Werte der für das Lemma geltenden Kategorien erfasst, in ‚inflectionClass‘ – analog zum Vorgehen bei der Wortart – die entsprechenden Werte des vorliegenden Vorkommens.³⁷ In ‚inflection‘ werden schließlich die Werte derjenigen Kategorien dargestellt, die nur für die vorliegende Wortform gelten.

Die Ebenen ‚clause‘ und ‚rhyme‘ gelten den die Wortform übergreifenden syntaktischen und ggf. metrischen Einheiten: Hier finden sich Angaben zur Satzart³⁸ und ggf., wie hier, die Auszeichnung des Reimworts (mit *E* für einen Endreim).

III

Die Annotation des ReA ist, wie in dem kurzen Beispiel deutlich geworden sein dürfte, formal ganz auf standardisierte Datenbank-Abfragen hin ausgerichtet. Für eine auf das sprachliche Textverständnis zielende Konsultation, die idealerweise

33 HiTS ist entstanden, indem das für das Neuhochdeutsche erstellte Tagset STTS (Stuttgart-Tübingen-Tagset) zur Annotation älterer Texte weiterentwickelt wurde, vgl. Stefanie Dipper u.a., HiTS: ein Tagset für historische Sprachstufen des Deutschen. In: Heike Zinsmeister, Ulrich Heid u. Kathrin Beck (Hgg.), Das Stuttgart-Tübingen Wortarten-Tagset – Stand und Perspektiven / The Stuttgart-Tübingen Part-of-Speech Tagset – Current Status and Plans (Journal for Language Technology and Computational Linguistics – JLCL 28–1/2013). Berlin 2013, S. 85–137. Auf S. 114–137 sind die HiTS-Tags für beide Ebenen aufgeführt und den STTS-Tags gegenübergestellt. http://www.jlcl.org/2013_Heft1/5Dipper.pdf (Zugriff: 12.12.2018).

34 *ADJD* steht für „Adjektiv, prädikativ oder adverbial“.

35 Dipper (Anm. 33), S. 92, Donhauser (Anm. 3), S. 41 sowie Sonja Linde, Manuelle Abgleichung bei automatisierter Vorannotation. Das Tagging grammatischer Kategorien im Referenzkorpus Altdeutsch. In: Armin Hoenen u. Thomas Jügel (Hgg.), Altüberlieferte Sprachen als Gegenstand der Texttechnologie / Ancient Languages as the Object of Text Technology (Journal for Language Technology and Computational Linguistics – JLCL 27–2/2012). Berlin 2012, S. 53–64, hier S. 54. http://www.jlcl.org/2012_Heft2/4Linde.pdf (Zugriff: 12.12.2018).

36 *ja,jo* steht für ‚ja-/jō-Stamm‘, *P* für ‚pronominale Flexion‘, *Pos_Masc_Sg_Nom_st* für ‚Positiv, Maskulinum, Singular, Nominativ, starke Flexion‘.

37 Linde (Anm. 35), S. 54f.

38 *CF_CS_U_M* steht für ‚Gliedsatz, finit, Unterbrechung, uneingeleitet, Hauptsatz‘.

im Zusammenhang mit der Benutzung von Wörterbuch, Grammatik und (wo vorhanden) einem wissenschaftlichen Textkommentar steht oder dazu anleiten sollte, ist sie anzupassen: Nicht alle Annotations-Arten/Ebenen sind dafür gleich wichtig, und es sollten für die morphologischen Angaben die herkömmlichen grammatischen Bezeichnungen gewählt werden, wie sie sich in den genannten wissenschaftlichen Hilfsmitteln finden. Wir haben uns daher dazu entschieden, im LeA zwar optional alle ReA-Annotations-Arten anzubieten, standardmäßig aber nur die folgenden einzublenden: Flexion (Bestimmung der flektierten Form), Flexionsklasse, Wortart, Lemma (Wörterbuch-Stichwort). Hinzugewählt werden kann eine zusätzliche Annotations-Art, die sich nicht im ReA findet, eine Normalisierung der Wortform im Standard-Althochdeutsch der Grammatik (idealisiertes Ostfränkisch auf ‚Tatian‘-Grundlage),³⁹ und es können auch die bereits vorgeführten weiteren Annotations-Arten des ReA eingeblendet werden.⁴⁰ Dazu gehören auch die Bedeutungsangaben des Wörterbuchs von SPLETT⁴¹ (im ReA unter translation, im LeA unter Wörterbuch-Bedeutung), die aber mit Vorsicht benutzt werden müssen, weil sie nicht auf die einzelne Textstelle bezogen sind. Für eine genaue Interpretation des jeweils vorliegenden Wortgebrauchs ist vor allem das AWB zu konsultieren, soweit es bereits erschienen ist; dies wird in Zukunft durch die Verknüpfung von LeA und AWB erleichtert.

39 Diese Annotations-Art war im ReA nur zur internen Qualitätskontrolle erzeugt worden, vgl. Roland Mittmann: Automated quality control for the morphological annotation of the Old High German text corpus. Checking the manually adapted data using standardized inflectional forms. In: Jost Gippert u. Ralf Gehrke (Hgg.), *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 5). Tübingen 2015, S. 65–76.

40 Peter Hinkelmanns, jetzt Mitarbeiter der MHDBDB Salzburg, zuvor im ZHistLex, danken wir für die Skripte zur Umwandlung der ELAN-Daten in die HTML-Darstellung.

41 Splett (Anm. 7).

Für das bereits genannte Textwort *muader* sieht die Gegenüberstellung mit der Annotation im ReA wie folgt aus:

Tabelle 3: Annotierte Beispiel-Wortform (standardmäßig nicht eingeblendete Angaben in Grau)

„Referenzkorpus Altdeutsch“ (ANNIS)		„Lesekorpus Altdeutsch“	
Ebenen	Werte	Werte	Annotations-Arten
edition	<i>múader</i>	<i>múader</i>	
inflection	<i>Pos_Masc_Sg_Nom_st</i>	<i>N. Sg. m.</i>	Flexion
inflectionClass	<i>P</i>	<i>st.</i>	Flexionsklasse
		<i>muodēr</i>	Normalisiert
inflectionClassLemma	<i>ja,jo</i>	<i>ja-St./jō-St.</i>	Flexionsklasse Lemma
posLemma	<i>ADJ</i>	<i>Adj.</i>	Wortart
pos	<i>ADJD</i>	<i>Adj., präd./adv.</i>	Wortart syntaktisch
lemma	<i>muodi</i>	<i>muodi</i>	Lemma
translation	<i>müde, ermüdet, erschöpft</i>	<i>müde, ermüdet, erschöpft</i>	Wörterbuch-Bedeutung
rhyme	<i>E</i>	<i>Endreim</i>	Reim
lang	<i>goh</i>	<i>ahd.</i>	Sprache
clause	<i>CF_CS_U_M</i>	<i>finiter Satz, Unterbrechung, uneingeleiteter Satz, Hauptsatz</i>	Satz

Das Ergebnis der Umformung, die neue Darstellung der Annotation im LeA, sei an einem kürzeren Abschnitt, dem Eingang von Otfrid 2,14 (Io 4: Jesus am Jakobsbrunnen / Christus und die Samariterin) demonstriert. Der Editionstext:

- 1 Sid tho thésen thingon fuar Krist zen héimingon,
in selbaz géwi sinaz; thio buah nénnent uns tház.
Thera férti er ward irmúait, so ofto fārantemo duit;
ni lazent thie árabeit es fríst themo wárlicho mán ist.
- 5 Fúar er thuruh Samárium, zī einera burg er thar tho quám,
in themo ágileize zi éinemo gisáze.
Tho gisaz er múader, so wir gizáltun hiar nu ér,
bí einemo brúnnen (thaz wir ouh púzzi nennen).

Abb. 1 zeigt einen Screenshot mit der standardmäßig eingeblendeten Annotation für diesen Abschnitt; am Kopf stehen die Check-Boxen, mit deren Hilfe die Auswahl aus den Annotations-Arten vorgenommen wird.

IV

Für Althochdeutsch-Lerner bietet die Annotation eine erhebliche Hilfestellung bei der Einübung der genauen sprachlichen Analyse: durch den Bezug auf die Wörterbuch-Normalform (unter ‚Lemma‘), die auch das Nachschlagen in den Wörterbüchern erleichtert, und durch die Angaben zur Wortart, Flexionsklasse und Flexionsform, die es ermöglichen, die syntaktischen Konstruktionen zu erkennen und ggf. in der Grammatik gezielt die entsprechenden Flexions-Paradigmen zu vergleichen. Der Unterricht und seine Vor- und Nachbereitung durch die Studierenden wird auf diese Weise beträchtlich entlastet und kann sich stärker auf die dann noch verbleibenden „höheren“ Schwierigkeiten des genauen sprachlichen Textverständnisses konzentrieren. Ein Vergleich der folgenden, nah am Original bleibenden Übersetzung – die das Ergebnis der gemeinsamen Lektürearbeit von Studierenden und Dozenten sein könnte – mit den vorhandenen Übersetzungen der Stelle zeigt, worin diese Schwierigkeiten bestehen können, zugleich aber auch den Nutzen einer genauen eigenständigen sprachlichen Durcharbeitung des Originaltextes:

Nach diesen Ereignissen dann zog Christus in seine Heimat, / in sein eigenes Land; die Schrift nennt uns seinen Namen [sc. Galiläa]. / Durch diese Reise wurde er ermüdet, wie es oft einem Reisenden ergeht. / Solche Anstrengungen erlassen das dem nicht, der wirklich ein Mensch ist. / Er zog durch Samaria, zu einer Stadt kam er dort dann, / in solcher Anstrengung, zu einem Ruheplatz. / Dann setzte er sich nieder, müde, wie wir es gerade eben hier erzählt haben, / an einer Wasserstelle (wozu wir auch Pütz [Ziehbrunnen] sagen).⁴²

Gefordert ist also z.B. die richtige Auffassung bestimmter Wendungen und Konstruktionen (z.B. **2** *thio buach* nicht ‚die Bücher‘, sondern ‚die Heilige Schrift‘;⁴³ **4** *frist lâzan* mit Dat. d. P. und Gen. d. S. ‚jmdm. etw. erlassen‘), die semantische Bestimmung des vorliegenden Wortgebrauchs (z.B. **4** *arabeit*⁴⁴ nicht ‚Arbeiten, Werke‘, sondern ‚[körperliche] Anstrengungen‘ und *man* nicht ‚Mann‘, sondern ‚Mensch‘; **8** *brunnen* ‚Quelle‘ wie auch ‚Brunnen‘), Kenntnis älterer Syntax (z.B. **4** *themo* in Doppelfunktion als Bezugswort und Relativum), vor allem aber auch, verbunden mit dem Vorigen, die richtige Auffassung des Textsinns überhaupt, so

⁴² Vgl. die Übersetzungen in: Johann Kelle, Christi Leben und Lehre besungen von Otfrid. Prag 1870 (ND Osnabrück 1966), S. 134–142; Schlosser 1980 (Anm. 1), S. 116–120; Stephan Müller (Anm. 1), S. 138–149; Heiko Hartmann, Otfrid von Weißenburg, Evangelienbuch. Aus dem Althochdeutschen übertragen. Bd. 2. Herne 2014, S. 38–41.

⁴³ Müller übersetzt ‚die Bücher‘ und sieht sich dann im Kommentar genötigt zu erklären, dass sich die Erzählung nur bei Johannes finde (Anm. 1), S. 341.

⁴⁴ Mit Elision des ausl. *i* vor Vokal, wie in der Normalisierung sichtbar wird.

O 2,14,5	Fúar	er	thuruh	Samárium	zi	einera	burg	er	thar	tho	quám
Flexion	3. Sg. Prät. Ind.	3. Sg. N. m.		A. Sg.		D. Sg. f.	D. Sg.	3. Sg. N. m.			3. Sg. Prät. Ind.
Flexionsklasse	st. 6			f., a-St.		st.	f., Wz-Nomen				st. 4
Wortart	Verb	Pers.-Pron.	Präp.	EN	Präp.	Indef.-Pron.	Subst.	Pers.-Pron.	Adv.	Adv.	Verb
Lemma	faran	ër	duruh	Samaria	zi	ein	burg	ër	där	dó	quëman

O 2,14,6	in	themo	ágileize	zi	éinemo	gisáze
Flexion		D. Sg. n.	D. Sg.		D. Sg. n.	D. Sg.
Flexionsklasse		n., a-St.	st.		st.	n., a-St.
Wortart	Präp.	Dem.-Pron.	Subst.	Präp.	Indef.-Pron.	Subst.
Lemma	in	dër	agaleigi	zi	ein	gisázi

O 2,14,7	Tho	gisaz	er	múader	so	wir	gizáltun	hiar	nu	ér
Flexion		3. Sg. Prät. Ind.	3. Sg. N. m.	N. Sg. m.		1. Pl. N.	3. Pl. Prät. Ind.			
Flexionsklasse		st. 5		st.			sw. 1a			
Wortart	Adv.	Verb	Pers.-Pron.	Adj.	Konj.	Pers.-Pron.	Verb	Adv.	Adv.	Adv.
Lemma	dó	gisizzen	ër	muodi	só	wir	gizellen	hier	nũ	ër

O 2,14,8	bi	einemo	brúnnen	(	thaz	wir	ouh	púzzi	nennen	)
Flexion		D. Sg. m.	D. Sg.		A. Sg. n.	1. Pl. N.		A. Sg.	1. Pl. Präs. Ind.	
Flexionsklasse		st.	m.					m., ja-St.	sw. 1a	
Wortart	Präp.	Indef.-Pron.	Subst.		Dem.-Pron.	Pers.-Pron.	Adv.	Subst.	Verb	
Lemma	bif	ein	brunno	dër	wir	ouh	píuzzi	nennen		

Abbildung 1: LeA-Textausschnitt mit standardmäßig eingeblendeter Annotation (Otfrid 2,14).

in 4 die Betonung der menschlichen Natur Christi.⁴⁵ Ein besonders kniffliger Fall ist v. 8, der wegen der Doppelbedeutung von *brunno* und dem standardsprachlich heute in der vorliegenden Bedeutung ohne Entsprechung bleibenden *puzzi* eigentlich unübersetzbar ist: Die geistliche Pointe der Erzählung erfordert im weiteren Textverlauf eine Unterscheidung wie von lat. *puteus* und *fons* im lat. Bibeltext, für die das Althochdeutsche aber mit *brunno* nur ein (Erb-)Wort hat, das für beides stehen kann. Otfrid führt daher schon in 8 in einer Parenthese für die Bedeutung ‚Ziehbrunnen‘ das Lehnwort *puzzi* (< lat. *puteus*)⁴⁶ zur Disambiguierung ein (das in dieser Bedeutung heute nur mundartlich als *Pütz* erhalten ist;⁴⁷ standardsprachlich setzt *Pfütze* das Lehnwort zwar fort, aber eben nicht in dieser Bedeutung). *puzzi* wird dann im Folgenden (29.34.45) in der Bedeutung ‚Ziehbrunnen‘ benutzt im Unterschied zu *brunno*, das jetzt nur noch in der Bedeutung einer ‚Quelle‘ (*springenter brunno*, 26; vgl. 37.39.44), deren Wasser das ewige Leben schenkt (*fons aquae salientis in vitam aeternam*, Io 8,15), gebraucht wird.⁴⁸

Viele der eben angedeuteten Schwierigkeiten behebt schon die Konsultation der Textwörterbücher (hier für Otfrid desjenigen von KELLE⁴⁹) und des AWB; die bereits vorbereitete Verknüpfung mit ihnen soll der Einstellung der annotierten Texte daher, wie schon gesagt, möglichst bald folgen. Auch dann wird das Lesekorpus Altdeutsch eine gründliche Einführung in die althochdeutsche Sprache und Literatur und in die Benutzung der Hilfsmittel zum genauen sprachlichen

45 Ganz verfehlt Übersetzungen von II,14,4 bei Kelle (Anm. 42), S. 134: „Er handelt als ein echter Mann, dem gönnt nicht Ruhe sein Beruf“; Schlosser 1980 (Anm. 1), S. 117: „Wer sich seinem Beruf ganz hingibt, dem gönnen seine Verpflichtungen keine Ruhe“; Hartmann (Anm. 42): „Wer ein echter Mann ist, dem lässt sein Werk keine Ruhe“. Richtiger nur Müller (Anm. 1), S. 139, doch mit trivialisierender Einfügung von „auch“ und mit „wurde“ statt „ist“ zur Verdeutlichung des Bezugs auf Christus: „Die Mühen lassen auch den nicht unbehelligt, der wahrlich Mensch wurde.“ – Vgl. zur richtigen Auffassung der Stelle schon Oskar Erdmann im Kommentar der großen Ausgabe, Halle/S. 1882, S. 402.

46 Vgl. AWB 7, 297–299 s. v. *phuzza* st. F. und *phuzzi* st. M. (zzt. noch nicht online).

47 Vgl. das Rheinische Wörterbuch 6, 1247f. s. v. ‚Pütz‘

(online im Wörterbuchnetz: <http://www.woerterbuchnetz.de/RhWB?lemma=puetz>; Zugriff: 12.12.2018).

48 Schlosser 1980 (Anm. 1) übersetzt *brunnen* mit ‚Brunnen‘, lässt *puzzi* unübersetzt in seinem Text und kommentiert S. 349 nicht völlig erhellend: „lat. *puteus* („Pfütze“) ‚gegrabener Brunnen‘ (d. h. keine frei sprudelnde Quelle)“; Kelle (Anm. 42), S. 135: „Bei einem solchen Brunnenquell, der Pfütze auch bei uns genannt“, Müller (Anm. 1), S. 139: „bei einem Brunnen, den wir auch ‚Pfütze‘ nennen“ und Hartmann (Anm. 42): „... an einer Wasserquelle... Wir nennen eine solche auch ‚Brunnen‘“) scheinen sich das Problem des Wortgebrauchs und seiner Übersetzung an dieser Stelle nicht recht klargemacht zu haben.

49 Kelle (Anm. 22).

Textverständnis selbstverständlich nicht ersetzen können; es soll sie unterstützen und ein daran anknüpfendes Selbststudium althochdeutscher und altsächsischer Texte erleichtern.