Das Mittelalter 2019; 24(1): 112–128

Christian Griesinger und Michael Stolz

Sprachwissenschaftliche Erschließungsmethoden für digitale Editionen mittelhochdeutscher Texte

https://doi.org/10.1515/mial-2019-0008

Abstract: This paper sheds light on the possibilities and perspectives of linking digital editions of Medieval German texts to each other and to other digital resources. Furthermore, it discusses some of the internal and technical conditions necessary to render this linkage meaningful, like lemmatisation, part-of-speech-tagging, and using standardised mark-up languages. Finally, the sustainability and reusability of digital editions are considered.

While in the past, editions of medieval texts were conceived as rather isolated scholarly works of individual editors, nowadays the collaboration and cooperation of greater working groups is essential in editing projects. Due to the complexity of editions consisting of multiple textual layers, e.g. apparatus entries, annotations, or facsimiles, the requirements for future digital editions have risen. The first approach to respond to these demands is to link the various textual layers to each other, enabling the users to navigate between these layers in a sensible way. The second step is to link the edited text to other resources, such as online dictionaries or other editions, allowing complex research networks to be created.

These goals are achieved by lemmatising and other tagging methods, ensuring the information being mapped to a normalised and idealised frame of reference. Common standards like Unicode or the TEI guidelines are of great importance for such purposes, as they assure the interchange and re-use of scientific data, as well as their sustainability.

Keywords: Digital editing, annotation, lemmatizing, linking of digital components, Medieval German studies

Kontakt: Christian Griesinger, M.A., Bergische Universität Wuppertal, Germanistik, Gaußstr. 20, 42119 Wuppertal, E-Mail: griesinger@uni-wuppertal.de
Prof. Dr. Michael Stolz, Universität Bern, Institut für Germanistik, Länggass-Str. 49, CH–3012 Bern, E-Mail: michael.stolz@germ.unibe.ch

source: https://doi.org/10.7892/boris.132536 | downloaded: 4.10.2021

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

1 Einführung

Digitale Editionsprojekte haben sich in den letzten Jahren als Standard der Editionswissenschaft etabliert. Nationale Förderinstitutionen wie die Deutsche Forschungsgemeinschaft (DFG) oder der Schweizerische Nationalfonds (SNF) binden ihre Unterstützung wissenschaftlicher Textausgaben mittlerweile an die Bereitstellung eines digitalen Editionskonzepts.[1] Normen der digitalen Erschließung, Auszeichnung und Verknüpfung von edierten Textbeständen bilden sich dabei freilich erst allmählich heraus und befinden sich konzeptionell im Fluss. Der vorliegende Beitrag gibt einen Überblick über die sich in diesem Bereich abzeichnenden Möglichkeiten und Erfordernisse. Er geht dabei zwar von praktischen Erfahrungen aus, die in digitalen Editionsvorhaben wie dem an der Universität Bern und Partnerinstitutionen durchgeführten ‚Parzival'-Projekt zur alltäglichen Arbeit gehören,[2] möchte jedoch nicht projektgebunden oder an konkreten Texten orientiert, sondern allgemein argumentieren.

In diesem Kontext stellen die in den folgenden Ausführungen vorgestellten Verfahren diverse Komponenten der Präsentation und Erschließung digital edierter Texte dar. Nicht alle der beschriebenen Methoden sind dabei in gleicher Weise für spezifische Projektbedürfnisse relevant. Und es ist mit den vorgestellten Verfahren auch nicht der Anspruch verbunden, diese mit letzter Konsequenz für ein bereits laufendes Vorhaben wie das ‚Parzival'-Projekt einzulösen. Es geht vielmehr darum, potenzielle Wege der Erschließung aufzuzeigen und diese in den Fachgemeinschaften der Digital Humanities, der Editionswissenschaft und der diversen Philologien als mögliche zukünftige, inhaltliche Standards zur Diskussion zu stellen. Wenngleich der Fokus dieses Beitrags auf der Edition mittelhochdeutscher Texte liegt, können die Erschließungsmethoden verallgemeinert und so auf andere Gebiete erweitert werden.

Dabei stehen im vorliegenden Beitrag sprachwissenschaftliche Aspekte im Vordergrund. Konkret sollen folgende Schwerpunkte in den Blick genommen werden: Verknüpfungsstrategien (editionsintern, bezogen auf textexterne Ressourcen wie Wörterbücher sowie in der Verbindung der edierten Texte zu größe-

1 „Editionsprojekte [stehen] unter dem Anspruch, die Textgrundlage der veröffentlichten Edition auch in standardisierter digitaler Form verfügbar zu halten, damit sie für verschiedene wissenschaftliche Fragestellungen – z. B. computerphilologische oder sprachwissenschaftliche – genutzt werden kann." Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft. www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/foerderkriterien_editionen_literaturwissenschaft.pdf (Zugriff: 23.11.2018).

2 Vgl. die Website www.parzival.unibe.ch (Zugriff: 23.11.2018); zur Methodik und Geschichte des Projekts die Abteilungen „Einführung" und „Projektpräsentationen".

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

ren Textcorpora; Kap. 2), einschlägige Erschließungsverfahren (über Indizes und Konkordanzen, mittels Lemmatisierung sowie spezifischen Annotationsverfahren wie *Part-of-Speech-Tagging*; Kap. 3), schließlich formale Verfahren der Langzeitarchivierung (einzelne Gesichtspunkte, bezogen auf Anliegen der langfristigen Lesbarkeit und Verwertbarkeit sowie der Standardisierung von Datei- und Datenformaten; Kap. 4).[3]

2 Verknüpfung und Vernetzung

2.1 Teile einer Edition untereinander

Editionen sind komplexe Gebilde, denn sie bestehen aus einer Vielzahl von Texten und Textsorten, die sinnstiftend miteinander in Beziehung gesetzt werden. Die Wichtigsten hiervon sind die edierten Texte und Transkriptionen (im Folgenden vereinfacht Haupttexte genannt) sowie Apparate, Kommentare bzw. Erläuterungen und nicht zuletzt die Editions- bzw. Transkriptionsprinzipien und Textzeugenbeschreibungen (im Folgenden als editorische Paratexte bezeichnet).[4]

Im Druck können wegen Platzmangels und aus Kostengründen nicht alle diese Komponenten, wie etwa Transkriptionen, präsentiert werden. Daher bleiben, abhängig von der Form und Zielsetzung der Edition, oft die Transkriptionen und Faksimiles ausgespart und es werden nur die Haupttexte, Apparate und die Editionsprinzipien veröffentlicht. Ein nicht unerheblicher Teil des erarbeiteten Materials bleibt somit in den Archiven der Editoren oder Arbeitsstellen zurück, sofern er nicht am Ende des Editionsprojekts vernichtet wird.

Dank der nahezu unbegrenzten Speicherkapazität der digitalen Welt bieten digitale Editionen für dieses Problem eine Lösung, indem sie es ermöglichen, alle Textebenen online zu veröffentlichen. Zwar stellt es schon einen Fortschritt dar, wenn Transkriptionen, Kommentare und Faksimiles online angesehen oder heruntergeladen werden können, ein einfaches Downloadangebot dieser Texte reicht

3 Die zu diesen Kernbereichen folgenden Überlegungen stehen in engem Bezug zu einem aktuellen Forschungsvorhaben über die Lemmatisierung mittelhochdeutscher Texte, das Christian Griesinger (Mitarbeiter am ‚Parzival'-Projekt von 2015–2017 und diesem weiterhin verbunden) derzeit durchführt.

4 Diese Differenzierung zwischen Haupt- und Paratexten ist unabhängig von editorischen Paradigmen zu verstehen und betont lediglich den Unterschied zwischen den edierten Texten selbst und allen sonstigen editorischen Beigaben. Es spielt für die Überlegungen dieses Beitrags weder eine Rolle, ob der bzw. die Haupttexte handschriftennah-diplomatisch, nach Leithandschriftenprinzip, historisch-kritisch rekonstruierend oder textgenetisch konstituiert werden oder ob nur eine oder synoptisch mehrere Fassungen eines Textes geboten werden.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

jedoch noch nicht aus; zu nützlichen Arbeitsinstrumenten werden digitale Editionen erst durch die sinnvolle Verknüpfung der zur Verfügung gestellten Inhalte.

So gewährleistet erst die Verbindung der Faksimiles der Textzeugen mit den Transkriptionen und den Haupttexten die (text-)kritische Überprüfung der Haupttexte. Wenn Handschriften, Transkriptionen und die am Ende des Prozesses stehenden Haupttexte simultan betrachtet werden können, werden die Genese der Edition nachvollziehbar und Herausgeberentscheidungen transparent. Während in gedruckten Editionen die Apparate Auskunft über die Abweichungen des hergestellten Textes von den Textzeugen geben müssen und so die editorischen Entscheidungen dokumentieren, kann diese Funktion in digitalen Editionen mittels der Verknüpfung der Texte erfüllt und die Apparate können folglich entlastet werden.[5]

Was muss eine solche Verknüpfung grundsätzlich leisten? Unabhängig von der eingesetzten Technik, unabhängig davon also, ob im Hintergrund XML-Dateien, relationale Datenbanken oder Graph-Datenbanken eingesetzt werden, sollten mindestens folgende Funktionen angeboten werden: Erstens müssen Benutzer den Weg vom Textzeugen zu den Haupttexten nachvollziehen können, d.h. nach Bedarf Faksimiles und bzw. oder Transkriptionen ansehen können. Zweitens muss es möglich sein, eine beliebige Auswahl der in einer Edition verarbeiteten Textzeugen neben- oder untereinander zu betrachten. ‚Beliebig' heißt in diesem Fall nicht, dass auf der Oberfläche beliebig viele Spalten nebeneinander angezeigt werden müssen, sondern dass frei entschieden werden kann, welche der Textzeugen miteinander verglichen werden sollen. Besonders relevant ist dies dann, wenn die Haupttexte z.B. mehrere Fassungen bieten und Benutzer Textzeugen verschiedener Fassungen miteinander vergleichen möchten. Drittens sollten die editorischen Paratexte, d.h. die Apparate und Kommentare so mit dem Text verknüpft sein, dass ein Springen zwischen diesen Ebenen möglich ist. Wenn z.B. ein Apparateintrag Lesarten aus einer oder mehreren Transkriptionen enthält, muss es möglich sein, über diesen Apparateintrag in den Volltext der Transkriptionen an die entsprechenden Stellen zu springen. Der Apparat hätte in diesem Fall die Aufgabe, das Bindeglied zwischen Transkription und Haupttext zu sein.

5 Die Überlastung oder Überfrachtung der Apparate in kritischen Editionen ist in editionswissenschaftlichen Beiträgen mittlerweile zu einem Gemeinplatz geworden. Die zum Teil überzogene Kritik, die den Apparat gar als „Variantengrab" betrachtet, soll hier nicht weiter ausgeführt werden. Vgl. den Tagungsband: Thomas Bein (Hg.), Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte (Beihefte zu editio 39). Berlin, Boston 2015, passim.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

2.2 Teile einer Edition mit externen Ressourcen

Doch nicht nur die wechselseitige Verknüpfung der einzelnen Editionsteile ist von Bedeutung: Enormes Potenzial liegt im Einbinden oder Verknüpfen externer Ressourcen, die dabei helfen, das edierte Material zu erforschen. Als eine besonders wichtige und nützliche Verlinkung darf für mittelhochdeutsche Texte jene zwischen Haupttext und Wörterbuch gelten. Wenn für alle Wörter der Haupttexte Lemmainformationen zu Grunde gelegt werden (Kap. 3.2), dann ist es möglich, dass die Benutzer unmittelbar auf die in Online-Wörterbüchern vorhanden Artikel, etwa in den Mittelhochdeutschen Wörterbüchern im Verbund: BMZ[6], ‚Lexer'[7], ‚Findebuch'[8] und MWB[9], zugreifen und damit ihre Lektüre mit lexikographischem Wissen fundieren können.

Aber nicht nur Wörterbücher, sondern auch (Fach-)Lexika und Enzyklopädien oder Kommentare sind für die Verlinkung attraktive Ziele, besonders dann, wenn komplizierte Sachzusammenhänge das Verständnis der mittelalterlichen Texte erschweren oder z. B. Hintergrundinformationen über zeitgenössische politische oder historische Vorgänge benötigt werden.

Ebenso ist im Bereich der Personen-, Werk- und Ortsbezeichnungen die Einbeziehung von Normdateien hilfreich. So ließen sich etwa in Artusromanen die beteiligten Figuren auf ein gemeinsames Figurenlexikon[10] abbilden oder nicht fiktionale Ortsnamen – soweit ihre modernen Entsprechungen ermittelbar sind – mit Geodaten verbinden. Netzwerke oder Ereignissen können so kartographisch visualisiert werden. Die der Edition zu Grunde liegenden Textzeugen könnten mit Katalogen und bibliographischen Portalen wie dem ‚Handschriftencensus'[11], ‚Manuscripta Mediaevalia'[12] oder Angeboten wie VD16, VD17 und

6 Wilhelm Müller u. Friedrich Zarncke, Mittelhochdeutsches Wörterbuch. Mit Benutzung des Nachlasses von Georg Friedrich Benecke. 3 Bde. Leipzig 1854–1866. woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=BMZ (Zugriff: 23.11.2018).

7 Matthias Lexer, Mittelhochdeutsches Handwörterbuch. 3 Bde. Leipzig 1872–1878. woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=Lexer (Zugriff: 23.11.2018).

8 Kurt Gärtner u. a., Findebuch zum mittelhochdeutschen Wortschatz. Mit einem rückläufigen Index. Stuttgart 1992. woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=FindeB (Zugriff: 23.11.2018).

9 Kurt Gärtner u. a. (Hgg.), Mittelhochdeutsches Wörterbuch. Im Auftrag der Akademie der Wissenschaften und der Literatur Mainz und der Akademie der Wissenschaften zu Göttingen. Stuttgart 2006 ff. mhdwb-online.de (Zugriff: 23.11.2018).

10 Z. B. auf eine noch zu digitalisierende Version des ‚Arthurian Name Dictionary' von Christopher W. Bruce. New York 1999.

11 www.handschriftencensus.de (Zugriff: 23.11.2018).

12 www.manuscripta-mediaevalia.de (Zugriff: 23.11.2018).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

VD18[13] verlinkt werden, sodass von dort aus unmittelbar Recherchen gestartet oder Handschriften- bzw. Druckbeschreibungen abgerufen werden können.

Auf diese Weise ginge die Editionswissenschaft engere Verbindungen mit der historischen Lexikographie und weiteren Fachgebieten ein und steigerte dadurch die Qualität und Benutzbarkeit der Editionstexte deutlich: Würden digitale Editionen von den Editionsprojekten z. B. so offen gestaltet, dass Forscher aus anderen Projekten oder Fachrichtungen, wie etwa Lexikographen, die Daten erweitern bzw. anreichern könnten, entstünden durch die Kollaboration ausgefeiltere Editionen.[14] Gerade mittelhochdeutsche Texte könnten von einem breiten, bereits existierenden Angebot an Online-Ressourcen profitieren, wie etwa dem ‚Wörterbuchnetz'[15] oder der ‚Mittelhochdeutschen Begriffsdatenbank' (MHDBDB).[16]

2.3 Editionen als Teil eines Textkorpus

Als Konsequenz dieser Überlegungen ergibt sich, dass die Möglichkeiten und Perspektiven des digitalen Mediums unser Verständnis von Editionen verändern: Betrachtete man früher Editionen als isolierte Leistungen in der Regel einzelner Herausgeber, so sind sie heute kollektive Errungenschaften, die zum Teil von mehreren Arbeitsstellen in Kollaboration ediert werden. Setzt sich dieser Trend in Zukunft fort und gehen wir den nächsten logischen Schritt in der Vernetzung von Editionen mit anderen Ressourcen, dann werden digitale Editionen wichtige Teile komplexerer Forschungsumgebungen werden.

Einen ersten Ansatz für diese Entwicklung können wir im ‚Hartmann-von-Aue-Portal'[17] erkennen, obwohl der oben skizzierte, notwendige Schritt einer sinnvollen Verknüpfung von Faksimile, Transkription und den anderen angebotenen Texten dort noch nicht durchgeführt ist. Jedoch werden an einer Stelle mehrere Werke gleichzeitig zur Verfügung gestellt und durch ein gemeinsames Kontextwörterbuch erschlossen. Diese Bündelung erzielt bereits einen Mehrwert. Umgekehrt sehen wir in der ‚Mittelhochdeutschen Begriffsdatenbank' die Verknüpfung der Texte weit fortgeschritten, weil hunderte mittelhochdeutsche Texte

13 www.vd16.de sowie www.vd17.de und www.vd18.de (Zugriff: 23.11.2018).

14 Vgl. hierzu unter dem Stichwort „distributed scholarly digital editions" Peter Boot u. Joris van Zundert, The Digital Edition 2.0 and The Digital Library. Services, not Resources. In: Christiane Fritze u.a. (Hgg.), Digitale Edition und Forschungsbibliothek (Bibliothek und Wissenschaft 44). Wiesbaden 2011, S. 141–152, bes. 143–145.

15 www.woerterbuchnetz.de (Zugriff: 23.11.2018). Siehe oben Anm. 6–8.

16 www.mhdbdb.sbg.ac.at (Zugriff: 23.11.2018).

17 www.hvauep.uni-trier.de (Zugriff: 23.11.2018).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

gleichzeitig als Textkorpus nach Wortformen, Lexemen oder Begriffen abgefragt werden können. Dort sind allerdings die einzelnen Texte nicht als Volltexte sichtbar und es gibt auch keine Transkriptionen oder Faksimiles.

Wird diese Entwicklung in aller Konsequenz weitergeführt und geben Editionen ihre isolierte Stellung zukünftig auf, so werden sie in größere Zusammenhänge eingebettet – die Editionen werden Teil eines umfassenderen Textkorpus. Denkt man sich zusätzliche Schnittstellen zwischen mehreren dieser Forschungsportale, beispielsweise zwischen dem Hartmann-von-Aue-Portal, der Edition des ‚Welschen Gastes'[18] und dem Berner ‚Parzival'-Projekt, so könnten Abfragen über mehrere Texte gleichzeitig gestellt werden: Gemeinsamer Wortschatz, intertextuelle Bezüge, ja sogar narrative Gemeinsamkeiten wie Motive und Stoffe könnten leichter erforscht werden, als das bei vereinzelten Druckeditionen der Fall wäre.[19]

Grundlage einer nach den obigen Überlegungen angestellten Verknüpfung ist das Internet, doch damit so verschiedenartige Ressourcen wie Editionen, Wörterbücher und Datenbanken aller Art reibungslos miteinander kommunizieren können, sind Anforderungen an die Schnittstellen zu formulieren.[20] In diesem Zusammenhang wird in Zukunft das Prinzip der *Linked Open Data* (LOD) zunehmend bedeutsamer. Hierbei geht es darum, mithilfe einfacher und standardisierter Abfrage- und Beschreibungssprachen[21] die Daten für eine maschinelle Verknüpfung und Abfrage aufzubereiten.

Die genannten Möglichkeiten sind allerdings an eine Reihe von Voraussetzungen gebunden: Es gilt innere, äußere und technische Anforderungen zu erfüllen. Zu den inneren Voraussetzungen gehört die Erschließung der Teile einer Edition mit Verfahren wie der Lemmatisierung oder dem *Part-of-Speech-Tagging*. Zu den äußeren Voraussetzungen, die in diesem Aufsatz nicht diskutiert werden können, gehören die forschungspolitischen, infrastrukturellen und finanziellen Rahmenbedingungen. Von den technischen Anforderungen werden in Kapitel 4 in Auswahl die Nachhaltigkeit und die Standardisierung der in Editionsprojekten erstellten Daten diskutiert.

18 digi.ub.uni-heidelberg.de/wgd (Zugriff: 23.11.2018).

19 Zu den Problemen, die bei der Zusammenführung unterschiedlicher Editionen entstehen, siehe Franz Fischer, Digital Corpora and Scholarly Editions of Latin Texts. Features and Requirements of Textual Criticism. In: Speculum 92/1 (2017), S. 265–287. DOI: 10.1086/693823.

20 Zum Verhältnis von Editionen und *Semantic Web* vgl. Jörg Wettlaufer, Semantic Web und digitale Editionen. In: Roland S. Kamzelak u. Timo Steyer (Hgg.), Digitale Metamorphose: Digital Humanities und Editionswissenschaft (Sonderband ZfdG 2). 2018. DOI: 10.17175/sb002_007.

21 Z.B. SPARQL und die *Web Ontology Language* (OWL).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

3 Erschließung von Editionen

3.1 Register: Indizes und Konkordanzen

Während die Lektüre einer Edition in der Regel das weitgehend lineare Lesen der Texte bedeutet, bieten Indizes und Konkordanzen einen systematischen Zugriff auf den in den Haupttexten oder Transkriptionen vorkommenden Wortschatz. Damit gehören sie zu den wichtigsten Erschließungsebenen von Texten, ohne die eine lexikologische oder lexikographische Auswertung der Texte kaum durchführbar ist.

Die erste und einfachste Ebene der Erschließung stellen die beiden Grundformen des nicht lemmatisierten Registers dar: Index und Konkordanz.[22] Indizes fassen die Wörter des Textes zu Wortformen zusammen, sortieren diese alphabetisch oder nach anderen formalen Kriterien[23] und geben die Belegstellen der Wortformen mit Referenzen an. Dies ist die am häufigsten anzutreffende Form von Registern. Gegebenenfalls können in Indizes auch die absoluten oder relativen Häufigkeiten der Wortformen gezählt und ausgewertet werden. Tritt zu den Belegstellen eine gegebenenfalls verkürzte Darstellung des Kontextes hinzu, liegt eine Konkordanz vor. Hier wird oftmals das Stichwort in der Mitte einer Zeile positioniert und der Kontext links und rechts davon ausgerichtet, so dass eine KWIC-Konkordanz (*Key-Word-In-Context*) entsteht, mit der die syntaktischen Zusammenhänge schnell erfasst werden können.

Bereits mit diesen einfachen Mitteln, die mithilfe elektronischer Datenverarbeitung innerhalb kürzester Zeit erzeugt werden können, lassen sich viele Fragen an den Text stellen und beantworten. Wenn z. B. von allen Textzeugen Register vorliegen, können beliebige Schnittmengen gebildet werden, wodurch sich die Verteilung des Wortschatzes auf die Textzeugen zeigen lässt. Dialektale Elemente etwa können auf diese Weise ermittelt werden. Mit alphabetischen und rückläufigen Registern ist ein Überblick über die Verteilung bestimmter Wortbildungsmittel wie Prä- und Suffigierungen möglich. Mit nach Häufigkeit oder Wortlänge sortierten Registern kann die quantitative Verteilung des Wortschatzes in den Blick genommen werden. Reimregister bilden, wo es sich um Verstexte handelt,

22 Diese Terminologie, bei der Register als Oberbegriff und Index sowie Konkordanz als Unterbegriffe definiert werden, folgt im Wesentlichen Kurt Gärtner u. Peter Kühn, Indices und Konkordanzen zu historischen Texten des Deutschen. Bestandsaufnahme, Typen, Herstellungsprobleme, Benutzungsmöglichkeiten. In: Hugo Steger u. Herbert E. Wiegand (Hgg.), Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung (HSK 2/1). 1. Teilband. 2. Auflage. Berlin, New York 1998.

23 Z. B. rückläufig alphabetisch oder nach Reimendungen.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

den zentralen Anfangspunkt für metrische, lautliche oder dialektologische Analysen.

Nicht lemmatisierte Register können jedoch nur die erste Erschließungsstufe bilden, da sowohl die Flexionsmorphologie als auch die vielfältigen Probleme der in den älteren Texten ungeregelten Orthographie dazu führen, dass die zu einem Lexem gehörigen Wortformen in einer alphabetischen Anordnung weit voneinander getrennt stehen. Ebenso sind Proklise, Enklise und weitere Formen der Wortverschmelzung ein großes Problem für diese Art von Registern. Diese Defizite lassen sich nur durch die Lemmatisierung der Texte beheben, die trotz aller Fortschritte in der elektronischen Datenverarbeitung für mittelalterliche deutschsprachige Texte noch nicht völlig zufriedenstellend automatisch erfolgen kann, sondern immer noch des Sachverstandes und der Sprachkenntnis des Bearbeiters bedarf, also immer manuell nachbearbeitet werden muss.

3.2 Lemmatisierung

Die Lemmatisierung, d. h. die Zuordnung der Wörter eines Textes zu einem Lexem unter einem bestimmten Lemma, gilt bereits seit vielen Jahren als „eine selbstverständliche Forderung an die mit dem Computer arbeitenden Lexikographen“[24] und die von ihnen erstellten Register. Wenn in Zukunft Register selbstverständlicher Teil von digitalen Editionen werden sollten, wird sich diese Forderung von den Lexikographen an die Editoren verlagern. In genuin digitalen Editionen ist es nicht sinnvoll, aus dem Text heraus ein Rohregister zu erzeugen und dieses erst in einem zweiten Schritt zu lemmatisieren. Stattdessen wird der edierte Text direkt mit Lemmainformationen angereichert (‚getaggt‘), auf deren Grundlage anschließend die Register erzeugt werden. Auch dieses Verfahren ist symptomatisch für das sich wandelnde Verständnis von Editionen.

Die Lemmatisierung dient jedoch nicht nur dem bereits oben skizzierten Zweck, verstreute Flexionsformen oder Schreibweisen eines Wortes zusammenzuführen. Lemmatisierung bedeutet immer auch zu einem gewissen Grad Normalisierung, da die Lemmata als Repräsentanten der Lexeme selbst der Normalisierung unterliegen müssen, um dem Benutzer ein systematisches Auffinden zu ermöglichen. Wenn ein Text lemmatisiert wird, impliziert dies also, dass der Text zugleich auf ein übergeordnetes Bezugssystem abgebildet wird. Im Bereich des Mittelhochdeutschen beispielsweise wird der Text auf das normalisierte Mittel-

24 Gärtner u. Kühn (Anm. 22), S. 715–743.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

hochdeutsch abgebildet, selbst wenn die Wörter des Textes selbst nicht in normalisierter Form ediert werden.[25]

Dieser Schritt potenziert die Verknüpfungsmöglichkeiten von edierten Texten mit externen Ressourcen, da nun über mehrere Texte hinweg nach Vorkommen eines bestimmten Lexems gesucht werden kann. Lemmatisierung bildet demnach die erste innere Voraussetzung für die Vernetzung von Texten. Ihre Anwendungsmöglichkeiten vergrößern sich dabei mit der Menge der lemmatisierten Texte.

In einer Edition, die mehrere Fassungen eines Textes bietet, könnten z.B. nicht nur für jede Leithandschrift oder Fassung separate Namensregister erstellt werden, sondern alternativ auch gemeinsame für eine beliebige Auswahl der Fassungen. So kann die handschriftliche Varianz im Bereich der Personen-, Orts- und sonstigen Eigennamen visualisiert und zugänglich gemacht werden. Bildlich gesprochen werden Editionen durch Lemmatisierung zu ‚Steinbrüchen' für zukünftige Wörterbuch- und Grammatikprojekte. Davon ausgehend können die erstellten Namensregister mit anderen lemmatisierten Namensregistern verbunden werden und so eine höherwertige Informationsquelle darstellen. Wenn auf diese Weise nicht nur der ‚Parzival' Wolframs von Eschenbach ausgezeichnet wäre, sondern auch die Artusromane Hartmanns, dann könnte man alle Belegstellen für die den verschiedenen Texten gemeinsamen Protagonisten wie beispielsweise Gâwân oder Îwein nebeneinander betrachten.[26]

Die Forschung im Bereich der Lemmatisierung fokussierte sich in den letzten Jahrzehnten auf die Automatisierung der Lemmatisierung mit dem Ziel, immer höhere Trefferquoten zu erzielen und den manuellen Nachbearbeitungsbedarf zu reduzieren. Aber auch hier kann mit einer Verschiebung des Forschungsschwerpunktes gerechnet werden, nämlich mit der Hinwendung zu standardisierten und in mehreren Projekten nutzbaren Lemmatisierungswerkzeugen. Damit rückten Kollaboration und Wiederverwertbarkeit von Forschungsleistungen in den Vordergrund. Voraussetzungen für solche Vorhaben sind zum einen die Verständigung der einzelnen Forschungsprojekte, das Festlegen bestimmter Standards und zum anderen die Auslagerung der für alle Projekte wichtigen Res-

25 Die bisherige editionswissenschaftliche Diskussion um die Normalisierung der edierten Texte ignorierte diesen tieferen Zusammenhang von Normalisierung und Lexikographie und betrachtete nur die Auswirkungen auf der Wortebene innerhalb konkreter Texte. Auf Lexemebene zeigt sich die Normalisierung als eine überaus nützliche Form von Standardisierung bzw. Abbildung auf Referenzwerke – in diesem Falle auf Wörterbücher. Das Potenzial der Normalisierung scheint also noch nicht vollständig ausgeschöpft zu sein.

26 Die Abbildung fiktionaler Figuren auf ein gemeinsames Nachschlagewerk kann Fragestellungen der Fiktionalitätsforschung und angrenzender Gebiete vereinfachen und erleichtern.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

sourcen.[27] So wäre es vermeidbar, dass jedes Projekt eine separate Lemmaliste verwalten müsste, stattdessen könnte eine zentrale Lemmaliste erschaffen werden, die allen interessierten Projekten zur Verfügung steht und sukzessive erweitert wird.[28]

3.3 Part-of-Speech-Tagging

Die neben den Registern und der Lemmatisierung dritte Ebene der Erschließung ist das *Part-of-Speech-Tagging* (PoS-Tagging), im Zuge dessen die kontextabhängige Wortart und gegebenenfalls Flexionsform jedes Wortes im Text analysiert und hinterlegt wird. Diese Form der Tiefenanalyse ist die Grundlage für alle weitergehenden syntaktischen Untersuchungen und bietet lexikologischen, lexikographischen sowie grammatikographischen Vorhaben wichtige Informationen. Obwohl Lemmatisierung und PoS-Tagging unabhängig voneinander betrieben werden können, lassen sie sich gewinnbringend zusammenführen: Die Disambiguierung homographer Wortformen im Bereich der Lemmatisierung lässt sich mit PoS-Daten teilweise lösen,[29] während Lemmainformationen für die Auflösung von Doppeldeutigkeiten auf der syntaktischen Oberflächenstruktur nützlich sind.

Während für neusprachliche Texte bereits hervorragende PoS-Tagger vorliegen, die Trefferquoten von weit über 90 % erreichen, sind für die historischen Sprachstufen des Deutschen auf Grund seiner orthographischen und dialektalen Vielfalt noch keine voll funktionsfähigen Werkzeuge verfügbar. Gute Ergebnisse wurden jedoch von einem Stuttgarter Team an Daten aus der MHDBDB erzielt, indem dieses Team den für neue Sprachen entwickelten *Tree-Tagger* an mittelhochdeutschen Texten trainierte.[30] Das wichtigste Beispiel für die Annotation historischer deutschsprachiger Texte sind jedoch zweifellos das ‚Referenzkorpus Mit-

27 Eine Vorreiterrolle kann hier dem Projekt ‚eHumanities-Zentrum für historische Lexikographie' (ZHistLex) zufallen, welches auf mehreren Ebenen zur Standardisierung beitragen kann. Siehe unter www.zhistlex.de (Zugriff: 23.11.2018).

28 In Sonderfällen wie dem ReM (siehe unten Kap. 3.3) hindert das nicht das Führen einer eigener Lemmaliste, sofern diese auf eine übergeordnete Lemmaliste bezogen ist. Als bislang aussichtsreichster Kandidat für eine projektübergreifende Lemmaliste kann aufgrund ihres Umfangs die des MWB betrachtet werden. Siehe unter www.mhdwb-online.de/lemmaliste.php (Zugriff: 23.11.2018).

29 Nämlich dann, wenn die Homographie wortartenübergreifend ist. Gehören die Homographen zur gleichen Wortart, hilft nur eine semantische Unterscheidung durch einen sachkundigen Bearbeiter.

30 Vgl. Nora Echelmeyer, Nils Reiter u. Sarah Schulz, Ein PoS-Tagger für das „Mittelhochdeutsche". www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/MHD_Tagger/paper.pdf (Zugriff: 23.11.2018).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

telhochdeutsch' (ReM)[31] und das darin eingeflossene ‚Korpus der mittelhochdeutschen Grammatik' (MiGraKo),[32] welches unter der Leitung von Thomas Klein mit PoS-Daten annotiert wurde.

Die über viele Jahre hinweg erarbeiteten Daten führten, wie die jüngst erschienenen Bände zur Flexionsmorphologie[33] zeigen, zu einer Präzisierung und Vertiefung unseres grammatischen Wissens über das Mittelhochdeutsche. Das ReM ist seit kurzem zwar online abfragbar und durchsuchbar,[34] aber leider wurde mit einer komplizierten, auf der Datenbank ANNIS aufsetzenden Abfragesprache[35] eine etwas unglückliche Wahl in Bezug auf die Nutzbarkeit getroffen. Da die Einarbeitung Zeit und Geduld benötigt und die Abfragen nicht intuitiv sind, kann es dazu kommen, dass einige Nutzer frustriert aufgeben, was dazu führt, dass das ‚Referenzkorpus' nicht so stark rezipiert wird, wie es das verdient hätte.

3.4 Sonstige Annotationen

Obwohl Lemmatisierung und PoS-Tagging zu den wichtigsten Erschließungsebenen gehören, die auf digitale Editionstexte angewendet werden können, sind sie keineswegs die einzigen: Metrische Analysen wie das Skandieren, literaturwissenschaftliche wie das Markieren von Topoi, Metaphern, intertextuellen Bezügen, narrativen Strukturen, um nur einige zu nennen, zeigen, dass es einige Perspektiven und Möglichkeiten gibt, deren Potenzial noch nicht ausgelotet ist und deren Kraft sich erst dann entfaltet, wenn die einzelnen Texte miteinander verknüpft sind.

Diese Annotationsebenen gehören zwar nicht zum editorischen Geschäft und müssen daher nicht von den Editoren selbst vorgenommen werden, jedoch sollten die Editionsdaten so eingerichtet werden, dass es anderen Nutzern möglich ist, weitere Annotationsebenen wie etwa die oben genannten anzulagern und so zu einer Bereicherung der Texte beizutragen. Diese Möglichkeit der Weiternutzung bzw. Nachnutzung ist ein für die Nachhaltigkeit der Forschungsdaten wichtiger Aspekt.[36]

31 www.linguistics.rub.de/rem (Zugriff: 23.11.2018).

32 Zu den Hintergründen und zur Dokumentation der Korpora siehe: Thomas Klein u. Stefanie Dipper, Handbuch zum Referenzkorpus Mittelhochdeutsch. www.linguistics.rub.de/bla/019-klein-dipper2016.pdf (Zugriff: 23.11.2018).

33 Thomas Klein, Hans-Joachim Solms u. Klaus-Peter Wegera (Hgg.), Mittelhochdeutsche Grammatik. Teil II. Flexionsmorphologie. 2 Bde. Berlin, Boston 2018.

34 www.linguistics.rub.de/annis/annis3/REM (Zugriff: 23.11.2018).

35 www.corpus-tools.org/annis/ (Zugriff: 23.11.2018).

36 An dieser Stelle sei darauf hingewiesen, dass für eine solche Vorgehensweise unsere bisherige Angewohnheit, die Texte ‚inline' auszuzeichnen, d.h. die Tags unmittelbar in den laufenden Text

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

4 Nachhaltigkeit und Standardisierung

4.1 Ausgewählte Aspekte der Nachhaltigkeit[37]

Der Begriff der Nachhaltigkeit hat in den letzten Jahrzehnten zwar vor allem durch gesellschaftspolitisch relevante Diskurse in der Ökonomie und Ökologie an Bedeutung gewonnen, er ist jedoch auch für die an Texten orientierten Wissenschaften wichtig, wenngleich mit anderem Schwerpunkt: Für die Wirtschaft ist der Aspekt der Regeneration zentral, d.h. es dürfen nicht mehr Ressourcen verbraucht werden, als regeneriert werden können; für die Textwissenschaften hingegen steht im Vordergrund, dass geleistete Forschungsarbeit einerseits eine lang anhaltende, dauerhafte Wirkung erzielt und andererseits leicht und unter nur geringem Aufwand zur Grundlage weiterer Forschungsarbeit gemacht werden kann. Gerade Editionen sind einerseits aufgrund ihrer hohen Erstellungskosten und andererseits wegen ihrer Ausrichtung als Grundlagenforschung keine Wegwerfprodukte: Es gilt, ihre Nachnutzbarkeit und damit ihre Nachhaltigkeit zu optimieren.

In diesem Zusammenhang liegt zunächst der Fokus auf der Langzeitarchivierung, d.h. der Frage, wie Forschungsdaten in den ständigen Wandlungen unterworfenen digitalen Medien dauerhaft verfügbar gemacht werden können.[38] Die vielen Datenträgertypen, die das 19. und 20. Jahrhundert hervorbrachte: Lochkarten, Magnetbänder, Disketten, CDs, DVDs und dergleichen, haben sich als nicht dauerhaft erwiesen. Die Datenträger verwitterten, die Daten wurden unlesbar, sofern sie nicht aufwändig von einem Datenträger auf den nächsten kopiert wurden, und so ging wert- und mühevolle Arbeit verloren. Die Speicherung von Forschungsdaten auf einzelnen magnetischen (Magnetband) oder optischen (CD, DVD) Trägern ist weitgehend abgelöst worden durch die elektronische Speicherung auf Festplatten, insbesondere auf mit dem Internet verbundenen Servern

zu schreiben, nicht besonders geeignet ist. Gerade wenn mehrere Annotationsebenen sich überlappen oder Mehrdeutigkeiten kodiert werden müssen, kommt auf XML basierendes *Inline-Markup* an seine Grenzen. Zu erwägen wäre daher die weitgehende Umstellung auf Stand-Off-Markup, welches größere Flexibilität bei der parallelen Auszeichnung konkurrierender Ebenen bietet.

37 Aufgrund der Komplexität des Nachhaltigkeitsbegriffs kann hier nur auf wenige ausgewählte Aspekte eingegangen werden. Rechtliche Fragen, wie Lizenzierung und Urheberrecht, technische Fragen, wie Versionierung und Zitierfähigkeit, und dergleichen mehr können hier nicht behandelt werden.

38 Zu verschiedenen Aspekten von Langzeitarchivierung vgl. Heike Neuroth u.a. (Hgg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf (Zugriff: 23.01.2019).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

(Stichwort: *cloud backup*).[39] Die Sicherung der Daten wird heutzutage dadurch gewährleistet, dass die Daten auf mehreren Servern vorgehalten und miteinander automatisch synchronisiert werden, sodass ein Ausfall eines oder mehrerer Computer durch die Sicherungskopien der jeweils verbleibenden Computer kompensiert werden kann. Dadurch ist aus technischer Sicht die Langzeitarchivierung prinzipiell gewährleistet – zumindest was die Hardware betrifft.

Einen zweiten Problembereich stellen die Lesbarkeit und Darstellbarkeit der Daten durch Software dar. Bis zur Einführung der *Standard General Markup Language* (SGML) in den 1980er Jahren wurden vorwiegend Programme eingesetzt, welche die Daten auf ganz bestimmte Weisen formatierten, sodass die Dateien nur durch Programme des gleichen Herstellers gelesen werden konnten (sogenanntes proprietäres Dateiformat),[40] wodurch eine nicht nur finanzielle, sondern auch essentielle Abhängigkeit von der verwendeten Software entstand. Wurde eine Software nicht mehr weiterentwickelt und ließen sich die Programme nicht mehr starten, waren die Dateien oft nicht mehr zu öffnen oder die Daten wurden beim Öffnen mit anderen Programmen beschädigt.

Um diesem softwareseitigen Datenverlust entgegenzuwirken, hat sich die deskriptive Annotation der Texte mithilfe softwareunabhängiger Auszeichnungssprachen (zunächst SGML, danach XML, heute z. T. JSON) eingebürgert. Die Daten werden als *plain text* gespeichert, sodass sie sich mit jedem beliebigen Textprogramm öffnen und lesen lassen, wobei die Daten nicht unmittelbar alle Angaben zur ihrer Darstellung enthalten müssen. Während anfänglich viele Auszeichnungssprachen wie HTML darauf abzielten, die Darstellungsinformationen durch die Tags auszudrücken, geht die Tendenz mehr und mehr zu einer semantischen Beschreibung der Textstruktur und einer davon getrennten Zuordnung der einzelnen Strukturelemente innerhalb sogenannter Stylesheets, d. h. Dateien, welche angeben, welches Element wie auf dem Bildschirm (oder auf dem Drucker) ausgegeben werden soll.[41]

So ist weitgehend sichergestellt, dass die Daten langfristig gelesen werden können, allerdings ist damit der zweite Aspekt der Nachhaltigkeit, die dauerhafte

39 Vgl. hierzu einen kurzen Überblick über die Entwicklung der Medien bei Jörg Hörnschemeyer, Textgenetische Prozesse in Digitalen Editionen. Diss. Köln 2013, S. 28–35.

40 Dieses Prinzip findet sich heute noch bei Office-Programmen wie Microsoft Office, weswegen sich diese Systeme nicht für die dauerhafte Speicherung, geschweige denn für die Erstellung von Forschungsdaten eignen. Trotz der Einführung der auf Office Open XML basierenden Dateiformate .docx, .xslx und .pptx, die einen besseren Austausch verschiedener Office-Suiten ermöglichen sollten, ist ein problemloser Datenaustausch nicht gewährleistet.

41 Zur Entwicklung des deskriptiven Markups vgl. Susan Schreibman, Ray Siemens u. John Unsworth (Hgg.), A Companion to Digital Humanities (Blackwell companions to literature and culture 26). Malden MA 2004, Kap. 17.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

Weiternutzung, noch nicht erreicht. Um nach Abschluss eines Editionsvorhabens sicherzustellen, dass die Texte weiterhin genutzt und in neue Zusammenhänge eingebunden werden können, ist es nötig, sie an bestimmte Standards anzupassen.[42]

4.2 Standardisierung

Zu den größten Herausforderungen der Textwissenschaften, also auch der Editionswissenschaft, gehört somit die standardisierte Auszeichnung ihrer Forschungsdaten wie zum Beispiel mithilfe von Unicode.[43] Wenn die Verknüpfung unabhängig voneinander ausgezeichneter Editionen zum Zwecke gemeinsamer Forschungsfragen gelingen soll, dann müssen die verwendeten Auszeichnungen miteinander vergleichbar oder zumindest aufeinander beziehbar sein. Die Etablierung eines Standards in diesem Zusammenhang soll daher zwar eine gewisse formale Normierung bewirken, doch sollen unterschiedliche Anschauungen der Forschenden sowie verschiedenartige Ansätze und Herangehensweisen weiterhin möglich sein, damit die Wissenschaft pluralistisch und vielfältig sein kann. Alle Standards in editionswissenschaftlichen Diskursen müssen deswegen einerseits stabil sein, um die langfristige Vergleichbarkeit der mit ihnen ausgezeichneten Daten zu gewährleisten, andererseits müssen sie sich sowohl den jetzigen als auch den zukünftigen Bedürfnissen der Forschungsgemeinschaft anpassen können.

Mit der Einrichtung der *Text Encoding Initiative* (TEI)[44] und den von dieser Organisation seit den frühen 1990er Jahren herausgegebenen *Guidelines* liegt ein mittlerweile umfangreiches System von Auszeichnungsmöglichkeiten vor, welches international Anerkennung gefunden hat und oftmals als De-facto-Standard betrachtet wird.[45] Um den oben skizzierten, divergierenden Anforderungen gerecht zu werden, bieten sie mehrere Möglichkeiten zur Auszeichnung bestimmter textueller Phänomene an; zudem werden die Richtlinien seit ihrer ersten Veröffentlichung kontinuierlich weiterentwickelt und erweitert. Auf diese Weise ermöglichen die TEI-Richtlinien zwar vergleichbare Auszeichnungen, ohne sie

42 Vgl. das Kapitel Digitale Edition. In: Fotis Jannidis, Hubertus Kohle u. Male Rehbein (Hgg.), Digital Humanities. Eine Einführung. Stuttgart 2017, S. 234–249.

43 www.unicode.org (Zugriff: 23.11.2018).

44 www.tei-c.org/index.xml (Zugriff: 23.11.2018).

45 Zum Thema TEI vgl. Lou Burnard, What is the Text Encoding Initiative? How to add intelligent markup to digital resources (Encyclopédie numérique 3). Marseille 2014.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

gleichzeitig zu erzwingen, aber gerade diese Dynamik und Offenheit hat ihre Schwierigkeiten.

Wenngleich sich die TEI-Richtlinien als ein nützliches Werkzeug herausgestellt haben, bereitet ihre Anwendung gewisse Probleme: Alle Projekte, die ein gemeinsames Netzwerk von Forschungsdaten erschaffen möchten, sind zu einem ständigen Dialog über die Anwendung der Richtlinien gezwungen, in dessen Verlauf sie Substandards entwickeln, d.h. eine Auswahl aus den angebotenen Möglichkeiten treffen und diese kanonisieren müssen. Andernfalls laufen die Projekte Gefahr, dass die jeweils gewählten Auszeichnungen nicht zu jenen anderer Projekte passen und langwierige Anpassungen und Nachbearbeitungen nötig sind, damit Vergleichbarkeit hergestellt werden kann.[46]

Die Erfahrungen aus dem ‚Parzival'-Projekt, das 2001 mit projektinternem Markup gestartet ist, zeigen zudem, dass eine nachträgliche Konvertierung von einem Markup in ein anderes wie TEI zwar möglich, aber zeitaufwändig ist.[47] Mitunter sind auch Kompromisse nötig, wenn die angewendeten Auszeichnungsmethoden nicht eins zu eins übertragen werden können. Darum sollte gerade in den Planungsphasen neuer Editionsprojekte unbedingt Zeit für die Ausarbeitung eines möglichst standardnahen Markups und für Überlegungen zur technischen Umsetzung investiert werden.

5 Fazit

Der vorliegende Beitrag hat gezeigt, dass sich Editionen und ihr Verständnis im Zuge der Digitalisierung verändern. Die Anforderungen an die edierten Texte steigen und ihre Komplexität nimmt zu; bedingt durch neue technische Möglichkeiten werden die Editionen vielschichtiger. Vor allem die Bereiche Verknüpfung und Erschließung zeigen vielversprechende Perspektiven auf, während Langzeitarchivierung und Standardisierung besondere Herausforderungen mit sich bringen. Indessen können diese Aufgaben nicht von einzelnen Projekten gelöst wer-

46 Auf einem TEI-Workshop zu Altgermanistischen TEI-Kodierungsstrategien, der 2017 vom Team der Neuedition des ‚Welschen Gastes' ausgerichtet wurde und an dem Forscher verschiedener Editionsprojekte teilnahmen, stellte sich heraus, dass Konsensfindungen in Bereich der Kodierungen oftmals schwierig sind. Es gibt noch einigen theoretischen und praktischen Klärungsbedarf, bevor gemeinsame Auszeichnungsempfehlungen, die speziell für mittelhochdeutsche Texte geeignet sind, ausgesprochen werden können.

47 Vgl. auch Gabriel Viehhauser, Standardisierung und proprietäre Annotation im Berner ‚Parzival'-Projekt. computerphilologie.digital-humanities.de/jg09/viehhauser.html (Zugriff: 23.11. 2018).

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40

den; sie bedingen vielmehr die Kooperation und den Dialog der Forschenden untereinander.

Durch den sich anbahnenden Wandel von Editionen als derzeit noch für sich stehenden Forschungsleistungen hin zu umfassenderen Forschungsumgebungen ist zu erwarten, dass die einzelnen Fachteile künftig engere Verbindungen eingehen werden: Editionswissenschaft, Lexikographie, Grammatikographie, Dialektologie, Computerphilologie, Informatik und andere Fächer werden als Digital Humanities gemeinsam an den Editionstexten arbeiten und dadurch einen neuen Blick auf die mittelalterlichen deutschen Texte, ihre Überlieferung und Sprache erlauben.

Angemeldet | michael.stolz@germ.unibe.ch Autorenexemplar
Heruntergeladen am | 11.08.19 12:40