

## Altgermanistische TEI-Kodierungsstrategien

**Veranstalter:** Judith Lange (Institut für Germanistik/Mediävistik, Universität Duisburg-Essen); Jakub Šimek (Sonderforschungsbereich 933 „Materiale Textkulturen“, Universität Heidelberg); Karin Zimmermann (Universitätsbibliothek Heidelberg); Deutsche Forschungsgemeinschaft

**Datum, Ort:** 16.03.2017–17.03.2017, Heidelberg

**Bericht von:** Judith Lange, Institut für Germanistik/Mediävistik, Universität Duisburg-Essen / Jakub Šimek, Sonderforschungsbereich 933 „Materiale Textkulturen“, Universität Heidelberg

In den Jahren seit der Gründung der Text Encoding Initiative (TEI) hat sich das gleichnamige, modular aufgebaute Dokumentenformat zur Kodierung und zum Austausch von Texten zum Standardmarkup innerhalb der Geisteswissenschaften entwickelt. Dennoch zeigte sich in den letzten Jahren zunehmend, dass die flexibel gehaltene Modularisierung, die eine sehr differenzierte Auszeichnung von Textmerkmalen ermöglicht, selbst innerhalb einzelner Fachdisziplinen zu unterschiedlichen, manchmal sogar gegenläufigen Nutzungen einzelner Elemente und Attribute führt und somit Standardisierungsprozessen auch im Wege stehen kann. Der zweitägige Workshop sollte daher dem wissenschaftlichen Austausch über Verfahren und Umgang mit TEI-Kodierungsstrategien in explizit altgermanistischen Projekten dienen. Rund 25 Expertinnen und Experten aus dem Bereich der digitalen altgermanistischen Editionswissenschaften kamen an der Universitätsbibliothek Heidelberg zusammen, um nach Lösungen für bestehende Schwierigkeiten zu suchen, voneinander zu lernen und einen Entwurf für ‚Empfehlungen zur Kodierung altgermanistischer digitaler Editionen in TEI‘ anzustoßen.

### TEI-Export aus proprietären Markups

Ein Fokus des Workshops lag auf dem Nutzen, aber auch auf den Schwierigkeiten der Arbeitsmethodik in Editionsprojekten, die sich zunächst eines proprietären Markups be-

dienen und in denen TEI erst am Ende als Export-Format zum Einsatz kommt. Anhand der beiden altgermanistischen Großprojekte ‚Parzival-Projekt‘ und ‚Lyrik des deutschen Mittelalters‘ (LDM) und der Vorträge von CHRISTIAN GRIESINGER (Bern) und SONJA GLAUCH (Erlangen) zeigte sich dabei, welche Auswirkungen Alter und Laufzeit eines Projekts auf die Versuche einer Anpassung an TEI haben können. Dass ein Export bzw. eine Konvertierung nach TEI vollautomatisch ablaufen müssten, betonten sowohl Glauch als auch Griesinger in ihren Werkstattberichten.

Die Anfänge des ‚Parzival-Projekts‘ und auch die seines proprietären Markups liegen in den frühen 2000er-Jahren und fallen somit in eine Zeit, bevor sich XML oder TEI als Standard durchsetzten. Diesem Umstand sei es laut Griesinger – mit Berufung auf Gabriel Viehhauser<sup>1</sup> – geschuldet, dass eine Umstellung auf die ausschließliche Verwendung der TEI-Richtlinien im Fall des Langzeitprojekts weder sinnvoll noch möglich sei. Als einen Problemfall stellte Griesinger die im proprietären Markup des Projekts gewählte Wort-für-Wort-Auszeichnung von Schreiber- und/oder Leser-Korrekturen vor, die zumindest in komplexeren Fällen nicht automatisch in das bevorzugte TEI-Format konvertiert werden könnten. Zwei Lösungsmöglichkeiten wurden erläutert: Zum einem könnten die Korrekturstellen in den bereits vorhandenen Transkriptionen nochmals aufgesucht und das Markup dahingehend modifiziert werden, dass auch die einzelnen korrigierten Zeichen erfasst würden. Damit entstünde eine Kompromisslösung, die mittels geeigneter Skripte ins TEI-Format konvertiert werden könnte. Allerdings müssten hierfür über 10.000 Einzelstellen im Parzival-Corpus bearbeitet werden, so dass der dabei anfallende Arbeitsaufwand diese Lösung ausschließe. Weniger zeitintensiv sei demgegenüber die Möglichkeit, die TEI-spezifischen Tags zu modifizieren und in neuen Kontexten zu verwenden. Trotz der Tatsache, dass durch solche Annotationen die ursprünglich inten-

---

<sup>1</sup>Gabriel Viehhauser: Standardisierung und proprietäre Annotation im Berner Parzival-Projekt. Online einsehbar auf: <http://computerphilologie.digital-humanities.de/jg09/viehhauser.html> (11.07.2017).

---

dierten Semantiken der TEI-Richtlinien nicht mehr gewahrt blieben, tendierte Griesinger zu dieser zweiten Lösung. Er betonte jedoch, dass eine solche Modifikation nur bei ausreichender Dokumentation im Header erfolgen dürfe – eine Aussage, die in der anschließenden Diskussion zu einer Debatte um die <editorialDecl> (editorial practice declaration) führte. Gewünscht wurde ein eher systematisierter Aufbau, der die bisher in Prosa anzugebenden Editionsprinzipien schematisiert. Angesichts der Komplexität, die TEI-Dokumente erreichen können, wies Griesinger darauf hin, dass die Auszeichnungsmethode des Stand-Off-Markups dabei helfen könnte, verschiedene Perspektiven auf ein Dokument anzuwenden und die Komplexität innerhalb des Dokuments zu verringern.

Der Vortrag von Sonja Glauch trug im Unterschied zu den Ausführungen Griesingers zur Konvertierung der Transkriptionsdateien des ‚Parzival-Projekts‘ der Tatsache Rechnung, dass die TEI im LDM-Projekt in einer einzigen Datei der Abbildung vielschichtiger Editionsprozesse (Transkription, leicht normierte Lesefassung und vollständig normalisierte Fassung) dienen soll. Als Ziel des LDM-Projektes formulierte Glauch die Vielfalt der mittelalterlichen Liedüberlieferung direkt für den Nutzer sichtbar und lesbar zu machen, wobei dank der Hilfe moderner digitaler Datenbankmöglichkeiten anders als in Printeditionen die Lesefreundlichkeit keineswegs gegen Überlieferungsnähe aufgewogen werden müsse. Die Problematik der direkten Umsetzung einer solchen Edition in TEI beschrieb Glauch damit, dass die Leistung der Auszeichnungssprache eigentlich darin liegen müsste, Arbeitsabläufe zu unterstützen und nicht, diese zu erschweren. XML-Strukturen zielten jedoch grundsätzlich auf maschinelle Weiterverarbeitung und seien für menschliche Leser eher „ungeeignet“ – folgerichtig stellte Glauch die unweigerliche aber nicht abschließend geklärte Frage nach dem „Nutzer“ reiner TEI-Dokumente und der Anwenderfreundlichkeit bei ihrer Erstellung. Da gerade Letzteres Wünsche offen ließe, wurde bei der Erarbeitung des proprietären Markups vor allem Wert auf die Anwenderfreundlichkeit gelegt. Gleichzeitig aber wurde laut Glauch auf eine Anlehnung an gän-

gige XML-Strukturen geachtet, so dass sich das Markup möglichst vollautomatisch in TEI exportieren lasse – als problematische Ausnahme benannte Glauch an dieser Stelle beschreibende Apparat-Einträge, die nicht ohne weiteres in schematische, strukturelle TEI-Schreibung überführt werden können. TEI als ‚output‘-Format aber werde, so auch Glauchs Einschätzung, grundsätzlich durch die vielfältigen Auszeichnungsmöglichkeiten (zum Beispiel von Abkürzungen, Korrekturen oder Initialen) erschwert, auf die auch Griesinger in seinem Vortrag explizit einging. Beide Vortragende stellten unterschiedliche Kodierungsmöglichkeiten vor, erwogen diese und stellten sie zur Diskussion. In der anschließenden Diskussion kamen zwei Punkte auf, deren Wichtigkeit im Verlauf des weiteren Workshops sich immer deutlicher herauskristallisierte: Die wünschenswerte, wenn nicht sogar notwendige Vermeidung von durch Tagging-Optionen verursachte Redundanzen sowie die Frage nach der Trennung unterschiedlicher Text- und Interpretationsebenen in einer Datei. Letzterem Punkt widmete JAKUB ŠIMEK (Heidelberg) in seinem Vortrag zum Projekt ‚Welscher Gast digital‘ besondere Aufmerksamkeit.

### **Genuin mit TEI ausgezeichnete altgermanistische Editionen**

Auch in diesem Werkstattbericht standen projektspezifische Auszeichnungsprobleme im Fokus. Anders als im ‚Parzival-Projekt‘ und LDM-Projekt arbeitet ‚Welscher Gast digital‘ jedoch direkt mit TEI als Basisformat für die Textauszeichnung. Die Besonderheit liegt darin, dass das Projekt mit einer physischen, die Handschriften beschreibenden „codexbasierten Sicht“ und einer textuellen „werkbasier-ten Sicht“ operiert. Beide Perspektiven werden in jeweils eigenen Dateien abgebildet und beide Orientierungen bergen dabei, so Šimek, unterschiedliche Problematiken. Die aus der Werk-Ebene automatisiert exportierte Codex-Orientierung ermögliche grundsätzlich weniger Anreicherung durch Interpretation, da sie zunächst auf eine Abbildung dessen, was im Codex sichtbar ist, fokussiert sei. Konkret sprach Šimek die Schwierigkeiten an, in der Codex-Orientierung Phänomene darzustellen, deren reguläre TEI-

Elemente als Inhalt von `<line>` (physische Zeile) nicht vorgesehen bzw. erlaubt sind, etwa Ersatzfälle mit dem `<subst>`-Element. Ebenfalls wurden Kodierungsvorschläge für eine Reihe von Zeichen erörtert, die in mittelalterlichen Handschriften häufig anzutreffen sind, etwa Umstellungs- und Einfügungszeichen (`<metamark>`), Paragraphenzeichen und Kapitelzahlen (zur Diskussion stellte Šimek `<pc>` und `<label>`). Intensiv diskutiert wurde im Anschluss an den Vortrag, wie die Zwitterfunktion von Initialen zwischen Buchschmuck und Gliederungseinheit am besten wiedergegeben werden könne (man plädierte für `<hi>` mit CSS-ähnlichen Spezifikationen im Attribut `@rend`).

Eine fächerübergreifende und dennoch sehr auf das Fach Germanistik bezogene Perspektive auf die TEI und ihren Einsatz in Editionen lieferte LEONHARD MAYLEIN (Heidelberg). Sein Blick richtete sich auf die Heidelberger Digitalisierungsinfrastruktur DWork und den Import aus unterschiedlichen Dateiformaten mittels XSLT ins Backend der Infrastruktur. Maylein bezeichnete vor allem proprietäre Markups als problematisch für eine gezielte Aufbereitung für die Visualisierung in bestehenden Infrastrukturen. Er betonte aber ebenso, dass auch zwischen den TEI-Daten einzelner Projekte wenig Kohärenz herrscht und sich mit der Visualisierung von XML-Daten befaste Informatiker beständig mit neuen Auszeichnungen auseinandersetzen hätten. Der längerfristige Status der TEI-Daten innerhalb der Infrastruktur sei noch ungeklärt; von der automatischen Langzeitarchivierung, die im Rahmen eines inzwischen etablierten Workflows für digitale Bilder angewandt wird, seien die TEI-Daten momentan noch ausgeschlossen. Konkret auf ‚Welscher Gast digital‘ bezogen erläuterte Maylein die starke Handschriftenfokussierung der Visualisierung (dargestellt werden können zum Beispiel Lagenpläne und unterschiedliche handschriftliche Darstellungsmodi), die die bereits von Šimek vorgestellte Codex-Orientierung der TEI-Auszeichnung erfordert (mit `<surface>`-Elementen für einzelne Seiten).

Das von ROBERT FORKE und MAE LYONS-PENNER (Stanford, USA) vorgestellte ‚Global Medieval Sourcebook‘ (GMS) er-

arbeitet seine Daten ebenfalls direkt in TEI; die Texte werden allerdings nur sehr rudimentär ausgezeichnet. Dass nur sehr basales TEI Verwendung findet, sei – so Forke und Lyons-Penner – darauf zurückzuführen, dass hier keine Textedition im Sinne der übrigen im Workshop diskutierten Projekte erstellt werde. Auf Basis älterer, Copyright-freier Editionen werden Übersetzungen bisher nicht in moderner englischer Übersetzung vorliegender mittelalterlicher Texte unterschiedlicher Kultur- und Sprachräume angefertigt und im Open Access auf einer Internetplattform verfügbar gemacht. Mit Blick auf die Benutzung des GMS im Rahmen universitärer Lehre sollen die Übersetzungen zeilen genau neben dem Editionstext präsentiert und durch Links zu Handschriftendigitalisaten, Textkommentare und Audioaufnahmen der Originaltexte ergänzt werden. Auf gewisse Kritik stieß die vom Projekt gewählte Methode, die originalen Editionstexte und die Übersetzungen innerhalb einer Datei mit `<lem>` (lemma) bzw. `<rdg>` (reading) auszuzeichnen. Dies suggeriere – in Anlehnung an die FRBR-Gedankenwelt – eine Beziehung zwischen einer ordnenden Entität und deren Varianten, die so zwischen Originaltext und Übersetzung streng gesehen nicht gegeben ist. Allerdings wurde zugegeben, dass die gewählte Lösung durchaus ihren Zweck erfüllt.

Zuletzt wurde der Vortrag HELMUT KLUGs (Graz), der aus Krankheitsgründen nicht persönlich anreisen konnte, als Video abgespielt. Als Altgermanist und Mitarbeiter am Grazer Zentrum für Informationsmodellierung beschrieb Klug zunächst seine Rolle als Vermittler zwischen den digitalen Geisteswissenschaften und den Fachwissenschaften, speziell der Mediävistik. Als Ziel seiner Arbeit benannte er den schrittweisen Aufbau von Strategien, die während der Editionsarbeit anfallenden Arbeitsabläufe an eine durchgehend digitale Arbeitsumgebung anzupassen. Die Grundlage der Editionsarbeit ist dabei grundsätzlich die ‚Grazer dynamische Editions-methode‘. Die Datenerfassung erfolgt, so Klug, über eine Mischung aus proprietärem Markup und TEI-Markup, die in Anlehnung an die Editions-methode ‚dynamische Modellierung‘ getauft wurde und deren Ziel die Darstellung der formalen

---

und optischen Erscheinung der Quelle ist. Drei Arbeitsansätze stehen dabei im Fokus: 1. die Transkription, 2. einzelne Forschungsaspekte (Schreiberhände, Textgenese) und 3. über die Transkription hinausgehende Annotationen. An vier Beispielen stellte Klug diese Arbeitsansätze dar, wobei er dezidiert auf Neuerungen gegenüber der bereits 2008 erstellten Hugo-von-Monfort-Ausgabe Wernfried Hofmeisters einging, deren `<editorialDecl>` noch eine genaue Beschreibung aller in einer Handschrift vorkommenden diakritischen Zeichen und Superskripte in Prosa umfasste. Im Unterschied hierzu werde in den neueren Projekten auf vollständige Zeichennotierung verzichtet und der Fokus auf das Sonderinventar einer Handschrift (Kürzungen, Korrekturen, Superskripte) gelegt. Eine Angabe schreiberhandspezifischer Leitgrahpe mittels Unicode-Standard in der `<charDecl>` ersetze das Graphinventar der Montfort-Ausgabe. In der Transkription werde nun auf normalisierte Formen zurückgegriffen, die mittels der in der `<charDecl>` festgelegten Markups spezifiziert würden (beispielsweise `<g ref="#u-acute">u</g>` für ú). Die Vorteile einer so erarbeiteten Ontologie liegen für KLUG in den Möglichkeiten formalisierter Beschreibung, Systematisierung, Semantisierung und Nachnutzung. Ebenfalls auf ein kontrolliertes Vokabular zielen die Grazer Bemühungen um die Entwicklung einer prozessorientierten Typologie mittelalterlicher Textrevisionsphänomene. Im Zentrum stehen hier die Modellierung und die Beschreibung sogenannter Textrevisionen in Schreibprozessen des MA mittels TEI. Klug stellte die aus dem Projekt erwachsene umfangreiche Typologie samt Vorschlägen zum Markup einzelner Phänomene vor. Im Zusammenhang zweier unterschiedlicher Beispiele für Transkriptions-Workflows warf Klug abschließend die schon von Sonja Glauch formulierte Frage auf, wieviel TEI den Mitarbeitern altgermanistischer Editionsprojekte zuzumuten sei. Klug kam zu dem Schluss, dass die Vor- und Nachteile beider Workflows sich gegenseitig aufhoben. Allerdings betonte er, dass gerade die mangelnde Übersichtlichkeit des XML-Codes bei der Arbeit direkt in Oxygen für weniger geschulte Mitarbeiter ein

Problem darstelle – auch wenn der Author-Modus eine gewisse Erleichterung vor allem bei Kollationierungsarbeit böte. Als Fazit zog er den Schluss, die TEI biete für alle Grazer Editionsprojekte ausreichend Spielraum für die Daten-Modellierung. Als wünschenswert wurde nochmals der Auf- und Ausbau fachwissenschaftlicher Vokabularien, Thesauri und/oder Ontologien geäußert. Obwohl sich das Plenum über den Nutzen einer einheitlichen Typologie für handschriftliche Korrekturvorgänge grundsätzlich einig war, wurde die von Klug vorgestellte Typologie in der Diskussionsrunde aufgrund ihrer stark interpretativen Auslegung ebenso wie einzelne Kodierungsvorschläge – besonders eine nicht TEI-konforme Verwendung von `<abbr>` – durchaus kontrovers diskutiert.

In der Abschlussdiskussion wurde darüber beraten, wie gemeinsame ‚Empfehlungen zur Kodierung altgermanistischer Texte in TEI‘ vielleicht auszusehen hätten bzw. was sie beinhalten müssten. Grundsätzlich war man sich einig, dass Empfehlungen sich in erster Linie auf eine ‚Basis‘-Kodierung von Transkriptionen zu beziehen hätten, die vor allem den handschriftlichen Befund wiederzugeben habe. Diese sollte dann durch Vorschläge zu interpretativen Element-Klammern ergänzt werden, die den Befund quasi als Container umschlössen und den Gegenstand konkretisierten. Auf Basis der gehörten Workshopberichte und Vorträge kristallisierten sich fünf maßgebliche Punkte heraus, die eine ausführlichere Beschäftigung in den geplanten ‚Empfehlungen zu altgermanistischen Kodierungsstrategien in TEI‘ unbedingt benötigen: 1. Korrekturen, 2. Abbriviaturen, 3. Initialen und `<metamark>s`, 4. Ebenentrennung in Editionen bzw. bei Textversionen und schließlich 5. Systematisierung der ‚editorial practice declaration‘.

#### **Konferenzübersicht:**

Jakub Šimek / Judith Lange: Begrüßung und Eröffnung

Sonja Glauch: TEI im Projekt „Lyrik des deutschen Mittelalters“

Jakub Šimek: TEI-Herausforderungen im Projekt „Welscher Gast digital“

Leonhard Maylein: Fächerübergreifende TEI-

Visualisierung im Rahmen der Heidelberger Digitalisierungsinfrastruktur DWork

Robert Forke / Mae Lyons-Penner: Mittelalterliche Texttraditionen digital zusammenführen mit TEI. Standfords „Global Medieval Sourcebook“

Christian Griesinger: Transformation proprietären Markups in TEI-konformes XML – Probleme und Perspektiven

Helmut Klug: Grazer Strategien zur TEI-Kodierung altdeutscher Texte [Videoübertragung]

Tagungsbericht *Altgermanistische TEI-Kodierungsstrategien*. 16.03.2017–17.03.2017, Heidelberg, in: H-Soz-Kult 29.08.2017.