

Datenmodell “Hyperdiplomatische Transkription”

Klug, Helmut W.; helmut.klug@uni-graz.at / Böhm, Astrid; astrid.boehm@uni-graz.at

Eine hyperdiplomatische Transkription versucht, die historische Quelle möglichst detailreich bis hin zur Teilzeichenebene (z. B. Superskripte) bzw. unter Berücksichtigung der Quellentopographie (Verortung der Informationseinheiten in einem digitalen Abbild der Quelle) in ein modernes Zeichensystem zu übertragen. In Grazer Projekten (*Mittelalterlabor*, *Cooking recipes of the Middle Ages*) wird nach einer hyperdiplomatischen Transkriptionsmethode gearbeitet, die auf die “Grazer dynamische Editionsreihe” (Hofmeister-Winter 2003) zurückgeht: Das Ziel ist eine graphematisch möglichst fein differenzierte Transkription, eine sogenannte “deskriptive Transkription” (Feigs 1979, Teil 1) bzw. “Basistransliteration” (Hofmeister-Winter 2003, S. 101), die sowohl sprach- und literaturwissenschaftlichen als auch geschichtswissenschaftlichen Ansprüchen genügt (Böhm/Klug 2020). Das im Anschluss vorgestellte Modellierungsmodell bildet diese Absicht ab.

Jegliches Zeicheninventar, das nicht mit Elementen der ASCII-Code-Chart dargestellt werden kann, wird mithilfe der TEI innerhalb der `<encodingDesc>` im Detail beschrieben und über das `<g>`-Element in den Fließtext des TEI-XML eingebunden. Mittels des TEI-XML-Textmodells wird die historische Quelle deskriptiv dargestellt: Seiten-, Spalten- und Zeilenlayout werden als makrostrukturelle Elemente der Quelle erhalten. Als Elemente der Mikrostruktur werden Überschriften, Initialen, Rubrizierungen, Strichelungen, Unterstreichungen etc. mit den entsprechenden TEI-Elementen und Attributwerten ausgezeichnet. Marginalien werden, wenn sie Teil des Haupttextes sind, ihrem Charakter nach annotiert (z. B. als Einfügung) und an der passenden Stelle in den Text integriert oder mit dem Element `<note>` kommentiert. Kustoden werden mit dem Element `<note>` modelliert und mit einem beschreibenden Kommentar versehen. Auf Wort- und Graphenebene werden Revisionen, Abkürzungen sowie editorische Rekonstruktionen mit den entsprechenden TEI-Elementen modelliert. Der Quellentext wird teilzeichengetreu wiedergegeben. Obwohl man eine Differenzierung der Modellierung sehr tiefgreifend anlegen könnte, sollte bei allen Projekten auf die Zweckmäßigkeit der Annotierung geachtet werden: Wenn gewisse Details (z. B. die konkrete Ausführung eines Superskripts tilde-, strich-, punktförmig usw.) nur eingeschränkt in die Transkription übernommen werden, sollten das Datenmodell sowie die Annotation zumindest das Auffinden und eine spätere Feindifferenzierung dieser Daten ermöglichen.

Im Rahmen der `<charDecl>` des TEI-Dokuments wird ein für die Quelle typisches Zeichensonderinventar angelegt: Die Zeichen in der *Character Declaration* werden auf Basis eines de facto-Community-Standards – der *Medieval Unicode Font Initiative* (MUFI) – beschrieben. Die Sonderinventarliste gliedert sich dabei in alphabetische Zeichen, Diakritika, Kürzungszeichen und Interpunk-

tionszeichen und folgt bei der Hierarchisierung der Einträge weitestgehend dem Beschreibungsmodell von Handschrift, wie es im Rahmen des DigiPal-Projekts entwickelt wurde. (Stokes 2011)

In der praktischen Umsetzung mit TEI-XML kann das Element `<charDecl>` als Kindelemente `<char>` haben; mithilfe des Elements `<glyph>` wird der Allograph näher beschrieben: So wird als *Character* das „LATIN SMALL LETTER I“ als `<char>`; geführt, als Allograph das „LATIN SMALL LETTER DOTLESS I“ `<glyph>`. Die unserem Modell zugrunde liegende Hierarchie kann aufgrund der Einschränkungen der TEI (`<char>`- und `<glyph>`-Elemente können nur auf gleicher Ebene und nicht verschachtelt geschrieben werden) nur linear und nicht hierarchisch realisiert werden. Eine Zeichengruppe wird in der *Character Declaration* folgendermaßen beschrieben:

```
<char xml:id="i">
  <charName>LATIN SMALL LETTER I</charName>
  <mapping/>
</char>
<glyph corresp="#i" xml:id="inodot" ana="allograph"
resp="https://bora.uib.no/handle/1956/10699" source="p.48">
  <glyphName>LATIN SMALL LETTER DOTLESS I</glyphName>
  <mapping type="normalized">i</mapping>
  <mapping type="transcription">i2</mapping>
  <mapping type="unicode_codepoint" subtype="LatExtA">0131</mapping>
  <mapping type="encoding" subtype="html_entity">&#305;</mapping>
  <mapping type="encoding" subtype="unicode_symbol">i</mapping>
</glyph>
```

In den Einträgen zu den einzelnen individuellen Allographenbeschreibungen werden mithilfe von XML-Attributen und den entsprechenden Attributwerten folgende Informationen abgebildet: Der Wert des `@corresp`-Attributs weist auf eine mögliche hierarchische Verknüpfung hin. Jeder Eintrag erhält einen eindeutigen Identifikator, der auf den *Entity Name* der MIFI *Character Recommendation* zurückgeht. Jeder Eintrag wird systematisch beschrieben als Allograph, Abbréviatur, Superskript oder Satzzeichen. Abbréviaturen wiederum werden nochmals in allgemeines Kürzungszeichen, Brevigraph oder Kontraktion differenziert. Die Benennung und Beschreibung der Zeichen beruht in der Regel auf der MIFI, deshalb wird mit entsprechenden Attributen (`@resp`, `@source`) auch auf diese Quelle mit Seitenreferenz verwiesen. Ist ein Zeichen oder eine Zeichenkombination dort nicht vorhanden, wird dessen Name analog zu den Unicode-Namen aufgebaut. Das Element `<mapping>` gibt die Darstellung eines Zeichens in unterschiedlichen Kontexten an: Aufgenommen sind die jeweiligen Werte des *Unicode Codepoint* und für die Darstellung die HTML-*Entity* und das Unicode-symbol. Zusätzlich sind für die automatisierte Verarbeitung noch das proprietäre Markup, das im Transkriptionsworkflow verwendet wird, und die normalisierte Zeichendarstellung hinterlegt. Zeichen und Zeichenkombinationen, die im

ASCII-Zeichensatz vorhanden sind, werden nicht im Detail beschrieben.

Der damit teilweise bis auf Teilzeichenebene modellierte und annotierte Text steht dann zur Weiterbe- und -verarbeitung bereit: Die entsprechenden Informationen zu den Zeichen werden mit dem `<g>`-Element und dem Attribut `@ref`, das als Wert die jeweilige `@xml:id` des Eintrags in der *Character Declaration* hat, im Text modelliert.

w`<g ref="inodot»i</g>ldw wildw`Für die Erstellung von hyperdiplomatischen Transkriptionen nach diesem Transkriptionsmodell gibt es einen fertig konzipierten und erprobten Workflow, der auf mithilfe von *Transkribus* und proprietären Markup erstellten Transkriptionen aufbaut und durch XSL-Transformationen TEI-XML-Dateien erstellt. Für eine Publikation derartiger Transkriptionen in GAMS gibt es Templates, mithilfe derer ein Archivprojekt umgesetzt werden kann.

Literatur:

- Die (hyper-)diplomatische Transkription und ihre Erkenntnispotentiale. URL: <https://www.hsozkult.de/event/id/termine-42210>
- Böhm, Astrid; Klug, Helmut W: Quellenorientierte Aufbereitung historischer Texte im Rahmen digitaler Editionen: Das Problem der Transkription in mediävistischen Editionsprojekten. In: Digitale Methoden und Objekte in Forschung und Vermittlung der mediävistischen Disziplinen. Akten der Tagung Bamberg, 08.-10. November 2018t: 2020, S. 51–72.
- Hofmeister-Winter, Andrea: Das Konzept einer „Dynamischen Edition“ dargestellt an der Erstausgabe des „Brixner Dommnesnerbuches“ von Veit Feichter (Mitte 16. Jh.). Göppingen: 2003.
- Feigs, Wolfgang: Deskriptive Edition auf Allograph-, Wort- und Satzniveau, demonstriert an handschriftlich überlieferten, deutschsprachigen Briefen von H. Steffens. Bern (u.a.): 1979.
- CfI: Die (hyper-)diplomatische Transkription und ihre Erkenntnispotentiale. URL: <https://dhd-blog.org/?p=12369>

Software:

Transcribo, Transkribus

Verweise:

Transkription, Diplomatische Transkription, Transkriptionsrichtlinien, Transkriptionswerkzeuge, Diplomatische Edition

Projekte:

CoReMA - Cooking Recipes of the Middle Ages, Mittelalterlabor, Medieval Unicode Font Initiative, <http://www.digipal.eu>

Themen:

Digitale Editionswissenschaft, Annotation und Modellierung

Zitiervorschlag:

Klug, Helmut W.; Böhm, Astrid. 2021. Datenmodell "Hyperdiplomatische Transkription". In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.50>