

SABINE TITTEL 

Die Editionen medizinischer und rechtssprachlicher Texte des Alt- und Mittelfranzösischen im Rahmen des Projekts *ALMA*

Keywords French Middle Ages; digital edition; specialist text; Handwritten Text Recognition; Linked Open Data

Projektbeteiligte

Team *ALMA*-Heidelberg 2025: Sara Aakrout, Gustavo Covolo Campedelli, Valentin Jablonowski, Viola Mariotti, Gloria de Meleghy, Ragini Menon, Xin Peng, Sophie Poser, Leo Reitmaier, Leonie Freiin von Schnurbein, Sabine Tittel (Projektleitung)

Institutionelle Anbindung

Heidelberger Akademie der Wissenschaften

Förderung

Akademienprogramm

Laufzeit

2022–2043 (Gesamtlaufzeit Projekt; Laufzeit Modul „Editionen“: 2022 – ca. 2030)

Kurzbeschreibung

Die Editionen der mittelalterlichen Medizin- und Rechtstexte werden im Rahmen des Langzeitforschungsvorhabens *ALMA – Wissensnetze in der mittelalterlichen Romania*; durchgeführt (HAdW, Bayerische Akademie der Wissenschaften, Akademie der Wissenschaften und der Literatur Mainz). Das Projekt verschränkt Methoden der historischen Lexikologie und Lexikographie, Textphilologie und Wissen(schaft)s-geschichte mit Herangehensweisen der Digital Humanities, des Semantic Web, Ontology Engineering und Linked-Open-Data-Modellierens.

ALMA ist nicht primär ein Editionsprojekt, sondern hat zum Ziel, die Wechselwirkung zwischen Sprache(n) und Wissen(schaften) (von ca. 1100 – ca. 1500) in der mittelalterlichen Romania zu untersuchen. Anhand von lexikalisch-semantischen Studien zu den Wörtern der zwei exemplarischen Domänen ‚Medizin‘ und ‚Recht‘ soll analysiert und beschrieben werden, wie mittelalterliches Italienisch, Französisch, Okzitanisch und Gaskognisch zu Wissen(schaft)ssprachen ausgebaut werden, die konzeptuell-begrifflich – parallel zur Entwicklung innerhalb der Fächer ‚Medizin‘ und ‚Recht‘ – an Komplexität gewinnen und zunehmend die schriftliche Fachkommunikation und Wissenstradierung vom Lateinischen in Richtung der Vernakularsprachen verschieben.

Als empirische Basis für die Analyse der zentralen Konzepte und ihrer Bezeichnungen in den vier Sprachen und damit für die Rekonstruktion der übereinzelsprachlichen Begriffsnetze der Wissensdomänen werden parallele Textkorpora

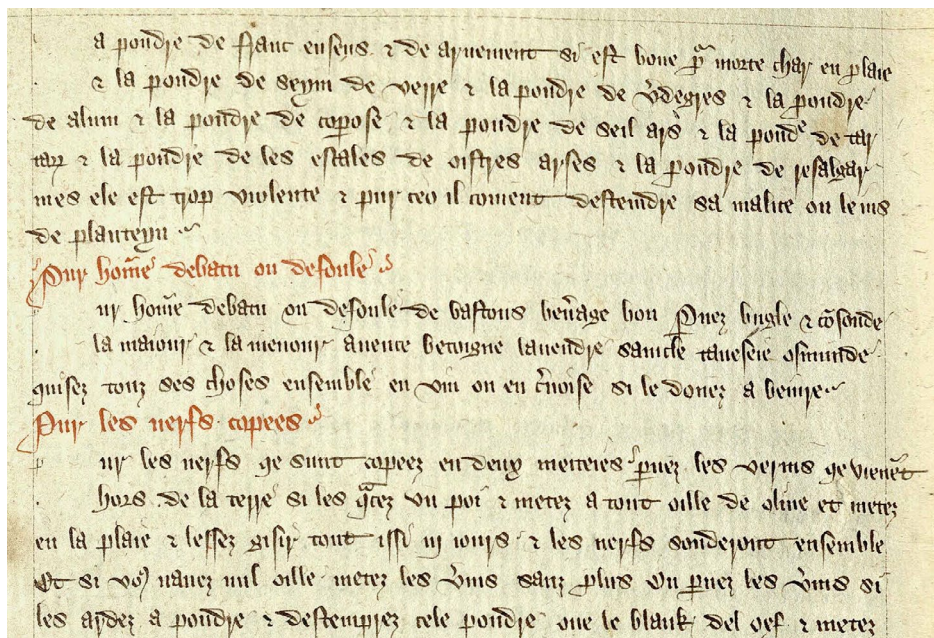


Abb. 1: Oxford, Bodleian Library, Handschrift Bodley 761 [ca. 1365], fol. 13^v.

erstellt. Sie enthalten Texte zur Medizin und zum Recht – hier sowohl theoretische Rechtstexte als auch historische Dokumente der Rechtsprechung. Die Korpustexte werden mit quantitativ-empiristischen Methoden der Korpuslinguistik ausgewertet, um erste Erkenntnisse zu relevanten Fachtermini und Schlüsselbegriffen zu generieren. Diese werden anschließend mit dem qualitativ-hermeneutischen Vorgehen der historischen Sprachwissenschaften analysiert (vgl. zu ALMA: PRIFTI/SCHWEICKARD/SELIG/TITTEL 2023; <https://www.hadw-bw.de/alma> [Stand: 02.06.2025])

Ein paar Schritte in *medias res*

Die Abbildung 1 zeigt Folio 13^v der Handschrift Oxford, Bodleian Library, Bodley 761 [anglonormannisch, ca. 1365]; diese enthält auf Folios 11^v–20^v und 86^{ro}–89^{ro} eine Sammlung von medizinischen Rezepten aus dem 13. Jh., die von ALMA unter dem Sigel RecMédBodl-ALMA ediert wird (<https://alma.hadw-bw.de/>; online im Laufe von 2025).

Abbildung 2 gibt einen Auszug aus der digitalen Edition in XML/TEI des Folios 13^v wieder. Dieser zeigt unter anderem, wie ALMA bei der Transkription mit Kürzeln und Sonderzeichen der mittelalterlichen Handschrift umgeht: ALMA setzt eine DTD (*Document Type Definition*, eine Schemadatei zum Validieren von XML) ein, die für diese Kürzel das Unicode-Zeichen gibt und dessen Übersetzung in einen

```

332 <pb n="13v" type="13" subtype="v"/>
333 <lb n="1"/><w><app><lem resp="#red">L</lem><rdg wit="#Bod"/><witDetail wit="#Bod">Lettrine d'attente
334 <hi rendition="em">l</hi>.</witDetail></app>a</w> poudre de franc ensens &et-tironien-non-barré; de
335 arvement si est bone &p-barré-par; morte char en plaie
336 <lb n="2"/>&et-tironien-non-barré; la poudre de seym de verre &et-tironien-non-barré; la poudre de
337 v&tilde-vertical-er;degres &et-tironien-non-barré; la poudre
338 <lb n="3"/>de alum &et-tironien-non-barré; la poudre de co&p-barré-par;ose &et-tironien-non-barré; la poudre
339 de seil ars &et-tironien-non-barré; la poude de <w>tar<lb n="4"/>tar&us-tironien-suscrit;</w>
340 &et-tironien-non-barré; la poudre de les escales de oistres arsés &et-tironien-non-barré; la poudre de
341 resalgar
342 <lb n="5"/>mes ele est trop violente &et-tironien-non-barré;; pur ceo, il coment destendre sa malice ou levis
343 <lb n="6"/>de planteyn.
344 <lb n="7"/>Pur hom&tilde-nasale-m;e debatu ou defoulé.
345 <lb n="8"/><w><app><lem resp="#red">P</lem><rdg wit="#Bod"/><witDetail wit="#Bod">Lettrine d'attente
346 <hi rendition="em">p</hi>.</witDetail></app>ur</w> hom&tilde-nasale-m;e debatu on defoule de bastons
347 bev&tilde-vertical-er;age bon &p-barré-per;nez ungle &et-tironien-non-barré; co&tilde-nasale-n;sonde
348 <lb n="9"/> la majour &et-tironien-non-barré; la menour, avance, betoigne, lavendre, sanicle, taneseie, osmunde
349 <lb n="10"/>quizez touz ses choses ensemble en vin ou en t&e-suscrit-re;noise, si le donez a beviere
350 <lb n="11"/>pur les nerfs copees.
351 <lb n="12"/><w><app><lem resp="#red">P</lem><rdg wit="#Bod"/><witDetail wit="#Bod">Lettrine d'attente <hi
352 rendition="em">p</hi>.</witDetail></app>ur</w> les nerfs qe sunt copeez en deux metteies &p-barré-per;nez
353 les vernis qe viene&tilde-nasale-n;t
354 <lb n="13"/>hors de la terre si les q&a-suscrit-ra;cez un por &et-tironien-non-barré; metez a tout oille de
355 olive et metez
356 <lb n="14"/>en la plaie &et-tironien-non-barré; lessez gisir tout issi .iij. jours &et-tironien-non-barré; les
357 nerfs souderont ensemble
358 <lb n="15"/>Et, si vo&us-tironien-suscrit; navez nul oille, metez les v&tilde-vertical-er;nis sanz plus. Ou
359 &p-barré-per;nez les v&tilde-vertical-er;niz, si
    
```

Abb. 2: Auszug aus der Edition von RecMédBodl-ALMA: Beginn Folio 13^{vo}.

lesbaren Text in XML kodiert, zum Beispiel (i) die tironische Note für lateinisch *et* (hier in der Variante ohne Querstrich, s. Abb. 3), ‚&et-tironien-non-barré;‘ (Z. 332; Abb. 1: Z. 1, Wort 6), XML/TEI <choice><expan><ex>et</ex></expan><abbr></abbr></choice>, Visualisierung in der Edition als kursiviertes *et*, oder (ii) der Buchstabe *m* mit Nasalstrich (*m̃*), ‚&tilde-nasale-m;‘ (Z. 339, Abb. 1: Z. 7, Wort 2), XML/TEI <choice><expan><ex>m</ex></expan><abbr>̃</abbr></choice>, Visualisierung in der Edition als *mm*. Die Kodierung und Darstellung der Sonderzeichen mittelalterlicher Texte in lateinischem Alphabet basiert auf und erweitert die *Medieval Unicode Font Initiative* (MUFI, <https://mufi.info/q.php?p=mufi> [Stand: 02.06.2025]).



Abb. 3: Tironische Noten für lat. et.

In Abbildung 2 ebenfalls zu sehen ist der Einsatz der TEI-Elemente zur Kodierung der fehlenden Initialen mit <app> mit <lem>, <rdg> und <witDetail> (zur Erklärung der Elemente, s. die TEI Guidelines, <https://www.tei-c.org/> [Stand: 02.06.2025]), s. z. B. Z. 332 (Z. 1 des Folios): <app><lem resp="#red">L</lem><rdg wit="#Bod"/><witDetail wit="#Bod">Lettrine d'attente <hi rendition="em">l</hi>.</witDetail></app>a poudre...

Relevanz des Gegenstands

Die von *ALMA* erarbeiteten Textkorpora erschließen einen bedeutenden Kulturraum der mittelalterlichen Romania. Sie bezeugen den konzeptuellen und damit verbundenen begrifflichen Ausbau der Volkssprachen, beides zentrale Analyseblickwinkel des Projekts, die gleichsam Indiz und Stimulans des Wandels der intellektuellen Landschaft auf dem Weg zur Renaissance sind. Die romanischen Sprachen und hier vor allem die wirkmächtigste Vernakularsprache, das Französische, sind bedeutende Träger des kulturellen Austauschs, und mit den Fach- und Wissenschaftssprachen wird ein wichtiger Bestandteil des kulturellen Erbes Europas in den Fokus gerückt.

Relevanz der Editionen

Der Blick in die bisherigen Forschungen zum Thema der mittelalterlichen Wissenskommunikation zeigt ein Desideratum: Vertiefte Untersuchungen der sprachlichen Entwicklungen in einzelnen Wissensdomänen basieren meist nur auf der Analyse einer Einzelsprache; sprachübergreifende Kulturräume geraten nur selten in den Blick, obwohl die untergeordnete Rolle von sprachlichen Grenzen im Mittelalter bekannt ist. Dazu kommt, dass linguistisch und fachlich wichtige Texte der beiden Domänen nicht, nur teilweise oder sprach- und wissenschaftshistorisch nicht belastbar ediert sind und darüber hinaus oft nicht für umfassende, digital getriebene Analysen bereitstehen.

Ein Schlüsseltext aus dem Bereich der Medizin ist z. B. die *Grande Chirurgie* des Gui de Chauliac (nach 1363), die bis ins 17. Jahrhundert als didaktisches Grundlagenwerk größten Einfluss auf die Lehre der Medizin und der Chirurgie hatte. Im Bereich des Rechts ist der *Grand Coutumier de Normandie* vom Ende des 13. Jahrhunderts zu nennen, der normannisches Gewohnheitsrecht und römisches Recht vereint und als eine der ältesten Rechtsquellen maßgeblich den Entwicklungsstand der Fachsprache des Rechts bezeugt.

ALMA erstellt daher paradigmatische, sprach- und wissenschaftsgeschichtliche Textkorpora zu den Bereichen ‚Medizin‘ und ‚Recht‘ in den vier Sprachen. In die Textkorpora fließen ein:

- 1) von *ALMA* erarbeitete, eigene Editionen in XML/TEI, zum Teil händisch transkribiert, zum Teil mit algorithmengestützter Transkription (*handwritten text recognition* / HTR mittels eScriptorium, <https://escriptorium.inria.fr/> [Stand: 02.06.2025]) auf Basis von eigens entwickelten Goldstandards und damit trainierten Algorithmen durchgeführt;
- 2) durch *ALMA* nachnutzbare digitale Editionen in XML/TEI, die in das XML/TEI-Format *ALMA*s überführt werden; hier profitiert *ALMA* von

Kooperationen mit Forschungsprojekten wie *Base de Français Médiéval* (<http://bfm.ens-lyon.fr/> [Stand: 02.06.2025]), CHrOMed (<https://www.arts.kuleuven.be/chromed/> [Stand: 02.06.2025]), *Anglo-Norman Dictionary* (<https://anglo-norman.net/> [Stand: 02.06.2025]), *École Nationale des Chartes* (<https://www.chartes.psl.eu/recherche/centre-jean-mabillon/projets-de-recherche/miroir-des-classiques> [Stand: 02.06.2025]);

- 3) durch ALMA nachnutzbare digitale Editionen proprietärer Formate älteren Datums, die in das XML/TEI-Format ALMAs überführt werden (z. B. machte Tony Hunt, Herausgeber zahlreicher medizinischer anglonormannischer Texte, seine Editionen in einem proprietären Druckformat der 1990er Jahre dem Projekt zugänglich);
- 4) in Buchform publizierte Editionen, die von ALMA digitalisiert und in das XML/TEI-Format ALMAs überführt werden.

Die so entstehende Forschungsgrundlage wird über die unmittelbaren Anforderungen der Projektarbeit hinaus eine wertvolle, vielseitig einsetzbare Basis für Anschlussforschungen zur Wissenskommunikation in der mittelalterlichen Romania bilden.

Darstellung und Aufbereitung

Die Editionen werden digital in XML/TEI P5 erstellt und mit einem Relax-NG-Schema validiert (exportiert aus einer ALMA-spezifischen ODD-Datei). Schemadateien (so auch die DTD, s. o.) und Transformationsskripte für die Weiterverarbeitung der Editionen (XML zu XML/TEI, ALTO XML zu XML/TEI, Word zu XML/TEI etc.; Tokenisierung, Lemmatisierung etc.) werden im Open Source auf dem GitLab des Projektes veröffentlicht (<https://gitlab.hadw-bw.de/alma/alma> [Stand: 02.06.2025]); publikationsfähige Editionen werden auf dem Projektserver <https://alma.hadw-bw.de> [Onlinestellung im Laufe von 2025] unter der Lizenz CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/> [Stand: 02.06.2025]) visualisiert und ebenfalls als XML-Dump zur Verfügung gestellt (die noch in Arbeit befindlichen Editionen werden auf einem nicht-öffentlichen Fork des GitLabs versioniert). Die Visualisierung der Editionen gehört zu den laufenden Arbeiten.

Probleme und Lösungsansätze

- 1) Der Projektkonzeption ALMA ist ein wissenschaftlich-ethisches Dilemma inhärent: Auf der einen Seite steht editionsphilologisch das Ziel, hochqualitative Texteditionen zu erarbeiten. Auf der anderen Seite steht der Umstand, dass das Projekt kein Editionsprojekt *per se* ist, dessen Zeitplanung

die Erstellung von Editionen fokussiert, sondern ein linguistisches Projekt mit einem (meta-)sprachlichen Gesamtziel: Editionen sind in *ALMA* einerseits Passion, andererseits aber nur Mittel zum Zweck. Die pragmatische Herangehensweise an dieses Dilemma vereint zwei Aspekte:

- a) HTR: Die Transkriptionsarbeiten werden mit automatisierten Verfahren beschleunigt, die eine möglichst gute Basis für eine kompetenzlinguistische, händische Nachbearbeitung in Form einer Edition schaffen.
 - b) Editionsart: Zunächst waren im Projektverlauf auch Editionen von Handschriftentraditionen mit Basishandschrift/ Variantenhandschriften angefertigt worden. Diese erforderten ein komplexes TEI-Vokabular, was wiederum maßgeblich die digitale Weiterverarbeitung erschwerte für (i) die Visualisierung inklusive des Variantenapparats (s. u. 3.), (ii) die korpuslinguistische Annotation (Tokenisierung, Part-of-Speech-Tagging, Lemmatisierung, semantische Disambiguierung) und (iii) die Implementierung von Exportpipelines, die die Wörter der Editionen aus eine XML-nativen eXist-Datenbank in eine relationale Datenbank für die Redaktion der lexikalisch-semantischen Studien überführen. Im Sinne des wissenschaftlichen Fokus *ALMA*s, der eben nicht auf Editionen als Selbstzweck liegt, zog das Projekt die Konsequenz und erzeugt nunmehr ausschließlich Editionen, die auf genau einer Handschrift basieren. Diese bilden eine belastbare Arbeitsbasis für die Textanalysen und entsprechen in ihrer Gesamtheit den Anforderungen, die *ALMA* an die Parallelkorpora zur mittelalterlichen Gallo- und Italo-romanischen Medizin und Recht stellt. Sofern jedoch (sprach-)historisch indiziert, werden für Texte, die in mehreren relevanten Überlieferungen vorliegen, entsprechend mehrere Überlieferungen ediert.
- 2) Bei der Digitalisierung bestehender Editionen müssen die Autoren- und Verlagsrechte beachtet werden; dies schließt in der Regel ein, dass ein Text nachgenutzt werden darf, Fußnoten und Variantenapparate aber nicht. Damit einerseits die Mitarbeiter *ALMA*s bei der Arbeit mit den Texten das Maximum an Informationen zum edierten Text digital vorliegen haben, sprich, die gesamten Fußnoten und Apparate einsehen können, andererseits diese dem Nutzer der Website aber nicht gezeigt werden, verwendet *ALMA* bei der Digitalisierung das XML-Element `<note resp=“...“>`, wo über das Attribut ‚resp‘ die Verantwortlichen für die `<note>` definiert werden (über Kürzel, die im `teiHeader` aufgelöst werden). Je nach Verantwortlichkeit wird die `<note>` dann digital ein- oder ausgeblendet.
 - 3) *ALMA*s Ziel ist es, Open-Source-Tools nachzunutzen, diese gegebenenfalls zu erweitern und als Webapplication in die Publikationsplattform des Projektserverns zu integrieren. Es zeigt sich, dass dies für die Publikation von Editionen mit komplexen Texteingriffen, mit `<note>` verschiedener Verantwortlichkeiten, mit Visualisierung von Text, respektive Textauszügen (je

nach Verlagsrecht) etc. schwierig ist: Tools, die kostenfrei und gemäß den Projektanforderungen die XML/TEI-Editionen in einen gut lesbaren Text umformen, scheint es nicht zu geben. ALMA hat daher eine freie JavaScript Library CETEIcean (<https://teic.github.io/CETEIcean/> [Stand: 02.06.2025]) zu diesem Zweck weiterentwickelt (Fertigstellung in Arbeit).

- 4) Um einen Zugang über die Recherchemöglichkeiten des Webservers hinaus zu bieten, modelliert ALMA die Editionen als LOD (*Linked Open Data*) in RDF (*Resource Description Framework*) im Sinne der F.A.I.R. Prinzipien (<https://www.go-fair.org/fair-principles/> [Stand: 02.06.2025]). Dafür stehen bereits Standard-Ontologien bereit: OntoLex-Lemon als Basisvokabular für die Beschreibung der Wörter (Lexem, Mehrwortverbindung, Suffix etc.), ihrer Bedeutungen, deren Konzepte etc., LexInfo (<https://lexinfo.net/> [Stand: 02.06.2025]) für die Zuweisung von Wortklasse, Genus und Numerus, lemonEty (<https://github.com/anasfkhan81/lemonEty> [Stand: 15.12.2025]) für Etyma etc. Darüber hinaus gibt es aber Lücken in der Forschungslandschaft, besonders in Bezug auf ALMAs *low-resourced languages*: Es fehlen etwa Sprachcodes für die Zuweisung von Lexemen zu einer historischen Sprachvarietät, z. B. Altgaskognisch, Altpikardisch, Altburgundisch (siehe Tittel 2024, 155–159; 359–380), eine Ontologie für die detaillierte Modellierung von Bedeutungswandel (Stand der Forschung, *ib.*, 159–163; Ontologie SemShift, *ib.*, 413–433). Darüber hinaus müssen historisierte Ontologien für die Domänen ‚Recht‘ und ‚Medizin‘ erarbeitet werden, die die Spezifik mittelalterlicher Erklärungsmuster berücksichtigen, statt für historische Konzepte (‚heiß‘ als Kardinalqualität der Humoralpathologie) moderne Pseudoäquivalente (‚heiß‘ als moderne Temperaturbeschreibung) anzubieten. Sie erlauben es, die Kontinuitätsbrüche zwischen mittelalterlichen und modernen Wissen(schaft)ssystemen sichtbar zu machen und nicht durch anachronistische Zuordnungen zu verdecken.

ORCID®

Sabine Tittel  <https://orcid.org/0000-0003-4746-7604>

Literaturverzeichnis

- Prifti, Elton/Schweickard, Wolfgang/Selig, Maria/Tittel, Sabine (2023), „Sprachdatenbasierte Modellierung von Wissensnetzen in der mittelalterlichen Romania (ALMA): Projektskizze“, in: *Zeitschrift für romanische Philologie* 139 (2), 301–322.
- Tittel, Sabine (2024), *Historische lexikalische Semantik und Linked Data*, Berlin/Boston.

Abbildungsnachweis

- Abb. 1** Folio 13vo der Handschrift Oxford, Bodleian Library, Bodley 761 [ca. 1365].
Abb. 2 RecMédBodI-ALMA (© Tittel).
Abb. 3 Wikimedia Commons, CC BY-SA 4.0.