

Clemens Beckstein, Robert Gramsch-Stehfest, Clemens Beck,  
Jan Engelhardt, Christian Knüpfer, Georg Zwillling

## Digitale Prosopographie

Die automatisierte Auswertung des *Repertorium Germanicum*,  
eines Quellenkorpus zur Geschichte geistlicher Eliten des  
15. Jahrhunderts

**Abstract:** This paper presents a new approach to the automatic evaluation of the Repertorium Germanicum (RG), a collection of registers on late medieval German ecclesiastical history, based on modern compiler construction technology. The RG, which is very large by medieval standards, uses an elaborate system of abbreviations, which makes the contents of the records accessible in a summarised form. For the automatic evaluation of these registers, it is necessary to record the semantics of the compressed text. With ANTLR, a parser generator for structured text from compiler construction, the individual components in the text can be identified by using a grammar. The creation of the grammar is explained with an example and its individual components are described. Subsequently, it will be considered how the results of such a parsing process can be used for the development of a general ontology for the prosopography of clerical elites, which will form the starting point for comprehensive data collection and its further analysis.

**Keywords:** Repertorium Germanicum, Vatican sources, historical network analysis, language processing, prosopography, Church history, Middle Ages

**Zusammenfassung:** In diesem Beitrag wird ein neuer Ansatz zur automatischen Auswertung des Repertorium Germanicum (RG), eines Regestenwerks zur spätmittelalterlichen deutschen Kirchengeschichte, auf der Basis moderner Compilerbau-Technologie vorgestellt. Um das für mittelalterliche Verhältnisse sehr umfangreiche Quellenmaterial zu erschließen, verwendet das RG ein elaboriertes System von Abkürzungen, auf dessen Basis die Inhalte der Akten komprimiert dargestellt werden. Zur automatischen Auswertung dieser Regesten ist es notwendig, die Semantik des komprimierten Textes zu erfassen. Dafür wird ANTLR, ein Parsergenerator für strukturierten Text aus dem Compilerbau, verwendet, mit dem die einzelnen Bestandteile im Text mittels einer von uns entwickelten Grammatik identifiziert werden können. An einem Beispiel werden die Erstellung der Grammatik erläutert und ihre einzelnen Bestandteile erklärt. Dar-

an schließen sich Überlegungen an, wie die Ergebnisse eines solchen Parserprozesses für die Entwicklung einer allgemeinen Ontologie zur Prosopographie geistlicher Eliten genutzt werden können, die die Ausgangsbasis für eine umfassende Datenerhebung und deren weitere Analyse bilden.

**Schlagwörter:** Repertorium Germanicum, Vatikanische Quellen, Historische Netzwerkanalyse, Language Processing, Prosopographie, Kirchengeschichte, Mittelalter

Denn auf diesem Ozean von Papier, wo keine Sterne den Weg zeigen, würde der suchende Forscher lange umherfahren können, ehe er die Inseln findet, von denen er nicht einmal mit Sicherheit voraussagen kann, ob sie überhaupt existieren.

Johannes Haller (1903)

Als der bekannte Mittelalter- und Papsthistoriker Johannes Haller 1892 frisch promoviert nach Rom kam, bewarb er sich am dortigen Königlich Preußischen Historischen Institut als „Hilfsarbeiter“ bei einem Projekt, das seine wissenschaftliche Laufbahn mitbestimmen und dem er selbst seinen Stempel aufdrücken sollte, dem *Repertorium Germanicum* (RG).<sup>1</sup> In dieses „Verzeichnis der in den päpstlichen Registern und Kameralakten vorkommenden Personen, Kirchen und Orte des deutschen Reiches“ sollten alle Aktenstücke aus dem kurz zuvor für die wissenschaftliche Forschung geöffneten Vatikanischen Geheimarchiv Aufnahme finden, welche sich auf die deutsche Geschichte zwischen 1378 und 1521 bezogen. Angesichts von tausenden dickleibigen Folianten, die sorgsam auf deutsche Betreffe durchzusehen, zu exzerpieren und in Regestenform zu bringen waren, handelte es sich um eine wahre Mammutaufgabe. Und so veranlassten die Erfahrungen aus diesem Projekt den mittlerweile zum Marburger Professor aufgestiegenen Haller elf Jahre später zu einer denkwürdigen Stellungnahme:

---

<sup>1</sup> Zur Person vgl. Benjamin Hasselhorn, *Johannes Haller: Eine politische Gelehrtenbiographie: Mit einer Edition des unveröffentlichten Teils der Lebenserinnerungen Johannes Hallers* (Göttingen: Vandenhoeck & Rupprecht, 2015), darin 47–57 zu seiner römischen Zeit. Die Geschichte des Editionsprojekts *Repertorium Germanicum* wird geschildert in: Dieter Brosius, „Das Repertorium Germanicum,“ in *Das Deutsche Historische Institut in Rom 1888–1988*, hg. v. Reinhard Elze und Arnold Esch (Tübingen: Niemeyer, 1990), 123–165; dazu auch Brigide Schwarz, „Das Repertorium Germanicum: Eine Einführung,“ *Vierteljahrsschrift für Sozial- und Wirtschaftsgeschichte* 90 (2003): 429–440. Zur Benutzung des RG immer noch unentbehrlich: Walter Deeters, „Über das Repertorium Germanicum als Geschichtsquelle. Versuch einer methodischen Anleitung,“ *Blätter für deutsche Landesgeschichte* 105 (1969): 27–43.

Mit der Quelle, aus der hier geschöpft wird, kann sich für diese Zeit keine andere in der Welt messen, sie übertrifft an Umfang und Mannigfaltigkeit schlechterdings alles, was aus dem ausgehenden Mittelalter überhaupt bekannt ist. [...] So reich der Inhalt, so schwierig ist der Zugang. Vor diesen rund 4000 Bänden steht der einzelne Forscher ratlos und hilflos da [...]. Hier kann man wohl finden, aber nicht suchen.<sup>2</sup>

Scharfsinnig analysierte Haller, wie diese unübersehbare Aktenfülle effizient für die Zwecke mediävistischer Forschung erschlossen werden könne. Sein visionärer Vorschlag war, den Stoff über ein höchst platzsparendes Indizierungssystem zu erschließen, das an eine moderne Datenbank erinnert: „[...] was wir vor allem brauchen, ist ein alphabetischer Index. [...] Wird zu den Namen (Personen und Orten) außer dem Datum auch noch mit kurzem Schlagwort die Natur der Erwähnung hinzugesetzt [...] so ersetzt für den weiteren Kreis der Benutzer der Index die Regestensammlung.“<sup>3</sup>

Extrem knappe, ja geradezu kryptische Regesten in der Quellsprache Latein, die auf engstem Raum alle wichtigsten Informationen beinhalten, dazu umfangreiche Indices: Dieser Aufbau wurde zum „Markenzeichen“ des *Repertorium Germanicum*, das mittlerweile auch in die digitale Sphäre umgezogen ist (RG Online).<sup>4</sup> Für mediävistische Verhältnisse handelt es sich regelrecht um „Big Data“ und so stellt sich die naheliegende Frage, ob dieses Material auch mit fortgeschritteneren Technologien automatisiert erschlossen werden könnte. Denn solange das Material nur von Einzelforschenden lesend im Buch oder am Bildschirm rezipiert wird, gilt noch immer das Wort von Johannes Haller, dass „auf diesem Ozean von Papier [...] der suchende Forscher lange umherfahren“ muss.<sup>5</sup> An dieser Stelle setzt unser Jenaer Forschungsprojekt Core-H („Collaborative Open Research Environment for the Humanities“) an, in dem sich die Erschließung der RG-Regesten als Fallstudie (*case study*) in den breiteren Kontext der Entwicklung computergestützter Workflows für die historische Forschung einordnet.<sup>6</sup>

---

<sup>2</sup> Johannes Haller, „Über das Repertorium Germanicum, Gutachten (Marburg, 5.4.1903)“, Archiv des DHI in Rom, Dienstakten A/8 (Acta generalia, Bd. VIII), 19 Bl., hier fol. 1. Zu dem Gutachten siehe auch Hasselhorn, *Johannes Haller*, 84–86.

<sup>3</sup> Haller, „Repertorium Germanicum“, fol. 10.

<sup>4</sup> Zugang über die Website des „Romana Repertoria“-Online-Portals des DHI in Rom: <http://www.romana-repertoria.net/993.html>, Zugriff am 06.10.2021. Vgl. dazu Jörg Hörnschemeyer, „Repertorium Germanicum Online“, in *Friedensnobelpreis und historische Grundlagenforschung. Ludwig Quidde und die Erschließung der kuralen Registerüberlieferung*, hg. v. Michael Matheus (Berlin: De Gruyter, 2012), 605–615.

<sup>5</sup> Haller, „Repertorium Germanicum“, fol. 10.

<sup>6</sup> Vgl. die Projektwebsite unter <https://www.mephisto.uni-jena.de/projekte/core-h> sowie die Blogbeiträge unter <https://www.mephisto.uni-jena.de/blog>, Zugriff am 06.10.2021.

Schauen wir uns einige typische Regesten aus dem RG Online (RG V) an – einen Ausschnitt aus der Pfründenvita<sup>7</sup> des Klerikers Johannes (Alberti) von Echte<sup>8</sup> aus den Jahren 1431 und 1437 (Abbildung 1).

 **RG V 04399**

zur Merkliste hinzufügen

Johannes de Echte **Johannes (Alberti) de Echte** Johannes de Eckt: l.b. iul. 1431 T 3 25v.

cler. Magunt. dioc.: de vicar. ad alt. ss. Laurentii, Cosme et Damiani atque Liberii in eccl. s. Crucis Northusen. Magunt. dioc. (4 m. arg. ) vac. p. resign. Johannis Wittenborn; n.o. disp. sup. def. nat. ( subdiac., c. ) 24. nov. 1431 S 272 81v.

**24. nov. 1431**

(...)

pres. in R. cur.: de vicar. ad alt. ss. Andree et Nicolai in colleg. eccl. s. Crucis Northusen. Magunt. dioc. (5 m. arg. ) vac. p. resign. Hermannii Rulen presb. vel p.o. Johannis de Bendeleyben; n.o. vicar. in capel. s. Spiritus e.m. op. Duderstat d. dioc. (2 m. arg. ), disp. sup. def. nat. ( subdiac., c. ) et quod pater ipsius J. cantor. c. can. et maior. preb. d. eccl. s. Crucis Northusen. obtin. 2. nov. 1437 S 341 224vs.

**2. nov. 1437**

**Abb. 1:** Ausschnitt aus der Pfründenvita des Klerikers Johannes (Alberti) von Echte im RG V, Nr. 4399 (Screenshot nach RG Online, RG IV 00133, URL: <<http://rg-online.dhi-roma.it/RG/4/133>> (Datum 14.10.2021)).

Im ersten Sublemma finden wir den vollständigen Namen des Geistlichen. Es folgt ein auf den 24. November 1431 datiertes Schreiben, in dem sich Johannes auf eine Vikarie am Altar St. Laurentius und weiteren in der Stiftskirche Heiligkreuz Nordhausen bewirbt, welche nach seiner Angabe ein jährliches Einkommen von vier Silbermark hatte. Weiter wird angegeben, dass die Pfründe durch den Verzicht (Resignation) eines gewissen Johannes Wittenborn frei (vakant)

**7** Pfründen sind mit konkreten Einkünften versehene kirchliche Ämter.

**8** Zu dessen Biographie vgl. Robert Gramsch, „Kurientätigkeit als ‚Berufsbild‘ gelehrter Juristen. Der Beitrag Roms zur Akademisierung Deutschlands im Spätmittelalter. Eine personengeschichtliche Betrachtung,“ *Quellen und Forschung aus italienischen Archiven und Bibliotheken* 80 (2000): 138–143. Eine Pfründenvita fasst alle Quellenbelege aus einem Pontifikat zusammen, die einem bestimmten Kleriker als Bittsteller (sogenannter Petent) zuzuordnen sind. Dieselbe Person kann darüber hinaus auch in den Pfründenviten anderer Petenten genannt sein, etwa als Konkurrent in einem Prozess. Im RG Online können Personen, Orte und so weiter über eine Stringsuche gefunden werden, die allerdings wegen variierender Schreibweisen nicht treffgenau ist.

geworden war, und dass Johannes von Echte eine Geburtsdispens besaß, eine Ausnahmegenehmigung, Pfründen erwerben zu dürfen, obwohl er aus der Verbindung eines Subdiakons und einer verheirateten Frau hervorgegangen war.<sup>9</sup> Das schließende Kürzel „S 272 81v“ gibt die Fundstelle im Vatikanischen Archiv an.

Im Regest zum 2. November 1437 erfahren wir, dass sich Johannes von Echte damals am Papsthof (*in Romana curia*) aufhielt und dass er sich um eine weitere Altarvikarie in Nordhausen bewarb. Diese war durch den Pfründenverzicht eines gewissen Hermann Rulen frei geworden, sowie (zuvor) durch den Tod des Johannes von Bendeleyben. Nach einem Semikolon folgt das Kürzel „n.o.“, welches den Beginn der sogenannten Nonobstanzenformel markiert, der Angabe, welche Pfründen Johannes von Echte bereits im Besitz hatte und deren Besitz dem Erwerb des erbetenen Altars nicht im Wege stehen sollte. Auch hier wird nochmals die Geburtsdispens genannt, diesmal konkretisiert durch die Angabe, dass Johannes' Vater Kantor am Heiligkreuzstift war.

Regesten dieser Art automatisiert beziehungsweise mit gewisser menschlicher Hilfe semiautomatisiert auszuwerten, wird durch mehrere Umstände entschieden vereinfacht: Zum einen sind es immer wieder dieselben Gegenstände, die den Inhalt der Regesten ausmachen: Bittschriften, päpstliche Gratialschreiben und weitere Aktenstücke, die sich um den Erwerb von kirchlichen Ämtern und Einkünften drehen.<sup>10</sup> Schon im spätmittelalterlichen Verwaltungsbetrieb der kurialen Behörden herrschten hier bürokratische Effizienz und Routine, wenn mit den immer gleichen Rechtsformeln ähnliche Sachverhalte, bei denen nur die Namen der beteiligten Personen und zur Disposition stehenden Pfründen variieren, dokumentiert wurden. Der Inhalt dieser Schreiben lässt sich somit durch den modernen Bearbeiter recht einfach auf den wesentlichen Gehalt reduzieren, hinreichend charakterisiert durch einige hundert *termini technici* (wobei freilich so manches individualisierende Detail wegfällt, das erst der Blick auf das Originaldokument enthüllt). Mit dem entsprechenden kirchenrechtlichen Hintergrundwissen lässt sich mithin aus vergleichsweise wenigen Worten, die zudem noch zu platzsparenden Abkürzungen ohne grammatikalische Finessen verkürzt worden sind, schon das Wesentliche erkennen. Das RG Online macht dieses Material digital abfragbar – allerdings durch den Flaschen-

---

<sup>9</sup> Zum Phänomen der Klerikerkinder, die wegen dieses „Geburtsdefekts“ (*defectus natalium*) einer päpstlichen Dispens bedurften, um in der Kirche Karriere machen zu können, vgl. Ludwig Schmutge, *Kirche, Kinder, Karrieren: Päpstliche Dispense von der unehelichen Geburt im Spätmittelalter* (Zürich: Artemis, 1995).

<sup>10</sup> Zu diesem Phänomen vgl. etwa Brigide Schwarz, „Römische Kurie und Pfründenmarkt im Spätmittelalter,“ *Zeitschrift für Historische Forschung* 20 (1993): 129–152.

hals einer Suchmaske, die nur Einzelabfragen und ein relativ mühseliges Datensammeln per *copy & paste* erlaubt. Hat man hingegen die Regestentexte digital zur Verfügung, so rückt die Möglichkeit eines automatisierten *Information Retrieval* in greifbare Nähe.

Gestützt auf den digitalen 10. Band (zum Pontifikat Sixtus' IV. 1471–1484) haben wir 2019 eine erste „Probepbohrung“ unternommen. Ihr Ziel war, jene Belege aufzufinden und automatisch auszuwerten, die den Geldtransfer zwischen Deutschland und Rom in Form der sogenannten Annaten zum Gegenstand haben. Diese wurden fällig, wenn ein Kleriker auf päpstliche Verleihung hin eine geistliche Pfründe erhielt, deren ersten halben Jahresertrag er als Gebühr an den Papst zu entrichten hatte. Uns interessierte vor allem, welche individuellen und institutionellen Akteure (Bankhäuser, aber auch Kurienpersonal) an diesen Transfers beteiligt waren. Eine solche Forschungsaufgabe wäre früher nur mit hohem Aufwand zu lösen gewesen, denn es hätte endlosen Nachschlagens in der Edition und einer peniblen Listenführung bedurft, um allein die Daten zu erheben. Unser Workflow ermöglichte es hingegen, die einschlägigen Regesten automatisch aus der Stoffmasse des RG herauszufiltern. Dafür wurden die Einträge in der TUSTEP-Datei des RG-Bandes mit einem Python-Skript nach der Zeichenkette „*solv. pro annat.*“ durchsucht. Zu einer so gefundenen Annatenzahlung wurden dann das Datum des Vorgangs, die Regestennummer sowie der Auftraggeber der Zahlung aus dem Lemma extrahiert. Weiterhin wurden der Übermittler der Zahlung, die zur Disposition stehende Pfründe und die Gebührenhöhe ermittelt. Diese Daten wurden zur weiteren Verarbeitung in eine CSV-Datei geschrieben. Mit der Programmiersprache R wurde anschließend das Personennetzwerk der Annatenzahlungen analysiert und visualisiert. Die Knoten in einem solchen Netzwerk sind die historischen Personen, eine gerichtete Kante verbindet den Gebührenzahler mit dem involvierten Finanzakteur (ein Bankhaus oder auch eine konkrete Person). Über Zentralitäts- und Clusteranalysen konnten so die wichtigsten Agenten in diesem Netzwerk identifiziert und Strukturmuster des damaligen kurialen Finanzsystems herausgearbeitet werden.<sup>11</sup>

Freilich hat sich der hier eingeschlagene Weg, über *regular expressions* den Inhalt der Lemmata zu erschließen, als problematisch für komplexere Auswertungen erwiesen. Benötigt wurde ein digitales Werkzeug, das in der Lage ist, die charakteristische Struktur eines jeden Regestentyps eindeutig zu identifizieren und dafür in einem ersten Schritt der späteren digitalen Auswertung einen kodierenden Parsebaum zu generieren. Eine Grammatik im informatischen Sin-

---

<sup>11</sup> Vgl. hierzu Robert Gramsch-Stehfest, „Von der Metapher zur Methode: Netzwerkanalyse als Instrument zur Erforschung vormoderner Gesellschaften,“ *Zeitschrift für Historische Forschung* 47 (2020): 20–28.

ne bildet dabei alle Optionen ab, in der ein bestimmter Regestentyp ausgeformt sein kann. Der Prozess der Erstellung einer solchen Grammatik setzt einen engen Feedbackloop zwischen den daran zwingend beteiligten Historiker:innen und Informatiker:innen voraus. Für die Erstellung unserer RG-spezifischen Grammatik und die sich darauf stützende Extrahierung der Informationen aus dem Text nutzen wir seit dem Frühjahr 2021 mit bemerkenswertem Erfolg ANTLR (ANother Tool for Language Recognition)<sup>12</sup>, ein ursprünglich für den Compilerbau entwickeltes, sehr mächtiges Werkzeug zur schnellen und komfortablen Implementierung domänenspezifischer Sprachen (DSLs), die sich über eine formale Grammatik charakterisieren lassen, also zum Beispiel von Programmiersprachen, Datenbeschreibungssprachen, Interpretern oder Textanalyse- und Texttransformationswerkzeugen jedweder Art.

ANTLR ist ein Programmgenerator, der mit Hilfe des rekursiven Abstieges, passend zu einer vorgegebenen Grammatik einen top-down Parser, also das eigentliche, später für die Textanalyse eingesetzte digitale Werkzeug erzeugt.<sup>13</sup> Bei diesem Verfahren kommt jedem Nichtterminalsymbol der Grammatik eine (von ANTLR automatisch generierte) Prozedur zu, die die Grammatikregel zu diesem Symbol für die (gegebenenfalls rekursive) Konkretisierung des zu diesem Symbol gehörenden Unterbaums im generierten Parsebaum benutzt. Die digitale Analyse und Transformation von strukturierten Texten (DSLs im oben genannten Sinne) erfolgt mit und durch ANTLR grob vereinfacht in drei aufeinanderfolgenden Schritten:

1. Es wird eine lexikalische Analyse des Textes vorgenommen, das heißt im Eingabetext werden die von der Grammatik zugelassenen elementaren Textbausteine (die Tokens, das heißt die „Wörter“ der von der Grammatik implizit definierten formalen Sprache) identifiziert, die dann im weiteren Analyseprozess die Grundlage für die Konstruktion von Parsebäumen bilden.
2. Der von ANTLR automatisch generierte Parser startet den Prozess der schrittweisen Parsebaumgenerierung, indem er die Prozedur zu derjenigen Grammatikregel aktiviert, die nach Einschätzung von ANTLR den größtmöglichen Teil des Eingabetextes bestmöglich abdeckt. Die von ANTLR er-

---

<sup>12</sup> Siehe <https://www.antlr.org/>.

<sup>13</sup> Vgl. Alfred V. Aho, Ravi Sethi und Jeffrey D. Ullman, *Compilers. Principles, Techniques, and Tools* (Reading, Massachusetts et al.: Addison-Wesley, 1986).



zeugten Parser verwenden dafür die ALL(\*) Parsertechnik, die Terence Parr, der Entwickler des Werkzeuges, zusammen mit Sam Harwell entwickelte.<sup>14</sup>

3. ANTLR baut nach von Nutzenden vorgebbaren Transformationsregeln (unter anderem über sogenannte *tree-walker*, siehe zum Beispiel Abbildung 3 unten) den im zweiten Verarbeitungsschritt berechneten „besten“ Parsebaum im Rahmen einer automatischen Traversierung in eine Form um, die für die weitere domänenspezifische Nutzung, Analyse und Visualisierung der in dem Baum kodierten Informationen optimal ist.

Diese für ANTLR typische Untergliederung des Verarbeitungsprozesses in drei jeweils für sich getrennt spezifizierbare Verarbeitungsschritte gestattet es, die eigentliche Analyse der Quelle weitgehend unabhängig von den sich daran anschließenden, domänenspezifischen digitalen Nutzungen des Analyseergebnisses zu halten: Ein und dieselbe ANTLR-Grammatik lässt sich so für die unterschiedlichsten Auswertungen von Quelltexten nutzen, sofern diese Texte nur korrekt durch diese Grammatik beschrieben werden. Und man ist bei diesem Ansatz nicht nur frei darin, welche Art von weitergehender Nutzung der von ANTLR generierten, sprachspezifischen Analyseergebnisse man realisieren möchte, sondern auch darin, wie man diese Nutzung unter Verwendung der dafür gegebenenfalls vorhandenen Werkzeuge und Infrastrukturelemente (Zielprogrammiersprachen, Visualisierer, Nonstandard-Datenbanken, Plattformen und vieles mehr), konkret realisiert. ANTLR ermöglicht so auf eine unkomplizierte und gerade auch für Nicht-Informatiker:innen nachvollziehbare Weise eine komfortable Entwicklung vergleichsweise komplexer sprachtechnologischer Werkzeuge.

Der für Anwender:innen herausforderndste Schritt in diesem Entwicklungsprozess ist natürlich die Identifikation der für die beiden ersten Verarbeitungsphasen benötigten Grammatik. Denn diese muss für die ersten beiden, primär quellenabhängigen Verarbeitungsschritte die Struktur der zu analysierenden Quelltexte angemessen charakterisieren. In unserem konkreten Anwendungsfall erfordert das insbesondere die Ausdifferenzierung der einzelnen Regestentypen, eine Aufgabe, die hauptverantwortlich durch Historiker:innen vorgenommen wird. Eine notwendige Voraussetzung dafür ist die genaue Kenntnis des *Repertorium Germanicum* und seiner standardisierten Abkürzungen sowie der Syntax und Semantik der Regesten selbst. Historiker:innen müssen dafür wissen, welche Abkürzungen zu welchem Rechtsakt gehören und welche von uns so genannten „Markerworte“ den entsprechenden Vorgang kennzeichnen.

---

<sup>14</sup> Terence Parr, *Language Implementation Patterns*, (Dallas, Texas et al.: Pragmatic Bookshelf, 2010).



Beispielsweise steht im dritten Band des RG ein *solv.* immer am Anfang einer Annatenzahlung oder ein *prov. super* immer am Anfang einer Provision (das heißt der Verleihung einer kirchlichen Pfründe durch den Papst).

Über ein systematisches Studium der Viten müssen zu Beginn des Prozesses zur Erstellung von Grammatikregeln nicht nur die Regelmäßigkeiten im Aufbau eines bestimmten Regestentyps herausgearbeitet werden, sondern auch, welche Anteile bei Vorliegen dieses Typs optional sind. Mittels einer Suche nach Markerworten im RG Online lässt sich dafür zunächst eine Auswahl von Regesten zusammenstellen, die zur gleichen Kategorie gehören. Dann lassen sich Regelmäßigkeiten innerhalb dieser Regesten identifizieren und mögliche variierende Ausprägungen ermitteln.

Erst durch das Zerlegen vieler gleichartiger Einträge nach dem gleichen Schema wird offensichtlich, welche Bestandteile dieses Schemas in den Regesten optional und welche obligatorisch sind. Auch hier ist historisches Wissen notwendig, um konkrete Abkürzungen, die an bestimmten Stellen eines Eintrages auftreten können, korrekt zu interpretieren. Beispielsweise können bei einer Provision nach dem Markerwort *prov. de* eine Pfarrkirche (*par. eccl.*) oder ein Kanonikat mit Pfründe (*can. et preb.*) stehen. Dies bedeutet, dass an jener Stelle wahrscheinlich alle Arten von kirchlichen Pfründen auftreten können. Anstatt nun in einer Grammatikregel alle möglichen Pfründen an jeder Stelle ihres Auftretens in der Grammatik aufzuzählen, wird für die Grammatik besser eine eigene Kategorie „Pfründe“ definiert, unter die sich unter Heranziehung des Abkürzungsverzeichnisses des RG Online alle Abkürzungen subsumieren lassen, die in diese Kategorie fallen.

Im nächsten Schritt wird die Grammatik anhand von weiteren Sublemmata getestet, die bei der eigentlichen Entwicklung der Regeln noch nicht in Betracht gezogen wurden, jedoch mit dem gleichen Markerwort beginnen. Immer wenn die Grammatik nicht in der Lage ist, das entsprechende Sublemma in einem Teilbaum korrekt abzubilden, müssen Anpassungen (ein *debugging*) der Grammatik vorgenommen werden. Im einfachsten Fall muss dazu nur eine bestimmte Ausprägung in einer schon bestehenden Kategorie ergänzt werden, in komplexeren Fällen muss die entsprechende Regel für einen Vorgangstyp grundsätzlich alteriert werden.

Im Folgenden wollen wir dies an einem Beispiel, nämlich dem oben bereits vorgestellten ersten Sublemma der Vita des Johannes von Echte (Abbildung 1) demonstrieren.

In einer mit XML (Extensible Markup Language) ausgezeichneten Version wird dieses Regest wie folgt dargestellt:

```
<lemma id="10504399" spStart="" spEnd=""><reg>
```

```
<head> Johannes de Echte Johannes (Alberti) de Echte </head>
```

```
<sublemma vol="5"> Johannes de Eck: <abk ref="415">l.b.</abk> <date
norm="14310700" year="1431">iu1. 1431</date> <fund norm="T:3:25v" l1="T"
l2="3" l3="25v">T 3 25v</fund>.</sublemma>
```

```
<sublemma vol="5"><abk ref="202">cler.</abk> <abk ref="33">Magunt.
</abk> <abk ref="296">dioc.</abk>: de <abk ref="792">vicar.</abk> ad <abk
ref="109">alt.</abk> <abk ref="726">ss.</abk> Laurentii, Cosme et Da-
miani atque Liberii in <abk ref="312">eccl.</abk> <abk ref="671">s.</abk>
Crucis Northusen. <abk ref="33">Magunt.</abk> <abk ref="296">dioc.</abk>
(4 <abk ref="438">m. arg.</abk>) <abk ref="787">vac.</abk> <abk
ref="521">p.</abk> <abk ref="658">resign.</abk> Johannis Wittenborn;
<abk ref="468">n.o.</abk> <abk ref="297">disp.</abk> <abk ref="743">sup.
</abk> <abk ref="267">def.</abk> <abk ref="469">nat.</abk> (<abk
ref="732">subdiac.</abk>, <abk ref="166">c.</abk>) <date
norm="14311124" year="1431">24. nov. 1431</date> <fund norm="S:272:81v"
l1="S" l2="272" l3="81v">S 272 81v</fund>.</sublemma>
```

Deutlich erkennbar ist eine Differenzierung nach Head und mehreren Sublemmata. Im Head stehen Angaben zu dem Antragsteller. Handelt es sich um eine natürliche Person (und nicht beispielsweise um eine juristische Person wie ein Kloster oder ein Domkapitel), dann werden regelmäßig ihr Name und gegebenenfalls weitere individualisierende Angaben wie beispielsweise der akademische Grad und die absolvierten Studienfächer, ihr Weihegrad, eine Ordenszugehörigkeit oder weltliche und geistliche Ämter, die die Person bekleidet, genannt. Diese persönlichen Angaben fassen wir in der Grammatik unter einer Regel mit dem (noch vorläufigen) Namen „Vita“ zusammen. In der ANTLR-Grammatik kann unsere Vita-Regel wie folgt Ausprägungen festlegen, die ihrerseits wieder über speziellere Unterregeln zu charakterisieren sind:

```
vita : ehestand? (funktion | inkardination | akadGrad | kirchAmt | herkunft
| stand | orden | familienstand | pfruende | kardinal | titel )+
```

Bei unserem Beispiel Johannes de Echte sind diese einleitenden biographischen Informationen sehr überschaubar. Angesichts der fehlenden biographischen Angaben im Head konzentrieren wir uns im Folgenden auf eine genauere Erläuterung des Sublemmas. Ein solches Sublemma wird in ANTLR zum Beispiel wie folgt dargestellt:

sublemma : SUBLEMMATAG vita? person? abks?

(provision | annatenVerkehr | dispense | gunstGP | schutzbrief | nachfolge  
| eid | expektanz | verlaengerung | nonObs | ausstellung | erlaubnis |  
gruendung)+

(date | fund)+ CLOSESUBLEMMATAG ;

Hier ist zunächst einmal allgemein festgehalten, dass in einem Sublemma biographische Informationen zum Petenten (im obigen Sinne), eine (weitere) Person oder eine Abkürzung auftreten können. Anschließend folgen in ihm regelmäßig einer oder auch mehrere der im RG vorkommenden Rechtsvorgänge. Dies wird im Sublemma durch das „+“ hinter dem geklammerten Ausdruck signalisiert, der die jeweils möglichen Ausprägungen des Sublemmas durch ein „|“ (es steht logisch für ein „oder“) voneinander getrennt aufzählt. Abgeschlossen wird jedes solche Sublemma durch Tokens des Typs Datum und Fundstelle. Da es sich um eine Arbeitsversion des Parsers handelt, ist diese Liste bei Weitem noch nicht vollständig. Endziel ist es, alle möglichen (respektive die häufigen) Ausprägungen von Sublemmata in jeweils eigenen Kategorien fassen zu können.

Das erste Sublemma der Vita Johannes von Echtes beinhaltet eine Provison. Bei dem Beispiel handelt es sich um eine „normale“ Provision, daher kürzen wir sie mit „nProvision“ ab. Ihr Sublemma ist wie folgt aufgebaut:

nProvision: intro (pfruede acquType? Schätzwert?)+;

Eine normale Provision erwähnt also nach dem typischen Intro, das im 5. Band des RG nur aus „de“ besteht, eine Pfründe, einen optionalen Erwerbstyp („AcquType“, das nachgestellte „?“ signalisiert, dass er auch fehlen kann) und den optionalen Schätzwert der Pfründe. Das „+“ nach dem geklammerten Ausdruck zeigt wieder an, dass sich der eingeklammerte Teil mehrmals wiederholen und dadurch beispielsweise auch der Schätzwert vor dem Erwerbstypen stehen kann. Die Erwerbstypen sind selbst wieder ausdifferenziert, je nachdem um welche Arten von Pfründen es sich handelt.

Da sich Johannes von Echtes um Altäre bewarb, leitet ANTLR den entsprechenden Teil der Regesten korrekt mit der Regel für Altar-Pfründen ab. Diese haben wir wie folgt definiert:

pfrundenAlt : altarIntro (patroName | 'et' | 'atque')\* #'situm'? instType  
patroName? location;

Das Altar-Intro kennzeichnet dabei die verschiedenen möglichen Stellen wie zum Beispiel die Altarvikarie (oder das Fehlen einer konkreten Angabe, wenn sich direkt um den Altar beworben wurde):

```
('<abk ref="791">vic.</abk>' | '<abk ref="792">vicar.</abk>')? 'ad'?
('<abk ref="109">alt.</abk>' | 'altare')
```

Deutlich zu sehen ist, dass im RG nach einem Altar-Intro ein Patrozinium (oder mehrere durch *et* oder *atque* verbundene Patrozinien) des Altars sowie die Institution (in der Regel eine Kirche), in der der Altar gelegen ist, stehen kann. Diese Institution kann auch selbst über ein Patrozinium verfügen, wobei die Optionaltät wieder durch das nachgestellte „?“ signalisiert wird. Der Institution folgen dann der Standort und die Diözese, was wir in unserer ANTLR-Grammatik in einer sogenannten Location-Regel zusammengefasst haben. Zuletzt muss der Schätzwert der Pfründe charakterisiert werden, um die Definition zu vervollständigen. Unserer Regel zufolge besteht er nur aus einem Integer-Wert und einer durch XML-Tags ausgezeichneten Abkürzung für eine Währung:

schätzwert : INT?

```
('<abk ref="354">fl.</abk>'
| '<abk ref="716">sol.</abk>'
| '<abk ref="277">den.</abk>'
| '<abk ref="354">fl.</abk> Renen.'
| '<abk ref="368">gr.</abk>'
| '<abk ref="438">m. arg.</abk>'
| 'fl')
```

Nach dem Schätzwert folgt entsprechend unserer Grammatikregeln der Erwerbstyp. Die im Beispiel des Johannes von Echte genannte Vakanz durch Resignation des Vorbesitzers, Johannis Wittenborn, ist nur eine mögliche Erwerbsausprägung. Weitere Vakanzgründe umfassen beispielsweise den Tod oder die Absetzung des Vorbesitzers; die Möglichkeit, dass der Vorbesitzer die Pfründe mit dem Antragsteller getauscht hat; die Heirat des Vorbesitzers, womit er aus dem geistlichen Stand ausgeschieden war; oder sogar, dass die Gründe der Vakanz unbekannt sind.

Nach dem Namen des Vorbesitzers können in einem Regest zusätzliche persönliche Informationen zu dieser Person stehen, beispielsweise deren Ordenszugehörigkeit oder weitere Pfründen und Ämter, die der Vorbesitzer bekleidet. Eine weitere wichtige (kirchenrechtlich und damit historisch relevante) Information ist der sogenannte Kurien-Status der Pfründe, da das Besetzungsrecht

auf Pfründen, die während eines Aufenthaltes des Vorbesitzers in Rom frei geworden waren, exklusiv beim Papst lag.

```
acquType: (tausch | tod | verzicht | klostereintritt | hochzeit |
translation | absetzung | uebertritt | befoerderung | widerruf | unbekannt
) natPerson (('de' | 'ad') (orden | pfruende))? kurienStatus? ('seu'
acquType)? /*imAngesicht?*/;
```

Ebenfalls zu berücksichtigen sind die unterschiedlichen Tempora für das Freiwerden der Pfründe. Im Beispiel des Johannes von Echte handelt es sich aufgrund der Formel „vac. p. resign.“ um einen Verzicht in der Gegenwart („verzichtistfrei“).

```
verzichtistfrei: '<abk ref="787">vac.</abk>'? ('per' | 'per.' | p. |
'post')? '<abk ref="658">resign.</abk>' ; //
```

Nach dem Erwerbstyp folgt in dem ersten Beispiel von Johannes von Echte das Kürzel „n.o.“, was den Beginn der Nonobstanzenformel markiert. Als Nonobstantien werden in der Regel vor allem Pfründen genannt, die der Petent bereits im Besitz hat. Hierfür enthält unsere Grammatik aus den vorab genannten Gründen bereits die entsprechende Kategorie „pfruende“. Als weitere mögliche Kategorie treten die Dispense hinzu – im vorliegenden Fall die Geburtsdispens. Somit lautet eine naheliegende Grammatikregel für die Nonobstanzenformel:

```
nonObs : '<abk ref="468">n. o.</abk>' ('et')? (pfruende | dispense))+';
```

Der auf der Basis dieser Grammatik durch ANTLR nach den ersten beiden Verarbeitungsschritten generierte Baum sieht wie folgt aus:

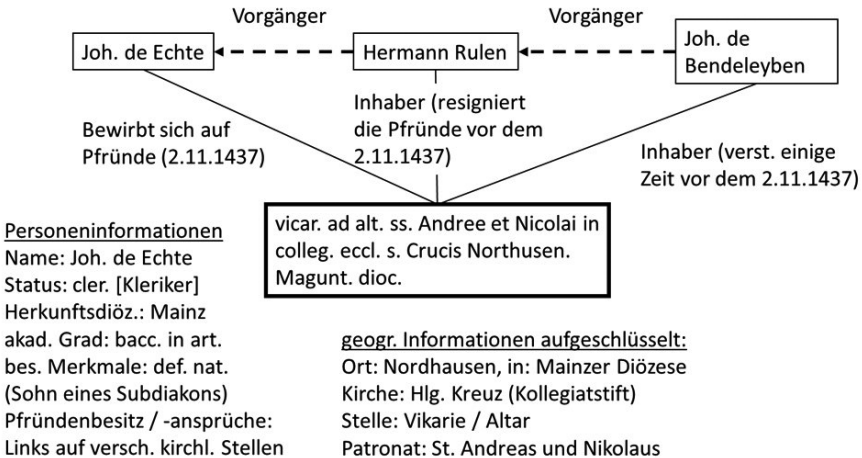


im *Repertorium Germanicum* enthaltenen biographischen und anderen Informationen in inhaltlich angemessener Weise kategorial zu ordnen und zu repräsentieren. Das Spezialwissen von Historiker:innen, wie bestimmte Quellenbefunde zu lesen und auszudeuten sind, wird somit explizit in das Datenmaterial inkorporiert, was dessen weitergehende Analyse nach unterschiedlichsten Fragestellungen gestattet. Ein Sublemma wie zum Beispiel das in der Vita des Johannes von Echte zum 2. November 1437 überlieferte, kann dann in ganz verschiedener Hinsicht personen- und institutionengeschichtlich oder auch netzwerkanalytisch ausgewertet werden (Abbildung 3).

pres. in R. cur.: de vicar. ad alt. ss. Andree et Nicolai in colleg. eccl. s. Crucis Northusen. Magunt. dioc. (5 m. arg. ) vac. p. resign. Hermann Rulen presb. vel p.o. Johanns de Bendeleyben; n.o. vicar. in capel. s. Spiritus e.m. op. Duderstat d. dioc. (2 m. arg. ), disp. sup. def. nat. ( subdiac., c. ) et quod pater ipsius J. cantor. c. can. et maior. preb. d. eccl. s. Crucis Northusen. obtin. 2. nov. 1437 S 341 224vs.

**Aus der Vita des Johannes von Echte im RG V**

2. nov. 1437



**Abb. 3:** Unterschiedliche Informationen aus einem Sublemma der Vita des Johannes von Echte (RG V 4399)

Die Abbildung zeigt, wie umfangreich und komplex die Sachverhalte sind, die schon in einem einzigen von zigtausenden RG-Einträgen dokumentiert sind. Diese Informationen möglichst erschöpfend automatisiert aus der Quelle herauszulesen, ist die Herausforderung, vor der wir stehen. Ermöglicht werden wird dies aber erst durch eine Ontologie, die nicht nur die entsprechenden *termini technici* enthält, sondern auch Abhängigkeiten zwischen diesen Begriffen beschreibt; damit kann sie einerseits als ein standardisiertes Vokabular dienen und andererseits interessante Arten von Schlussfolgerungen auf den Quelltext-



ten ermöglichen.<sup>15</sup> So ist es letztlich ein Fernziel unseres Core-H-Projektes, exemplarisch eine auf die Themenstellung der Prosopographie spätmittelalterlicher gelehrt-klerikaler Eliten spezialisierte Ontologie zu entwickeln, die aus Gründen der Generalisierbarkeit und Verknüpfbarkeit mit anderen Datenbeständen als Spezialisierung entsprechender Konzepte des bekannten CIDOC-CRM-Standards formalisiert ist. Diesem Ziel sind wir durch die Arbeit an der ANTLR-Grammatik bereits ein gutes Stück nähergekommen.

All dies ist beileibe kein einfacher Stoff. Es gibt eine Reihe von Spezialaufsätzen, die man studieren muss, wenn man die Regesten des *Repertorium Germanicum* richtig lesen und verstehen will. Die Herausforderung, dieses Wissen in ein digitales Begriffs- und Regelsystem zu überführen, ist entsprechend groß. Dennoch erscheint die Aufgabe, diese konkreten Verstehensvorgänge der Historiker:innen zu einem wesentlichen Teil zu algorithmisieren, grundsätzlich lösbar. Die Entwicklung einer entsprechenden Ontologie und die darauf aufbauende Erschließung von digitalen Quellenkorpora in einer mehr und mehr umfassenden Wissensbasis, würden historisches Wissen in einer völlig neuen Form verfügbar machen. Am Ende winkt schließlich jene Verheißung, die Johannes Haller schon vor über hundert Jahre mit Blick auf die Akten des Vatikanischen Archivs aussprach: „So bietet, bei aller Unscheinbarkeit im einzelnen, dieses Material, im großen genommen, ein Hilfsmittel zur Belebung des geschichtlichen Bildes dar, wie es reicher und zuverlässiger nicht leicht gedacht werden kann.“<sup>16</sup>

## Bibliographie

- Aho, Alfred V., Sethi, Ravi und Jeffrey D. Ullman. *Compilers. Principles, Techniques, and Tools*. Reading, Massachusetts et al.: Addison-Wesley, 1986.
- Brosius, Dieter. „Das Repertorium Germanicum.“ In *Das Deutsche Historische Institut in Rom 1888–1988*, hg. v. Reinhard Elze und Arnold Esch, 123–165. Tübingen: Niemeyer, 1990.
- Deeters, Walter. „Über das Repertorium Germanicum als Geschichtsquelle. Versuch einer methodischen Anleitung.“ *Blätter für deutsche Landesgeschichte* 105 (1969): 27–43.
- Gramsch, Robert. „Kuriertätigkeit als ‚Berufsbild‘ gelehrter Juristen. Der Beitrag Roms zur Akademisierung Deutschlands im Spätmittelalter. Eine personengeschichtliche Betrachtung.“ *Quellen und Forschung aus italienischen Archiven und Bibliotheken* 80 (2000): 117–163.

---

<sup>15</sup> Vgl. hierzu etwa Wolfgang Hesse, „Ontologie(n),“ in: *Informatiklexikon der Gesellschaft für Informatik*, 2005. <https://gi.de/informatiklexikon/ontologien> (Zugriff am 13.10.2021) sowie allgemein John F. Sowa, *Knowledge Representation. Logical, Philosophical, and Computational Foundations* (Pacific Grove, California: Brooks/Cole Publishing, 2000).

<sup>16</sup> Haller, „Repertorium Germanicum,“ fol. 9.

- Gramsch-Stehfest, Robert. „Von der Metapher zur Methode: Netzwerkanalyse als Instrument zur Erforschung vormoderner Gesellschaften.“ *Zeitschrift für Historische Forschung* 47 (2020): 1–39.
- Hasselhorn, Benjamin. *Johannes Haller: Eine politische Gelehrtenbiographie: Mit einer Edition des unveröffentlichten Teils der Lebenserinnerungen Johannes Hallers*. Göttingen: Vandenhoeck & Rupprecht, 2015.
- Hesse, Wolfgang. „Ontologie(n).“ In: *Informatiklexikon der Gesellschaft für Informatik*, 2005. <https://gi.de/informatiklexikon/ontologien>. Zugriff am 13.10.2021.
- Hörnschemeyer, Jörg. „Repertorium Germanicum Online.“ In *Friedensnobelpreis und historische Grundlagenforschung. Ludwig Quidde und die Erschließung der kurialen Registerüberlieferung*, hg. v. Michael Matheus, 605–615. Berlin: De Gruyter, 2012.
- Parr, Terence. *Language Implementation Patterns*. Dallas, Texas et al.: Pragmatic Bookshelf, 2010.
- Schmugge, Ludwig. *Kirche, Kinder, Karrieren: Päpstliche Dispense von der unehelichen Geburt im Spätmittelalter*. Zürich: Artemis, 1995.
- Schwarz, Brigide. „Römische Kurie und Pfründenmarkt im Spätmittelalter.“ *Zeitschrift für Historische Forschung* 20 (1993): 129–152.
- Schwarz, Brigide. „Das Repertorium Germanicum: Eine Einführung.“ *Vierteljahrsschrift für Sozial- und Wirtschaftsgeschichte* 90 (2003): 429–440.
- Sowa, John F. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Pacific Grove, California: Brooks/Cole Publishing, 2000.

