

Mittelalterliche Schreiber und sprachliche Eigenheiten: maschinelle Auswertung deutschsprachiger Urkunden des 13. und 14. Jahrhunderts

Fiebig, Annegret

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Fiebig, A. (1995). Mittelalterliche Schreiber und sprachliche Eigenheiten: maschinelle Auswertung deutschsprachiger Urkunden des 13. und 14. Jahrhunderts. *Historical Social Research*, 20(2), 278-283. <https://doi.org/10.12759/hsr.20.1995.2.278-283>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Mittelalterliche Schreiber und sprachliche Eigenheiten. Maschinelle Auswertung deutschsprachiger Urkunden des 13. und 14. Jahrhunderts

*Annegret Fiebig (Berlin)**

Die Voraussetzung zu den folgenden Ausführungen sind Überlegungen zur Entstehung mittelalterlicher Originalurkunden. Im Unterschied zu Büchern, also zu Handschriften mit im engeren Sinne literarischem Inhalt, handelt es sich bei Originalurkunden nicht um Abschriften von einer Vorlage. Urkunden wurden zwar unter Verwendung von Formeln und wiederkehrenden Wendungen verfaßt, und ältere Vorlagen haben sicherlich als Muster gedient. Der Text wurde aber jeweils neu formuliert, weshalb Urkunden als frühe Vertreter mittelalterlicher >Originale< gelten können.

Die Methode zur computergestützten Sprachuntersuchung wurde für deutschsprachige Urkunden aus Österreich aus dem 13. und frühen 14. Jahrhundert entwickelt und bildet einen Teil meiner Dissertation. Ich verfolge damit zwei Ziele: Zum einen sollen bestimmte sprachliche Konstanten und Varietäten ermittelt werden. Neben der Analyse des Schreibdialekts werden die syntaktischen Konstruktionen, die Mittel für die Textgliederung sowie >innere Urkundenmerkmale< ausgewertet. Der zweite Schritt besteht darin, aus den Ergebnissen weiterführende Schlüsse zu ziehen: Läßt sich aus den erarbeiteten Daten der Nachweis führen, daß bestimmte sprachliche Merkmale für einen individuellen Schreiber oder für ein Skriptorium konstitutiv sind? Die Auffindung sprachlicher Abweichungen und Übereinstimmungen könnte somit einen erweiterten >Diktatvergleich< darstellen und Aufschluß über Schreibzusammenhänge und Schriftorganisation geben.

Auswahl des Untersuchungsmaterials

Um die Tragfähigkeit einer Untersuchungsmethode zu erweisen, muß von einer vergleichbaren Textbasis ausgegangen werden. Etwa 470 Urkunden aus der

* Protokoll des 61. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 2. Juli 1994.

Region Niederösterreich bilden die Untersuchungsgrundlage. Es handelt sich dabei vor allem um Urkunden, die im Zusammenhang mit den >Landherren<, den aufstrebenden österreichischen Ministerialen im 13. Jahrhundert, stehen. Als erstes Landherrngeschlecht in Österreich urkundeten die einflußreichen Herren von Kuenring auf deutsch (1281). An Leutold von Kuenring, dem Inhaber des Schenkenamts unter Rudolf von Habsburg und Anführer des Aufstands gegen den Herzog, ist die Zeitspanne für die Untersuchung orientiert: 1281, der Beginn österreichischer Beurkundung in deutscher Sprache, und 1312, das Todesjahr Leutolds von Kuenring, bilden die Eckdaten. Das Kriterium für die Auswahl war eine Beteiligung am Beurkundungsgeschäft - als Aussteller, Empfänger, Zeuge oder in sonstiger Funktion. Zum Vergleich wurden Urkunden der Herren von Ebersdorf aufgenommen, deren gesamtes mittelalterliches Archiv im Niederösterreichischen Landesarchiv überliefert ist. Eine dritte Urkundengruppe stellen die städtischen Urkunden Wiens dar.

Aufbereitung des Materials bei der Computererfassung

Die deutschsprachigen Urkunden des 13. Jahrhunderts sind in diplomatischem Abdruck im *Corpus der altdeutschen Originalurkunden* ediert. Sie wurden über das Texterkennungssystem >Optopus< in Tübingen eingelesen. Hinzu kommen lateinische Urkunden sowie Urkunden aus dem 14. Jahrhundert, die in österreichischen Archiven vom Original transkribiert wurden. Dabei wurde der Text mit folgenden Codierungen ausgezeichnet:

Anfang und Ende einer Urkunde erhielten eigene Kennungen. Der Beginn der einzelnen Formulareile wurde bezeichnet. Orts- und Personennamen wurden mit Codierungen versehen, in denen gleichzeitig Kürzel zur jeweiligen Funktion in der Urkunde angegebenen sind (beispielsweise >Person< als Aussteller (pa), als Zeuge (pz), >Ort< als Ausstellungsort (oaj usw.).

Zusätzlich wurde der aus dem handschriftlichen Original erschlossene Befund eingearbeitet:

Die Zeilenumbrüche des Originals wurden vermerkt.

Initialen, Majuskeln und *litterae notabiliores*, also auf irgendeine Weise hervorgehobene Buchstaben, erhielten Kennungen.

Abkürzungen wurden - mit Ausnahme von Nasalstrichen - in normalisierten Schreibungen aufgelöst, dabei jedoch markiert.

Außerdem wurde der gesamte Text auf der Grundlage des *Wörterbuchs der mittelhochdeutschen Urkundensprache* lemmatisiert. Zuerst wurde ein Wortformenregister (ohne Kontext) erstellt, so daß jedes Wort in einer eigenen Zeile ausgegeben wurde. Aus der dem Wörterbuch zugrundeliegenden Lemmaliste (8776 Stichwörter) wurden Kennungen für die einzelnen Einträge von Hand in

den Index eingefügt, so daß über das Kommando #KORREKTURAUSFUHRE die Lemmata dann in den Text eingeschrieben werden konnten.

Kriterien für die Untersuchung. Untersuchungsschritte

Vom gesamten Textcorpus wurde ein Wortindex sowie ein KWIC-Index erstellt. Für weitere Untersuchungsschritte kann als einzig sichere Annahme vorausgesetzt werden, daß jede Urkunde für sich genommen ein eigenes Schreibsprachsystem darstellt. Ein Schreiber hat mit einem bestimmten Diktat oder von einer bestimmten Vorlage zu einem bestimmten Zeitpunkt eine Urkunde, in der ein bestimmtes Rechtsgeschäft beurkundet wird, geschrieben. Da für jede andere Urkunde davon auszugehen ist, daß sich zumindestens einer der genannten Faktoren verändert hat, ist es notwendig, jeden Text gesondert auszuwerten.

1. Gliederungszeichen und Layout

Der Aufbau einer Urkunde und die Unterteilung des Textes spielen eine wichtige Rolle beim Textvergleich. Hierbei ist zu unterscheiden zwischen >äußeren Urkundenmerkmalen<, der Ausführung von Initialen und Majuskeln, Schrifttrand etc. und den Zeichen zur Textgliederung, die als sprachliche Merkmale zu bezeichnen sind. Zur inhaltlichen Untergliederung eines Textes werden Interpunktionszeichen verwendet, Initialen, Majuskeln oder litterae notabiliores. Auch Zeilenwechsel können mitunter die Funktion von sinnhaften Einschnitten haben.

Interpunktion

Es wird zwischen dem Inventar der Zeichen in einer Urkunde und ihrer Funktion unterschieden. Insgesamt erscheinen in den Urkunden die Zeichen *punctus* in verschiedenen Positionen auf oder über der Zeile, *virgula suspensiva*, *virgula plana*, *punctus elevatus*, Komma, Apostroph, Kolon, Doppelpunkt, Bindestrich und *paragraphus*. Hinzu kommen diverse Schlußzeichen, deren zum Teil sehr aufwendige Ausführung die maschinelle Erfassung nicht sinnvoll erscheinen läßt. In jeder Urkunde tritt eine bestimmte Auswahl dieser Zeichen in jeweils unterschiedlicher Zusammenstellung auf.

Das Inventar ist über den Wortindex leicht zu erfassen. Dabei ist die Angabe der relativen Häufigkeit der Zeichen wichtig, denn die Frequenz der Interpunktionszeichen schwankt in den einzelnen Urkunden beträchtlich.

Der KWIC-Index gibt eine Übersicht über die Funktionen: Interpunktionszeichen können zur Abgrenzung der einzelnen Formulareile verwendet werden, in Aufzählungen, als Referenzpunkte, zur Hervorhebung von Zahlen oder als Abschluß vor neuen Zusatzbestimmungen.

Für die Gewinnung von Daten über die Frequenz der Interpunktionszeichen, ihre Relevanz für die Prosodie und über ihre Bedeutung als Zäsur in der münd-

lichen Rede ist es für die maschinelle Auswertung notwendig, zu vereinfachen. Hier erwies sich die Zahl der Wörter zwischen einzelnen Zeichen als tragfähiges Kriterium. Über ein #KOPIERE-Programm werden die Wörter zwischen Interpunktionszeichen zunächst ausgezählt und aus den Zahlen der empirische Mittelwert errechnet. Da dieser Wert allein keine tragfähige Aussage erlaubt, wurde darüber hinaus die empirische Standardabweichung ausgerechnet. Diese rein quantitativen Werte enthalten keine exakten Informationen über den Sprachrhythmus und inhaltlich betonte Schwerpunkte in einer >Redeeinheit<. Die Zahlen können jedoch als Näherungswerte interpretiert werden.

Im Anschluß ergab die Diskussion, daß die einzelnen Formulareile gesondert ausgewertet werden müssen. Denn besonders an den Zeugenlisten wird offensichtlich, daß hier, aufgrund der Abgrenzung der Namen durch Zeichen, eine viel höhere Frequenz von Interpunktionszeichen vorliegen muß als in den anderen Teilen einer Urkunde.

2. Wortfolgen

Die Dispositio beginnt in den meisten Fällen mit einem daz-Satz. Danach treten in der größten Zahl der Urkunden bestimmte Wortfolgen zur Anbindung von rechtlichen Zusatzbestimmungen auf. Neue Sinneinheiten setzen häufig mit Konjunktionen oder pronominalen Subjekten ein. An zweiter Stelle folgt meistens das finite Verb: man sol, vn hat, ich *han ouch* und ähnliches. Mit folgenden Konjunktionen und Pronomina werden neue Bestimmungen häufig angebunden: *man, ouch, also, so, vnd*. Speziell in rechtssetzenden Urkunden treten außerdem konditionale Fügungen auf. Der Stil der Urkunden unterscheidet sich unter anderem durch die Wortwahl und -position beim Einsatz neuer Bestimmungen. Über den KWIC-Index ist die Analyse der Funktion möglich: Es ist schnell zu ersehen, ob beispielsweise *und* in erster Linie in Aufzählungen oder mit nachfolgenden Verben als Einleitung von neuen Sätzen oder Bestimmungen auftritt. (In der Urkundensprache werden häufig zusätzliche rechtliche Informationen und Grundlagen als neue Redeeinheiten formuliert, ohne daß das vorher genannte Subjekt wiederholt werden würde.)

3. Wortstellung

Das finite Verb entscheidet über die Art der Rechtshandlung - über Verkauf, Stiftung, Schenkung, Testament, Privilegienbestätigung, Schlichtung, Einigung oder Gesetzgebung. Dabei variieren beispielsweise *verkoufen - koufen - ze koufenne geben, (ver)lihen - ze lihenne geben*. Ist die Wahl der Vokabel für den Urkundentyp wichtig, so im Hinblick auf die Schreiberpraxis ihre Position: Das Verb kann schon bald nach Beginn der Dispositio gesetzt sein, unmittelbar hinter dem Subjekt. Im Unterschied dazu kann es jedoch auch erst viel später, nach Dativobjekten der Person, Akkusativobjekten der Sache, möglicherweise auch nach adverbialen Wendungen und

präpositionalen Ergänzungen erscheinen. Der Satzrahmen kann noch beliebig erweitert werden durch den Einsatz von adverbialen und präpositionalen Ergänzungen. Entsprechend entsteht eine geringe oder hohe Satzspannung. Obwohl aufgrund der Lemmatisierung die Position der Verben mit rechtlicher Bedeutung maschinell ausgezählt werden könnte, enthält ein solches Verfahren eine zu hohe Fehlerquote. Die Formulierungen mit Verben in Sperrstellung variieren zu sehr, außerdem erschweren anakoluthische Konstruktionen die exakte maschinelle Zählung. Sinnvoller erscheint es, den Beginn der Dispositio festzuhalten und bei der Lektüre selbst Wahl und Position des Verbs festzustellen.

4. Abkürzungen

Die Markierung der Abkürzungen bei der Eingabe ermöglicht es, rasch festzustellen, für welche Silben und Buchstaben Kürzel verwendet wurden und wie oft. In deutschen Texten erscheinen wesentlich weniger Abkürzungen als in lateinischen Urkunden, weshalb die auftretenden Unterschiede der Verwendung weniger stark ins Gewicht fallen. Sie sind aber dennoch von Bedeutung für die Schreibpraxis. Hinsichtlich der Ausführung von Kürzeln muß im Einzelfall das handschriftliche Original herangezogen werden.

5. Lexikalischer Befund, Vokabular.

6. Schreibungen, Graphie

Die Lemmatisierung ermöglicht eine objektive und vom Eindruck der Schreibungen unabhängige Auswertung des Vokabulars. Auch bei der Analyse der einzelnen Schreibungen kann auf die normalisierten Wortformen rekurriert werden. Auf diese Weise werden die Wörter nach >normalmittelhochdeutschen< Lang- und Kurzvokalen, Diphthongen und den Konsonantenschreibungen im An- und Inlaut sortiert.

7. Analyse der Formulareile

Zusätzlich zu den sprachlichen Auswertungen wird das Urkundenformular untersucht. Die Kennungen geben an, welche Teile in einer Urkunde enthalten sind und in welcher Reihenfolge sie erscheinen. Zu den >inneren Merkmalen< einer Urkunde gehört außerdem die Formulierung von Protokoll und Eschatokoll, besonders die Datierungsangabe sowie möglicherweise Corroboratio und Renuntiationsformel. Eine objektive Vergleichsmöglichkeit entsteht dadurch, daß der normalisierte Text der einzelnen Formulareile untereinander geschrieben und verglichen wird.

8. Orts- und Personennamen

Ergänzend zum Wortindex wird auf der Grundlage der normalisierten Schreibungen ein Register der Orts- und Personennamen erstellt. Die Liste enthält Angaben über die jeweilige Funktion beim Beurkundungsvorgang: Es kann

sich um Aussteller, Empfänger, Zeugen oder Urkundsparteien in sonstiger Funktion handeln, um Ausstellungsorte, verhandelte Liegenschaften, Ort der mündlichen Rechtsabmachung o.ä.

9. Äußere Urkundenmerkmale, paläographischer Befund

Zur Überprüfung der Untersuchungsergebnisse ist die paläographische Auswertung und Überprüfung der äußeren Urkundenmerkmale unerlässlich. Übereinstimmungen oder Abweichungen des paläographischen und sprachlichen Befundes erlauben Schlüsse auf die Tragfähigkeit der Methode.

Bisherige Ergebnisse

Bisher lassen Einzelergebnisse erkennen, daß der sprachliche und der paläographische Befund einander entsprechen. Beispielsweise konnten fünfzehn Urkunden unterschiedlichen *Typs* aufgrund von Übereinstimmungen der Interpunktion, der Wortstellung und der Schreibungen zu einer Gruppe zusammengestellt werden. Alle Urkunden sind der gleichen Hand zuzuordnen, die als Kuenringer-Schreiber >Kuenr. K< identifiziert werden konnte (Sigle von ZAWREL).

Eine zweite Gruppe von sechs Urkunden stammt ebenfalls von derselben Hand aus dem Bereich der Wiener städtischen Kanzlei.

Edition und Erschließung eines frühmittelalterlichen Gesetzbuches:

Die Kapitulariensammlung des Abtes Ansegis

*Gerhard Schmitz (München)**

Kapitularien sind in Kapitel gegliederte legislative Texte höchst verschiedener Art, die von karolingischen Herrschern - insbesondere von Karl dem Großen und Ludwig dem Frommen - erlassen wurden. Thematisch decken sie ein weites Spektrum ab: Teils werden durch sie die einzelnen Volksrechte ergänzt oder verändert, teils werden wichtige politische Dinge geregelt, teils aber auch ganz konkrete Einzelfälle entschieden. Sie beziehen sich gleichermaßen auf weltliche und kirchliche Angelegenheiten, in ihnen spiegeln sich soziale und wirtschaftliche Verhältnisse. Mit dieser Spannweite bilden sie eine der wichtigsten Quellengruppen für die karolingische Zeit: keine andere verrät mehr über

* Protokoll des 61. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 2. Juli 1994.

die konkrete Administration des karolingischen Reiches, über die Verfassungswirklichkeit (im weitesten Sinne verstanden) dieser Zeit und über die Versuche der Karolinger, ihren Willen bekanntzugeben und durchzusetzen.

Anders als Urkunden haben Kapitularien keine bestimmte Form: Es kann sich um durchformulierte Stücke handeln, die in mancher Hinsicht den Urkunden nahekommen, es kann sich aber auch - und dies ist die Mehrzahl der Fälle - um mehr oder weniger formlose Erlasse handeln, bis hin zu stichwortartig formulierten Aide-mémoires, die den königlichen Amtsträgern als Erinnerungsstütze bei ihren Reisen in bzw. durch ihre Amtsbezirke dienten.

Über die Verbreitung der Kapitularien ist nicht allzu viel bekannt: Vor allem zur Zeit Ludwigs gibt es zwar vereinzelt Vorschriften über eine geordnete Vervielfältigung und Bekanntgabe, aber es ist schwer zu sagen, ob diese Anordnungen eingehalten wurden. Im Normalfall wird es so gewesen sein, daß Bischöfe, Äbte und weltliche Amtsträger selbst Abschriften eines Kapitulars nehmen mußten, schon weil die Schreibkapazität der königlichen Kanzlei kaum ausgereicht haben dürfte, innerhalb kurzer Zeit eine genügende Anzahl von Einzelexemplaren herzustellen. Noch weniger wissen wir über die Archivierung der erlassenen Gesetze und Vorschriften am karolingischen Hof. Man muß annehmen, daß dort wenigstens von den wichtigeren Kapitularien authentische Fassungen aufbewahrt wurden, aber ganz sicher geschah dies nicht in einer geordneten Form: Ein offizielles, »amtliches« Gesetzbuch hat es - anders etwa als bei den Langobarden - im fränkischen Reich der Karolinger nicht gegeben.

In diese Lücke stieß 827 der Abt Ansegis von St. Wandrille: Er hat alle Kapitularien, derer er habhaft werden konnte, gesammelt und in vier Büchern und drei Appendices zusammengefaßt. Das Material versuchte er nach kirchlichen und weltlichen Gesichtspunkten zu ordnen: In den Büchern I und II sollten die kirchlichen Gesetze Karls bzw. Ludwigs, in den Büchern III und IV die weltlichen Verordnungen zu finden sein. Tatsächlich hat er dieses Gliederungsprinzip mehrfach durchbrochen, so daß von einer konsequenten Durchführung seines Plans nicht die Rede sein kann. Immerhin: durch sein Werk und die beigegebenen Kapitelverzeichnisse machte Ansegis die Kapitularien in einer ganz anderen Weise verfüg- und zitierbar, als dies bis zum Erscheinen seines Werkes möglich war. Den Fortschritt, den dies für die Administration des Reiches bot, hat man am Hof sofort erkannt. Schon 829 hat Ludwig der Fromme die Sammlung des Ansegis in seinen Wormser Kapitularien zitiert und nicht zuletzt dadurch dem *liber legiloquus* einen offiziellen Charakter verliehen. Dies hat zu der enormen Beliebtheit des Werkes, das wohl bald schon in nahezu jeder besseren Bibliothek des Frankenreiches zu finden gewesen sein dürfte, erheblich beigetragen: Noch heute läßt sich erkennen, daß mit den Wormser Gesetzen von 829 ein enormer Verbreitungsschub einsetzte.

Die im Rahmen der Monumenta Germaniae Historica erscheinende Neuedition der Sammlung wird mit TUSTEP vorbereitet. Auf Grund verschiedener Umstände konnte TUSTEP leider nicht von Anfang an, sondern erst nach Ab-

schluß der ersten Kollation und Erstellung eines ersten Editionstextes eingesetzt werden. Die vielfältigen Möglichkeiten des Programms boten von diesem Zeitpunkt ab eine willkommene Unterstützung: eine effiziente Nachkollation z. B. ließ sich dadurch erreichen, daß lediglich jeweils die Varianten einer Handschrift oder Handschriftengruppe ausgedruckt wurden. In Anbetracht der zahlreichen in die Edition einbezogenen Überlieferungen bot sich hier die Möglichkeit einer wirksamen Fehlerminimierung. Dank der Flexibilität des Programms und seiner Kapazität im ganzen war es möglich, die äußeren Besonderheiten einer MGH-Edition auch im Erscheinungsbild genau einzuhalten. Dies gilt auch für das Register, dem als Basis ein KWIC-Index zugrundelag. Durch die Verwendung spezieller Sortierkriterien war es möglich, die innerhalb eines Lemmas gebotenen Wortverbindungen nach sachgerechten Kriterien zu ordnen. Weil die so gewonnenen Registereinträge wieder in die Quelldatei eingespielt werden können, konnte ein fertiges Register von Personen, Orten, Sachen und Wörtern erstellt werden, ohne daß ein endgültiger Umbruch des zu edierenden Textes mit definitiven Seitenzahlen vorlag - ein im Vergleich zu herkömmlichen Prozeduren der Registererstellung arbeits- und zeitsparendes Verfahren.

Nach Abschluß der Edition wird ein komplettes Programmpaket vorliegen, das auch anderen Editoren zur Verfügung gestellt werden soll (und partiell bereits heute von anderen benutzt wird). Es wird allerdings nötig sein, die einzelnen Programmteile noch einmal zu überarbeiten und zu einem in sich stimmigen Baukasten zusammenzufügen. Nach Auffassung des Referenten sind nämlich viele mit Editionsprojekten befaßte Kollegen nicht in der Lage, umfangreichere und vom Ablauf her optimale TUSTEP-Programme selbst zu schreiben. Sie arbeiten daher meist mit Textverarbeitungsprogrammen, die sie zwar selber einigermaßen kennen, die aber für Editionsprojekte gar nicht konzipiert wurden und deshalb dafür letztlich auch nicht geeignet sind. Wo nicht, wie in Tübingen, eine unkomplizierte und direkte Kontaktmöglichkeit mit den entsprechenden Mitarbeitern des ZDV besteht, ist eine eigene TUSTEP-Programmierung oft zu kompliziert und für den weniger Erfahrenen auch zu zeitaufwendig. Da sich zudem absehen läßt, daß für Editionen lediglich ein Teil der von TUSTEP gebotenen Möglichkeiten genutzt werden muß, sieht der Referent in der Bereitstellung von aufeinander abgestimmten Programmbausteinen bzw. von Musterdateien einen vernünftigen Weg, eine ganz bestimmte Anwendergruppe instand zu setzen, ihre Edition mit professionellen Mitteln zu bearbeiten, ohne selbst Programmprofi zu sein.