

Nutzungsperspektiven elektronischer Texte in der historischen Sprachwissenschaft -- Eine Skizze

von THOMAS GLONINIG, Marburg

Die meisten Editionen historischer Texte nehmen heute ihren Weg über eine elektronische Textfassung. Die elektronische Fassung ist entweder eine handwerklich notwendige Zwischenstufe beim traditionellen Verfahren des Drucks von Texten oder aber die elektronische Fassung ist die Edition. In der vorliegenden Skizze möchte ich in erster Linie erläutern, welchen Wert elektronische Texte für die historische Sprachwissenschaft haben können, im Anschluß daran möchte ich für die Retrodigitalisierung historischer Quellen werben und das Problem der Nutzung bereits vorhandener elektronischer Texte ansprechen, die dem elektronischen Satz und Druck zugrundeliegen, aber nicht in den Pool der allgemein zugänglichen Forschungsressourcen eingehen.

Editoren historischer Texte und historische Sprachwissenschaftler sind wechselseitig auf einander angewiesen, im Idealfall sind die entsprechenden Kenntnissysteme in einer Person vereinigt. Der Editor braucht sprachwissenschaftliche Kenntnisse, um das Prinzip der *sprachadäquaten* Edition erfüllen zu können, z.B. bei der Herstellung des Textes, bei der Beurteilung von Fehlern und Schreibweisen, bei der Lokalisierung und der Datierung von Texten, aber auch bei der kommentierenden Erschließung des Textes in Anmerkungen und Glossaren; umgekehrt stellt der edierte Text neues Material für den Sprachgeschichtler dar, und die Forschungsergebnisse des Editors (als eines guten Kenners eines bestimmten Texts) fließen zurück in den sprachwissenschaftlichen Wissensbestand. Mittlerweile sind eine ganze Reihe interessanter Modelle für die elektronische Edition neuer Texte und auch für die Retrodigitalisierung vorhandener Editionen zu erkennen. Bevor ich zum Wert und zu den Nutzungsperspektiven elektronischer Texte als der Herzstücke von Editionen komme, möchte ich kurz die Frage beleuchten, welche Darstellungsformen und Vernetzungsmodelle sich eignen, diesen wechselseitigen Zusammenhang zwischen editorischen Entscheidungen/ Zielsetzungen und sprachwissenschaftlichen Kenntnissen/ Interessen darzustellen. Ich werfe zunächst einen Rundblick auf Darstellungselemente und Vernetzungsformen.

Die Darstellungselemente, die in Editionen verwendet werden, gliedern sich hauptsächlich in vier Gruppen:

- (i) *Textelemente* (kritischer Text, normalisierte/ diplomatische Textfassungen, Textstufen, Dokumentation von Varianten)
- (ii) *Bilder* der Überlieferung
- (iii) Elemente der *Textvermessung* (z.B. kanonische Referenzen)
- (iv) *Erschließende Elemente* (z.B. Glossareinträge, Stellenverzeichnis-Einträge, Konkordanz-Einträge, Parallelstellen-/ Quellenhinweise, systematische Register-einträge/ Markierungen; sprachliche, textkritische, sachliche Anmerkungen, ggf. auch darauf bezogene Rechtfertigungstexte).

Zu den Formen der Vernetzung gehören vor allem Hyper-Links unterschiedlicher Art, synoptische Darstellungen, Fenster- und Rahmentechniken, Farbleitsysteme und der Gebrauch von Textvermessungs-Elementen als Absprung- und Zielpunkte. Im Hinblick auf die Organisation linearer Abfolgen im elektronischen Text eignen sich vor allem sog. ‚geführte Touren‘. Von besonderem Wert für die historische Sprachwissenschaft sind die Darstellungselemente der ersten Gruppe, vor allem die Texte und für bestimmte Fragestellungen auch die dokumentierten Varianten. Ich möchte nun erläutern, worin der Wert einer allgemein zugänglichen Sammlung elektronischer Texte für die historische Sprachwissenschaft bestehen kann.

Einige sprachwissenschaftliche Nutzungsperspektiven elektronischer Dokumente

Elektronische Dokumente und Editionen sind Arbeitsmittel, die nicht in unmittelbarer Konkurrenz zum Buch und zu anderen gedruckten Texten stehen. Sie sind eher Erscheinungsweisen von Texten, die ihre Nachteile haben, die in bestimmten Arbeitszusammenhängen aber eine Reihe von Vorteilen gegenüber gedruckten Texten haben; einige dieser Vorteile für die historische Sprachwissenschaft sind seit längerem bekannt, z.B. ihr Nutzen bei der Produktion von Wörterbüchern. Andere Nutzungsperspektiven sind -- soweit ich sehe -- relativ neu, z.B. die Rolle, die sie bei der Untersuchung und Beurteilung der äußerst komplexen Phänomene des Sprachwandels spielen können. Ausgangspunkt der folgenden Überlegungen zu Vorteilen und Nutzungsperspektiven sei folgendes Szenario.

(A) Der historische Sprachwissenschaftler A hat im Jahre 1980 eine persönliche Handbibliothek mit alten Texten, die ihn besonders interessieren, eine Instituts- und eine Universitätsbibliothek, deren Bestände ganz gut sind, die aber wie überall die Zufälligkeiten der Institutsgeschichte und der Beschaffungspolitik spiegeln. Dann gibt es noch die Bestände, die per Fernleihe erreichbar sind und die Altbestände, die man sich auf Mikrofilmen oder durch Reisen in die betreffende Bibliothek oder das betreffende Archiv beschaffen muß.

(B) Die historische Sprachwissenschaftlerin B hat im Jahre 2000 dieselben Voraussetzungen mit ein paar Editionen mehr, die in den 20 Jahren dazugekommen sind. Darüber hinaus hat sie auf ihrer Festplatte eine kleine Sammlung elektronischer Texte angelegt, teils selbst erfaßt, teils im Austausch mit Kolleginnen und Kollegen bekommen.

(C) Die historische Sprachwissenschaftlerin C hat im Jahr 2000 immer noch dieselben Voraussetzungen an Büchern und Zeitschriften, wieder sind in den 20 Jahren ein paar Editionen dazugekommen. Darüber hinaus hat C auf ihrer Festplatte oder auf Datenträgern, von deren Beschaffenheit wir heute noch nichts wissen, das ‚Historische Nationalcorpus der deutschen Sprache‘. Es umfaßt alle wesentlichen Texte der deutschen Sprachgeschichte in zuverlässigen Ausgaben, eine leistungsfähige Suchsoftware, die z.B. auch die Möglichkeit vorsieht, in Untercorpora zu recherchieren (z.B. in Zeitungen, Chroniken, Kochbüchern). Die Frage ist nun: wie verändern sich die Forschungsmöglichkeiten von A zu C?

Welche Vorteile hat die Arbeitsumgebung von C?

Zugänglichkeit

Elektronische Editionen sind in manchen Fällen ziemlich genaue Gegenstücke zu gedruckten Editionen, und auch bei der Anlage von elektronischen Texten kann es sogar ein leitender Gesichtspunkt sein, eine vorhandene Edition zu ‚spiegeln‘. Auch als reine Abbilder von bereits vorhandenen, gedruckten Editionen sind solche elektronischen Versionen wertvoll, denn sie fördern die Zugänglichkeit und die Verbreitung von Material, das nicht überall oder bisweilen sogar nur schwer zugänglich ist, z.B. wenn an entlegener Stelle gedruckte Editionen digitalisiert und im Internet angeboten werden.

Betrachten wir folgendes Beispiel: ein Editor oder ein Wortschatzforscher sitzt über einem Text des 15. Jahrhunderts und begegnet einer Wortform, die aussieht wie „iulank“ oder „inlank“. Aus dem Kontext ergibt sich, daß es sich um eine Fischbezeichnung handelt. In den Wörterbüchern und Hilfsmitteln ist aber nichts zu sehen von einer solchen Fischbezeichnung. Anton Birlinger, dem diese Form bei einer Kochbuchedition begegnet ist, hat hier zu einer Konjektur gegriffen und gebessert. Die Form „iulank“ als Fischbezeichnung steht aber auch in den Aufzeichnungen der Ravensburger Handelsgesellschaft, die Schulte herausgegeben hat. Die Stelle aus diesem Werk ist ein mehr oder weniger zufälliger Lektürefund von Wissenschaftler A. Aber er gibt A zu denken: denn man vermutet, daß man bei geduldiger Lektüre möglicherweise weitere Belege für diese Form in dieser Verwendungsweise finden könnte. Wissenschaftlerin C dagegen kann nach Recherchen im elektronischen Nationalcorpus über Einzelprobleme dieser Art mit ganz anderer Autorität befinden: sie könnte sagen, ob, wie oft und wo ein Wort wie „julank“ mit seinen Varianten vorkommt und was ggf. die plausibleren Konjekturen für eine bestimmte Stelle sind. Dasselbe Problem geht natürlich auch den Lexikographen an, der Belegstellen braucht, um beurteilen zu können, welche Gebrauchsaspekte eines Wortes konventionalisiert und etabliert sind bzw. welche Gebrauchsaspekte nur zufällige Aspekte des individuellen Sprachgebrauchs sind. Natürlich kann man Probleme wie das „julank“-Problem auch durch Lektüre lösen, aber wer hat die Zeit, für ein einzelnes Problem dieser Art eine ganze Bibliothek durchzulesen, auf Verdacht hin zahlreiche Textkandidaten per Fernleihe zu bestellen oder eine Vielzahl von Indices durchzusehen in der Hoffnung, auf die betreffende Wortform zu stossen. Die elektronische Textsammlung ist somit ein Schlüssel zu Editionen, die für ein bestimmtes Problem Material bzw. Aufschluß enthalten.

Zugriffsmittel

Ein verwandter, aber etwas anders gelagerter Gesichtspunkt ist die Möglichkeit, elektronische Dokumente als ein Zugriffsmittel zu nutzen, das die Arbeit mit einer bereits gedruckten Edition erleichtert, z.B. bei der Einzelwort- und der Wortformensuche in längeren Texten, wie sie etwa für historische Semantiker, Wortbildungsforscher und Syntaktiker bei bestimmten Fragestellungen nützlich ist. Ein Beispiel für diese Art der Nutzung elektronischer Textsammlungen sind etwa die Modalverben des Deutschen, die eine ganz außergewöhnliche Bedeutungsgeschichte haben. Modalverben kommen in

jedem deutschen Text in recht großer Zahl vor. Das Problem ist also weniger die Frage: in welchen Texten findet man sie, sondern die Frage: wie bekomme ich eine Liste aller Belegstellen zum Beispiel für das Modalverb *sollen* aus einem bestimmten Text oder den Texten einer bestimmten Textsorte. Wenn man zum Beispiel annimmt, daß eine Verwendungsweise von *sollen* als Kennzeichnung der Quellenperspektive und des Berichts aus zweiter Hand dient („der Gesandte *soll* angekommen sein“), dann kann man nach dieser Verwendungsweise vor allem in Texten suchen, in denen Quellen und Berichte aus zweiter Hand vorkommen, z.B. in Chroniken, in Zeitungen oder auch in älteren wissenschaftlichen Texten. Elektronische Texte sind in dieser Nutzungsart ein Arbeitsmittel, mit dem sich Teilindices, Teilkonkordanzen oder Belegsammlungen zu einzelnen Wörtern bzw. Wortformen erstellen lassen. Diese Produkte haben in sich noch keinen besonderen Aussagewert, sie sind aber eine wertvolle Arbeitsgrundlage, mit der sich die Beleglage besser und effektiver überblicken läßt. Man kann vielleicht sagen: Eine elektronische Belegsammlung ist sehr wertvoll, aber wenn man sie hat, fängt die eigentliche Arbeit erst an. Man kann weiterhin sagen, daß die Arbeit mit elektronischen Corpora die Lektüre von Texten keinesfalls ersetzt, denn Hypothesen über brauchbare und fruchtbare Fragestellungen bekommt man zunächst durch Lektüre, auch wenn sich Fragestellungen und Hypothesen später bei der Arbeit mit Corpora verfeinern oder modifizieren lassen.

Erweiterte Dokumentation der Text- und Überlieferungsgeschichte

In bezug auf die Text- und Überlieferungsgeschichte ist vor allem attraktiv, daß im elektronischen Medium keine Raumbeschränkungen gelten, die im gedruckten Medium dazu führen, daß man bestimmte Dokumentationsleistungen unterläßt oder zumindest verkürzt. Zu den Mehrleistungen gehören unter anderem Abbildungen der Überlieferung oder von Teilen der Überlieferung und eine erweiterte Dokumentation des ‚Flusses‘ von Texten im Laufe ihrer Benutzungs- und Rezeptionsgeschichte. Hinzu kommt die Möglichkeit, die Varianz von Texten nicht in Form zerstückelter Apparat-Einträge dokumentieren zu müssen, sondern die Varianten in ihrem jeweiligen Textzusammenhang zu belassen. Dieser Punkt wurde vor allem in der Diskussion um Cerquiglinis ‚Eloge de la variante‘ und um die New Philology hervorgehoben.

Ein erstes, elementares Beispiel für diese Nutzungsperspektive. Sie lesen in einer Transkription eines Kochbuchs aus dem 17. Jahrhundert eine Überschrift wie „Höcht auff bugerisch“. Das „bugerisch“ kommt Ihnen seltsam vor: sollte das eine Ableitung von einem Eigennamen sein? Aber das wäre im Kontext von Rezepten ungewöhnlich. Ein gewisses Mißtrauen stellt sich ein und ein Bedürfnis, das wohl jeder Benutzer von Editionen kennt, das Bedürfnis, die Handschrift selbst zu sehen. In elektronischen Umgebungen kann man solchen Bedürfnissen entgegenkommen und Abbildungen der überlieferung anbieten. In unserem Beispiel würde sich ergeben, daß „bugerisch“ ein Lesefehler für „vngerisch“ ist. Das „v“ hatte links eine hochgezogene Schleife, so daß die Gestalt an ein „b“ erinnert, daß „u“ und „n“ um 1600 leicht verwechselt werden

können, ist bekannt. Ein zweites Beispiel: Nehmen wir an, Wissenschaftlerin C hat im Jahr 2020 auch einige Texte in ihrem Corpus, von denen es mehrere Fassungen gibt, die sich regional oder durch Textumbildungen der Textnutzer (Auslassungen, Ergänzungen, Änderungen) ergeben. In dieser Lage ist zu erwarten, daß diese Forscherin sich den Fragen der regionalen Varianz oder der Frage nach den sprachlichen Umbildungen im Lauf der Textgeschichte (z.B. Modernisierung im Wortschatz) viel eher zuwenden wird, als wenn sie sich für jeden der Texte zunächst Mikrofilme usw. besorgen muß.

Solche breiten Dokumentationen zur textgeschichtlichen Überlieferung können zum Beispiel dadurch zustandekommen, daß diejenigen Arbeitsstellen, wo die Mikrofilme und Abzüge sowieso vorhanden sind, elektronische Abbilder davon herstellen und in einen gemeinsamen Pool einspeisen. D.h.: die Arbeitsstelle, wo an der ‚Küchenmeisterei‘ gearbeitet wird, digitalisiert die dort vorhandene Überlieferung der alten Drucke von 1485 bis 1690. Die Arbeitsstelle, die an den Benediktinerregeln arbeitet, tut dasselbe für die Handschriften mit Benediktinerregeln, und so verfährt auch die Arbeitsstelle oder der Forscher zu Romanen, medizinischen Texten usw. Das setzt natürlich ein Maß von Koordination für die Infrastruktur der text- und sprachgeschichtlichen Forschung voraus, das heute nicht erkennbar ist, nicht einmal in Ansätzen.

Das Corpusprinzip in der historischen Sprachwissenschaft

Elektronische Dokumente, ihre Markierung, Kodierung und Auswertung für bestimmte Fragestellungen stellen heute eine wichtige Konkretisierung des Corpusprinzips in der historischen Sprachwissenschaft dar, z.B. in der Lexikographie, der historischen Syntax, der historischen Textlinguistik oder der historischen Grammatik. Das Prinzip besagt, daß Aussagen über ältere Sprachstadien, über ihre Regelmäßigkeiten, aber auch über die typische Charakteristik möglicher Varianten auf ein Corpus von Texten bezogen werden soll, das in regionaler Hinsicht, im Hinblick auf Textsorten, Textnutzer usw. repräsentativ ist. Ein Beispiel für corpusgestützte Arbeit ist zum Beispiel die Grammatik des Frühneuhochdeutschen, die auf dem sog. Bonner Corpus zum Frühneuhochdeutschen beruht, das seit einiger Zeit auch im Internet verfügbar ist.

In den verschiedenen Arbeitsbereichen sind die Anforderungen an die Bearbeitung und Kodierung der Texte sehr unterschiedlich. Während ein Wortschatzforscher, der die Möglichkeiten der Varianz historischer Wortformen und Schreibweisen ‚im Kopf‘ hat, schon mit unlemmatisierten elektronischen Texten ganz gut arbeiten kann, ist für umfangreichere lexikographische Unternehmungen eine Lemmatisierung der Texte wünschenswert. Im Bereich der Grammatik ist der Kodierungsaufwand in der Regel sehr hoch. Aber während der Wortschatzkundler sehr große Corpora benötigt, ist den grammatischen Projekten auch mit kleineren Textcorpora gedient. Hier als Beispiel für die Form der grammatischen Codierung eine Zeile aus einem Text des Bonner Corpus zum Frühneuhochdeutschen. Von jedem Text, der Bestandteil des Corpus ist, wurden auf diese Weise jeweils 40 Standardseiten grammatisch kodiert und ausgewertet.

|Z210 leben+ / sitten+ vnd gebra<euch+ / war#zunem<-en\$/ sonder auch
 @sg_gebra<euch @sp_ge= @sl_brauch @sk_020 @sv_a<eu @sz_
 @sg_leben @sl_leben @sk_213 @sv_e @sz_
 @sg_sitten @sl_sitte @sk_220 @sv_i
 @vg_leben @vl_leben @vk_5000 @vs_20
 @vg_war#zunem<-en @vp_wahr= @vl_nehmen @vk_5000 @vs_14a

Große elektronische Text-Corpora und Sprachwandel-Forschung: Innovationen und komplexe sprachliche Verbreitungsvorgänge

In der Theorie des Sprachwandels unterscheidet man sprachliche Innovationen von den sich daran anschließenden Verbreitungsvorgängen. Viele Phänomene des Sprachwandels beruhen zwar auf den sprachlichen Äußerungen einzelner Individuen, aber der Sprachwandel selbst ist nicht Folge einzelner Äußerungen, sondern die kollektive Folge sehr vieler Einzeläußerungen. Verdeutlicht werden solche Phänomene zum Beispiel am Modell des Trampelpfades. Jeder Einzelne, der quer über eine Wiese läuft, um schneller zur Mensa zu kommen, hat nur die begrenzte Absicht, Zeit und Energie zu sparen. Aber die kollektive Folge des Handelns vieler Einzelner ist, daß im Lauf der Zeit ein Trampelpfad entsteht.

Ein Beispiel für Sprachwandel ist etwa der Bedeutungswandel des Wortes *billig*. Das Wort wurde anfangs im Sinne von (1) ‚berechtigt, angemessen‘ verwendet, dann im Sinne von (2) ‚niedrig im Preis‘, schließlich im Sinne von (3) ‚von schlechter Qualität‘. Wie ist so etwas zu erklären? Man geht davon aus, daß sich der Wandel von (1) nach (2) zuerst in Kontexten wie „das ist ein billiger Preis“ vollzogen hat, um zu sagen, daß ein Preis angemessen, fair, berechtigt ist; aus der Sicht eines Käufers ist ein angemessener Preis ein eher niedriger Preis. Schließlich konnte man das Wort nicht nur auf den Preis beziehen, sondern auch auf die Waren, für die der Preis bezahlt wurde („die Kartoffeln sind billig“). Den Wandel von (2) zu (3) erklärt man sich aus dem stereotypen Wissen der Sprecherinnen und Sprecher, daß billige Waren oft Waren von schlechter, minderer Qualität sind. Soweit die rekonstruierte Kette der Entwicklungsschritte dieses Bedeutungswandels, den man in Ansätzen auch heute schon zeitlich verankern kann. Aber: was die Details der Innovation und der Verbreitung, und was die damit verbundenen Äußerungen betrifft, liegt vieles noch im Dunkeln: z.B. die Frage, wann und in welchen Zusammenhängen der Übergang des Gebrauchs von *billig* in bezug auf einen Preis zum Gebrauch in bezug auf die Ware selbst vonstatten ging und wie und wann dieser neue Gebrauch sich etabliert hat. Einsichten in Mechanismen dieser Art sind nicht zuletzt Einsichten, wie die Sprache als Verständigungsmittel ‚funktioniert‘.

Generell gilt, daß es in bezug auf historische Zeitstufen bislang kaum möglich ist, eine hinreichend große Zahl von Verwendungen eines Ausdrucks so zu dokumentieren, daß sich eine plausible und auch detaillierte Innovations- und Verbreitungsgeschichte z.B. einer semantischen Neuerung daraus erarbeiten läßt. Große elektronische Textcor-

pora werden für diese Fragen nach Innovationen und nach Verbreitungsgeschichten einen *qualitativen* Umschlag bringen: man wird in diesem Zusammenhang nicht nur Dinge schneller tun können, die man auch bisher tun konnte, sondern man wird Dinge tun können, die wir zur Stunde gar nicht können: eben *detaillierte* Innovations- und Verbreitungsgeschichten von Neuerungen erarbeiten.

E-Texte, E-Wörterbücher und die Komplexität der Wortschatzarchitektur

Eine weitere Anwendungsperspektive elektronischer Dokumente und darauf bezogener Markierungen sind Probleme der Komplexität. Die Architektur des Wortschatzes einer Sprache ist ein typisches Beispiel für ein solches Komplexitätsproblem. Die Übersicht über den Aufbau und die Zusammenhänge im Wortschatz einer natürlichen Sprache übersteigt die Fähigkeiten des menschlichen Kopfes beträchtlich, denn die Struktur eines Wortschatzes ist gekennzeichnet durch eine überaus große Zahl von Wörtern, eine sehr große Zahl an Verwendungsweisen von Wörtern, durch eine Mehrzahl von Strukturierungshinsichten (z.B. Bedeutung, Sachgebietszugehörigkeit, Wortbildungstyp, Wortfamilienzugehörigkeit, Textsortentypik, Herkunft von Fremdwörtern usw.) und schließlich durch komplexe Zusammenhänge zwischen diesen Elementen und Strukturierungshinsichten. Markierungssysteme in elektronischen Wörterbüchern und in den darauf bezogenen elektronischen Texten dienen dazu, diese Komplexität ‚befragbar‘ zu machen. Der menschliche Bearbeiter kodiert im elektronischen Medium z.B. Textwörter wie *obgemelt* als Querverweisausdrücke oder Verwendungsweisen wie antizipieren im Sinne von ‚eine rechtlich bestehende Geldforderung vor dem eigentlichen Termin gütlich erbitten‘ z.B. als Fremdwort, das zum Sachgebiet des Handels und der Finanzpolitik gehört. Die Stärke des Computers besteht darin, aufgrund solcher Markierungen auch komplexe Anfragen beantworten zu können, wie z.B. ‚Wirf alle Fremdwort-Verwendungsweisen samt den zugehörigen Belegen aus, die zum Sachgebiet des Militärs gehören und als Ereignisbezeichnungen gelten können, sofern Text-Belege für den Zeitraum von 1570 bis 1630 vorhanden sind‘. Mehrere solcher Anfrage-Ergebnisse kann ein menschlicher Bearbeiter wiederum benutzen, um eine Bild von der komplexen Wortschatzorganisation zu einer bestimmten Zeit zu zeichnen. Der Mehrwert der elektronischen Umgebung besteht darin, solche Probleme der sprachlichen Komplexität behandelbar und bearbeitbar zu machen.

Ich breche diesen Aufriß einiger Nutzungsperspektiven elektronischer Dokumente in der historischen Sprachwissenschaft hier ab. Es dürfte deutlich geworden sein, daß die Vorteile auf einer Skala liegen. Es gibt Dinge, die man bisher schon konnte, die man nun aber schneller tun kann. Dann gibt es Dinge, die man bisher zwar konnte, die man aber wegen des Aufwandes kaum in Angriff genommen hat und die man jetzt eher angreift, weil man sie leichter und schneller tun kann. Schließlich gibt es Fragestellungen, deren Bearbeitung mit großen elektronischen Corpora überhaupt erst möglich werden.

Ich möchte nun einige Überlegungen anstellen, was wir in den nächsten Jahren tun können und müssen, um den skizzierten Nutzungsperspektiven näher zu kommen.

Historisches Nationalcorpus: Retrodigitalisierung und Nutzung von Satzdaten

Unsere gegenwärtige Lage, die etwa der Lage von B in unserem Szenario entspricht, ist eine unangenehme Zwischenlage. Wir wissen, was sich mit großen Textcorpora machen ließe, aber wir haben sie noch nicht. Auf der anderen Seite wissen wir auch, daß die Voraussetzungen für ein großes Nationalcorpus zur Geschichte der deutschen Sprache eigentlich gegeben sind. Texte liegen in großer Zahl auf den einzelnen Festplatten von Editoren, Arbeitsstellen und Verlagen, und dort, auf den einzelnen Festplatten, liegen auch zahlreiche elektronische Dokumente, die auf der Grundlage von vorhandenen Editionen retrodigitalisiert wurden. Aber sie liegen auf einzelnen Festplatten, und wir wissen wenig von einander. Ansätze zu einer Koordination, die über persönliche Initiativen oder die Initiativen einzelner Arbeitsstellen (z.B. der TITUS-Server an der Universität Frankfurt; das IdS-Corpus in Mannheim) hinausgehen, sind mir nicht bekannt geworden. Es gibt auch noch keine ‚Belohnungen‘ und keine Anreize für die Bereitstellung elektronischer Dokumente.

Zur Zwischenlage im Jahr 2000 gehört es auch, daß wir im Moment sehr viel Zeit auf die Herstellung elektronischer Dokumente verwenden müssen, die von der Zeit, die wir den eigentlichen Fragestellungen widmen sollten, abgeht.

Und schließlich sind wir im Hinblick auf Erfassungsstandards und Software in der unangenehmen Lage, daß wir zwar von den Vorteilen einer SGML- oder XML-gemäßen Markup-Umgebung überzeugt sind, daß aber für die alltägliche Praxis der Pflege elektronischer Dokumente von historischen Texten keine bequeme und erschwingliche Software zur Verfügung steht. Ähnliches gilt auch für die Recherche- und Retrieval-Funktionen.

Was können wir (auch Sie als Historikerinnen und Historiker) in dieser Lage tun?

-- Man kann hoffen und darauf warten, daß in absehbarer Zukunft eine Institution (z.B. eine Akademie, die DFG) die Koordination der Arbeiten für ein Nationalcorpus zur Geschichte der deutschen Sprache übernimmt.

-- Man sollte überprüfen, ob diejenigen elektronischen Fassungen, die innerhalb der eigenen Arbeitsprojekte erstellt wurden, auch in anderen Arbeitszusammenhängen nützlich sein können und zum Beispiel über das Internet angeboten werden können. Chroniken, Urkunden oder Rechnungen sind auch wertvolles Material für die historische Sprachwissenschaft! Solange es keine zentrale Koordination gibt, scheint die Beteiligung am dezentralen Angebot der Internet-Texte die beste Lösung zu sein.

-- Man sollte auf jeden Fall dafür sorgen, daß diejenigen elektronischen Fassungen, die innerhalb der eigenen Arbeitsprojekte anfallen, in einer zukunftsicheren Version archiviert sind und daß zumindest im Kreis der Fachgenossen auch bekannt ist, daß es die betreffenden Texte in elektronischer Fassung gibt.