



Marco Scheider

Einladung zur Enthauptung?

Elektronische Revolution und traditionelle Lexikographie

Gab es tatsächlich eine Zeit, in der Wörterbücher nur in gedruckter Form vorlagen? Und ist das wirklich weniger als zehn Jahre her? Der Umgang mit elektronischen Nachschlagewerken und den einschlägigen Hilfsmitteln ist so selbstverständlich geworden, dass die Erinnerung an die Zeit schwerfällt, da nur dickleibige Wälzer Abhilfe in sprachlichen Fragen versprochen. Wörterbücher und andere Nachschlagewerke sind prädestiniert für den elektronischen Zugriff, und heute stellt sich nur noch die Frage, *welche* lexikographischen Informationen für *welche* Endgeräte *wie* aufbereitet werden sollen. Man kann gute Gründe dafür benennen, Wörterbücher, jedenfalls die meisten von ihnen, weiterhin in gedruckter Form anzubieten. In jedem Fall aber müssen sie *auch* im Netz verfügbar sein – übrigens zu beiderseitigem Vorteil: Die Benutzer profitieren von dem leichteren und in der Regel kostenlosen Zugang, und so manches Wörterbuch, das zuvor eher ein Schattendasein fristete, kann mit einem Webauftritt sein Google-Ranking, also seinen Bekanntheitsgrad, rasant steigern. Der Zeitpunkt dürfte nicht mehr fern sein, an dem man sagen kann, dass ein Wörterbuch, das nicht im Netz abrufbar ist, praktisch nicht existiert. Wenn die geläufige Einschätzung stimmt, nach der 90 Prozent aller Nutzer ein Wörterbuch aufschlagen (nun sagt man wohl besser: benutzen), weil sie nicht wissen, wie man ein Wort schreibt oder was es bedeutet, so lässt sich die Aufgabe einer elektronisch strukturierten Lexikographie *prima vista* klar umreißen: Die Informationen müssen so schnell und so gut wie möglich ins Netz. Diese Forderung bezieht sich *zum einen* auf die Retrodigitalisierung älterer Wörterbücher: Bei dieser lassen sich gleich noch praktische Hilfsmittel für die Benutzung einbauen – die Gliederungsübersichten der zum Teil sehr langen Artikel in der digitalisierten Erstausgabe des Deutschen Wörterbuchs (DWB)¹ oder die erst in der elektronischen Fassung vorgenommene chronologische Sortierung der Belege im Deutschen Rechtswörterbuch (DRW)² stellen

hierfür anschauliche Beispiele dar. *Zum anderen* gilt sie für die sukzessive entstehenden Artikel in Bearbeitung befindlicher Wörterbücher, etwa für die Artikel des neuen Mittelhochdeutschen Wörterbuchs (MWB)³, die nach Ablauf einer mit dem Verlag vereinbarten Frist im Netz zugänglich gemacht werden.

Wirken sich nun die von der EDV hervorgebrachten Veränderungen auch auf die lexikographische Analyse selbst aus? Inwieweit hat die elektronische Revolution also diesen philologischen Alltag durchdrungen und zu einer Verbesserung geführt? Hier sind die Befunde ambivalent, wie sich unter drei Aspekten zeigen lässt.

1) Die Durchsetzung des PCs als Arbeitsgrundlage kam in der Philologie und insbesondere in der Lexikographie mit zahlreichen Verheißungen daher, deren erste und naheliegendste die der *Beschleunigung* war, was sich wiederum in ehrgeizigen Projektplänen niederschlug und noch -schlägt. Das Digitale Wörterbuch der deutschen Sprache (DWDS)⁴, der Prototyp einer ebenso radikal wie konsequent auf die neuen Möglichkeiten setzenden Methode, begrüßte seine Besucher denn auch mehrere Jahre auf der Startseite mit einem Zitat von Hans Magnus Enzensberger: »Endlich ist es möglich, lexikographische Arbeit so zu beschleunigen, dass ihre Ergebnisse in kurzer Zeit verfügbar werden.« Wie steht es im Jahr 2009 um diese Zuversicht? Das Zitat ist von der Homepage verschwunden, das DWDS, als Arbeitsgruppe gegründet, gehört inzwischen zu den Langzeitvorhaben im Akademienprogramm. Hier (aber noch längst nicht überall) ist die Einsicht gereift, dass die Verfügbarkeit großer Datenmengen zwar eine neue Dimension von Textcorpora ermöglicht hat – laufende Textwörter werden inzwischen nach Milliarden berechnet –, deren semantische Analyse aber nach wie vor das menschliche Gehirn leisten muss, weil auf absehbare Zeit keine Software diese Aufgabe zu übernehmen auch nur annähernd in der Lage ist, erst recht nicht bei historischen Texten, deren größere Varia-

es was einer mit nammen Arnutus/ein angenẽmer
mensch vnder den so die feylen ding auszschrau-
wen/

bilität bezüglich Orthographie und Grammatik den maschinellen Einsatz schon auf einer weniger komplexen Ebene wie der Stichworterkennung von vornherein begrenzt. Um ein einfaches Beispiel zu nehmen: Wenn Caspar Hedio in seinem »Josephus deutsch« schreibt: *es was einer mit nammen Arnutus / ein angenẽmer mensch vnder den so die feylen ding aufschrauwen*, so kann ein geschulter und geübter Exzerptor erkennen, dass am Ende des Satzes das Verb *ausschreiben* steht, Lemmatisierungsprogramme gehen an dieser Stelle in die Knie. An eine nennenswerte Beschleunigung der wesentlichen Arbeitsschritte ist also nicht zu denken.

2) Was die *Verfügbarkeit großer Datenmengen* im Netz angeht, so stehen sich einstweilen zwei unterschiedliche Perspektiven gegenüber. (Computer-)Linguisten, die ihre Untersuchungen bis vor wenigen Jahren mangels vorhandener Corpora im Wesentlichen an selbst gebildeten Beispielen vornehmen mussten, zeigen sich naturgemäß enthusiastisch ob der Möglichkeit, mit einer nicht nur umfangreichen, sondern oft auch im Hinblick auf Ausgewogenheit bzw. Repräsentativität kontrolliert zusammengestellten Datenbasis zahlreiche Fragestellungen wie zum Beispiel die nach Häufigkeitsverteilungen oder nach Kollokationen viel fundierter als bisher angehen und von der Möglichkeit ganz neuer und gezielter Abfragen Gebrauch machen zu können. Dass dies aber nur die eine Seite der Medaille ist und letztlich die Art der Fragestellung über die Bewertung der Corpora entscheidet, wird ihnen von Editoren und »traditionellen« Lexikographen entgegengehalten, die sich auf ihre in Jahrhunderten etablierten philologischen Methoden berufen und darauf verweisen, dass sie schon viel länger mit großen Datenbeständen arbeiten, nur eben nicht mit Corpora, sondern mit Archiven und Zetteln, deren Zahl nicht selten in die Millionen geht. Von dieser Seite gerät in erster Linie die unter textkritischen Aspekten völlig unzureichende Qualität der im Netz verfügbaren Digitalisierungen etwa beim Projekt Gutenberg⁵ oder bei der auf den entsprechenden CDs der Digitalen Bibliothek beruhenden Website www.zeno.org⁶ in den Blick, die nur selten wissenschaftlichen Anforderungen entsprechen. Auch Googles Digitalisierungsprogramm ist in dieser Hinsicht eher zwiespältig: Einstweilen spielen besonders in Deutschland, wo es bekanntlich noch kaum entsprechende Vereinbarungen mit Verlagen gibt, Copyright-Fragen bei der Auswahl der Digitalisate eine entscheidende Rolle, weswegen häufig veraltete Ausgaben der modernen Klassiker

eingelesen werden; das wirkt nun wie ein Hohn auf die avancierten Editions-methoden und die hoch subventionierten Textausgaben besonders der zweiten Hälfte des 20. Jahrhunderts und nährt das Vorurteil, dass die elektronische Form häufig wichtiger genommen wird als der philologische Standard. Wo höherwertige Textsammlungen vorhanden sind, werden sie hingegen meist als geistiges Eigentum behandelt und entsprechend abgeschottet. Die Berlin-Brandenburgische Akademie der Wissenschaften hat zwar mit dem DFG-Projekt Deutsches Textarchiv (DTA)⁷ einen Versuch unternommen, das, was in einigen anderen Ländern schon realisiert ist, auch hierzulande zu etablieren: ein für alle Interessierten frei zugängliches Nationalcorpus mit einer hochwertigen Textgrundlage. Beim DTA werden Erstausgaben sorgfältig ausgewählt, digitalisiert und annotiert. Am Ende des auf sechs Jahre angelegten Projekts werden aber maximal 1500 Texte aus der Zeit von 1650 bis 1900 zur Verfügung stehen, und für ein wirkliches Nationalcorpus muss die Textmenge natürlich viel größer sein. Ziel des DTA ist deswegen auch eine umfassende Integration von bzw. Vernetzung mit weiteren Textcorpora.

3) Die umfassende *Vernetzung* ist als dritte große Verheißung naturgemäß die Grundidee des WWW. Für die philologische Recherche ist diese Idee aber noch längst nicht in befriedigender Weise umgesetzt: Übergreifende Abfragen in größeren Corpora sind bis jetzt, von den oben erwähnten problematischen Sammlungen abgesehen, lediglich beim DWDS und beim Institut für deutsche Sprache (IdS)⁸, das heißt eigentlich nur für das 20. Jahrhundert möglich; für die vor dieser Zeit liegende Literatur muss, wer darauf angewiesen ist, Dutzende von Abfragen auf horrend vielen Websites durchführen. Allein die Linkübersicht »Alte Drucke digital«⁹, die ausschließlich von deutschen Institutionen digitalisierte Drucke vor 1800 erfasst, listet (Stand März 2009) 69 verschiedene Adressen auf, unter denen man entsprechend unterschiedlich aufbereitete und abfragbare Textsammlungen findet. Bei der Vernetzung lexikographischer Informationen wiederum sieht es nur dort besser aus, wo retrodigitalisierte Wörterbücher miteinander verknüpft werden, also beim Trierer Wörterbuchnetz¹⁰ und (noch) in geringerem Umfang beim Wörterbuchportal der Berlin-Brandenburgischen Akademie der Wissenschaften und der Heidelberger Akademie.¹¹ Wo hingegen die automatisierte Vernetzung selbst Bestandteil der lexikographischen Konzeption ist, wirken die Ergebnisse bis



jetzt noch eher unausgereift. Wer zum Beispiel – wie jüngst bei der Neubearbeitung des entsprechenden DWB-Artikels – in Trier Informationen über das Stichwort ›Beding‹ sucht, bekommt dort vier Artikel aus dem DWB, dem Goethe-Wörterbuch sowie aus zwei Mundartwörterbüchern, dem Rheinischen und dem Pfälzischen Wörterbuch; dazu kommt die Abfragemöglichkeit beim verlinkten DRW. Als Ergebnis dieser Abfrage kennt man vier Bedeutungen und hat Beispiele vom 15. bis zum 20. Jahrhundert gelesen. Diese Artikel sind allerdings schon älteren Datums und stammen aus Wörterbüchern, deren Anlage und Informationsstruktur als nicht mehr zeitgemäß gilt. Was weiß demgegenüber ein neuartiges lexikalisches Informationssystem wie das am IdS betriebene Online-Wortschatz-Informationssystem Deutsch (OWID)¹², das schon sehr viele Stichwörter kennt, aber nur wenig eigene Analysen durchführt und deswegen mit vernetzten Informationen arbeitet? Man erfährt dort, dass ›Beding‹ *Beding* geschrieben und *Be-ding* getrennt wird, für alle weiteren Angaben wird man auf canoo.net verwiesen, mit einem Link, der auf das Verb ›bedingen‹ führt, was in der Regel den Abbruch der Recherche nach sich ziehen dürfte. Auch hier muss also noch einige Arbeit geleistet werden, bis die Ergebnisse befriedigen können.

Für den Lexikographen ergibt sich insgesamt eine schwierige Lage: Auch wenn die Bereitschaft zur weiteren Umstellung auf EDV-basierte Verfahren überhaupt nicht mehr infrage zu stellen ist, erschweren doch die immer noch kursierenden unrealistischen Beschleunigungswünsche der Projektverantwortlichen oder Geldgeber, die noch zu selten den philologischen Standards entsprechenden Digitalisate und die bis jetzt fehlende fachgemäße und effektive Vernetzung der Ressourcen den Umgang mit dem neuen Medium. So wird der Paradigmenwechsel von nicht wenigen Mitarbeitern nur anhand der skizzierten Unzulänglichkeiten bewertet; die schon jetzt unbestreitbaren Errungenschaften treten vielfach in den Hintergrund. Geduld und das gegenseitige Verständnis aller Beteiligten sind gefordert, vielleicht auch eingedenk Pnins Respekt vor der *avant la lettre* formulierten computerlinguistischen Vision, wonach die neue Methode »wie ein von Fels zu Fels platschender Wasserfall aufhört, ein Medium verständiger Navigation zu sein, aber in sagenhafter Zukunft vielleicht dazu dienen mag, esoterische Dialekte – Basales Baskisch und dergleichen – hervorzubringen, welche nur von komplizierten Ma-

schinen gesprochen werden«¹³. Die Gegenwart jedenfalls ist gekennzeichnet vom Spagat zwischen der begründeten Hoffnung auf komfortablere Lösungen in der Zukunft und aktuellen Produktions- bzw. Publikationsnöten, für die Visionen von einem kommenden Goldenen Zeitalter keine Rolle spielen. Ein Ausweg aus diesem Dilemma ist vorerst nicht in Sicht.

1 germazope.uni-trier.de/Projects/DWB

[letzter Zugriff hier wie im Folgenden 10. 3. 2009]

2 drw-www.adw.uni-heidelberg.de/drw/

3 www.uni-trier.de/index.php?id=14644

4 www.dwds.de/

5 gutenberg.spiegel.de/

6 www.zeno.org/

7 www.deutsches-textarchiv.de/

8 www.ids-mannheim.de/

9 wiki.netbib.de/coma/AlteDruckeDigital

10 www.woerterbuchnetz.de/

11 www.woerterbuch-portal.de/

12 www.owid.de/

13 V. Nabokov: *Pnin*. Deutsch von Dieter E. Zimmer. Rowohlt Verlag⁶2007, S. 11