Erschienen in: Lenz, Friedrich/Schierholz, Stefan J. (Hrsg.): Corpuslinguistik in Lexik und Grammatik. - Tübingen: Stauffenburg, 2008, 2. Auflage. S. 41-59. (Stauffenburg Linguistik 37)

Carolin Müller-Spitzer

Erstellung lexikografischer Daten aus Korpora. Eine neue Art elektronischer Wörterbücher

Electronic corpora play an ever growing role in lexicography. On the one hand, new access to linguistic usage is made possible through the use of text corpora and intelligent corpus-based query tools; however, the final results are still interpreted and described by lexicographers. In this case corpora are used for data acquisition. On the other hand, there are also projects that provide purely automatically acquired data in the form of "dictionaries". Lexicographers play only a minor role here. This latter type of corpus use creates a completely new kind of electronic dictionary. This article addresses the questions as to what extent these dictionaries differ from lexicographic tradition and whether they must be considered in metalexicography. Starting from previously compiled electronic dictionary typologies, we try to supplement the formulation of lexicographic data as a distinguishing feature. Finally, based on the findings of the project *elexiko* (Institute for the German Language – IDS), we demonstrate that the distinction between electronic versus man-made lexicographic data is also relevant to lexicographical practice.

1. Vorbemerkung: Korpora und Lexikografie

Elektronische Textkorpora spielen in der Lexikografie eine immer größere Rolle. Zwar gilt die Feststellung de Schryvers – „Far from being science fiction, electronic corpora have been used in lexicography for at least three decades, and no serious compiler would undertake a large dictionary project nowadays without having one (and preferably several) at hand." (Schryver 2003:167) – mehr für den angloamerikanischen Raum, doch ist auch in der deutschen Wörterbuchlandschaft das grundsätzliche Quellen- und Belegprinzip ein zentrales Thema lexikografischer Arbeit. Mit der weiten Verbreitung des neuen Mediums und dem damit ermöglichten Aufbau umfangreicher elektronischer Textkorpora ist allerdings eine neue Art der lexikografischen Arbeit möglich geworden.

Dass Lexikografen Texte als empirische Basis verwenden, ist selbstverständlich nicht neu. Als methodisches Ideal ist dies nach Bergenholtz (Bergenholtz 1996:255) zur Zeit der späten Junggrammatiker aufgestellt worden. Besonders bekannt und interessant ist dabei das Vorwort zu Otto Behaghels *Deutscher Syntax* (Behaghel 1923:VIIff.), welches auch für die heutige Korpuslinguistik erstaunlich aktuell scheint. Behaghel beschreibt in diesem Vorwort sein Vorgehen bei der Erforschung der deutschen Syntax und grenzt sein Verfahren von anderen ab:

> Tomanetz hat einmal gemeint, eine deutsche Syntax könne nur das Ergebnis einer umfassenden, durch ein ganzes Leben hindurch fortgesetzten Sammelarbeit sein. [...] Wer nach dem Rezept von Tomanetz verfährt, wird vorzugsweise das Auffallende buchen; die großen Massen des Regelmäßigen und seine Wandlungen entziehen sich dem Sammler, und Eines kann er überhaupt nicht verzeichnen: das Fehlen bestimmter Erscheinungen. (Behaghel 1923:VIII)

Dem gegenüber schlägt Behaghel folgende Methode vor:

> Bei diesen Untersuchungen habe ich das Verfahren beobachtet, das sich mir bei meiner Arbeit über die Zeitfolge bewährt hat, das Verfahren der Stichproben, das gewisse Stücke gewisser Denkmäler vollständig auszubeuten sucht. Wer danach andere Stücke und andere Quellen durchmustert, wird vielleicht wertvolle Ergänzungen bieten können, aber das von mir Gefundene kaum gänzlich umwerfen. (Behaghel 1923:VIII)

Statt gezielt nach einem bestimmten Phänomen zu suchen, schlägt Behaghel also eine vollständige Analyse bestimmter Texte vor. Zwar würde man nach heutigen korpuslinguistischen Erkenntnissen die ‚gewissen Stücke gewisser Denkmäler' zur Erforschung des sprachlichen Usus problematisieren, doch im Prinzip nimmt Behaghel hier schon eine methodische Trennung vor, die sich aktuell bei Kathrin Steyer und Cyril Belica in der Unterscheidung von Konsultations- vs. Analyseparadigma wiederfindet. Denn auch hier wird das Konsultationsparadigma charakterisiert als die Vorgehensweise, in der man ein Ausgangsproblem hat und dazu das Korpus befragt. (Belica 1998:31f.) Problematisch sei dabei jedoch: „Man sieht nur, was man sehen will bzw. man findet auf diesem Wege nur genau das, wonach man konkret sucht." (Steyer 2004:93). Im Analyseparadigma ist das Vorgehen dagegen anders: „Es werden systematisch große Sprachausschnitte auf der Suche nach usuellen sprachlichen Phänomenen analysiert." (Steyer 2004:93). Durch große elektronische Textkorpora und vor allem intelligente mathematisch-statistische Analysemethoden wird es also möglich, auf breiter Basis den sprachlichen Usus anhand von Massendaten zu untersuchen. Dabei bleibt jedoch nach Steyer die „deutende und interpretierende Hand des Linguisten für viele Zwecke letztlich immer unabdingbar" (Steyer 2004:90). Allerdings gibt es Projekte, die diesen Schritt nicht machen, die also die automatischen Analyseergebnisse unbearbeitet dem Benutzer zur Verfügung stellen. Diese Form – vertrieben meist im Internet und betitelt als Wörterbücher oder Lexika – stellen eine ganz neue Art elektronischer Wörterbücher dar. Denn bei diesen Produkten ist nicht allein der Weg anders, auf dem die Lexikografen zu ihren Daten gelangen, sondern Lexikografen spielen selbst (fast) gar keine Rolle mehr.

In diesem Beitrag soll es in Bezug auf die Unterscheidung elektronischer Wörterbücher daher nicht allgemein um korpusbasierte elektronische Wörterbücher gehen, sondern um einen Sonderfall innerhalb dieses Bereiches: um die (teil-)automatische Erstellung lexikografischer Produkte, die allerdings erst durch den Einsatz von Korpora möglich geworden sind. Denn diese Produkte stellen für die Lexikografie und Wörterbuchforschung eine besondere Herausforderung dar.

2. Ein Ausgangsbeispiel: das *Wortschatz-Lexikon*

Noch 1997 stellte Helmut Feldweg fest, dass es sich bei dem damaligen elektronischen Wörterbuchmarkt weitgehend um eine „Widerspiegelung des Papierwörterbuchmarktes auf reduziertem Niveau" (Feldweg 1997:110) handelte, da die meisten elektronischen Wörterbücher ein Abbild ihrer gedruckten Pendants waren und die Möglichkeiten des neuen Mediums nicht nutzten. Eben ist jedoch schon angedeutet worden, dass diese Feststellung jetzt – sieben Jahre später – nicht mehr zu halten ist, da der Einsatz des Computers in der Herstellung von Wörterbüchern und auch die Nutzung des elektronischen Mediums als Publikationsmedium wesentlich vorangetrieben worden sind. Außerdem steht mit dem Internet eine neue und auch preiswerte Publikationsmöglichkeit zur Verfügung, wodurch Produkte auf den Markt kommen, die so nicht gedruckt worden wären und auch in der Form nicht gedruckt werden könnten.

Wenn es sich aber bei diesen neuen Produkten nicht um ‚alten Wein in neuen Schläuchen' (so der Titel von Feldwegs Aufsatz), d.h. um Bekanntes in neuem Gewand handelt, stellen diese Produkte eine Herausforderung für die Wörterbuchforschung, besonders für die Wörterbuchkritik dar. Denn dann stellen sich Fragen wie: Können Bewertungsmaßstäbe, die sich für gedruckte Wörterbücher etabliert haben, auf elektronische übertragen werden?[1] Und selbst wenn man diese erste Frage bejaht, ist im Anschluss die Frage zu stellen, ob dieses „Ja" für alle Arten elektronischer Wörterbücher gilt oder wie terminologische Unterscheidungsmöglichkeiten aussehen könnten, damit innerhalb dieses Bereiches besser differenziert werden kann. Auch andere Fragen sind zu stellen im Zusammenhang mit der automatischen Gewinnung von Daten aus Textkorpora, z.B.: Sind Produkte, deren Daten rein automatisch aus Korpora gewonnen sind, überhaupt als Wörterbücher zu bezeichnen?

Inwiefern die automatische Erstellung von wortschatzbezogenen Daten eine besondere Herausforderung für die Wörterbuchkritik darstellt, soll zunächst an einem Ausgangsbeispiel, dem *Wortschatz-Lexikon* der Universität Leipzig, veranschaulicht werden. Bei diesem Projekt handelt es sich nach Projektinformationen um ein „umfangreiches Vollformenwörterbuch des Deutschen", in dem „die typischen Inhalte und Funktionen unterschiedlicher Wörterbuch- und Lexikontypen [...] zur Verfügung stehen (Nachschlagen von Begriffen; Querverweise; morphologische, syntaktische, semantische und pragmatische Information; statistische Daten; Einarbeitung von Ontologien) und durch die zusätzlichen Möglichkeiten des elektronischen Mediums (automatisch linguistische Analyseverfahren, Recherche, Hypertextualisierung, automatische Generierung unterschiedlich strukturierter Einträge, Visualisierung von Relationen zwischen Einträgen) ergänzt

---

[1] So schreiben z.B. Stefan Engelberg und Lothar Lemnitzer in *Lexikographie und Wörterbuchforschung*: „Auch für elektronische Wörterbücher gelten deshalb zunächst die Bewertungskriterien und -maßstäbe, die wir weiter oben für Printwörterbücher aufgezählt und kommentiert haben." (Engelberg/Lemnitzer 2001:194).

werden“ (Quasthoff/Wolff 1999:1). Auf der Webseite[2] wird das Projekt auch allgemein als „Nachschlagewerk für Wörter und ihren Gebrauch“ bezeichnet.

Interessiert man sich nun für seinen Nachnamen, beispielsweise für „Meier“, so erhält man den in Abbildung 1 gezeigten Eintrag.

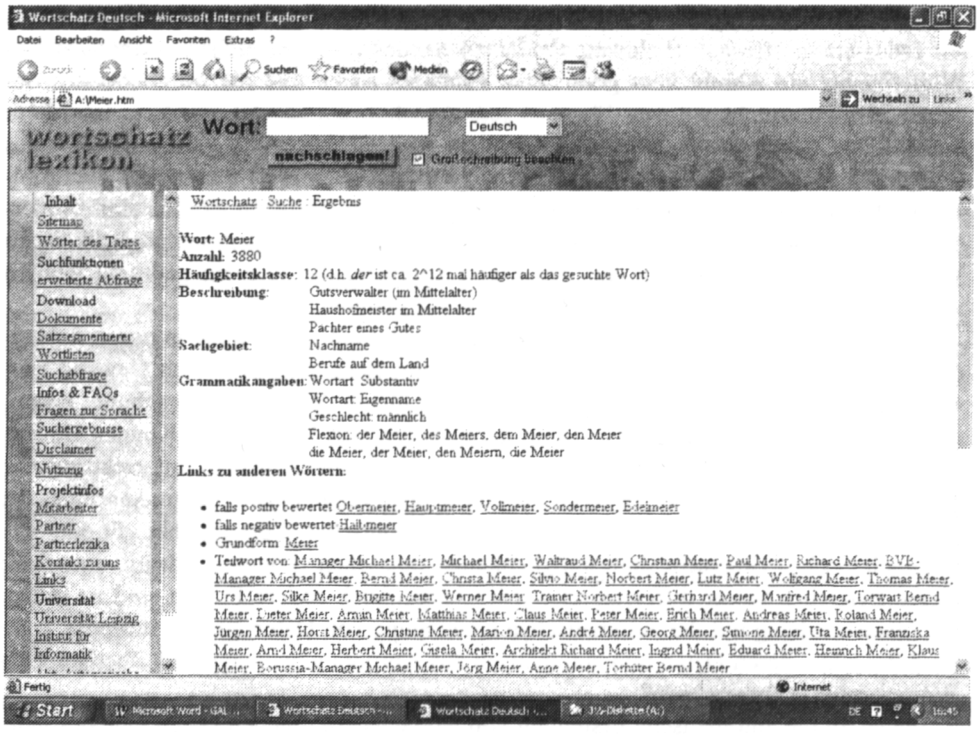

Abb. 1: Eintrag „Meier“ aus dem Wortschatz-Lexikon

Die Lemmatisierung von Eigennamen kann für allgemeine, einsprachige Wörterbücher als unüblich angesehen werden; in der gängigen Praxis werden sie nur aufgenommen, wenn sie als Appellativa gebraucht werden. Das *Wortschatz-Lexikon* wird von den eigenen Mitarbeitern jedoch als „Vollformenwörterbuch“ bezeichnet; wobei anscheinend Eigennamen zu den Vollformen gerechnet werden.

Besonders interessant wird es aber unter der Rubrik „Links zu anderen Wörtern“. Zum Beispiel steht da zu „Meier“: „falls positiv bewertet: Obermeier, Vollmeier, Hauptmeier, Sondermeier, Edelmeier“ oder „falls negativ bewertet: Halbmeier“. Verfolgt man eine dieser Angaben und klickt z.B. auf „Edelmeier“, so erhält man den in Abbildung 2 gezeigten Eintrag.

Als Sachgebiet für „Edelmeier“ wird „Nachname“ angegeben, das einzige Beispiel, was aufgeführt wird, stammt aus dem Telefonbuch, welches anscheinend in das Korpus

---

2 http://wortschatz.uni-leipzig.de.

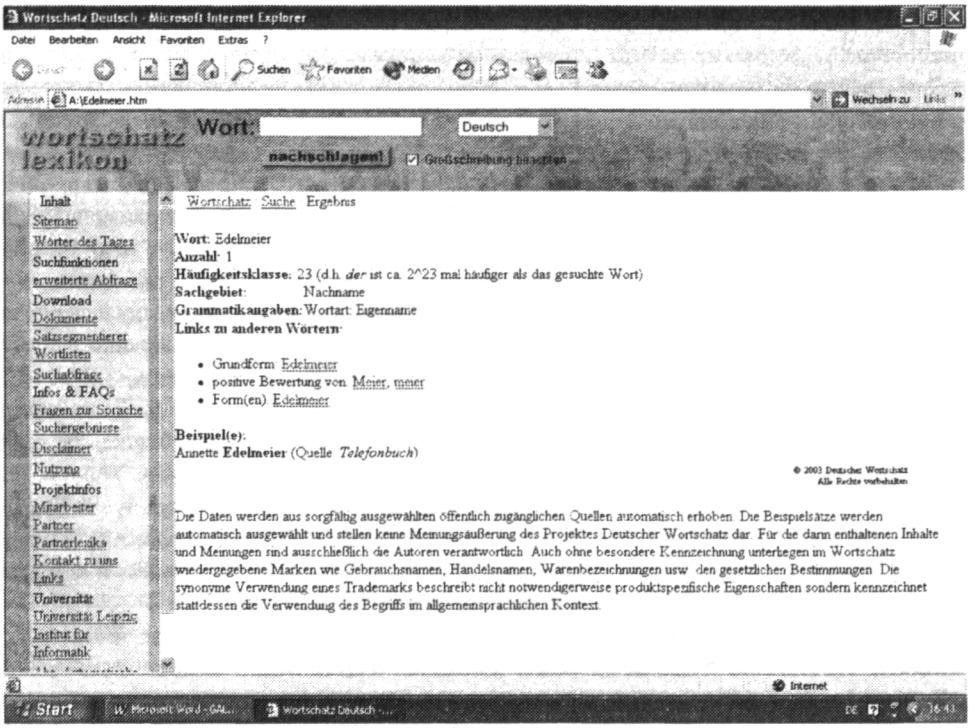

Abb. 2: Eintrag „Edelmeier" aus dem Wortschatz-Lexikon

aufgenommen wurde. Wie es zu dieser Vernetzung von „Meier" zu „Edelmeier" kommt, ist nicht zu erkennen.[3]

Dass die Daten rein automatisch erstellt werden, sieht man auch am Eintrag „Kohl". Unter „Beschreibung" wird „deutscher Bundeskanzler, deutscher Politiker, deutscher Staatsmann" angegeben, als „Sachgebiet" findet sich allerdings neben „Nachname" auch „Ackerbau" oder „Literarische Motive Stoffe Gestalten", unter „ist Synonym von" werden „Koks" und „Kraut" genannt, unter „Links zu anderen Wörtern" wird „Kohl" als „ein(e): Blattgemüse, Gemüseart" eingeordnet, unter „Teilwort von" finden sich dagegen wiederum nur Beispiele wie „Helmut Kohl, Hannelore Kohl, [...] Georg Kohl GmbH" und unter den Unterbegriffen sind alle möglichen Kombinationen von „Helmut Kohl" über „Sibylle Steinkohl", „Rosenkohl", „Alkohl" [?], „Anti-Kohl" bis hin zu „Jelzin-Kohl" aufgeführt.

Es geht hier allerdings nicht um das *Wortschatz-Lexikon* an sich, sondern darum, was diese Beispiele in diesem Themenzusammenhang demonstrieren können. Die hohen Zugriffszahlen auf das *Wortschatz-Lexikon* zeigen schließlich, dass ein bestimmter

---

[3] Auf ein solches Beispiel angesprochen kündigte ein Projektmitarbeiter des *Wortschatz-Lexikons* auf einem Kolloquium im Februar 2003 am IDS an, dass diese Vernetzung bei einer neuen Version verbessert werden solle.

Benutzerkreis gerne das *Wortschatz-Lexikon* besucht. Das Produkt ist aber ein gutes Beispiel dafür zu fragen, ob man wirklich die Bewertungsmaßstäbe, so wie man sie für gedruckte Wörterbücher anlegt, hier auch anwenden kann. Denn wenn man dies tun würde, müsste die Kritik an die eben gezeigten Einträge und vor allem deren Vernetzung untereinander verheerend ausfallen, denn sie ist aus lexikografischen Gesichtspunkten in keiner Weise nachvollziehbar.

Aber ist es überhaupt legitim, diese Maßstäbe anzulegen? Dafür ist der Passus unter den Einträgen zu beachten, in dem steht: „Die Daten werden aus sorgfältig ausgewählten öffentlich zugänglichen Quellen automatisch erhoben." Die Fragen, die man bei der Bewertung dieses Produktes stellt, sind also vom Ansatzpunkt nicht mit denen vergleichbar, die man sonst an Wörterbücher stellt. Man kann nicht fragen, nach welchen Kriterien die Lexikografen bestimmte Angaben angesetzt haben, warum sie eine Angabe so und nicht anders formuliert haben etc., sondern es stellen sich Fragen wie: Mit welchen Analysemethoden werden die Textkorpora analysiert? Wie sind die Textkorpora zusammengesetzt? etc. Diese Fragen stimmen daher nicht mit denen überein, die man üblicherweise an lexikografische Produkte stellt, z.B. an eine elektronische Version des *Großen Wörterbuchs der deutschen Sprache* aus dem Dudenverlag (vgl. auch Engelberg/Lemnitzer 2001:194ff.). Die Bewertungsmaßstäbe, die sich für gedruckte Wörterbücher etabliert haben, können daher nicht grundsätzlich auf elektronische übertragen werden, zumindest nicht auf alle Arten elektronischer Wörterbücher.

Wie vorhin zitiert wurde, heißt es in der Projektbeschreibung zum *Wortschatz-Lexikon*, dass es die Funktionen eines „üblichen Wörterbuchs" erfülle. Dazu gehört allerdings auch – auf jeden Fall in der wissenschaftlichen Lexikografie (i.S.v. Wiegand 1998:40ff.) – der Anspruch an die Zuverlässigkeit der lexikografischen Daten. Dies auch v.a. deshalb, da bekannt ist, dass Wörterbuchbenutzer oft Angaben in Wörterbüchern glauben. Wie viele Leute gibt es beispielsweise, die meinen, ein Wort, welches nicht im Duden steht, gebe es auch nicht? Dies wird von Lexikografen und Wörterbuchforschern zwar immer wieder kritisiert oder relativiert. Trotzdem ist bei einem Benutzer, der mit automatisch erstellten Daten arbeitet, generell eine andere und mehr eigene Reflexion nötig als bei einem Wörterbuchbenutzer, der ein normales, d.h. lexikografisch erarbeitetes Produkt benutzt. Denn um bei dem eben gezeigten Beispiel zu bleiben: Es ist gravierender zu glauben, dass man Nachnamen allgemein durch eine Vorsilbe positiv bewerten kann, als wenn in einem allgemeinen, einsprachigen Wörterbuch die Bedeutungsparaphrase nicht den Kern der aktuellen Bedeutung erfasst.

Dies soll allerdings nicht generell die Wertigkeit automatisch erstellter Daten in Zweifel ziehen. Im Gegenteil: Automatisch erstellte Daten erlauben z.T. einen unverfälschten Blick auf die Sprache, und sie werden immer wertvoller, je besser die zu Grunde gelegten Korpora und v.a. die angewendeten Analysemethoden sind. Trotzdem ist auf Seiten der Benutzer mehr Kenntnis vonnöten, um rein automatische Ergebnisse richtig einordnen zu können. Deshalb sollte man auch vorsichtig mit der Behauptung sein, das *Wortschatz-Lexikon* erfülle die Funktionen eines üblichen Wörterbuchs. Das *Wortschatz-Lexikon* ist daher ein Beispiel dafür, dass es nicht reicht, es nur als elektronisches Wörterbuch zu bezeichnen. Man braucht verschiedene Einordnungs- und Benennungsmöglichkeiten für auf der einen Seite ein solches Produkt und auf der anderen Seite

irgendeine elektronische Version eines normalen allgemeinen einsprachigen Wörterbuchs, damit man analog zu diesen unterschiedlichen Benennungen den Produkten unterschiedliche Bewertungskataloge zuordnen kann. Solche terminologischen Unterscheidungen bieten dabei selbst noch keine Bewertungsmöglichkeiten, sondern sollen mögliche Ordnungsvorstellungen liefern, nach denen dann Bewertungskriterien entwickelt oder schon entwickelte gruppiert werden können. Bevor nun die Form der Erarbeitung der lexikografischen Daten in eine mögliche Unterscheidung elektronischer Wörterbücher eingearbeitet wird, sollen zunächst kurz einige grundlegende Typologisierungsmöglichkeiten elektronischer Wörterbücher referiert werden.

3. Unterscheidungsmöglichkeiten elektronischer Wörterbücher: bereits vorgelegte Typologien

Die Bewertung und Einordnung elektronischer Wörterbücher ist für die Wörterbuchforschung ein neuer wichtiger Bereich. Als Beiträge in der deutschsprachigen Wörterbuchforschung sind u.a. die Aufsätze von Andrea Lehr (Lehr 1996) und Annette Klosa (Klosa 2001) zu nennen. Den aktuellsten Vorschlag einer Typologie mit umfangreicher Zurkenntnisnahme der internationalen (meta-)lexikografischen Literatur bietet de Schryver in seinem schon einmal zitierten Aufsatz „Lexicographer's Dreams in the Electronic-Dictionary Age" (Schryver 2003). De Schryver konstatiert darin, dass die bisherigen Typologien für die heutige Variabilität elektronischer Wörterbücher nicht hinreichend seien. Er schlägt daher eine eigene 3-stufige Typologie vor, die nach einer dreigliedrigen Grundfrage entwickelt ist:

> We would therefore like to suggest a typology based on one main, rigid criterion: the way in which dictionaries are *accessed*. More particularly, in designing this typology, we had one question in mind: 'Who accesses what where?' The resulting three-step typology is thought to be flexible enough to cater for future innovations. (Schryver 2003:147)

In einem ersten Schritt wird demnach gefragt, wer auf das Wörterbuch zugreift. Hier bestehen nach dieser Typologie zwei grundsätzliche Möglichkeiten: Ein Mensch oder eine Maschine greift auf die Daten zu.[4] Der zweite Teil der Frage ist, auf was zugegriffen wird – „What is accessed?" (Schryver 2003:149) –, d.h., wie das physikalische Objekt beschaffen ist. Im dritten Schritt wird dann geklärt, wie die Daten gespeichert sind, d.h.: „Where does one access the dictionary data?" (Schryver 2003:149). Diese Frage zielt damit auf den „type of storage" (Schryver 2003:149). Ein elektronisches

---

[4] De Schryver versteht damit unter einem „electronic dictionary" sowohl ein Produkt für Menschen als auch für Maschinen, auch wenn er an einer Stelle sagt, dass „the term ‚ED' will therefore stand for ‚human-oriented electronic dictionary'" (Schryver 2003:146). Dies unterscheidet sich von der hier vorgeschlagenen Auffassung, dass unter einem elektronischen Wörterbuch nur ein Produkt für Menschen verstanden werden soll (siehe Kapitel 4).

Wörterbuch kann mit dieser Typologie nach vielen Aspekten klassifiziert und eingeordnet werden. [5]

Zusätzlich zu dieser Typologie kann nach de Schryver eine von Lehr vorgeschlagene metalexikografische Bewertung vorgenommen werden: „In addition, for each of the EDs one could add what Lehr (1996) termed a 'metalexicographic evaluation'." (Schryver 2003:150). Dies soll kurz erläutert werden: Zur Eröffnung der neuen Lexicographica-Rubrik „Electronic Dictionaries" wurde von Andrea Lehr die sinnvolle und für Rezensionen schon mehrfach angewendete Unterscheidung zwischen papierorientierten vs. innovativ gestalteten elektronischen Wörterbüchern getroffen:

> In (meta-)lexikographischer Hinsicht müssen wir zwischen elektronischen Wörterbüchern, die auf ein Papierwörterbuch zurückgehen und solchen, die Neuentwicklungen sind, unterscheiden. Erstere lassen sich außerdem danach subklassifizieren, ob sie eine wesentliche Veränderung bezüglich der Erscheinungsform ihrer Wörterbuchartikel erfahren haben oder nicht, und letztere danach, ob bei der Gestaltung der Wörterbuchartikel an traditionelle lexikographische Formen angeknüpft oder ob ein neuer Weg beschritten wurde – wir sprechen beide Male von *papierorientierten* vs. *innovativen* elektronischen Wörterbüchern. (Lehr 1996:314)

Diese Art der Bewertung kann daher als eine weitere Folie über die Typologie de Schryvers gelegt werden.

Genauso sollen die im Folgenden vorgestellten terminologischen Klärungen als eine Ergänzung zu der eben ausgeführten Typologie gesehen werden, da hier ein Aspekt aufgegriffen wird, der bisher noch nicht in Unterscheidungsmöglichkeiten elektronischer Wörterbücher eingearbeitet ist.[6]

4. Automatisch erstellte vs. menschlich bearbeitete Wortschatzinformationssysteme: Vorschlag einer terminologischen Differenzierung

Bei der Bewertung und Einordnung elektronischer Wörterbücher spielen – wie bereits gezeigt wurde – sehr viele Fragen eine Rolle. Eine ist m.E. die Art der Erarbeitung der lexikografischen Daten. Denn auch in diesem Bereich müssen begriffliche Differenzierungen eingeführt werden, um die Vielfältigkeit der heute auf dem elektronischen Wörterbuchmarkt befindlichen Produkte hinreichend differenziert betrachten zu können. Dies wurde bereits am *Wortschatz-Lexikon* demonstriert. Die Entwicklung von solchen terminologischen Klärungen dient dabei nicht als Selbstzweck für diejenigen, die Rezensionen schreiben und die sich mit Terminologie aus diesem Bereich beschäftigen, son-

---

[5] Für einen grafischen Überblick über de Schryvers Typologie siehe Abbildung 2 in Schryver (2003:150).

[6] Die hier vorgestellten terminologischen Differenzierungen werden auch, z.T. formal ausführlicher, in Müller-Spitzer (2004) behandelt.

dern sie sollen dazu dienen, den Blick für die neue Vielfältigkeit im Zusammenhang mit elektronischen Wörterbüchern zu schärfen und diese Differenziertheit auch den Wörterbuchbenutzern vermitteln zu können. Im Bereich der elektronischen Wörterbücher hat sich gegenüber den gedruckten so viel verändert, dass den Benutzern diese Veränderung deutlich gemacht werden muss, wenn sie nicht die gleichen traditionellen Erwartungen an völlig anders geartete Produkte stellen sollen.

In einem ersten Schritt wird nun eingegrenzt, was hier unter einem elektronischen Wörterbuch verstanden werden soll und was nicht.

4.1 Was ist ein elektronisches Wörterbuch?

Die Bezeichnung „elektronisches Wörterbuch" ist generell eigentlich nicht passend für den Bezugsgegenstand, also für die elektronische Publikation lexikografischer Daten, da ein elektronischer Datenträger kein *Buch* ist. Die Übertragung des Begriffs Wörterbuch auf die Publikation lexikografischer Daten auf einem elektronischen Datenträger kann aber dann hilfreich sein, wenn damit auch bestimmte Eigenschaften des gedruckten Wörterbuchs auf das elektronische Wörterbuch übertragen werden können, d.h., wenn damit zur kommunikativ adäquaten Verwendung (im fachsprachlichen Kontext) beigetragen wird. Dies gilt m.E. allgemein für Metaphern im Bereich einer neuen Technologie.

Zur kommunikativ adäquaten Verwendung kann beigetragen werden, wenn die übertragenen Eigenschaften dem Bezugsgegenstand entsprechen. Um nun genauer zu prüfen, für welche Art von elektronischen lexikografischen Daten die Bezeichnung elektronisches Wörterbuch sinnvoll ist, ist daher zunächst zu fragen, wie ein gedrucktes Sprachwörterbuch zu definieren ist.[7] Ausgehend von diesen Begriffsbestimmungen soll daraufhin eine Definition des Begriffs „elektronisches Wörterbuch" vorgenommen werden.

Um eine Definition für ein Sprachwörterbuch zu entwickeln, bestimmt Wiegand zunächst, was ein Nachschlagewerk ist:

Def. 1 Ein *Nachschlagewerk* ist ein *Buch* (hier verstanden als etwas Gedrucktes) mit wenigstens einer definierten äußeren Zugriffsstruktur, dessen genuiner Zweck darin besteht, daß ein potentieller Benutzer aus den lexikographischen Textdaten Informationen zum Gegenstandsbereich des Nachschlagewerkes gewinnen kann. (Wiegand 1998:58)

Darauf aufbauend definiert Wiegand ein Sprachwörterbuch:

Def. 2 Ein *Sprachwörterbuch* ist ein *Nachschlagewerk*, dessen genuiner Zweck darin besteht, daß ein potentieller Benutzer aus den lexikographischen Textdaten Informationen zu sprachlichen Gegenständen gewinnen kann. (Wiegand 1998:58)

---

[7] Der folgende Versuch von Begriffsklärungen lehnt sich an die Wörterbuchforschung von Herbert Ernst Wiegand (1998) an, da diese Ausführungen meiner Kenntnis nach die ausführlichsten und genauesten Überlegungen dazu darstellen, was ein gedrucktes Sprachwörterbuch ist.

Folgt man diesen Festlegungen (und sie treffen m.W. mit der üblichen Verwendung dieser Begriffe im Fachkontext überein), dann ist die Verwendung des Terminus „elektronisches Wörterbuch" demnach nur dann für eine kommunikativ adäquate Verwendung im fachsprachlichen Kontext hilfreich, wenn diese Festlegungen folgendermaßen übertragen werden können:

Def. 3 Ein *elektronisches Nachschlagewerk* ist eine *elektronisch verfügbare Datensammlung* mit wenigstens einer definierten äußeren Zugriffsstruktur, dessen genuiner Zweck darin besteht, dass ein potenzieller Benutzer aus den lexikografischen Textdaten Informationen zum Gegenstandsbereich des elektronischen Nachschlagewerkes gewinnen kann.

Def. 4 Ein *elektronisches Sprachwörterbuch* ist ein *elektronisches Nachschlagewerk*, dessen genuiner Zweck darin besteht, dass ein potenzieller Benutzer aus den lexikografischen Textdaten Informationen zu sprachlichen Gegenständen gewinnen kann.

Nun bezeichnet aber Wiegand z.B. auch Komponenten eines Sprachübersetzungssystems, welche als eine Art Wörterbuch fungieren, als Wörterbuch, genauer als Maschinenwörterbuch, bei dem es entweder keine Benutzer gibt oder sozusagen der Computer der Benutzer ist. Auch in anderen metalexikografischen Forschungen werden hier keine grundsätzlichen begrifflichen Unterschiede gemacht. Ist es sinnvoll, auch in diesem Fall von einem Wörterbuch zu sprechen?

Dazu wieder zurück zu den Definitionen: Der Benutzer in den Definitionen ist ein Handelnder, der „Informationen gewinnen" kann; er muss daher ein Aktor mit kognitiven Fähigkeiten sein. Der Computer ist jedoch kein Aktor mit kognitiven Fähigkeiten, sondern eine Maschine, die vom Menschen dazu programmiert werden kann, Daten zu verarbeiten; ein Computer kann aus Daten keine Informationen gewinnen (vgl. Wiegand 1998:171).

Davon ausgehend ist es sinnvoll, elektronische Wörterbücher nur dann als „Wörterbücher" zu bezeichnen, wenn sie für einen menschlichen Benutzer gemacht sind. Nur dann können grundsätzliche Eigenschaften eines gedruckten Wörterbuchs sinnvoll auf das elektronische Wörterbuch übertragen werden. Ist der Computer als Benutzer gedacht, sollte besser von „lexikalischen Ressourcen für sprachtechnologische Produkte" (basierend auf Engelberg/Lemnitzer 2001:230) gesprochen werden, z.B. ein ‚automatisches Übersetzungsprogramm basierend auf einer lexikalischen Ressource'.

Die Definitionen sollten demnach dahingehend ergänzt werden, dass mit „potenziellem Benutzer" immer ein Mensch gemeint ist. Diese Einengung des Begriffs ist deshalb angemessen, weil nur so Kriterien, die Wörterbücher ausmachen und die über die Präsentation in verschiedenen Medien Bestand haben, sinnvoll übertragen werden können.

Nun wurde der Begriff „(elektronisches) Wörterbuch" dahingehend eingegrenzt, dass als Benutzer immer ein Mensch vorausgesetzt wird; d.h., dass es ein Produkt *für Menschen* ist. Das Beispiel aus dem *Wortschatz-Lexikon* hatte jedoch mehr zum Gegenstand, inwieweit die Daten *von Menschen* erarbeitet wurden. Letzteres soll später in die Unterscheidung von automatisch erstellten vs. lexikografisch, d.h. menschlich bearbeiteten,

elektronischen Wörterbüchern aufgenommen werden. Zunächst soll jedoch noch die Frage verfolgt werden, ob man nicht grundsätzlich eine alternative Benennung zu „elektronisches Wörterbuch" finden könnte.

4.2 Eine alternative Bezeichnung: Wortschatzinformationssystem

Wie vorhin herausgestellt wurde, ist die Benennung „elektronisches Wörterbuch" genau genommen nicht passend, da mit dem bezeichneten Bezugsgegenstand kein Buch vorliegt. Von der Definition her ist die Benennung zwar gerade so eingegrenzt worden, dass sie zumindest hilfreich für die kommunikativ adäquate Verwendung im fachsprachlichen Kontext sein kann; trotz dieser Eingrenzung bleibt das o.g. Defizit bestehen.

Wie könnte also ein wortschatzbezogenes elektronisches Nachschlagewerk genannt werden, wenn man den Begriff Wörter*buch* darin nicht verwenden will? Mein Vorschlag lautet: Als grundsätzliche Bezeichnung für sprachbezogene elektronische Nachschlagewerke soll *Wortschatzinformationssystem* dienen. Diese Benennung hat zwei wesentliche Vorteile:

1. Sie bedient sich nicht der Buchmetapher.
2. Die Einordnung als *Informationssystem* kommt von der Bezeichnung her der neuen Art des Arbeitens mit einem elektronischen Sprachnachschlagewerk näher, da sie mehr Dynamik in der Datenabfrage vermuten lässt.[8]

Der Terminus Wortschatzinformationssystem kann in der metalexikografischen Forschung synonym mit dem Terminus elektronisches Wörterbuch verwendet werden, auch wenn eventuell andere Konnotationen damit verbunden sind. Analog zu elektronischen Wörterbüchern können darüber hinaus auch nähere Einordnungen eines Wortschatzinformationssystems vorgenommen werden, z.B. ein ‚allgemeines, einsprachiges Wortschatzinformationssystem' oder ein ‚zweisprachiges Wortschatzinformationssystem französisch-deutsch'.[9]

4.3 Automatisch erstellte vs. lexikografisch bearbeitete Wortschatzinformationssysteme

Im Ausgangsbeispiel wurde die Frage aufgeworfen, ob und wie man Wortschatzinformationssysteme, deren Daten rein automatisch aus Textkorpora erhoben wurden, von solchen Wortschatzinformationssystemen unterscheiden kann, die von Lexikografen erarbeitet wurden. In dem in Abschnitt 2 genannten Beispiel spielen die Menschen im Erarbeitungsprozess zwar auch eine wichtige Rolle, aber eine *andere* als üblicherweise

---

[8] Mehr zum Begriff des *Informationssystems* und den dazugehörigen Definitionen vgl. Müller-Spitzer (2004).

[9] Ein Nachteil dieser Benennung ist, dass sie sehr lang ist. Eine Mehrworteinheit wie „wortschatzbezogenes Informationssystem" hätte jedoch den gravierenden Nachteil, dass die o.g. näheren Charakterisierungen schwierig wären, wie „wortschatzbezogenes zweisprachiges Informationssystem polnisch-deutsch"; und „Informationssystem" allein ist zu unspezifisch, da die Art der Information nicht charakterisiert wird.

Lexikografinnen, nämlich v.a. in der Entwicklung der Korpusanalysemethoden und auch in der Zusammenstellung der zu Grunde gelegten Korpora. Der Unterschied zwischen rein automatisch erstellten vs. menschlich bearbeiteten Wortschatzinformationssystemen ist, dass die automatisch gewonnenen Daten nicht von Lexikografen sortiert und bewertet werden. Dabei sind diese Formen der Datenerarbeitung keine sich gegenseitig ausschließenden Erstellungsformen. Im Gegenteil: Sie können sich aneinander anschließen oder sogar im Zuge der Erarbeitung in einem wechselseitigen Prozess immer wieder angewendet werden. Hier interessieren jedoch nicht vordringlich der Verlauf und die Phasen der automatischen oder automatisch unterstützten Erstellung von Wortschatzinformationssystemen, sondern der Status der Daten in einem Wortschatzinformationssystem, wenn es publiziert, also z.B. im Internet der Öffentlichkeit zugänglich gemacht wird.

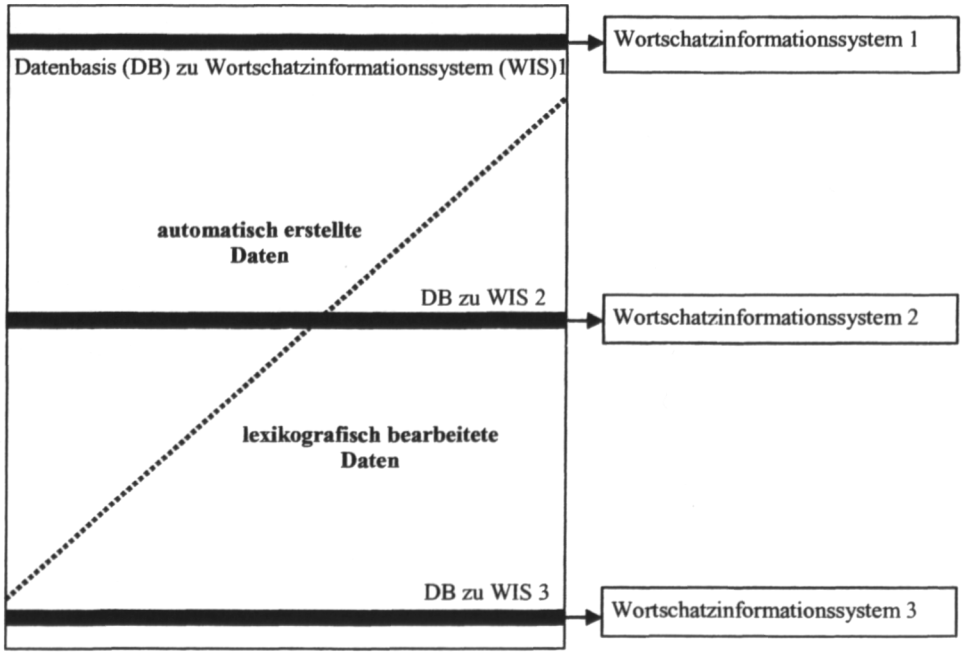

Abb. 3: Arten von Wortschatzinformationssystemen unterschieden nach ihrer Datenbasis „⟶" bedeutet soviel wie: „aus dieser Datenbasis wird entwickelt"

Auf welcher Grundlage diese Unterscheidung genau erfolgt, zeigt die Abbildung 3. Die Abbildung zeigt einen fiktiven Datenpool, der aus automatisch erstellten und lexikografisch bearbeiteten Daten besteht. Aus diesem fiktiven Datenpool wird eine konkrete Datenbasis herausgegriffen, die durch die graue Linie veranschaulicht wird; aus dieser konkreten Datenbasis wird ein Wortschatzinformationssystem entwickelt. Diese einzelnen Produkte können aufgrund der Art ihrer Datengrundlage unterschieden werden. Das erste Wortschatzinformationssystem besteht nur aus automatisch erstellten Daten, das zweite sowohl aus automatisch erstellten als auch aus lexikografisch bearbeiteten Daten

und das Wortschatzinformationssystem 3 besteht nur aus lexikografisch bearbeiteten Daten. Dieser fiktive Datenpool kann in einem konkreten Projekt – von oben nach unten betrachtet – den zeitlichen Verlauf eines Projektes darstellen, wenn zunächst eine Menge automatisch erstellter Daten vorliegt, die dann sukzessiv lexikografisch bearbeitet wird. Es ist aber auch ein Projekt denkbar, bei dem nur ein Wortschatzinformationssystem der ersten Art entsteht, so wie es beim *Wortschatz-Lexikon* augenscheinlich der Fall ist.

Unter lexikografischer Bearbeitung soll dabei jede Art der reflektierten menschlichen Bearbeitung der automatisch erstellten Daten verstanden werden, vom Überprüfen über das Umsortieren bis hin zum Kommentieren. Es ist also mit lexikografischer Bearbeitung nicht unbedingt gesagt, dass das objektsprachliche Material durch Kommentierungen o.Ä. angereichert wird. Die lexikografische Bearbeitung steht allein dafür, dass die Daten von Experten reflektiert, gesichtet und überprüft wurden und darüber hinaus eventuell kommentiert sind. Denn allein diese Überprüfung ist in ihrem Wert nicht zu unterschätzen. Beispielsweise würde bei jeder Überprüfung der oben gezeigte Eintrag zu „Kohl" sicherlich zumindest so umsortiert werden, dass die Angaben, die „Kohl" als Eigenname betreffen, von denen abgegrenzt werden, die sich auf „Kohl" als Gemüse beziehen. Außerdem würden augenscheinliche Fehler entfernt.

Wichtig ist dabei aber auch: Diese verschiedenen Arten von Wortschatzinformationssystemen sollen nicht als qualitative Stufen verstanden werden, d.h. ein Nachschlagewerk von der ersten Art ist ein Schlechtes und dann wird es Stufe für Stufe besser. Innerhalb jeder dieser Arten kann es große graduelle Unterschiede in der Qualität geben. Sind z.B. bei einem zunächst automatisch erstellten Wortschatzinformationssystem schon das zu Grunde gelegte Textkorpus und v.a. die Analysemethoden nicht gut, dann kann das Produkt trotz weit reichender Überarbeitung kaum wirklich überzeugend werden. Was hier aber Gegenstand der Betrachtung ist und zu der vorgeschlagenen Aufteilung in drei Arten führt, ist eine Klassifikation nach dem Status der Daten im Wortschatzinformationssystem zur Zeit der Publikation. Dieser ist in jeder dieser Arten unterschiedlich, egal, von welcher Qualität die Daten sind.

Der unterschiedliche Status der Daten sollte in der Benennung der Produkte deutlich werden. Es ist daher notwendig, den Wortschatzinformationssystemen eine Zusatzbezeichnung zu geben, die diesen Unterschied deutlich macht. Dies soll – wie es hier auch schon praktiziert ist – über Attribute geschehen, die die einzelnen Formen von Wortschatzinformationssystemen spezifizieren.

Die Typen, die in der Abbildung 3 unterschieden wurden, werden in Abbildung 4 mit folgenden Attributen in ihrer Benennung unterschieden:

1. Automatisch erstelltes Wortschatzinformationssystem
2. Semiautomatisch erstelltes Wortschatzinformationssystem
3. Lexikografisch bearbeitetes Wortschatzinformationssystem

Die Unterscheidung in automatisch erstellte vs. lexikografisch bearbeitete Daten ist hier auf der Ebene des gesamten Produktes angesetzt. Bei Wortschatzinformationssystemen, in denen die Daten automatisch erstellt und z.T. menschlich bearbeitet sind, also bei einem semiautomatisch erstellten Wortschatzinformationssystem, muss diese Unterscheidung bei genauer Betrachtung jedoch auf einzelne Einträge und vielleicht sogar auf

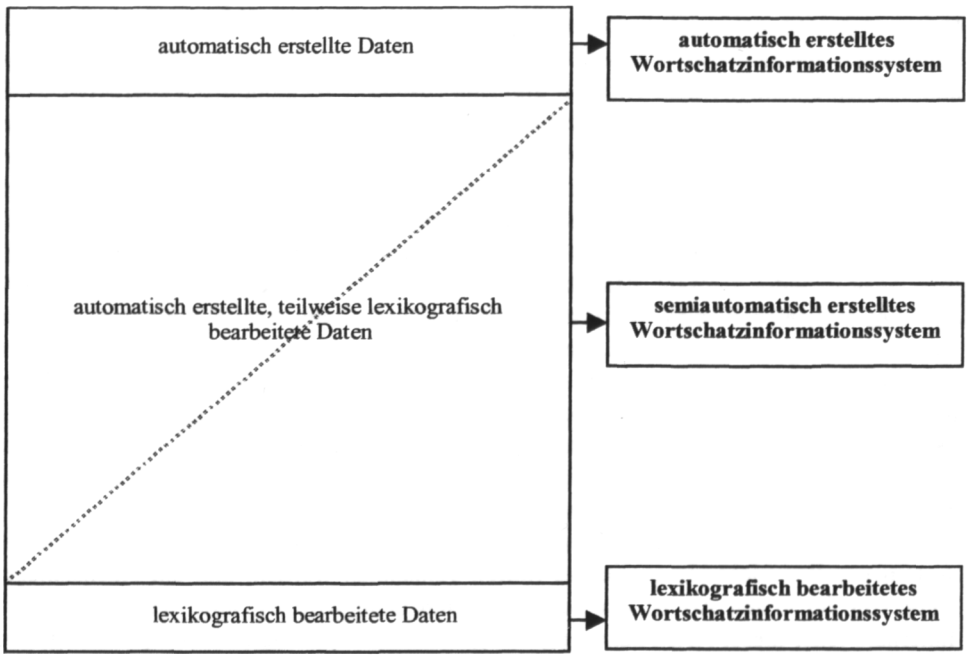

Abb. 4: Arten von Wortschatzinformationssystemen unterschieden nach ihrer Datenbasis „⟶" bedeutet soviel wie: „aus einer solchen Datenbasis wird entwickelt"

die Ebene einzelner Angaben bezogen werden. Beispielsweise können Teile der Lemmastrecke rein automatisch erstellt sein, andere Teile sind bereits lexikografisch bearbeitet (siehe auch Kapitel 5).

Damit sind Vorschläge für terminologische Grundunterscheidungen für Arten von Wortschatzinformationssystemen analog zu der Art ihrer Datenbasis getroffen und die dazugehörigen Benennungen präzisiert. Dem Benutzer eines Wortschatzinformationssystems sollte jederzeit transparent sein, welchen Status die Daten haben, die er vor sich hat und wie verlässlich daher z.B. die Angaben sind. Und auch um ein elektronisches Sprachnachschlagewerk zu beurteilen, sollte dies geklärt werden, da für verschiedene Arten von Wortschatzinformationssystemen jeweils andere Beurteilungsmaßstäbe herangezogen werden müssen.

4.4 Gehören automatisch erstellte Wortschatzinformationssysteme zum Gegenstandsbereich der Lexikografie und Wörterbuchforschung?

Bei der automatischen Erstellung von Wortschatzinformationssystemen spielen die Menschen im Erarbeitungsprozess eine andere Rolle als üblicherweise Lexikografen. Allein die menschliche Bearbeitung der Daten soll deshalb explizit als lexikografische Bearbeitung gelten.

Es stellt sich daher die Frage: Gehören automatisch erstellte Wortschatzinformationssysteme überhaupt zum Gegenstandsbereich der Lexikografie? Zwei wesentliche Gründe sprechen dafür:

1. Der genuine Zweck dieser Gebrauchsgegenstände (also sowohl automatisch erstellter wie lexikografisch bearbeiteter Wortschatzinformationssysteme) gleicht sich auf einer allgemeinen Ebene: Sie werden erstellt, damit sie benutzt werden können, um aus den lexikografischen Textdaten Informationen über den jeweiligen Gegenstand des Sprachnachschlagewerkes zu gewinnen.
2. Man kann daher vermuten, dass Wörterbuchbenutzer zwischen diesen Produkten keine grundsätzlichen Unterschiede sehen, auch wenn z.B. die Verlässlichkeit der dargebotenen Daten stark divergiert.

Da die Wörterbuchbenutzer in der Lexikografie jedoch im Zentrum der Überlegungen stehen und stehen sollten – da Wörterbücher eben Gebrauchsgegenstände sind – müssen m.E. auch automatisch erstellte Wortschatzinformationssysteme in den Gegenstandsbereich der Lexikografie und Wörterbuchforschung einbezogen werden.

Die Eigenschaften von Lexikografie als eigenständiger kultureller und wissenschaftlicher Praxis (vgl. Wiegand 1998:41), so wie sie sich in der Printlexikografie entwickelt haben, sind zu großen Teilen jedoch nur auf lexikografisch bearbeitete Wortschatzinformationssysteme übertragbar. Rein automatisch erstellte Wortschatzinformationssysteme sollten daher deutlich davon abgegrenzt werden. Zwar gilt auch für Letztere, dass „Lexikografie aus menschlichen Handlungen und ihren Ergebnissen" (Wiegand 1998:52) besteht, denn die Menschen spielen im Erarbeitungsprozess eine Rolle. Die Art dieser menschlichen Handlungen unterscheidet sich jedoch sehr stark von bisherigen lexikografischen Tätigkeiten und demnach unterscheiden sich auch die Maßstäbe, wie die Ergebnisse dieser Handlungen beurteilt werden können. Aus diesem Grund wurden hier die entsprechenden begrifflichen Unterscheidungen eingeführt. Denn nun können im Anschluss – z.B. für Wörterbuchrezensionen aber auch bei der Entwicklung von Produkten – Fragen gestellt werden wie:

Welche Erwartungen kann man an automatisch erstellte Wortschatzinformationssysteme stellen? Welche gerade nicht? Wo liegen die Stärken der verschiedenen Erarbeitungsarten? Für welche Benutzergruppe bzw. welche Benutzungssituationen sind welche Arten der Erarbeitung Erfolg versprechend? etc.

In diesem Sinne sollen die hier vorgestellten Überlegungen als eine ordnende Vorarbeit für eine weitere Beschäftigung mit diesem Thema in Lexikografie und Wörterbuchforschung dienen.

5. Ein Anwendungsbeispiel: das Projekt *elexiko*

Elexiko ist die Abkürzung für „elektronisches, lexikologisch-lexikografisches, korpusbasiertes Wortschatzinformationssystem" (früher auch bekannt unter *Wissen über Wör-*

ter).[10] Ziel des Projektes ist eine Beschreibung des deutschen Wortschatzes nach Regeln wissenschaftlicher Lexikografie; die gesamte Lemmaliste wird etwa 300 000 Einträge umfassen, um einen Eindruck von der Größe des Projektes zu vermitteln. Das Projekt mündet jetzt von der Planungs- in die Realisierungsphase, d.h., noch kann man in diesem Wortschatzinformationssystem nicht umfassend recherchieren.

Korpusbasiert heißt in *elexiko*, dass automatische Methoden eine sehr wichtige Rolle bei der Beschreibung des Wortschatzes spielen werden, da der sprachliche Usus anhand von Massendaten untersucht werden soll und so eine empirisch sehr gut fundierte Beschreibung des Wortschatzes angestrebt wird. Hier soll daher das in der Vorbemerkung erwähnte ‚Analyseparadigma' das leitende Prinzip sein. (vgl. Steyer 2004:93) Dies ist am IdS auch deshalb besonders gut möglich, da hier die weltweit größten elektronischen Textkorpora zur deutschen Sprache zur Verfügung stehen und elaborierte Recherchewerkzeuge entwickelt worden sind, wie beispielsweise die Kookkurrenzanalyse[11]. Außerdem ist der Einsatz automatischer Methoden bei der Größe des Projektes und demgegenüber bei der Anzahl der Projektmitarbeiter unerlässlich.

Da im Projekt demnach auf der einen Seite eine Anreicherung der gesamten Lemmastrecke um bestimmte Informationen in absehbarer Zeit nur in automatischer Weise möglich ist und auf der anderen Seite aber auch umfangreiche Wörterbuchartikel lexikografisch erarbeitet werden sollen, sprechen wir von einer horizontalen und vertikalen Füllung dieses Wortschatzinformationssystems.

Horizontale Füllung soll bedeuten, dass über die gesamte Lemmastrecke nach und nach automatisch erstellte Angaben ergänzt werden. Dabei gibt es zwei verschiedene Ansatzpunkte: Zum einen die gezielte automatische Erarbeitung bestimmter Typen von Angaben, z.B. orthografischer Angaben (siehe unten), bei der individuell geprüft werden muss, auf welche Art bzw. mit welcher Software diese automatische Erstellung sinnvoll ist, und auf der anderen Seite die Weiterentwicklung der bereits genannten automatischen Methoden, die generell die Grundlage für die Erarbeitung der Wörterbuchartikel sein soll.

Bisher ist diese horizontale Füllung bereits realisiert worden im Bereich der orthografischen Angaben. Die gesamte Lemmaliste wurde zunächst aus den IDS-Korpora erstellt, d.h. die Lemmazeichengestaltangaben waren zunächst meist in alter Rechtschreibung angesetzt, da diese Formen wesentlich häufiger in den Korpora vorkommen.[12] Konsens im Projekt war jedoch, dass alle Lemmazeichengestaltangaben in neuer Rechtschreibung angesetzt sein sollten. Aus diesem Grund wurden zwei Softwareprogramme eingesetzt, die eine Konvertierung der Angaben in die neue Rechtschreibung vornahmen; gleichzeitig wurden in diesem Schritt orthografische Varianten ergänzt.[13] Auch Silbenangaben

---

[10] Für Informationen zum Projekt siehe www.elexiko.de.

[11] Computerprogramm: „C. Belica: Statistische Kollokationsanalyse und Clustering. COSMAS Analysemodul, © 1995. Institut für Deutsche Sprache. Mannheim".

[12] Verantwortlich hierfür ist Ulrich Schnörch. Zu näheren Informationen, wie die Stichwortliste entwickelt wurde, vgl. die Projekthomepage.

[13] Verantwortlich für diesen Bereich ist Annette Klosa. Nähere Informationen unter www.elexiko.de.

wurden automatisch erstellt. Dementsprechend sind z.B. beim Eintrag „Jogurt“ zurzeit unter lesartenübergreifenden Angaben folgende Informationen vorhanden:

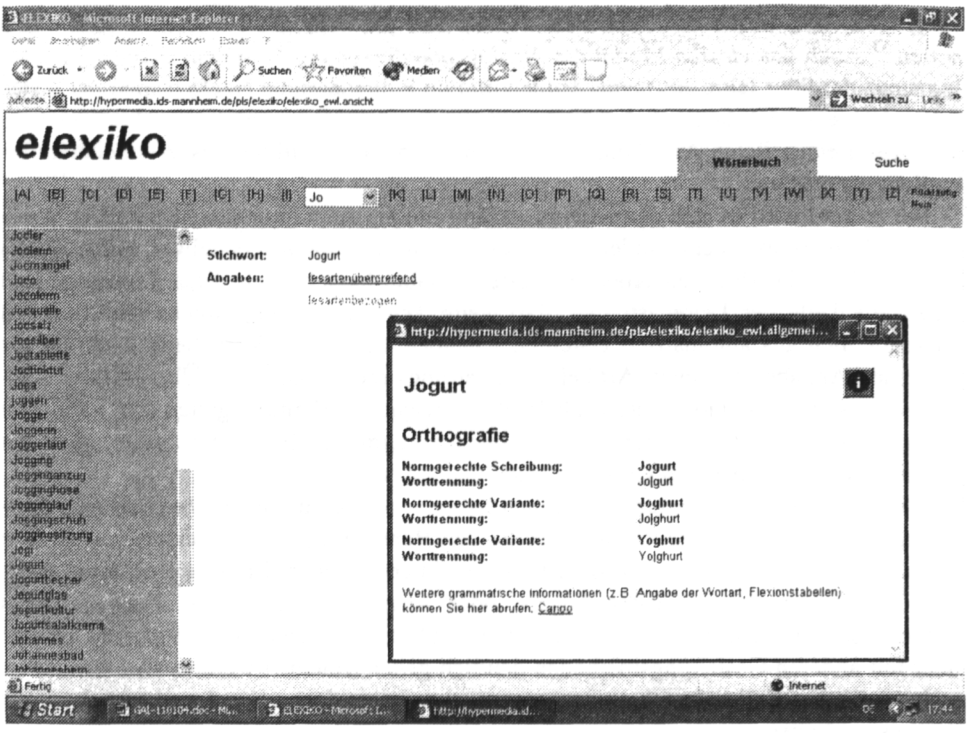

Abb. 5: orthografische Angaben zu „Jogurt“ aus elexiko

Auf diese Art sollen im Folgenden auch morphologische Angaben ergänzt werden.

Die andere Seite – die systematische Anwendung der mathematisch-statistischen Methoden als Zugang zu sprachlichem Usus – wird in *elexiko* v.a. im Modul *Usuelle Wortverbindungen*[14] geleistet. Dies geschieht in enger Kooperation mit der AG Korpustechnologie, die die automatischen Korpusanalysemethoden weiterentwickelt. Steyer formuliert dies als wechselseitigen „Prozess von automatischer Erkennung sprachlicher Ereignisse und menschlicher Interpretation dieser beobachteten Ereignisse“; „ein wechselseitiger Prozess, der zu einer Optimierung der Rechenprozesse führen soll“ (Steyer 2004:94). Dabei wird angestrebt, möglichst viel dem Rechner überlassen zu können und auch in diesem Bereich das Wortschatzinformationssystem automatisch zu füllen. Auch in diesem Modul ist jedoch daneben eine lexikografische Aufarbeitung der automatisch gewonnenen Angaben geplant, und zwar in der Erarbeitung systematisierter Kookkurrenzcluster auf unterschiedlichen Komplexitätsstufen und mit unterschiedlichem Explizierungsgrad.

---

[14] Siehe www.ids-mannheim.de/lexik/UsuelleWortverbindungen.

Diese lexikografische Bearbeitung ist daher der vertikalen Füllung des Wortschatzinformationssystems zuzurechnen. Vor allem in der *elexiko*-Kerngruppe werden dabei umfangreiche Wörterbuchartikel lexikografisch erarbeitet, die Angaben zu Semantik, Pragmatik, Grammatik etc. enthalten werden und die sehr gezielt recherchierbar sein sollen.[15] Auch das *elexiko*-Modul der *Neologismenforschung*[16], in dem bereits die Neologismen der 90er Jahre erarbeitet wurden und im Folgenden die Neologismen der ersten Jahre des neuen Jahrtausends bearbeitet werden, ergänzt ein gesamtes Produkt aus lexikografisch erarbeiteten Angaben, d.h. trägt zur vertikalen Füllung des Wortschatzinformationssystems bei.

Bei *elexiko* wird es sich demnach insgesamt um ein semiautomatisch erstelltes Wortschatzinformationssystem handeln, welches z.T. aus automatisch erstellten, teilweise aus lexikografisch bearbeiteten Angaben besteht. Dabei bezieht sich diese Trennung zum einen auf Teile der Lemmastrecke, d.h. weite Teile der Lemmastrecke bestehen fast ausschließlich aus automatisch erstellten Angaben, auf der anderen Seite auch auf die einzelnen Angaben in einem Artikel. Da allerdings die menschliche Überprüfung automatisch erstellter Angaben bereits als lexikografische Bearbeitung gelten soll, kann man davon ausgehen, dass in jedem Artikel, der lexikografisch bearbeitet wurde, auch nur Angaben von diesem Status enthalten sind.

Welchen Status die einzelnen Angaben haben – ob sie automatisch erstellt oder lexikografisch be- bzw. erarbeitet sind – sollte den Benutzern jederzeit kenntlich sein. Wie genau diese Kennzeichnung erfolgt, ist noch nicht entschieden: Zum Beispiel kann durch eine Kurzinformation zu jeder Angabe oder Angabegruppe auf die Erarbeitungsart hingewiesen werden, die Angaben könnten farblich unterschiedlich dargestellt werden oder mit anderen Gestaltungsmethoden deutlich gekennzeichnet werden. Den Benutzern sollte auf jeden Fall jederzeit transparent sein, welchen Status die Angaben haben, mit denen sie arbeiten.

Insofern ist *elexiko* ein Beispiel dafür, wie die unterschiedliche Erarbeitungsart der im Wortschatzinformationssystem vorhandenen Angaben auch in der Praxis eine wichtige Rolle spielen kann.

6. Literatur

Behaghel, Otto (1923): *Deutsche Syntax. Bd. 1: Die Wortklassen und Wortformen*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Belica, Cyril (1998): „Statistische Analyse von Zeitstrukturen in Korpora". In: Teubert, Wolfgang (Hg.): *Neologie und Korpus*. Tübingen: Narr (= Studien zur deutschen Sprache, 11), 31-42.

Bergenholtz, Henning (1996): „Korpusbasierte Lexikographie – Bericht über ein Symposium in Kopenhagen am 10. und 11.2.1996". In: *Lexicographica*, 12, 255-259.

---

[15] Siehe www.elexiko.de/Organisation.html und www.elexiko.de/Recherche.html.

[16] Siehe www.ids-mannheim.de/lexik/Neologie/.

Engelberg, Stefan/Lothar Lemnitzer (2001): *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.

Feldweg, Helmut (1997): „Wörterbücher und neue Medien: Alter Wein in neuen Schläuchen?". In: *Zeitschrift für Literaturwissenschaft und Linguistik*, 107, 110-123.

Lehr, Andrea (1996): „Zur neuen Lexicographica-Rubrik ‚Electronic Dictionaries'". In: *Lexicographica*, 12, 310-317.

Klosa, Annette (2001): „Qualitätskriterien der CD-ROM-Publikation von Wörterbüchern". In: Lemberg, Ingrid/Bernhard Schröder/Angelika Storrer (Hgg.): *Chancen und Perspektiven computergestützter Lexikographie*. Tübingen: Niemeyer (= Lexicographica. Series Maior, 107), 93-101.

Müller-Spitzer, Carolin (erscheint 2004): „Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen". In: *Lexicographica*, 19.

Quasthoff, Uwe/Christian Wolff: „Korpuslinguistik und große einsprachige Wörterbücher", *Linguistik online*, 3/2, <www.linguistik-online.de/2_99/quasthoff.html>.

Schryver, Gilles-Maurice de (2003): „Lexicographer's Dreams in the Electronic-Dictionary Age". In: *International Journal of Lexicography*, 16/2, 143-199.

Steyer, Kathrin (2004): „Kookkurrenz: Korpusmethodik, linguistisches Modell, lexikografische Perspektiven". In: Steyer, Kathrin (Hg.): *Wortverbindungen – mehr oder weniger fest*. Berlin/New York: de Gruyter (= Jahrbuch des Instituts für Deutsche Sprache 2003).

Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd.. Berlin/New York: de Gruyter.