



Korpuslexikographie und Flexionsmorphologie

Farø, Ken Joensen

Published in:

Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen

Publication date:

2005

Document version

Også kaldet Forlagets PDF

Citation for published version (APA):

Farø, K. J. (2005). Korpuslexikographie und Flexionsmorphologie. I H. Gottlieb, J. Erik Mogensen, & A. Zettersten (red.), *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen* (s. 217-226). Tübingen: Max Niemeyer Verlag.

Ken Farø

Korpuslexikographie und Flexionsmorphologie

1. Einleitung

Die Angaben zur Flexionsmorphologie gehören zu den weniger beachteten Informationskategorien in der Lexikographie,¹ sowohl vom Benutzer- als auch vom metalexikographischen Gesichtspunkt her (vgl. Mugdan 1989: 519). In den relativ wenigen Fällen, wo die Flexionsmorphologie im Mittelpunkt einer lexikographischen Untersuchung steht, wie z.B. in Mugdan (1983), wird, von seltenen Ansätzen abgesehen (op.cit.: 203 und 208f.), meist dem Verhältnis zwischen lexikographischer Behandlung und sprachlicher *Empirie* wenig Aufmerksamkeit zuteil. Die Flexionsparadigmen werden im Allgemeinen als etwas schon Gegebenes betrachtet,² ob dies nun mit Normtreue zusammenhängt oder womöglich andere Gründe hat (siehe u.a. Fahl et al. 2002: 228ff.).³

Vor dem Hintergrund der Arbeit mit der Flexionsmorphologie im neuen korpusbasierten "Dänischen Wörterbuch"⁴ (Den Danske Ordbog, abgekürzt: DDO) der Dänischen Sprach- und Literaturgesellschaft (Det Danske Sprog- og Litteraturselskab, abgekürzt: DSL) möchte ich hier einige Schwierigkeiten beschreiben, die mit der Herausarbeitung der Flexionsangaben in einem Wörterbuch dieses Typus verbunden sind. Und zwar werden Probleme behandelt, die sich in einem Konfliktgebiet befinden, das aus *Empirie*, *Norm* und *lexikographischen Ansprüchen* gebildet wird.

Das Problem kann in diesem Rahmen nur ansatzweise beschrieben werden, obwohl es, da sich ja die korpuslinguistische⁵ Methode in der lexikographischen Praxis immer mehr verbreitet, eine viel umfassendere Untersuchung verdient hätte.

¹ Vorgesehen ist jedoch ein Artikel mit dem Titel „Lexicography“ im geplanten *HSK 17* (Morphologie), Bd. 2 (siehe Literaturverzeichnis: Booij et al.).

² Eine wichtige Ausnahme davon bildet Karlsson (2000), der sich jedoch leider nicht mit den Konsequenzen für die Lexikographie beschäftigt (siehe unten).

³ Trap-Jensen (2002), der aus Gründen der Editionspraxis zwar vor diesem Artikel erschienen ist, jedoch später vorgelegt wurde, beschäftigt sich ebenfalls mit Problemen der lexikographischen Norm und Empirie, wenn auch in einem weiteren Kontext, indem sowohl Orthographie, Bedeutungen, Lemmaselektion und Flexion einbezogen werden.

⁴ Als sechsbändiger moderner Nachfolger des 28-bändigen "Ordbog over det Danske Sprog" aus den Jahren 1919-1956 konzipiert.

⁵ Zur Kritik dieses Terminus: siehe Bergenholtz (1996: 7f.).

2. DDO und seine Flexionsangaben

“In comparison to other grammatical characteristics, corpus-based analyses of morphological characteristics are relatively straightforward” (Biber et al. 1998: 58). Dies ist eine Haltung, der man in der korpuslinguistischen Literatur öfters begegnet.

In der Tat schien die Aufgabe relativ unkompliziert zu sein: die Flexion der Lemmata des Dänischen Wörterbuchs zu beschreiben. Dabei sollten zwei verschiedene Quellen als Hauptgrundlage dienen: teils die Empirie, die aus dem 40 Millionen Textwörter umfassenden Textkorpus des Wörterbuchs besteht (siehe unten), teils die Angaben des offiziellen dänischen Rechtschreibwörterbuchs, das vom dänischen Sprachrat herausgegeben wird (siehe Dansk Sprognævn 2001), einer Instanz, die u.a. die für die dänischen Behörden und Schulen verbindliche Orthographie festlegt. Diese Quellenheterogenität rührt vom Wunsch her, das Wörterbuch nicht bloß als eine wissenschaftlich interessante Dokumentation der zeitgenössischen dänischen Lexik zu konzipieren, sondern gleichzeitig auch als ein Werkzeug zur Sprachproduktion, womit der Wert des Wörterbuchs für die Öffentlichkeit erhöht werden sollte (Norling-Christensen/Asmussen 1998: 224) – schließlich wurde das Projekt zur Hälfte von öffentlichen Mitteln getragen.

Wo Diskrepanzen zwischen den beiden Quellentypen registriert würden, sollten die normwidrigen Formen, falls im Korpus in signifikanter Menge auftretend, im Wörterbuch ebenfalls beschrieben werden, jedoch mit dem Vermerk, dass sie vom Sprachrat *nicht* sanktioniert sind – ein in der dänischen Lexikographie bisher nicht völlig unbekanntes Phänomen (siehe unten).

Im Folgenden werden – nach Typen aufgeteilt – Beispiele solcher inoffiziellen, im Sprachgebrauch aber frequenten, Formen angeführt:

Pluralflexive:

albums (-mer/-Ø im offiziellen Rechtschreibwörterbuch und anderswo – Korpusbelege: 71 (*albums*) gegenüber 3 und 9)

baguettes (-r im offiziellen Rechtschreibwörterbuch und anderswo – Korpusbelege: 4 (*baguettes*) gegenüber 2; Google-Belege: 252 (*baguettes*) gegenüber 138)

bjørnekløer (Pflanze) (*bjørneklo/bjørnekloer* im offiziellen Rechtschreibwörterbuch und anderswo – Korpusbelege: 3 (*bjørnekloer*) gegenüber 0 und 0; Google-Belege: ca. 25 (*bjørnekloer*) gegenüber (?) und 12) (Sg.: *bjørneklo*)

champignons (-er im offiziellen Rechtschreibwörterbuch und anderswo – Korpusbelege: 33 (*champignons*) gegenüber 172)

Determinativflexive:

balsamen (*balsammen* im offiziellen Rechtschreibwörterbuch und anderswo) – Korpusbelege: 1 (*balsamen*) gegenüber 0; Google-Belege: 97 (*balsamen*) gegenüber 14)

bonusen (*bonussen* im offiziellen Rechtschreibwörterbuch und anderswo – Korpusbelege: 3 (*bonusen*) gegenüber 1; Google-Belege: 413 (*bonusen*) gegenüber 330)

Die beiden bei weitem umfangreichsten Kategorien inoffizieller Flexion bei den Substantiven im Korpus sind:

a) Fremdwörter aus dem Englischen und aus den romanischen Sprachen, die in der Empirie ein -s im Plural haben, was vom Dänischen Sprachrat in vielen Fällen nicht anerkannt wird (etwa 50 % der Fälle).

und

b) Wörter, davon auch einige fremde, die bis vor kurzem, d.h. bis zur dritten Ausgabe des Rechtschreibwörterbuchs (Dansk Sprognævn 2001), in ihren deklinierten Formen normgerecht ohne Konsonantengemination geschrieben werden konnten, was in der Empirie tatsächlich sehr häufig vorkommt. Mit der neuen Ausgabe des offiziellen orthographischen Wörterbuchs wurden diese Formen in der Norm abgeschafft. Damit wurden auf einmal sehr viele Flexive inoffiziell. Einerseits will Das Dänische Wörterbuch zwar darauf aufmerksam machen, doch andererseits sollte auch der Sprachgebrauch beschrieben werden, der im Korpus zu finden ist, was die primäre Aufgabe des Wörterbuchprojekts ist (ungefähr 33 % aller Fälle).

Fälle wie die oben dargestellten konnten meistens ohne große Schwierigkeiten registriert und beschrieben werden.

Das Verfahren, inoffizielle Formen ins Wörterbuch aufzunehmen, ist nicht einzigartig in der dänischen Lexikographie. So haben mittlerweile sowohl das Bedeutungs- als auch das Rechtschreibwörterbuch aus dem Verlag Politiken (= Becker-Christensen et al. 1999 bzw. Appel et al. 2001) häufig vorkommende inoffizielle Formen kodifiziert und diese als solche markiert. Doch hier scheint es sich ausschließlich um orthographische und nicht um flexionsmorphologische Probleme zu handeln. Außerdem liegt dem Verfahren dieser Wörterbücher, vom dänischen Sprachrat nicht sanktionierte Wortformen zu lemmatisieren, offenbar der Wunsch zugrunde, dem Benutzer zu helfen, die orthographisch normierte Form im Wörterbuch überhaupt zu finden, falls er bloß eine (frequente) “unkorrekte” Form kennt (siehe op. cit., Vorwort S. 7 bzw. 8).

3. Das Verhältnis zur offiziellen Rechtschreibung

Das Dänische Wörterbuch ist nicht in erster Linie ein Rechtschreibwörterbuch – deren gibt es im Dänischen mittlerweile vier, was zu gewissen prinzipiellen Kompetenzproblemen führen kann – und tatsächlich führt (siehe Fahl et al. 2002). Weil aber das Wörterbuch auch als Werkzeug der normgerechten Sprachproduktion dienen soll, so wird die offizielle Rechtschreibung immer berücksichtigt, wenn auch nicht in jedem Fall kritiklos

übernommen – zumindest nicht auf dem Gebiet der Flexionsmorphologie. Wenn das Wörterbuch somit der offiziellen dänischen Rechtschreibung folgt, so gilt dies mit gewissen Einschränkungen, die mit dem Hauptziel des Wörterbuchs zusammenhängen, den Kernwortschatz des Gegenwärtigen auf empirischer Grundlage zu dokumentieren.

Die Einschränkungen sind folgende:

a) Das Wörterbuch enthält Wörter, vor allem neuere, die vom offiziellen dänischen Rechtschreibwörterbuch, aus welchen Gründen auch immer, (noch?) nicht behandelt worden sind. Obwohl immer versucht wurde, Empfehlungen des Sprachrats einzuholen und in einigen Fällen auch Analogieschlüsse zu ziehen, vor allem von Komposita auf tatsächlich beschriebene Simplicia im Rechtschreibwörterbuch, hat die Flexion dieser Wörter im Prinzip keinen offiziellen Status, solange diese nicht im orthographischen Wörterbuch des Sprachrats kodifiziert sind. Dies ist ein prinzipielles Problem und hätte vielleicht im DDO expliziter gehandhabt werden können. Jedenfalls ist die jetzige Situation die, dass im Dänischen Wörterbuch nur Flexive als „unoffiziell“ („inoffiziell“) markiert werden, die zu den im offiziellen Rechtschreibwörterbuch gegebenen Formen in *expliziter* Opposition stehen.

b) Das andere Problem betrifft Wörter, die im Rechtschreibwörterbuch zwar enthalten sind, dort aber Flexionsparadigmen besitzen, die von der Korpusempirie gar nicht verifiziert werden konnten – das, was Karlsson (2000: 647) „theoretically possible word forms“ nennt (siehe auch Fahl et al. 2002: 228ff.). Obwohl Carstairs-McCarthy (1998: 324) meint, dass „paradigm inconsistency never, or almost never arises“, so haben die Erfahrungen aus dem DDO-Projekt gezeigt, dass dies eine zu schemenhafte Verallgemeinerung ist, die von der Korpusarbeit nicht bestätigt werden kann. In der Tat trifft zu, was Karlsson (2000: 652) treffend formuliert hat: „Defectivity turns out to be a gradient phenomenon“.

Im Korpus des DDO gibt es somit eine Menge von Wörtern, die entweder das Phänomen der „absolute defectivity“ (aufgrund grammatischer Restriktionen) aufzeigen oder die aus eher semantischen und pragmatischen Gründen von „relative defectivity“ geprägt sind (Karlsson 2000: 651) und die gleichzeitig von der normierenden Instanz in nicht wenigen Fällen „mechanisch“ ein rein grammatisch legitimes Flexionsparadigma zugeteilt bekommen haben.⁶ Diese werden im Dänischen Wörterbuch reduziert, z.B. wenn ein Wort im Korpus nie im Plural vorkommt und auch dem Interpreten zu sonderbar erscheint. Dadurch glauben wir der Zielsetzung der Empirie-treuen Beschreibung näher gekommen zu sein. In manchen Fällen ist sogar das ganze Paradigma, das in den traditionelleren Wörterbüchern angegeben wird, aufgrund von Korpusanalysen weggelassen worden, weil es im Sprachgebrauch nur unflektiert vorzukommen scheint.

Diese Reduktionen absolut und relativ defektiver Paradigmen stellen übrigens wohl kaum eine Einschränkung der Benutzerfreundlichkeit des Wörterbuchs als *Textprodukt*

⁶ Zwar hat mit der neuen Ausgabe des Rechtschreibwörterbuchs (Dansk Sprogævn 2001) die Flexionsmorphologie des Wörterbuchs einen Schritt in Richtung einer mehr empirischen Beschreibung gemacht. Doch gleichzeitig ist dadurch, weil diese Neuerung keineswegs konsistent durchgeführt ist (siehe Fahl et al. 2002: 228ff.), das Gesamtbild heterogener geworden. Und eine wichtige Frage ist in diesem Zusammenhang natürlich auch, in welchem Maß ein präskriptives Wörterbuch sich überhaupt nach empirischen Daten richten sollte.

tions-werkzeug dar, denn der Benutzer braucht die ausgelassenen Formen sicherlich gar nicht.

4. Probleme der praktischen Umsetzung

Im Folgenden werden ein paar Beispiele für die Schwierigkeiten gezeigt, die während der Arbeit auftauchen. Es handelt sich um die *Korpusgröße*, sowie um das Problem der *sondersprachlichen Flexive*.

4.1. Zum Problem der Korpusgröße

Dass zur Lösung verschiedener lexikographischer Probleme unterschiedliche Korpusgrößen erforderlich sind, wurde schon mehrmals in der Forschung erwähnt (siehe z.B. Bergenholtz 1996: 7). Doch wie groß muss ein Korpus sein, um eine adäquate Grundlage für die Beschreibung der Flexionsmorphologie eines Wörterbuchs zu bilden? Diese Frage ist nur pauschal beantwortbar. Erstens hängt ihre Antwort von der Größe des zu beschreibenden Sprachauschnitts ab, zweitens ist sie vom Ehrgeiz der Herausgeber abhängig: Wie detailliert möchte man die Flexionsmorphologie beschreiben?

Das Korpus des Dänischen Wörterbuchs besteht aus 40 Mio. Textwörtern. Auf der Grundlage von mehreren Untersuchungen wurde dies als hinreichend betrachtet, um den Kernwortschatz des Gegenwärtigen unter seinen wesentlichsten semantischen, morphologischen, syntaktischen und phraseologischen (u.a.m.) Aspekten relativ gründlich zu beschreiben.

Von einem so genannten „totalsprachlichen“ Wörterbuch, das heißt eines, das auch Fachsprachen decken sollte, wurde in einer anderen Untersuchung von 50 Mio. Textwörtern ausgegangen (siehe Bergenholtz 1996: 9).

Ob diese Zahlen nun für sämtliche lexikographischen Detailaufgaben hinreichend sind, kann sicherlich diskutiert werden. Obwohl im Korpus für die allermeisten Untersuchungen reichliche, ja überwältigende Mengen an Material zur Verfügung stehen, gibt es einzelne Problembereiche, für die externe empirische Quellen einbezogen werden mussten, wenn die Grundlage der Entscheidungen nicht Introspektion oder nonempirische Quellen sein sollte. Die Flexionsmorphologie gehört zu diesen Bereichen.

Zwar ist es in den meisten Fällen möglich, die Flexionsmorphologie aufgrund der Korpusdaten zufriedenstellend zu beschreiben. Doch vor allem bei Wörtern, die zu den weniger frequenten im Korpus gehören, besteht das Problem, dass z.B. im Korpus keine Pluralformen vorkommen. In einigen Fällen ist sogar überhaupt kein Flexionsparadigma vorhanden. Manchmal hängt dies wirklich mit dem Sprachgebrauch zusammen, wo das Wort nur im Singular oder immer gänzlich unflektiert vorkommt. In anderen Fällen aber wird bei der Zuhilfenahme zusätzlicher empirischer Quellen deutlich, dass die Datenlage des Korpus in diesem Punkt einfach nicht groß genug ist, und dass die Formen durchaus in der Allgemeinsprache im Gebrauch sind.

Natürlich kann man kritisieren, dass externe Quellen in die Analyse Eingang finden, denn in diesem Punkt wird ja damit die Operationalisierbarkeit der Arbeit angetastet, ein Dilemma, das in der Korpuslinguistik zu den altbekannten Problemen gehört. Man sollte aber stets vor Augen haben, was in solchen Ausnahmefällen wichtiger ist: die methodologische *Konsequenz*, der zuliebe man auf weitere Untersuchungen verzichtet und somit dem Benutzer Details verschweigt, oder aber die *empirische* Sprachbeschreibung an sich. Wie auch immer sie durchgeführt wird.

4.2. Zum Problem der anderssprachlichen Flexive

Wie verfährt man in den Fällen, in denen eine Flexionsform in allgemeinsprachlichen Texten nicht vorkommt, weil die Flexionsform als *fachsprachlich* anzusehen ist? Dies ist ein Problem, das für die meisten Wörterbuchtypen von Belang ist, nicht nur für das Korpuswörterbuch, obwohl es in diesem Fall ohne Zweifel deutlicher zu Tage tritt, als in den traditionelleren Wörterbüchern. Mugdan (1983: 208f.) hat gezeigt, dass in der Wörterbuchpraxis auf diesem Gebiet große Heterogenität herrscht. Dies ist immer noch feststellbar, zumindest in der dänischen und deutschen Lexikographie. Vor allem bei so genannten „Stoffbezeichnungen“ (siehe z.B. Helbig/Buscha 1991: 229) wird man mit dem Problem konfrontiert, wie beispielsweise bei den Wörtern *beton* und *cement* (deutsch: *Beton* und *Zement*, siehe unten). In der neuesten Ausgabe des offiziellen dänischen Rechtschreibwörterbuchs (Dansk Sprognævn 2001) haben diese beiden Wörter ein volles Flexionsparadigma. Dies ist auch in sämtlichen konkurrierenden orthographischen Wörterbüchern (Appel et al. 2001, Axelsen et al. 2001, und Bergenholtz et al. 1999), sowie im großen Bedeutungswörterbuch (Becker-Christensen 1999) der Fall. Doch ist diese Situation so selbstverständlich? Man findet den Plural im Korpus nicht und in anderen empirischen Quellen nur in genuinen Fachtexten wie den folgenden, die im Internet gefunden wurden:

beton ('Beton'):

Hærdeegenskaber af cementfattige betoner – effekter ved delvis substitution af cement med mikro-silica og flyveaske
www.civil.auc.dk/i6/forskpr/materials.html

Disse betoner er ikke underlagt DBC. Der anvendes sand i materialeklasse A, sten i materialeklasse M, samt dansk cement og puzzolaner
www.bb-beton.dk/betonpris.html

cement ('Zement'):

Bro Cement kombinerer fordelene ved ægte glasionomer og resinbaserede cementer. Glasionomerdelen giver fluorfrigørelse og kemisk binding til tandsubstans
www.3m.com/market/dental/dmc/dk/prod_html/products/dk_prod_krone.html

Hurtighærdnende cementer (styrkeklasse 52,5) kan anvendes til vinterbyggeri, hvor en hurtig hærdning af mørtlen har betydning
www.mur-tag.dk/muc/materialer/Cement.MLB.html

Aus diesem Grund hat die 1955-Ausgabe des dänischen Rechtschreibwörterbuchs wohl auch bei *beton* die Angabe: *ohne Plural*.

Hat eine veränderte empirische Lage dazu geführt, dass der Sprachrat seine Beschreibung ebenfalls verändert hat – seit der 1986-Ausgabe (Dansk Sprognævn 1992)? Wahrscheinlich nicht, jedoch hat man inzwischen sicherlich Belege für diesen Plural gefunden. Doch um welche Art von Plural handelt es sich? In der Literatur (z.B. in Mugdan 1983: 208) wird sie manchmal unter der Bezeichnung „Sortenplurale“ geführt⁷.

Im Vergleich zur bisher ziemlich homogenen Behandlung in der dänischen Lexikographie gibt es in der deutschen divergierende Haltungen dazu, wie man diese Plurale handhaben soll. Im Duden Universalwörterbuch wird die Pluralform zwar gegeben, jedoch mit der Präzisierung, dass es sich um „Arten“ handelt (Klosa et al. 2001).

Im Wahrig Deutsches Wörterbuch steht wie in der älteren Ausgabe des dänischen Rechtschreibwörterbuchs „ohne Plural“ (Wahrig-Burfeind et al. 2000).

Es gibt in der lexikographischen Praxis somit mindestens vier verschiedene Verfahrensweisen, mit diesem Problem umzugehen, und zwar:

- a) die Pluralform wird unkommentiert gegeben
- b) die Pluralform wird kommentiert gegeben
- c) die Pluralform wird unkommentiert ausgelassen
- d) die Pluralform wird kommentiert ausgelassen

Was ist nun am adäquatesten? Diese Frage kann wohl nur relativ zur Hauptfunktion des Wörterbuchs beantwortet werden. Handelt es sich um ein Wörterbuch, das primär Hilfe bei der *Textproduktion* geben möchte, dann kann man im Prinzip für ein maximierendes Paradigma argumentieren, d.h. hier kann der Plural gegeben werden. Allerdings würde dann die Hinzufügung von Nutzen sein, dass es sich um keine gewöhnliche Pluralfunktion handelt, sondern um eine, die „Arten“ indiziert und die außerdem fachsprachlich ist. Handelt es sich aber um ein *dokumentierendes* Wörterbuch, das nur den Kernwortschatz und die Kernbedeutungen dieser Wörter beschreiben kann oder will, dann sind diese Pluralformen nicht besonders relevant. Genauso wie von einem Wörterbuch des Kernwortschatzes nicht gefordert werden kann, dass es sämtliche Bedeutungen eines Wortes beschreibt, so kann man auch nicht fordern, dass es Flexionsparadigmen berücksichtigen soll, die in der Allgemeinsprache nicht vorkommen.

Man sollte sich dessen bewusst sein, dass man sich mit Randformen und -bedeutungen befasst, wenn man diese fachsprachlichen Plurale in einem gemeinsprachlichen Wörterbuch

⁷ An dieser Stelle soll auch auf den Artikel *Mass and Collection* im geplanten 2. Bd. vom HSK 17 (Morphologie) hingewiesen werden (siehe *Booij et al.* im Literaturverzeichnis).

berücksichtigt. Wie Mugdan (1983: 208) über diese peripheren Pluralformen geschrieben hat: "Wer sie aus irgendwelchen Gründen bilden und verwenden mag, sollte diese Freiheit haben". In ein Wörterbuch des Kern- und Allgemeinwortschatzes gehören sie aber kaum.

5. Zusätzliche Fragen

Es könnten in diesem Zusammenhang weitere Probleme diskutiert werden, z.B. das Problem des flexionsmorphologischen Verhältnisses von *Simplex* und *Kompositum* in der Korpuslexikographie: Besteht hier ein Abhängigkeitsverhältnis, oder sollte man Simplex und Kompositum als zwei isolierte Lexeme in ihrem eigenen Recht behandeln? Auch das Problem der *lexikalischen Paradigmen* gehört in diesen Zusammenhang: ist es eine berechnete Forderung, Bezeichnungen für Schulfächer, Krankheiten, Münzeinheiten usw. flexionsmorphologisch möglichst *gleich* zu behandeln, oder sollten hier einzig und allein die Korpusdaten das Sagen haben?

Schließlich sollte die Frage von diachronischen⁸ Veränderungen in der Flexionsmorphologie (morphologische Innovation) und inwiefern diese in die Lexikographie integriert werden bzw. integriert werden sollten, genauer untersucht und intensiver diskutiert werden. Denn obwohl die Flexionsmorphologie im Vergleich zu anderen grammatischen Kategorien generell konservativ ist (Wurzel 1984: 194), so können auch auf diesem Gebiet Schwankungen und eigentliche Wandlungen registriert werden, wovon auch Das Dänische Wörterbuch Zeugnis ablegt.

Wir können auf diese und weitere Fragen, die zum Problemkomplex Korpuslexikographie und Flexionsmorphologie gehören, an dieser Stelle jedoch nicht weiter eingehen. Doch sollte in diesem Zusammenhang betont werden, dass es sich um ein prinzipielles und für die zukünftige (Korpus-)Lexikographie weiterhin schwieriges Problem handelt, das tiefer gehende Untersuchungen erfordert.

6. Zusammenfassung

In mehreren neueren Arbeiten wurde hervorgehoben, dass Monofunktionalität eines Wörterbuchs erstens ein seltenes Phänomen und zweitens auch nicht wünschenswert sei (siehe dazu z.B. Tarp 1995 und Bergenholtz 1997). Dieser Einwand hat gegen die "naive" Auffassung, es bestehe zwischen Form und Funktion eines Wörterbuchs eine 1:1-Beziehung, ohne Zweifel seine Berechtigung. Dennoch muss man vor Augen haben, dass Polyfunktionalität die konsistente Herausarbeitung der einzelnen Informationskategorien des Wörterbuchs erheblich erschwert. Und dies gilt nicht zuletzt für die Flexionsmorphologie.

⁸ Für eine allgemeinere (nicht-lexikographische) Darstellung, siehe jedoch Joseph (1998).

Die Fragen und Schwierigkeiten bei der Erarbeitung der Flexionsmorphologie im Rahmen eines Korpuswörterbuchs mit wesentlichen Nebenfunktionen wie z.B. Hilfe für die Sprachproduktion sind vor allem folgende:

- Ist das Korpus groß genug, um auch die Flexion der weniger frequenten Lemmata empirisch zu erfassen? Einiges deutet darauf hin, dass dieses Detailproblem ein noch größeres Korpus erfordert, als man im Allgemeinen glaubt.
- Wie sollte sich das Wörterbuch gegenüber der offiziellen *Norm* verhalten – falls vorhanden – , wenn das von der Norm Beschriebene in der Empirie kaum oder gar nicht vorkommt und gleichzeitig unwahrscheinlich wirkt?
- Wie sollte man das Problem lösen, dass einige Flexive an bestimmte, anderssprachliche Bedeutungen gebunden sind? Sollte man sie auslassen oder mitnehmen, und dies gegebenenfalls dies explizit kommentieren?
- Wie sollte das Verhältnis zwischen Simplicia und Komposita sein? Sollten diese flexionsmorphologisch völlig getrennt beschrieben werden, oder zieht die Behandlung des Simplex eine bestimmte Beschreibung der Komposita nach sich?

und schließlich:

- Ist die Forderung nach Konsequenz, die in der Metalexikographie, darunter nicht zuletzt in den Rezensionen, oft zu hören ist, eine berechnete Forderung, wenn die *Empirie* alles andere als homogen ist?

Auch auf einem relativ leicht überschaubaren und anscheinend objektivierbaren Gebiet wie der Flexionsmorphologie herrscht in einigen Punkten letztendlich eine gewisse Willkür in der Korpuslexikographie. Und dies kann wohl gar nicht anders sein, denn „there are many unclear and borderline cases“ (Karlsson 2000: 649). Trotzdem stellt die Beschreibung der Flexionsmorphologie im Dänischen Wörterbuch hoffentlich einen Fortschritt dar, weil stets die Empirie die Grundlage der Herausarbeitung bildet, wodurch es sich von den herkömmlichen dänischen Wörterbüchern unterscheidet. Diese haben aber auch nicht unbedingt das gleiche Ziel wie Das Dänische Wörterbuch.

Will man auf dem Gebiet der Flexionsmorphologie Konsequenz im strengsten Sinne des Wortes, dann muss man sie *schaffen*, indem man ein Wörterbuch ohne Rücksicht auf den Sprachgebrauch *konstruiert*. Die Empirie kann sie nicht liefern. Denn die Sprache ist bei weitem nicht immer so logisch strukturiert, wie es für die Linguisten wünschbar wäre.

Literatur

Appel, Vibeke; Christian Becker-Christensen, Iben Rasmussen et al. (2001): *Politikens Retskrivningsordbog*. Kopenhagen: Politikken.

- Axelsen, Jens; Grethe Hjorth, Thomas Ingemann, Pia Vater (2001): *Retskrivningsordbog*. Kopenhagen: Gyldendal.
- Becker-Christensen, Christian et al. (1999): *Politikens Nudansk Ordbog med etymologi*. Kopenhagen: Politiken.
- Bergenholtz, Henning (1996): Korpusbaseret leksikografi. In: *LexicoNordica* 3: 5-17.
- (1997): Polyfunktionale ordbøger. In: *LexicoNordica* 4: 15-29.
- (1999): *Danskordbogen*. Aarhus: Systime/Center for Leksikografi.
- ; Joachim Mugdan (1979): *Einführung in die Morphologie*. Stuttgart: Kohlhammer.
- Biber, Douglas; Susan Conrad, Randi Reppen (1998): *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Booij, Geert; Christian Lehmann, Joachim Mugdan et al. (Hg.) (2000): *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung* (= HSK, Bd. 17.1). Berlin, New York: de Gruyter.
- Carstairs-McCarthy, Andrew (1998): Paradigmatic Structure: Inflectional Paradigms and Morphological Classes. In: A. Spencer/A. N. Zwicky (Hg.): 322-334.
- (2000): Lexeme, Word-form, Paradigm. In: G. Booij, C. Lehmann, J. Mugdan (Hg.): 595-606.
- Dansk Sprognævn (1980 [1955]): *Retskrivningsordbog*. Kopenhagen: Gyldendal.
- (1992 [1986]): *Retskrivningsordbogen*. Kopenhagen: Gyldendal.
- (2001): *Retskrivningsordbogen*. Kopenhagen: Dansk Sprognævn und Aschehoug.
- Fahl, Laurids; Ken Farø, Jan Heegaard, Henrik Lorentzen (2002): Review of: *Retskrivningsordbogen*, 3. udgave, Dansk Sprognævn og Aschehoug, 2001, *Retskrivningsordbog*, Gyldendal, 2001, *Politikens Retskrivningsordbog*, 1. udgave. In: *Danske Studier* 99: 222-237.
- Helbig, Gerhard und Joachim Buscha (1991): *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Berlin: Langenscheidt.
- Hjorth, Ebba und Kjeld Kristensen et al. (2003f.): *Den Danske Ordbog*. Kopenhagen: Gyldendal und DSL.
- Joseph, Brian D. (1998): Diachronic Morphology. In: A. Spencer/A. N. Zwicky (Hg.): 351-373.
- Karlsson, Fred (2000): Defectivity. In: G. Booij, C. Lehmann, J. Mugdan (Hg.): 647-653.
- Klosa, Annette; Kathrin Kunkel-Razum, Werner Scholze-Stubenrecht, Matthias Wernké (2001): *Duden Deutsches Universalwörterbuch*. Mannheim: Duden.
- McEnery, Tony; Andrew Wilson (1997) [1996]: *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mugdan, Joachim (1983): Grammatik im Wörterbuch: Flexion. In: *Germanistische Linguistik* 1-4/1982: 179-237.
- (1989): Information on Inflectional Morphology in the General Monolingual Dictionary. In: Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta (1989-1991): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter.
- Norling-Christensen, Ole and Jørg Asmussen (1998): The Corpus of the Danish Dictionary. In: *Lexikos* 8: 223-242.
- Spencer, Andrew and Arnold N. Zwicky (Hg.) (1998): *The Handbook of Morphology*. Oxford: Blackwell Publishers.
- Stump, Gregory T. (1998): Inflection. In: A. Spencer/A. N. Zwicky (Hg.): 13-43.
- Tarp, Sven (1995): Wörterbuchfunktionen: Utopische und realistische Vorschläge für die bilinguale Lexikographie. In: *Germanistische Linguistik* 127-128: 17-61.
- Trap-Jensen, Lars (2002): Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases. In: A. Braasch, C. Povlsen (Hg.): *Proceedings of the Tenth EURALEX International Congress*. Vol. II. Kopenhagen: CST: 503-508.
- Wahrig-Burfeind, Renate et al. (2000): *Wahrig Deutsches Wörterbuch*. Gütersloh/München: Bertelsmann Lexikonverlag.
- Wurzel, Wolfgang Ullrich (1984): *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademieverlag.

Franziskus Geeb

Perspektiven der Rechneranwendung in der Lexikographie: Lexikographische Datenmodellierung und Datenstrukturierung in relationalen Datenbanken

1. Vorüberlegungen

1.1. Perspektiven

Die Perspektiven der Rechneranwendung in der Lexikographie verstanden als die Facetten der möglichen Technologien sind vielfältig und Gegenstand umfassender Forschung (Lemberg, Schröder & Storrer 2001). Eine besondere Perspektive der Rechneranwendung ist die Nutzung der hiermit gegebenen Möglichkeiten als Modellierungswerkzeug.¹ So ist eine Modellierung und damit Begründung von lexikographischen Informationen gerade durch diese Werkzeuge denkbar und bereits vollzogen.² In der Regel werden in der lexikographischen Produktion jedoch neben Texteditoren der verschiedensten Art auf Grund ihrer weiten Verbreitung auch relationale Datenbanken eingesetzt. Diese Datenbanken ermöglichen durch den Entwurfsschritt Entity-Relation-Ship-Diagramm und teilweise durch die Schritte der Normalisierung entscheidende Vorüberlegungen zur Konzeption eines lexikographischen Nachschlagewerkes³. Die wesentliche Perspektive der Rechneranwendung in der Lexikographie, die im Folgenden diskutiert werden soll, ist damit die Nutzung des Rechners als Hilfe zur Modellierung und Modelldarstellung von lexikographischen Daten und Informationen.

1.2. Lexikographische Information

Auf der Grundlage einer linguistischen Zeichendefinition lässt über die lexikographischen Relevanzkriterien (z.B. die Bestimmung des Benutzers in der Benutzermatrix: Geeb 1998 39ff;) eine Fortschreibung von lexikographischen Daten in lexikographische Information vertreten (Geeb 2001, 27 ff, Geeb 2002). Die lexikographische Information entsteht damit erst und ausschließlich durch die dem Lexikographen bewusste oder auch unbewusste Bin-

¹ Datenmodellierung sei verstanden als der Entwurf von (lexikographischen) Dateneinheiten und ihren funktionalen Beziehungen ohne Rücksichtname auf bereits bestehendes Datenmaterial. Datenmodellierung (in Datenbanken und in Auszeichnungssprachen) ermöglicht dabei einen erheblich höheren Freiheitsgrad als Datenstrukturierung. Datenstrukturierung sei im folgenden verstanden als die Tätigkeit, bei der eine Struktur auf bestehendem (lexikographischen) Material aufgesetzt wird. Hauptkennzeichen der Datenstrukturierung ist damit die Analyse bestehenden Datenmaterials und dann die folgende Erstellung von strukturellen Einheiten, die in einer Struktur funktional verbunden werden.“ (Geeb 2001, 29)

² Für Auszeichnungssprachen (eXtensible Markup Language) z.B. Geeb 2001.

³ Geeb 1998, 19 ff; auch: Kunze & Wagner 2001, 229; Simonsen 2001.