ULRIKE HASS

Textkorpus und Belege. Methodologie und Methoden

1.	Zum Verhältnis von Praxis und Theorie	214
1.1.	Der Wörterbuchtyp	214
1.2.	Der methodische Rahmen	217
2.	Die Wörterbuchbasis	221
2.1.	Grundlagen der Funktionsbestimmung von Primärkorpora	221
2.1.1.	Die Grundgesamtheit	222
2.1.2.	Die elementaren Einheiten	223
2.1.3.	Die Relation zwischen Korpus und Grundgesamtheit	225
2.1.4.	Das wissenschaftliche Handlungsziel	227
2.2.	Die Funktionen lexikographischer Primärkorpora	227
2.2.1.	Korpus versus Sprachkompetenz des Linguisten	229
2.2.2.	Korpus und Verstehenskompetenz	231
2.3.	Organisationsformen von Primärkorpora	234
2.3.1.	Sample-Korpora	235
2.3.2.	Projektspezifische Korpora	235
2.3.3.	Monitor-Korpora	236
2.4.	Die Funktionen lexikographischer Sekundärkorpora	239
2.5.	Informationsprogramm und Korpuskonzept	241
2.5.1.	Ein Informationsprogramm für den Typ Kulturschatzwörterbuch	241
2.5.2.	Konzept eines Korpus für den Typ Kulturschatzwörterbuch	243
3.	Die Korpora des IDS als monitor-Korpus	245
3.1.	Die Teilkorpora. Kurzcharakterisierung	245
3.2.	Technische Organisation	247
4.	Verfahren der Korpusanalyse	249
4.1.	Computer-Recherchen	250
4.1.1.	Wortformen-/ Wortbildungsparadigmen	250
4.1.2.	KWIC-Belege	254
4.1.3.	Satzbelege	256

	213
4.2. Einbeziehung nicht-computerverfügbaren Materials	258
4.3. Analyse-Beispiel	259
4.4. Analyse-Ergebnisse im Hinblick auf die einzelne Informationsarten	264
5. Beispielbelege	272
5.1. Zu Funktion und Definition von Beispiel und Beleg	272
5.2. Auswahl der Beispielbelege	274
5.3. Relation zwischen Beispielbelegen und Informationsarten	276
5.4. Vom Korpusbeleg zum Beispielbeleg: Bearbeitungsverfahren	281
5.4.1. Der Belegblock	281
5.4.2. Belegschnitt	283
5.4.3. Belegergänzung und -kommentierung	284
6. Literatur	287

1. Zum Verhältnis von Praxis und Theorie

Der nachfolgende Beitrag versucht Antworten auf die Fragen zu geben, wie die in Wörterbüchern vermittelten linguistischen Aussagen zustandekommen und inwiefern diesen Aussagen intersubjektive Gültigkeit zugesprochen werden kann. Es ist insofern zu verstehen als Teil der „Rekonstruktion lexicographischer Praxis als einer [...] von [...] Vorannahmen über den Objektbereich geleiteten, nach bestimmten methodischen Prinzipien organisierten komplexen Handlung“ (Schader 1981, 13). Dabei geht es nicht nur um die Darstellung von Methoden einer einmaligen lexicographischen Praxis, nämlich der des Lexikons „Brisante Wörter“ (Strauß/Haß/Harras 1989), sondern auch in noch anzugebendem Maße verallgemeinert um die diese Praxis leitenden Prinzipien, deren Herleitung und Begründung. Zum Gegenstandsbereich dieses Beitrags gehört auch ein Teil der (lexicographischen) Methodologie, verstanden als „Verfahren der Wissensbildung“ und als „Verfahren der Sicherung der Geltung wissenschaftlicher Aussagen“ (Mittelstraß 1984, 2, 887). Das Wissen, das es mithilfe methodologischer Reflexion zu sichern gilt, ist – allgemein formuliert – das semantisch-pragmatische Wissen über ausgewählte Klassen lexikalischer Einheiten des Deutschen.

1.1. Der Wörterbuchtyp

Die im folgenden behandelten methodischen Prinzipien sollen soweit verallgemeinerungsfähig formuliert werden, daß sie auf alle Wörterbücher des Typs angewendet werden könnten, dem das Lexikon „Brisante Wörter“ als ein Exemplar zuzurechnen ist. Es handelt sich um einen bisher relativ selten realisierten Typ des einsprachigen Bedeutungswörterbuchs,¹ der durch die Kombination folgender, in Abweichung vom „Standardtyp“ des „allgemeine(n) einsprachige(n) Definitionswörterbuchs“ (Hausmann 1985, 372ff) formulierter Merkmale bestimmt werden kann:

- dominant synchron (ohne Ausschluß der Diachronie)
- ausschließlich gegenwartsbezogen (nicht historisch)
- dominant linguistisch (ohne Ausschluß philologischer Dokumentation)
- ausschließlich standardsprachlich (nicht regionalsprachlich)
- ausschließlich kontrastiv differenzbezogen (nicht gesamtwortschatzbezogen)
- dominant gemeinsprachlich (ohne Ausschluß der Fachsprachen)

¹ Bezeichnenderweise läßt sich der hier zu bestimmende Typ mithilfe der Wörterbuchtypologie von Kühn 1978, die ausschließlich auf der Klassifikation vorhandener Wörterbücher basiert, nicht hinreichend kennzeichnen.

- bildungssprachbezogen (vgl. Strauß/ Zifonun 1985, Bd. 1, 86ff)
- nicht idiolektbezogen
- auf markierte Wortschatzbereiche wie Fremdwörter, Neologismen, Schlagwörter vor allem des kulturspezifischen und politisch-ideologischen Wortschatzes, Jargonwörter, „schwere Wörter“ (im Sinne der *hard words*), fachextern gebrauchte Fachwörter bezogen.

Es handelt sich also um einen Wörterbuchtyp, dessen primäres Charakteristikum die Beschränkung auf einen in politischer, sozialer, kultureller und historischer Kommunikation relevanten Teilwortschatz ist, der zum großen Teil durch subsystemtranszendente Polysemien² ausgezeichnet ist und der deshalb nicht nur im Bereich der Standardvarietät, sondern varietätenübergreifend beschrieben wird. Außer diesen phänomenologischen sind noch funktionale, d.h. darstellungs- und adressatenbezogene Typmerkmale (Hausmann 1989, 970) anzugeben; diese sind für korpus- und belegmethodische Fragen unmittelbar relevant:

- überwiegend diskursiv (nicht abkürzend und elliptisch)
- textdokumentierend
- informierend und sprachreflexiv für erwachsene L1-Sprecher mit relativ hoher Allgemeinbildung
- nicht für L1- oder L2-Lernende, nicht ausschließlich für Fachwissenschaftler.

Außer dem Lexikon „Brisante Wörter“ könnten als weitere (zum Teil erst noch zu realisierende) Exemplare desselben Typs gelten: Europäismen-, Internationalismen- und Fremdwörterbücher, bildungssprachliche und interdisziplinär-wissenschaftssprachliche Wörterbücher (soweit sie diskursiv und textdokumentativ sind). Alle realisierten oder denkbaren Exemplare dieses Typs werden in diesem Beitrag abkürzend als *Kultu r w o r t s c h a t z* wörterbücher bezeichnet, wobei *Kultur* hier im weitesten Sinn zu verstehen ist.

Das dem speziellen Typmerkmal 'kontrastiv auf markierte Wortschatzteile bezogen' übergeordnete Typmerkmal 'teilwortschatzbezogen' ist das in korpusmethodologischer Hinsicht relevanteste Merkmal. Viele der bisherigen Überlegungen zu Status, Funktion und Nutzen lexikographischer Korpora³ – sie wurden bisher überwiegend von praktisch arbeitenden an-

² Terminus von Strauß/Zifonun 1985, Bd. 2, 330ff.

³ Stellvertretend seien genannt: Rey-Debove 1971, Pfeffer 1979, Hartmann 1983, Landau 1984 (vgl. besonders die Liste der Wörterbücher S.X), Renouf 1984, Ilson 1985.

gelsächsischen und französischen Lexikographinnen und Lexikographen⁴ angestellt – sind gesamtwortschatzbezogen, d.h. sie beziehen sich mehr oder weniger ausdrücklich auf die möglichst vollständige Beschreibung des standard- bzw. gemeinsprachlichen Kernwortschatzes eventuell zuzüglich direkt daran anschließender („konzentrischer“) Varietätenwortschätze.⁵ Damit beziehen sie sich auf einen ganz anderen Wörterbuchtyp als den hier zu behandelnden, nämlich auf den Typ des Gebrauchs- und Lernerwörterbuchs. Dieser Typ ist allenfalls insofern teilwortschatzbezogen zu nennen, als er sich auf den zentralen Wortschatzteil einer Einzelsprache bezieht. Im Unterschied dazu kann der hier zu behandelnde Wörterbuchtyp als sektoral teilwortschatzbezogen gekennzeichnet werden. Dieser Unterschied führt hinsichtlich einer ganzen Reihe von methodologischen, das Korpus und die Darstellung von Beispielbelegen betreffenden Fragen für Kulturwortschatzwörterbücher zu grundsätzlich anderen Antworten, als sie die Lexikographen von Gebrauchs- und Lernerwörterbüchern mit einem mehr oder weniger explizit erhobenen Anspruch Allgemeingültigkeit formuliert haben.

Worin die Relevanz der verschiedenen typspezifischen Merkmale für die methodischen Entscheidungen hinsichtlich Korpus und Belegen im einzelnen besteht, läßt sich nur im Rahmen einer das lexikographische Handeln insgesamt begründenden Methodik bestimmen. Von den in der allgemeinen wissenschaftstheoretischen Methodologie unterschiedenen Methodiken kommt, wie Schaefer begründet, für die Lexikographie nur die als „primär-induktiv“ oder „empirisch-deskriptiv“ gekennzeichnete infrage (Schaefer 1981, 45ff). Damit haben lexikographische Korpora von vornherein einen ganz anderen Status und andere Funktionen als Korpora für grammatische bzw. grammatikographische Forschungen, die vor allem auf „primär-deduktiven“ Methoden basieren.⁶ Bei der primär-induktiven Methodik handelt es sich um „eine in ihrem Aufbau geordnete und als Anleitung für das praktische Vorgehen geeignete Gesamtheit begründet aufeinander bezogener Methoden“ (Schaefer 1981, 46), die arbeitstechnisch in vier Teilschritte untergliedert werden können.

⁴ Schätzungsweise gibt es in diesem Beruf inzwischen mehr Frauen als Männer. Zur Vermeidung doppelter Bezeichnungen wird im Folgenden die Form *Lexikograph(en)* als sexusneutrale Form verwendet.

⁵ Z.B. Pfeffer 1979, 308, Landau 1984, 279ff.

⁶ Dieser grundlegende Unterschied wurde bzw. wird in der Diskussion um korpus- versus kompetenzgestütztes Vorgehen z.T. ignoriert. Siehe unten Abschnitt 2.2.1.

1.2. Der methodische Rahmen

Der folgende Abschnitt soll – immer mit Bezug auf den genannten Wörterbuchtyp – Antworten auf die Frage nach den Vorbedingungen geben, von denen die lexikographischen Verfahren der Korpusorganisation (Auswahl, Zusammenstellung), der Korpusanalyse und der Behandlung der Beispielbelege (Auswahl, Anordnung, Schnitt) ebenso abhängen wie die übrigen Verfahren der Bedeutungserläuterung oder des Artikelaufbaus; auf die zuletzt genannten Verfahren soll hier aber nicht näher eingegangen werden.⁷

Zur Orientierung lassen sich die von Schaefer (1981) dargelegten vier methodischen Teilschritte heranziehen.⁸ Sie sollen im Folgenden konkretisiert und schwerpunktmäßig auf ihre Relevanz für die Aspekte des zu behandelnden Themas 'Korpus und Belege' geprüft werden.

Der erste Teilschritt der primär-induktiven Methode ist die Explizierung der Vorannahmen über den Gegenstandsbereich Sprache im allgemeinen und das einzelne Forschungsobjekt im besonderen. Sie lassen sich für den genannten Wörterbuchtyp folgendermaßen spezifizieren: Die Vorannahmen über den Gegenstand Sprache sind pragmalinguistischer, handlungssemantischer Art.⁹ Gemäß dieser Auffassung sind lexikalische Ausdrücke und zwar besonders die vorthoretisch so genannten Schlag- und Schlüsselwörter, Grund- und Leitbegriffe wichtige Faktoren im Vollzug der gesellschaftlichen Kommunikation. Sie gehören zum mehr oder weniger bewußt gehandhabten Instrumentarium von Sprechern, die mittels solcher Ausdrücke Weltbilder und Meinungen konstituieren, signalisieren und reproduzieren. Deshalb sind bei der Formulierung von Verwendungsregeln im Wortartikel Sprecher bzw. Sprechergruppen in der Agensrolle wiederzugeben.

Das – zunächst – lexikologisch-semantische Erkenntnisinteresse, auf das hin ein Korpus organisiert und interpretiert wird, konzentriert sich beim Lexikon „Brisante Wörter“ als einem Exemplar des genannten Wörterbuchtyps auf solche Wörter,

⁷ Vgl. die Beiträge von G. Strauß und G. Harras in diesem Band.

⁸ Wiegand (1983) – in diesem methodischen Teilbereich weniger stark differenzierend als Schaefer – nennt zwei Schritte: erstens die Datenerhebung („Dies ist eine Theorie darüber, wie eine Wörterbuchbasis zu erstellen ist, d.h.: es geht erstens um die Gewinnung, den Aufbau, die Repräsentativität und die Funktion von lexikographischen Korpora relativ zu Wörterbuchtypen. [...] Zweitens muß die Rolle der Sekundärquellen in der Wörterbucharbeit geklärt werden.“ (Wiegand 1983, 102)), und zweitens die Datenverarbeitung.

⁹ Natürlich ist es möglich, daß Wörterbücher des genannten Typs einer anderen Semantikauffassung verpflichtet sind; es scheint allerdings besonders plausibel, soziokulturell relevante Wortschatzbereiche in einer stärker die Sprecheraktivität als etwa die Lexikonsystematik betonenden Weise zu beschreiben.

„deren Gebrauchsweisen die sprachliche Verständigung erschweren oder sogar stören können, also Wörter, die unter verschiedenen Gesichtspunkten erklärungsbedürftig sind. Wichtige Ursachen für eine solche Erschwernis bzw. Störung können v.a. sein:

- unterschiedliche Wertsetzungen, die mit einem Wort zum Ausdruck gebracht werden können, z.B. *Konservatismus*, *reaktionär*;
- unterschiedliche Objekte, die mit einem Wort bezeichnet werden können, z.B. politische Systeme mit den Wörtern *Faschismus*, *Totalitarismus*;
- beschönigende oder übertragene, bildliche Gebrauchsweisen, die typischerweise mit einem Wort verknüpft werden, z.B. mit *Entsorgung*(*spark*);
- die Vagheit von Mode- und Schlagwörtern wie z.B. *Postmoderne*, *Zeitgeist*, *Konservatismus* und *Radikalismus*;
- die Verwobenheit eines Ausdrucks mit einer fachsprachlichen Nomenklatur bei gleichzeitig davon verschiedenen, zum Teil wertenden, Gebrauchsweisen in der öffentlichen Kommunikation, z.B. *Atom-/Kernenergie*, *Totalitarismus*, *Pluralismus*;
- der mit dem Gebrauch vieler bildungssprachlicher Ausdrücke verknüpfte Prestigeanspruch, z.B. *subtil*, *kulinarisch*, *elitär*;
- unterschiedliche Arten von Welt- und Kontextwissen, die zum Verständnis vieler Begriffe erforderlich sind, z.B. *Anarchismus*, *Restrisiko*, *Manierismus*.“ (Strauß/Haß/Harras 1989, 9)

Im Hinblick auf den größeren Allgemeinheitsgrad, den die Formulierung des gemeinsamen Gegenstands aller Exemplare des Typs Kulturwortschatzwörterbuch aufzuweisen hätte, kann die Liste der Wortschatzcharakteristika unabhängig von Schwerverständlichkeit, ideologischer Mehrdeutigkeit u.ä. um alle politisch, sozial und historisch relevanten¹⁰ Aspekte erweitert werden. Unter relevant fällt dann auch so etwas wie die (meist negativ wertende) Charakterisierung entlehnter lexikalischer Einheiten als „Fremd“-Wörter. Das zusammenzustellende Korpus soll die Lexikographen in die Lage versetzen, die kommunikative Rolle solcher, einen Teilwortschatz mit Systemcharakter bildenden Wörter zu beschreiben und zu

¹⁰ Koselleck verdeutlicht die Relevanz der historischen Aspekte je aktueller Verwendungsweisen mit Hilfe von Metaphern: es sind die „diachron verschieden tief gestaffelten Vorgaben“, die – in synchroner Perspektive – den „Argumentationshaushalt ebenso sehr erschließen wie begrenzen, bzw. eine „diachrone[n] Schubkraft [...] über die die Sprecher nicht willkürlich, keinesfalls souverän verfügen“ (Koselleck 1989, 666).

erklären. Auf welche Weise dieses Ziel die Zusammenstellung und Organisation eines Korpus leitet, wird in Abschnitt 2 dargestellt.

Der zweite Teilschritt der primär-induktiven Methode, die Beschreibung, umfaßt nach Schaefer (1981, 48) mehrere Teilaktivitäten: Beobachtung, innersprachliches Vergleichen, Identifizieren, Delimitation oder Segmentierung sowie Klassifikation. Sie gewährleisten die projektspezifisch erneute Reflexion von z.T. altbekannten lexikographischen Problemen. Z.B. wird das Problem der Identifizierbarkeit mehrerer Wortverwendungen (Aktualisierungen) als einunddieselbe Lexikoneinheit, das sich bei der Lemmatisierung stellt, traditionell und pragmatisch-praktisch einerseits über die Gleichheit der Ausdrucksseite der nichtflektierten Form gelöst, andererseits mittels struktureller Verfahren zur Abgrenzung von Homonymie und Polysemie. Für den hier interessierenden Wörterbuchtyp können solche Probleme als eher peripher vernachlässigt werden.

Ein anderes Problem stellt sich jedoch unter den Prämissen einer Handlungssemantik in neuer Weise: Gegenstand jeder einzelnen Beschreibungshandlung ist nicht ein Wort mit meist mehreren Bedeutungen, die man sich aus Komponenten aufgebaut vorstellt, sondern die Weisen der Aktualisierung eines Worts durch sozial, historisch usw. bestimmte Sprecher. Mithin muß die Frage, die in der Terminologie der strukturellen lexikalischen Semantik die Frage nach der „Gliederung des semasiologischen Feldes“ bzw. nach Zahl und Anordnung der Einzelbedeutungen genannt wird, zweifach neu formuliert werden. Einmal für die Ebene der lexikologischen Beschreibung: Worin besteht die Identifizierbarkeit einundderselben Verwendungsweise (d.h. die Identifizierbarkeit des Aktualisierungstyps) eines Worts bzw. nach welchen Kriterien wird die Menge der Verwendungen bestimmten Verwendungsklassen zugeordnet? Zum zweiten für die Ebene der lexikographisch darstellenden Beschreibung: Welche Art und welcher Abstufungsgrad der Klassifizierung hat für die Wörterbuchadressaten hohen Erklärungswert, weil fein differenziert, und hohe Plausibilität, weil auch für Nicht-Linguisten nachvollziehbar? Oder salopp formuliert: Welche Artikelgliederung ist sprachtheoretisch konsequent und gleichzeitig weder banal noch haarspalterisch?

Die Antworten auf diese Fragen gehören nicht unmittelbar zum Thema dieses Beitrags, sondern eher zur Methodologie der semantischen Analyse und des Artikelaufbaus. Dennoch ist festzuhalten, daß die Fragen nach dem, was mittels des zusammenzustellenden Korpus eigentlich genau beschrieben werden soll, schon die Entscheidungen bei der Auswahl der Korpus-texte und erst recht die Verfahren der Korpusanalyse beeinflussen. Ein häufiges und wesentliches Klassifikationskriterium von Verwendungsweisen eines Ausdrucks ist z.B. die fachspezifische durch Fachexperten im Unterschied zur nichtfachlichen Verwendungsweise durch Nicht-Experten. Sollen beide Verwendungsweisen und eventuell auch ihr kommunikativer Zusam-

menhang erst beobachtet und dann beschrieben werden, so müssen in das Korpus hinreichend viele Texte von Experten und von Nicht-Experten über das Thema, in dessen Zusammenhang der fragliche Ausdruck steht, aufgenommen werden. Hingegen braucht z.B. auf ein areales Spektrum der Korpus-texte nicht geachtet werden, wenn areale Verwendungsunterschiede nicht zum definierten Gegenstandsbereich des Wörterbuchs zählen.¹¹

Als dritten Teilschritt enthält Schaeders Modell der primär-induktiven Methode (hier in enger Orientierung an Toulmin) die Argumentation, die vom Singulären (Daten aus dem Korpus) zum Verallgemeinerten (lexikographische Aussagen) führt.¹² Das Wesentliche an diesem Schritt ist, daß er die Möglichkeit intersubjektiver Nachweisbarkeit bietet. Nicht nur durch die Datenbasis (das Korpus) selbst, aus der in einem Abstraktionsschritt eine (qualifizierte) Behauptung aufgestellt wird, sondern auch durch die ebenfalls empirischen Schritte 'Rechtfertigung' und 'Stützung' wird das gesamte Verfahren mehrfach an ein Korpus gebunden. Zwar schützt dieses Modell m.E. nicht hinreichend vor einer (zirkulären) Vermischung von Behauptung, Rechtfertigung und Stützung, jedoch kann für die Kontrolliertheit des Argumentationsverfahrens als Ganzem festgehalten werden, daß durch Abstraktion gebildete Behauptungen gleich welcher Art nicht auf einem einzelnen sprachlichen Datum basieren dürfen, sondern auf mehreren, empirisch möglichst eindeutig voneinander unterscheidbaren Daten basieren müssen. Nur so kann der jeder induktiven Methode innewohnenden Tendenz zu mangelnder Allgemeingültigkeit der Ergebnisse entgegengewirkt werden. Diese sich methodologisch ergebende Forderung geht als Prämisse in die in Abschnitt 4 beschriebenen konkreten Verfahren der Korpusanalyse ein.

Im letzten Teilschritt der primär-induktiven Methode nach Schaefer muß projektspezifisch beschrieben und begründet werden, welche Schritte von der Analyse der Daten zu „Beschreibung und Erklärung“ (Schaefer 1981, 52ff) im Wortartikel führen. Einerseits ist die Unterscheidung zwischen „Beschreibung“ und „Erklärung“ bloß hypothetisch. Die Explizierung der theoretischen Vorannahmen im ersten Teilschritt hat bereits deutlich gemacht, daß die Beschreibung von Wortverwendungen immer schon auf die einer bestimmten Semantikauffassung verpflichteten Erklärung hin

¹¹ Im Lexikon „Brisante Wörter“ wird gelegentlich auf DDR-spezifischen Wortgebrauch hingewiesen; dies geschieht allerdings nicht unter arealem, sondern unter ideologischem Aspekt.

¹² Das von Schaefer (1981, 50f) referierte und von Toulmin (1969) entwickelte Schema enthält die Hauptschritte: „1. Daten (D); 2. Behauptung (B), die unter Umständen auch Konklusion (K) heißt; 3. Rechtfertigung (R); 4. Stützung (S) als weitere Untermauerung; 5. (modale) Qualifizierung (Q); 6. Ausnahmeforderungen (A)“ und hat die Form: „D – weil R – aufgrund von S – deshalb Q, B (K) – außer wenn (A)“.

angelegt ist. Andererseits verweisen die Unterschiede in den Alltagssprachlichen Konzepten von *Beschreibung* und *Erklärung* auf eine Differenz, die auch im lexikographischen Verfahren zum Tragen kommt: Zwar vermittelt schon die linguistische Beschreibung dessen, „was der Fall ist“, dem Wörterbuchleser mehr, als er sprachbetrachtend selbst zu erkennen in der Lage ist, aber eine Erklärung dessen, „was der Fall ist“, vermittelt noch darüber hinaus etwas: Sie stellt Zusammenhänge her zwischen den beschriebenen Phänomenen und verweist vor allem auf Ursachen sprachlicher Phänomene im außersprachlichen Kontext, d.h. im gesellschaftlichen, politischen, historischen Rahmen. Eine über die Beschreibung hinausgehende Erklärungsintention ist also gerade für solche Wörterbücher konstitutiv, die kulturell relevanten Wortschatz darstellen, sich dabei auf eine handlungsemantisch-pragmatische Sprachauffassung beziehen und sie lexikographisch konsequent umsetzen wollen. In Abschnitt 5 dieses Beitrags soll im einzelnen dargelegt und begründet werden, welcher Anteil den Beispielbelegen eines Wortartikels an Beschreibung und Erklärung zugewiesen wird und welche methodischen Konsequenzen sich daraus für die Bearbeitung von Korpussegmenten zu Beispielbelegen (Auswahl, Anordnung, Schnitt, Ergänzungen) ergeben.

Mit der wörterbuchtypspezifischen Beschreibung der vier Teilschritte der primär-induktiven lexikographischen Methode ist nun der methodologische Rahmen angegeben, innerhalb dessen die nachfolgenden Überlegungen zu Status, Funktion und Organisation des Wörterbuchkorpus als auch zur Analyse- und Darstellungspraxis angestellt werden.

2. Die Wörterbuchbasis

Die Datenbasis eines Wörterbuchs setzt sich zusammen aus einem Text- und/oder Belegkorpus (siehe 2.1 bis 2.3) und einem Korpus aus Sekundärliteratur (siehe 2.4).

2.1. Grundlagen der Funktionsbestimmung von Text- und/oder Belegkorpora

Die Definitionen bzw. Funktionsbestimmungen linguistischer Korpora sind kontrovers. Ungeachtet einer eher lexikographischen oder einer eher grammatikographischen Ausrichtung waren und sind sie eng verknüpft mit der Empiriediskussion im Anschluß an die generative Grammatik.¹³ Den unterschiedlichsten Funktionsbestimmungen von Korpora in der lexikographischen Literatur ist eine Reihe von Parametern gemeinsam, die für jede (projektspezifische) Korpusdefinition konkretisiert werden müssen:

¹³ Vgl. Schaefer 1981, 70f, der verschiedene Textkorpus-Definitionen diskutiert.

- (i) Grundgesamtheit
- (ii) elementare Einheiten
- (iii) Relation zwischen Korpus und Grundgesamtheit
- (iv) wissenschaftliches Handlungsziel¹⁴

2.1.1. Die Grundgesamtheit

Die Grundgesamtheit bildet den Forschungsgegenstand, über den mit Hilfe des Korpus Erkenntnisse gewonnen bzw. Aussagen gemacht werden sollen, „z.B. die Menge aller Aktualisierungen deutscher Sprache“ (Schaefer 1981, 70). Für ein englisches Wörterbuch wird diese Grundgesamtheit beispielsweise mit „given language“ angegeben (nach Schaefer 1981, 72); das Quellenkorpus des Frühneuhochdeutschen Wörterbuchs geht von der „Grundgesamtheit von Sprachhandlungen frühneuhochdeutscher Zeit“ aus (Reichmann 1986, 59).

Nach der Korpusdefinition von Bungarten (1979, 34) ist die Grundgesamtheit die „natürliche(n) Sprache L1“ bzw. derjenige „Sprach- und Kommunikationsbereich“, der durch das Korpus „repräsentiert“ wird. Diese Definition ist, vor allem wenn sie auf Korpora des hier infrage stehenden (teilwortschatzbezogenen) Typs des Kulturwortschatzwörterbuchs angewandt wird, zirkulär, denn der Parameter (i) 'Grundgesamtheit' wird als Funktion des Parameters (iii) 'Relation zwischen Korpus und Grundgesamtheit' bestimmt. Damit besteht die Gefahr, den Forschungs- bzw. Wörterbuchgegenstand als den „gemeinsamen Nenner“ aller Texte irgendeiner vorkorporalen Sammlung zu beschreiben, dessen Zusammenstellung unbekannt oder nicht wissenschaftlich begründet ist und zumindest nicht in einem methodologischen Zusammenhang mit dem gesetzten Forschungsziel steht. Dies könnte im Extremfall bedeuten, daß der Forschungs- bzw. Wörterbuchgegenstand in Abhängigkeit von der computertechnischen Verfügbarkeit irgendwelcher Texte gerät.

Die jeweilige Grundgesamtheit muß definiert sein (Schaefer 1981, 72), damit die Verfahren der Korpuserstellung als begründet gelten können. Eine Grundgesamtheit läßt sich allerdings nur in Form einer Abstraktion beschreiben, mit der alle (realisierten und möglichen) Sprachhandlungen eines einzelsprachlich, regional, historisch, sozial und medial definierten Sprach- und Kommunikationsbereichs mit dem Ausdruck *Sprach- und Kommunikationsbereich X* zusammengefaßt werden. Hingegen läßt

¹⁴ Der Ausdruck *wissenschaftliches Handlungsziel* (anstelle von *Forschungsziel*) wird gewählt, um die von vorneherein spezielle Ausrichtung korpusmethodologischer Überlegungen auf eine konkrete wissenschaftliche Textsorte, z.B. Wörterbücher, deutlich zu machen; relevant ist dies vor allem für die Funktion des Korpus, Beispielbelege für die Ziel-Textsorte Wörterbuch zu „liefern“.

sich das die Grundgesamtheit repräsentierende¹⁵ Korpus auf einer niedrigeren, „linguistisch greifbareren“ Abstraktionsstufe, etwa als eine Menge von Sätzen oder Texten, beschreiben.

2.1.2. Die elementaren Einheiten

Die das Korpus konstituierenden elementaren Einheiten sind die Menge der Sprachdaten, für die im einzelnen festzulegen ist, ob es sich um (medial fixierte) sprechsprachliche oder um schriftsprachliche Daten handelt. Je nach zugrundegelegtem Sprachzeichenbegriff können die elementaren Korpuseinheiten aufgefaßt werden als: Wortformen (tokens), objektsprachliche Äußerungen, Sätze, „Kommunikationsakte“ (Bungarten 1979, 34) oder als Texte, hier verstanden als komplexeste Art von Sprachzeichen, die die Sprachteilhaber eines Kommunikationsbereichs intentional verwenden.

Die Bestimmung dessen, was als elementare Korpuseinheit gilt, begründet den Unterschied zwischen Beleg- und Textkorpus. Die Einheit des „Belegs“ umfaßt in der Regel mindestens eine Satzgliedphrase, meist einen Satz, mit einer nicht festgelegten Menge an Kontext und an Kontextinformationen. Die Herstellung eines Belegs und die Organisation der Belegsammlung orientiert sich an nicht mehr als einem Lemmakandidaten pro Beleg. Lemmatisierungsentscheidungen fallen beim Belegkorpus also viel früher als bei einem Textkorpus. Ebenso verhält es sich mit der Formierung der späteren im Wortartikel anzuführenden Beispielbelege: Während in einem Belegkorpus hier bereits Vorentscheidungen über Art und Umfang des Kontextes getroffen werden, die später kaum noch revidiert werden können, bleibt die Entscheidungsfreiheit hinsichtlich der Belegbearbeitung (siehe 5.4) bei einem Textkorpus unbeeinträchtigt.

Andererseits bekommen bei einem Textkorpus die Methoden der Selektion der elementaren Einheiten, d.h. der Textzusammenstellung größeres Gewicht als es das – kaum systematisch selektiv verfahrenende – Sammeln von Belegen besitzt. Beim Textkorpus ist Selektion und Organisation der elementaren Einheiten der „point of no return“, von dem die Qualität aller nachfolgenden lexikographischen Arbeitsschritte entscheidend abhängt.

Ein Korpus bildet in einer noch näher zu bestimmenden, „exemplarischen“ Weise diejenigen sprachlichen Phänomene ab, deren Erforschung – und vorgängig deren nähere Bestimmung – durch das wissenschaftliche Handlungsziel vorgegeben ist. Textkorpora unterscheiden sich in diesem Punkt nicht prinzipiell – wie Bungarten (1979, 35) meint – von Belegsammlungen.¹⁶ Auch die Datenbasis eines Textkorpus ist bereits unumkehrbar

¹⁵ Zur „Repräsentation“ siehe 2.1.3.

¹⁶ „Die Belegsammlung orientiert sich in der Suche nach Datenmaterial an einem morphologisch, syntaktisch oder semantisch genau definierten, konkreten

durch mehr oder weniger eng gefaßte und spezifische Fragestellungen vorbestimmt. Der tatsächliche und wesentliche Unterschied zwischen Textkorpora und Belegsammlungen liegt m.E. in der Segmentierung der elementaren Einheiten (die bei Belegsammlungen traditionell eher weit unterhalb des Textranges vorgenommen wird) und der bei Textkorpora, vor allem wenn sie maschinell verfügbar sind, weitgehend uneingeschränkten Möglichkeit, den transphrastisch-textuellen Zusammenhang lexikalischer und syntaktischer Phänomene und komplexere Einheiten (Texte) in die Reihe der Erkenntnisziele mit aufzunehmen.

Für die elementaren Einheiten eines Korpus gilt die Bedingung, daß sie „nicht manipuliert“ (Bausch 1979, 78) oder „vorfabriziert“ (Schaefer 1981, 73) sein dürfen, d.h. sie müssen eine Teilmenge aller die Grundgesamtheit konstituierenden Äußerungen sein und müssen zum jeweiligen Kommunikationsbereich gehören, in dem in aller Regel objektsprachliche Äußerungen und metasprachliche Äußerungen von Linguisten, wie z.B. Grammatikalitätsurteile oder semantische Regelformulierungen, *n i c h t* ununterscheidbar zusammengefaßt und vermischt werden. Sprachkommentierende (metasprachliche) Äußerungen können als Grundgesamtheit eines gesondert anzulegenden (Sekundär-)Korpus bestimmt werden (siehe unten 2.4).

Die Festlegung der Menge der elementaren Einheiten, d.h. des Korpusumfangs ist abhängig vom wissenschaftlichen Handlungsziel, speziell von der Sprachebene, der die Untersuchungsgegenstände angehören. Gegenstände auf den unteren Sprachebenen, z.B. graphematische oder morphematische, kommen in einer konstanten Textmenge ungleich häufiger vor als lexematische. Während für die ersteren der maximale Korpusumfang berechnet werden kann (vgl. Hoffmann 1984, 676), ist dies für die Untersuchung z.B. (varietätenübergreifender) lexikalischer Bedeutungen oder gar für die Untersuchung kommunikativer Strategien mittels lexikalischer Einheiten und für alle „metasprachlichen Daten“ (Hoffmann 1984) nicht möglich, d.h. es ist nicht möglich anzugeben, von welchem Umfang an das Korpus nur noch zu rekurrenten und redundanten Ergebnissen bezüglich dieser für den Typ des Kulturwortschatzwörterbuchs zentralen Untersuchungsgegenstände führen würde und damit groß genug sei. Daraus folgt, daß für einige Wörterbuchtypen, nämlich für die, deren Beschreibung auf die höheren und komplexeren Sprachebenen zielt, der Umfang des Korpus nicht fixiert werden muß (vgl. unten 2.3.3).

Sprachphänomen, für das in der Performanz Exempla gesucht werden. Eine Belegsammlung kann daher immer nur für einige wenige, begrenzte Fragestellungen als Datenbasis fungieren. Das Korpus demgegenüber wird nicht nach einer isolierten sprachlichen Erscheinung strukturiert, sondern entsprechend der Intention, umfassende Sprach- und Kommunikationsbereiche vollständig oder ausschnittsweise durch eine repräsentative oder exemplarische Stichprobe abzubilden.“ (Bungarten 1979, 35).

2.1.3. Die Relation zwischen Korpus und Grundgesamtheit

Für die Relation zwischen Korpus und Grundgesamtheit gilt zunächst, wie in der Statistik, die Forderung nach Adäquatheit der „Stichprobe“ inbezug auf die „Grundgesamtheit“, genauer: die Verfahren der Erstellung und Organisation des Korpus sollen gewährleisten, daß es „ein adäquates Modell der Grundgesamtheit darstellt“, aus der es stammt (Schneider 1981, 70).

Adäquates Modell der Grundgesamtheit kann ein Korpus auf verschiedene Weise sein. Die bisher metalexikographisch meistdiskutierte Art der Relation ist die der *R e p r ä s e n t a t i v i t ä t*; der Repräsentativitätsbegriff gilt im Anschluß an Rieger jedoch als für linguistische Korpora unangemessen.¹⁷ Sieht man von der grundsätzlich möglichen Relation der *I d e n t i t ä t* ab, die z.B. bei Korpora zu Textwörterbüchern¹⁸ ganz und bei Korpora zu Autorenwörterbüchern¹⁹ fast ganz erreicht wird und die letztlich in der bloßen Umorganisation, z.B. Alphabetisierung des gesamten Wortschatzes, einer als Grundgesamtheit gesetzten Sprachdatensmenge besteht, so bleibt nur die Relation der im Einzelfall mehr oder weniger konkret bestimmten *E x e m p l a r i z i t ä t*.

Welche der möglichen Arten exemplarischer Repräsentation können herangezogen werden, um das Verhältnis zwischen Korpus und Grundgesamtheit zu bestimmen? Davon ausgehend, daß exemplarische Repräsentationen – oder in anderer Formulierung: daß paradigmatische Denkmuster – zu den grundlegenden universal-anthropologischen Erkenntnisverfahren gehören, lassen sich unter Zuhilfenahme geistesgeschichtlicher und philosophischer Traditionen nach Scheuerl (1969, 39-86) (mindestens) sieben Typen exemplarischer Repräsentation unterscheiden.²⁰ Zwei von ihnen, *Exemplar* und *Pars pro toto*, sollen hier als typologische Aspekte (denn sie schließen einander nicht, wie Typen im strikten Sinne, aus) herangezogen werden, um 'Exemplarizität' von Korpora näher zu bestimmen.

Zunächst ist sicherzustellen, daß die Ausdrücke *Exemplarizität*, *exemplarisch* inbezug auf Korpora als semantisch kongruent mit *Exemplar*, nicht

¹⁷ Die Diskussion zum Problem der Repräsentativität eines Textkorpus einschließlich der Argumentation Riegers referiert Schneider (1981, 74ff).

¹⁸ Vgl. z.B. die in Amsterdam erscheinende Reihe: *Indices verborum* zum alt-deutschen Schrifttum.

¹⁹ Z.B. Goethes sämtliche schriftlichen Texte als Grundgesamtheit eines Belegkorpus oder einer exhaustiven Wortkonkordanz.

²⁰ Die folgenden paradigmatischen Formen: *Paradigma*, *Exemplar/Exempel*, *Typus*, *reiner Fall*, *Muster/Modell/Gleichnis*, *Pars pro toto* und *Analogie* bezeichnet Scheuerl (1969, 86) ausdrücklich als nicht abgeschlossene Reihe.

mit *Exempel* im Sinne von 'normativer, vorbildlicher Fall',²¹ verstanden werden, obwohl in der Redeweise von „exemplarischen Wortartikeln“ und „Exemplarizität des Lexikons“²² unüberhörbar jenes „Pathos der Wesentlichkeit“ mitschwingt, „das dem Herausgegriffenen eine besondere Würde verleiht“ (Scheuerl 1969, 84). Exemplarisch ist ein Korpus insofern, als es eine aus der Grundgesamtheit aller Äußerungen herausgehobene Teilmenge von Äußerungen enthält, wobei eine strenge Merkmalsidentität beider Mengen vorausgesetzt wird (Scheuerl 1969, 48). Von dieser Voraussetzung hängt die Verallgemeinerbarkeit der linguistischen Aussagen ab, die zwar allein auf der Basis eines - begrenzten - Korpus gewonnen werden, aber für die Grundgesamtheit allgemein gültig sein sollen. Ein Korpus ist inbezug auf die Grundgesamtheit also nicht Typus, Modell oder Analogie, denn bei diesen Repräsentationstypen konstituieren nur bestimmte, wesentliche und nicht sämtliche Merkmale die paradigmatische Relation.

Die in einem Korpus enthaltenen Äußerungen sind in ihrer Gesamtheit ein Besonderes, das das Allgemeine, d.h. die Grundgesamtheit, aus dem es herausgehoben worden ist, vertritt, ohne daß jenes Besondere dabei ins Gewicht fällt. Es wäre durchaus möglich und methodologisch unerheblich, weitere, in Bezug auf die Grundgesamtheit gleichwertige Teilmengen (Korpora) zu bilden, weil das distinktiv Spezifische der Sprachdaten eines konkreten Korpus nicht zum Forschungsgegenstand gehört.²³ Die Exemplarizität des Korpus „liegt also darin, daß man gar nicht eigentlich auf es selber blickt, sondern auf einen Komplex von Merkmalen, die wandelbar und vertauschbar sind [...]“ (Scheuerl 1969, 47).

Als herausgehobene Teilmenge der Grundgesamtheit läßt sich das Korpus auch als der paradigmatische Typ des *Pars pro toto* auffassen. „Der Teil vertritt das Ganze, insofern er die gleichen Merkmale trägt wie alle übrigen Teile.“ (Scheuerl 1969, 69). Darüberhinaus enthält die Vorstellung vom Korpus als *Pars pro toto* ein dynamisches Moment: die Ergänzung des Teils zum Ganzen. Unter diesem Aspekt kann das Korpus als prinzipiell zur Grundgesamtheit hin auffüllbare Teilmenge von Sprachdaten aufgefaßt werden. „Das Ganze, auf das die Ergänzung zu zielen hat, ist eine Gesamtheit, ein durch Einfügen immer weiterer Summanden zustande kommendes Aggregat.“ (Scheuerl 1969, 69). Für die Annahme dieses dynamischen Moments als konstitutivem Merkmal linguistischer Korpora spricht

²¹ Zu Herausbildung und Relativierung des normativen Aspekts von *Exempel*, *exemplarisch* in bestimmten historischen Kontexten siehe Buck 1971.

²² Die Exemplarizität des Lexikons „Brisante Wörter“ wird in der Einleitung bestimmt als „selektiv und zugleich systematisch-erschöpfend“ (Strauß/Haß/Harras 1989, 10).

²³ Mit der schon erwähnten Ausnahme von Textwörterbüchern, für die aber auch nicht Exemplarizität, sondern – zumindest – quantitative Identität besteht.

ihr Bezug auf das wissenschaftliche Handlungsziel, das zumindest im Falle lexikographischer und auch grammatikographischer Endprodukte die Subsummierung weiterer, außerhalb des Korpus realisierter oder möglicher Äußerungen desselben Kommunikationsbereichs, d.h. die weitergehende „Anwendung“ der aufgrund und anhand des Korpus erarbeiteten „Regeln“ durch z.B. Wörterbuch-Adressaten nicht nur zuläßt, sondern geradezu intendiert.

Zusammengefaßt: Die Relation zwischen Korpus und Grundgesamtheit ist exemplarisch in dem Sinne, daß ein Besonderes ein Allgemeines so vertritt, daß dessen Besonderheiten in den Hintergrund treten, und sie ist exemplarisch in dem Sinne, daß ein Teil ein Ganzes so vertritt, daß das Erkenntnisverfahren auf die Ergänzung hin angelegt ist. Damit sind auch offene Korpuskonzepte, wie das unten vorgeschlagene *monitor corpus*-Konzept, methodologisch legitimiert.

2.1.4. Das wissenschaftliche Handlungsziel

Das jeweilige wissenschaftliche Handlungsziel bestimmt den Rahmen, innerhalb dessen sich die heuristische Funktion des Korpus und die anzuwendenden Analyseverfahren erst hinreichend konkretisieren lassen. Die Frage, wozu ein Korpus dienen soll, wird in den diversen Korpusdefinitionen auf sehr unterschiedlichen Abstraktionsstufen beantwortet. Bungarten (1978, 34) gibt als Handlungsziel sehr allgemein „operationalisierende linguistische Untersuchungen“ an, für die ein Korpus die „empirische Datenbasis“ darstellt. Nach Bergenholtz/Schaeder (1985, 278) dienen Textkorpora im „Rahmen einer primär-induktiv bzw. empirisch-deskriptiv verfahrenen Linguistik“ allgemein formuliert dem „Zweck der Dokumentation und/oder als Material für literatur- oder sprachwissenschaftliche Untersuchungen“ (ebenda 281). Je allgemeiner die Angabe des Handlungsziels, desto allgemeiner und prinzipieller auch die Bestimmung der die Parameter (i) bis (iv) miteinander verknüpfenden Korpusfunktion(en). Das Folgende beruht hingegen auf der Auffassung, daß eine differenziertere Zielbestimmung, etwa nach lexikologisch-lexikographischen, grammatisch-grammatikographischen oder soziolinguistischen Zielen, zu einer je unterschiedlichen und spezifischeren Bestimmung von Status und Funktion eines Textkorpus führt und daß somit verschiedene Funktionstypen des Korpus existieren, ohne daß diese in theoretischer Konkurrenz zueinander stünden.

2.2. Die Funktionen lexikographischer Primärkorpora

Bergenholtz/Schaeder (1985) geben für den lexikographisch relevanten Rahmen einer primär-induktiven, empirisch-deskriptiven Linguistik eine Reihe von fünf Funktionen an, die ein Textkorpus gleichzeitig erfüllen kann; praxisrelevant sind davon m.E. lediglich drei:

„1. als Material zur Formulierung von Hypothesen bzw. Prämissen über den Objektbereich; [...]

4. als Material für die Bestätigung bzw. das Verwerfen von Hypothesen sowie zur Überprüfung der Adäquatheit der aus bestimmten Prämissen gefolgerten Konklusionen;

5. als Sammlung von Belegen zur Illustration bestimmter Gebrauchsweisen von Wörtern und Syntagmen.“ (Bergenholtz/ Schaefer 1985, 281)

Diese Funktionsbestimmungen gelten aber nur für die Erkenntnisintention einer primär-induktiv bzw. empirisch-deskriptiven Linguistik, insofern sie „auf die Entdeckung einer in der aktuellen Sprachverwendung repräsentierten Systematik [zielt]“ (ebenda 280). Im Unterschied dazu richtet sich die Erkenntnisintention einer primär axiomatisch-deduktiv verfahrenen Linguistik „auf die Konstituierung eines Systems möglicher Sprachverwendungen“ (ebenda) und muß damit auch zu einer anderen Bestimmung der heuristischen Funktion des Korpus führen.

Der „prototypische“ Fall der primär-induktiven, empirisch-deskriptiven Linguistik, für die die o.g. Funktionen gelten sollen, ist die Lexikographie – das hat Schaefer auch in neueren Arbeiten immer wieder hervorgehoben. Hingegen wird die Frage nach der Funktion eines Textkorpus für grammatische oder grammatikographische Forschungsziele meist anders beantwortet. Aus dieser Differenz konnte ein z.T. polemische Formen annehmender Prinzipienstreit²⁴ nur deshalb entstehen, weil die Bindung der dem Korpus zugewiesenen Funktionen an das jeweils mehr oder weniger explizit gemachte wissenschaftliche Handlungsziel nicht hinreichend berücksichtigt wurde.

Einen Grenzfall stellt die zwischen Grammatikographie und Lexikographie angesiedelte Valenzlexikographie dar, die für die neben Schaefer gründlichsten korpus- und beispieltheoretischen Überlegungen von Ballweg-Schramm (1978) den wissenschaftlichen Handlungsrahmen bildet. Entstanden aus der Praxis des „Verben in Feldern“-Projekts resultieren die „Prinzipien“ von Ballweg-Schramm wegen der stark grammatikographischen Ausrichtung dieses Wörterbuchs aus der Entscheidung für eine primär-deduktive Methodik und unterscheiden sich daher zwangsläufig von diesbezüglichen Überlegungen im Rahmen einer primär-induktiv verfahrenen Lexikographie, ohne daß dieser Bezug auf ein konkretes valenzlexikographisches Projekt in allen Konsequenzen für eine Methodologie

²⁴ „[...] we must face the necessity of giving up the rather simple-minded nineteenth-century concept of objectivity which pretended that the lexicographer not merely could but did derive his meanings from the citations.“ (Quirk 1986,4).

berücksichtigt würde.²⁵

Während für Bergenholtz/Schaeder (1985) ein Textkorpus die durch nichts zu ersetzende *Basis* für die Formulierung von *Hypothesen* über den Objektbereich (Grundgesamtheit) darstellt (Funktion 1) und erst in einem zweiten Schritt erneut zur Überprüfung der Hypothesen herangezogen wird (Funktion 4), sieht Ballweg-Schramm in einem Textkorpus vor allem das Korrektiv für die methodisch *vorgängige Hypothesenbildung*, d.h. ausschließlich als ein Instrument der Hypothesenüberprüfung. Die Bildung der Hypothesen geschieht hingegen auf der Basis der Sprachkompetenz, über die Linguisten als Sprecher der zu beschreibenden Sprache verfügen:

„Unter textcorpusorientierung verstehe ich die sichtung von textcorpora mit dem ziel, die hypothesen zu überprüfen, die man als sprachsubjekt aufgestellt hat und seinem „alltäglichen sprecherbewußtsein“ verdankt.“ (Ballweg-Schramm 1978, 2).

Mit „sprecherbewußtsein“ (von anderen Autoren auch bezeichnet als „idiolektale (oder Idio-)Kompetenz“, „introspection“ und „Intuition“ des Linguisten) wird ein methodologisch einzuordnender Faktor bezeichnet, der in Bezug auf die Frage, *was als primäre Datenbasis irgendeiner linguistischen Untersuchung herangezogen wird*, in einem Konkurrenzverhältnis zum Korpus steht. Die Hintergründe der bestehenden Kontroverse sind wissenschaftsgeschichtlicher Art. Nachfolgend wird eine Lösung der Kontroverse im Hinblick auf kulturwortschatzbezogene Lexikographie vorgeschlagen.

2.2.1. Korpus versus Sprachkompetenz des Linguisten

Hoffmann (1984, 670) skizziert vier Stufen, in denen sich die z.T. heute noch vertretenen Positionen hinsichtlich „Funktion und Methode der Materialgrundlegung“ entwickelt haben:

(1) „Die klassische Sprachwissenschaft bis zum Beginn des 20. Jahrhunderts verwendete das Sprachmaterial durchweg als Belegsammlung“, ohne den Schritt von da zu verallgemeinernden Aussagen theoretisch zu fundieren.

(2) „Der taxonomische Strukturalismus führte [...] die strenge Korpus-Orientierung ein als einzige Grundlage der Ermittlung und Beschreibung

²⁵ Z.B. belegen der Titel des Aufsatzes von Ballweg-Schramm (1978) („Zur prinzipienlehre der lexikographie“) und der einleitende Satz („Dieser beitrage beschäftigt sich mit einigen grundlegenden lexikographischen problemen.“) eher einen Anspruch auf Allgemeingültigkeit der methodologischen Aussagen als das Bewußtsein ihrer Bindung an das spezielle Handlungsziel eines semantisch fundierten Valenzwörterbuchs des Typs „allgemeines wörterbuch für den kognitiven (selbst-)unterricht“ (ebenda 4).

von Sprachstrukturen".

(3) Die generative Grammatik vertrat das „Konzept des idealen Sprecher-Hörers und der Intuition des Linguisten als einziger Beurteilungsinstanz für sprachliche Phänomene".

(4) „Die Anstöße der Soziolinguistik und der Sprachsoziologie, die die systematische Variabilität von Sprache und Sprachgebrauch im Zusammenhang mit sozialen, situativen und psychischen Kontexten nachzuweisen suchen, haben dann zu einer positiven Neubewertung von empirischer Sprachdatenerhebung geführt."

Von diesen vier Positionen haben sich vor allem die erste und die dritte in lexikographischen Projekten und Methodendiskussionen niedergeschlagen. Ein zum Korpuskonzept der generativen Grammatik alternatives und der vierten Position korrespondierendes Korpuskonzept speziell für die Lexikographie ist erst in Ansätzen (dazu gehört das Lexikon „Brisante Wörter") entwickelt. Für diese Korrespondenz spricht, daß die Erkenntnisziele einer im weitesten Sinne kulturwortschatzbezogenen Lexikographie denen der Soziolinguistik verwandter sind als denen einer eher systematisch-formalen Grammatik. Es gilt also ein Textkorpuskonzept zu entwickeln, das jene „positive Neubewertung" (Hoffmann) der empirischen Fundierung speziell einer handlungssemantischen Lexikographie möglich macht.

Bausch (1979, 75ff) referiert die Diskussion über die Intuition als Datenbasis ausgehend von Chomsky ausführlicher; dabei wird deutlich, daß linguistische Intuition von denen, die ihr methodologische Priorität einräumen, immer ausschließlich an grammatischem Regelwissen, nicht an lexikalischem Wissen exemplifiziert worden ist, ohne die Frage nach der Anwendbarkeit und Geltung der Intuition als primäre Datenbasis auf die Lexikographie zu stellen.²⁶

Im Rahmen der lexikologisch-lexikographischen Handlungsziele besonders des hier infrage stehenden Wörterbuchtyps spricht gegen die Intuition des Linguisten als primäre Datenbasis die interne funktionale Differenzierung einer Sprachgemeinschaft, d.h. die mit Hoffmann (1984, 670) „systematische Variabilität von Sprache und Sprachgebrauch" und damit des Lexikons. Die Sprachkompetenz des Linguisten als Mitglied einer Sprachgemeinschaft erstreckt sich auf diejenigen funktionalen Varietäten, an die er aufgrund der eigenen Sozialisation und Biographie gebunden ist, niemals jedoch auf alle funktionalen Varietäten gleichzeitig und in aktiver Weise. Für Wörterbücher des hier verhandelten Typs, die Bedeutungen varietätenübergreifend analysieren und beschreiben wollen und die sich gerade auf die kommunikativen Probleme konzentrieren, die durch das Nebeneinander varietätenspezifischer Bedeutungen entstehen, kann nur ein

²⁶ So auch Steger 1980, 353.

Korpus, nicht die Idiokompetenz des Linguisten als Basis der Hypothesenbildung dienen. Wenn in der semantischen Analyse der Intuition das methodologische Primat zugewiesen wird, wird die Heterogenität²⁷ des anvisierten Kommunikationsbereichs überdeckt oder zumindest in Richtung auf eine bestimmte, sozial und intellektuell markierte Sprechergruppe (die der Linguisten) homogenisiert. Korpora hingegen können so organisiert werden, daß sie die spezifische Heterogenität des als Forschungsobjekt gesetzten Kommunikationsbereichs widerspiegeln. Die Kompetenz selbst einer mehrköpfigen Gruppe von Linguisten kann nicht als 'exemplarisch' im oben dargestellten Sinne in bezug auf eine Grundgesamtheit 'öffentliche Diskussion' o.ä. gelten.

Hingegen können Vertreter einer Position, die dem Korpus das methodologische Primat zuweist, in der Kompetenz des Linguisten, d.h. in dessen Subjektivität entweder einen Störfaktor sehen oder aber ihre Rolle im Prozeß der Korpusanalyse positiv zu bestimmen suchen. Zu den ersten gehört Greimas; er ist in seiner Korpusdefinition auf das Problem von Subjektivität und Zirkularität eingegangen, das sich bereits bei der Korpuskonstitution, erst recht bei Analyse und Beschreibung der Korpusdaten stellt, und hat daraus die Forderung nach selbstreflektierenden „Vorsichtsmaßnahmen und praktischen Ratschlägen“ (Greimas 1971, 129) abgeleitet.

2.2.2. Korpus und Verstehenskompetenz

Die andere Möglichkeit besteht darin, die individuelle, besser: sozial und historisch gebundene Sprachkompetenz des Linguisten als *V e r s t e h e n s* kompetenz gegenüber dem Textkorpus aufzufassen, die die Hypothesenbildung aus der Datenbasis allererst ermöglicht und in bestimmter Weise steuert. Mit der Wahl des Ausdrucks *Verstehenskompetenz* soll auch der Tatsache Rechnung getragen werden, daß sich linguistische Sprachkompetenz von alltäglichem, partiell reflexivem Sprecherbewußtsein vor allem hinsichtlich ihrer Komplexität (vgl. Schmidt 1988, 336f) unterscheidet, d.h. daß Linguisten die verschiedenen historischen, sozialen und formalen Dimensionen von Äußerungen gleichzeitig und in ihren Interrelationen wahrnehmen und reflektieren können und daß eine diesbezügliche Reflexion systematisch erweiterbar und auch auf vergangene Sprachstadien ausdehnbar ist. Die umfassende Textverstehenskompetenz des Lexikographen beruht auf der mehrdimensionalen, zusätzliche Lesarten erschließenden Lektüre von (auch gegenwartssprachigen) Korpustexten. Sie ist die Vorraussetzung dafür, daß die semantisch-pragmatischen, in besonderer Weise (kon)textgebundenen Verhältnisse des zu beschreibenden Wortschatzes, (text)analytisch erfaßt werden können. Das hier vertretene

²⁷ Zur Heterogenitätsproblematik siehe Bausch 1979, 73ff.

Modell lexikographischen Handelns steht im Gegensatz zu dem von Neubauer/ Petöfi (1981) geforderten Verfahren, in dem die Möglichkeit der Textinterpretation von der vorhergehenden Konstruktion eines Lexikons abhängig gemacht wird.²⁸

Lexikographisches Handeln läßt sich tendenziell eher mithilfe des hermeneutischen Modells verstehen, in dem die verschiedenen methodischen Schritte der praktischen Lexikologie (Korpusanalyse) und der Lexikographie (Darstellung im Wörterbuch) als 'Verstehen' und 'Auslegen' notwendig aufeinander bezogen werden.²⁹ Der besonderen Bedeutung der auslegenden Sprache in der Hermeneutik entspricht dann die besondere Bedeutung der lexikographischen Darstellungsmittel, insbesondere die Kommentar- und Beschreibungssprache. D.h. die Zuhilfenahme des hermeneutischen Modells impliziert ferner, daß es einen systematischen und methodologisch zu begründenden Zusammenhang zwischen Korpusprinzip und Darstellungsweise samt Beschreibungssprache gibt, dem hier jedoch nicht weiter nachgegangen werden kann.

Das Korpus ist in diesem Modell erstens Objekt des Verstehens und zweitens dasjenige Instrument, durch das das Verstehen des Linguisten (als Voraussetzung des expliziten Auslegens in der Wörterbuchdarstellung) intersubjektiv kontrollierbar gemacht wird (Vgl. Steger 1980, 353). Solange das Korpus allgemein zugänglich ist, gilt die intersubjektive Kontrollierbarkeit grundsätzlich auch dann, wenn es sich nicht um ein textdokumentatives, d.h. Beispielbelege enthaltendes Wörterbuch handelt. Die Datenbasis des introspektiv, sich primär auf die eigene Idiokompetenz stützenden Lexikographen ist hingegen nicht intersubjektiv verfügbar, da sie nicht konkret in Form der Angabe von Texten oder Kommunikationsbereichen mitgeteilt ist.

In Stegers Modell ist das Objekt linguistischen Verstehens anders als hier vorgeschlagen – nicht das Korpus, sondern unmittelbar das „konkrete(n) Sprechen(s) in einem sozialen Kontext“ (Steger 1980, 352). Hingegen entspricht es lexikographisch-praktischer Erfahrung, „das Verstehen von T e x t e n “ [Hervorhebung von mir] als „die wichtigste Bedingung für das Erkennen von Bedeutungen und Gebrauchsbedingungen von Wörtern“ (Schmidt 1988, 354) zu begreifen.

²⁸ „müssen weitere Versuche unternommen werden, Lexika für Interpretationszwecke zu konstruieren. Wir sind davon überzeugt, daß Fortschritte auf dem Gebiet der Lexikographie und auch der allgemeinen Semantik nur über die Konstruktion solcher Lexika [...] möglich werden.“ (Neubauer/Petöfi 1981, 186).

²⁹ „Auslegung ist nicht ein zum Verstehen nachträglich und gelegentlich hinzukommender Akt, sondern Verstehen ist immer Auslegung, und Auslegung ist daher die explizite Form des Verstehens.“ (Gadamer⁴ 1975, 291). Auch Steger (1980, 353) beruft sich auf die „Hermeneutiktradition“.

Die Funktionsbestimmung des Korpus nach Bergenholtz/ Schaeder (siehe oben) enthielt als letzten Punkt: „[das Korpus dient als] Sammlung von Belegen zur Illustration bestimmter Gebrauchsweisen von Wörtern und Syntagmen.“ (Bergenholtz/ Schaeder 1985, 281). Genereller formuliert dient das Korpus als Quelle für alle objektsprachlichen Angaben im Wortartikel, d.h. nicht nur für die Beispielbelege, sondern auch für paradigmatische („sinnverwandte/bedeutungsähnliche Ausdrücke“) und syntagmatische („typische Verbindungen“) Angaben.³⁰ Nach dem bisher Gesagten ist der die Beispielbelege betreffende Funktionsaspekt des Korpus für jede der Positionen in der Kontroverse um Korpus versus Idiomkompetenz in anderer Weise zu präzisieren.

Wird der Idiomkompetenz das Primat bei der Hypothesenbildung und dem Korpus folglich die Funktion eines Regulativs und Korrektivs zugewiesen, ist die Funktion der aus dem Korpus gewonnenen Beispielbelege (gleichgültig ob diese die Form von Syntagmen, Sätzen oder Text(segment)en haben) diejenige der Illustration einer generell formulierten „Regel“.³¹ In diesem Sinne präzisiert Ballweg-Schramm (1978) den von Henne eingeführten Ausdruck „Textorientierung“ als „hinwendung zum text als einem/dem mittel der exemplifizierung linguistischer aussagen [...] spezieller: die adäquatheit der semantischen und morphosyntaktischen erklärung eines wortes soll veranschaulicht werden, indem man die verwendung des wortes in texten mit hilfe von texten demonstriert.“ (Ballweg-Schramm 1978, 3)

Für das Beispiel bzw. den Beispielbeleg (ob in einer Grammatikdarstellung oder in einem Wörterbuch) folgt daraus, daß es das Sprachübliche und Typische zeigen muß; es soll nicht oder nur im Ausnahmefall das Besondere und von der Regelformulierung Abweichende einer Ausdrucksweise zeigen: Das Besondere, das von Landau 1984 und auch von anderen im Hinblick auf Gebrauchswörterbücher als „rubbish“ heruntergestuft wird,³² kann für einen anderen Wörterbuchtyp gerade von herausragendem Wert sein. Auch für Ballweg-Schramm (1978, 4) stehen die Nachteile „authentischer [Beispiel-] sätze“ im Vordergrund, unter anderem deswegen, weil in ihnen „fast nie die maximalvalenz höherwertiger verben aktualisiert wird“. Das Untypische, das Besondere, das hier mit Bezug auf einen bestimmten Wörterbuchtyp (Valenzwörterbuch) ausgeschlossen wird, zeichnet sich

³⁰ Zu den Artikelpositionen (8) und (9) im Informationsprogramm des Lexikons „Brisante Wörter“ siehe unten Abschnitt 2.5.1.

³¹ Vgl. Harras 1990. Zur Illustratio als einem von mehreren in der Lexikographiegeschichte herausgebildeten Funktionstypen des Beispiel(beleg)s siehe Haß 1990.

³² „Some citation files are filled with rubbish: nonlexical phrases, advertising slogans, foreign words, ephemeral abbreviations [...]“ (Landau 1984, 153).

aber beispielsweise sehr häufig durch Sprechergruppen-, Ideologie- oder Situationsspezifika oder dadurch aus, daß in ihm Tendenzen des Sprachwandels sichtbar werden, d.h. durch Aspekte, die für kulturell dokumentierende Wörterbücher gerade von herausragendem Interesse sind.

Für die Vertreter der Position, die dem Korpus das Primat bei der Hypothesenbildung zuweisen, dienen die dem Korpus entnommenen Beispielbelege in erster Linie dem Nachweis und der Nachvollziehbarkeit der in Form von „Regeln“ formulierten Ergebnisse der Korpusanalyse. Von Textdokumentativität der Beispielbelege im strengen Sinne kann überhaupt nur bei Korpusfundierung gesprochen werden, denn intuitiv oder introspektiv gewonnene Urteile sind per se nicht dokumentierbar, sondern nur illustrierbar.

(Text-)dokumentative Beispielbelege sind auch nicht wie illustrative darauf beschränkt, das Typische und Allgemeingültige des Objektbereichs zu zeigen; mit ihnen können und sollen gerade jene Verwendungsspezifika sowohl nachgewiesen als auch illustriert werden, die im Zentrum der kulturpädagogischen Intention des hier behandelten Wörterbuchtyps stehen.³³

2.3. Organisationsformen von Korpora

Zu Beginn der Computerisierung lexikographischer Arbeit herrschte hinsichtlich der Wörterbuchbasis noch – nicht anders als schon in den wissenschaftlichen Wörterbüchern des 19. Jahrhunderts – die Vorstellung von durch Quellenexzerption entstandenen Belegsammlungen vor. Die Möglichkeiten der neuen Technik wurden lediglich in quantitativer Hinsicht wahrgenommen,³⁴ d.h. man fragte nach den Auswirkungen der Tatsache, daß man mit Hilfe des Computers noch mehr Belege („citations“) verwalten können. Die Möglichkeiten, ganze Texte anstelle von Textsegmenten in Form von Beleg- oder Exzerptzetteln maschinell zur Verfügung zu halten und so zu einer qualitativen Veränderung wortsemantischer Analysen durch Einbeziehung textlinguistischer (textgrammatischer) und kontextueller Aspekte zu gelangen, wurden lange nicht gesehen.

Ein Großteil der computerkorpus-theoretischen Überlegungen fand vor oder ungeachtet der pragmatischen Wende in der Linguistik statt.

Obwohl in den meisten dieser Arbeiten der Anspruch erhoben wird, ganz allgemein für alle denkbaren linguistischen und lexikographischen Vorha-

³³ Z.B. fordert Wiegand (1985, 370) für „bestimmte Wörter“, die sich „nicht dazu eignen, einfach lexikalisch paraphrasiert zu werden [...] wie *Freiheit*, *Gerechtigkeit*, *Faschismus*, *Kapitalismus*, *Bürger*, *Revolution*, *Radikalenerlaß* usw.“, mit ihnen müsse „textdokumentativ und historisch erläuternd verfahren werden“.

³⁴ Vgl. z.B. Sherman 1979, 137.

ben gültig zu sein,³⁵ herrscht eine mehr oder weniger bewußte Festlegung auf Gebrauchs- und Lernerwörterbücher vor.³⁶ Dies zeigt sich beispielsweise in den Versuchen, das „Ungewöhnliche“ und „Seltene“ in der Datenbasis mithilfe des Computers in den Hintergrund zu drängen (Sherman 1979, 142). Es zeigt sich auch in der allgemein starken Fixierung auf Wortfrequenz, die allenfalls zur Ermittlung von Stichwortlisten für Grundwortschatzwörterbücher u.ä. sinnvoll ist.

Es existieren drei verschiedene Konzepte von Computerkorpora. Das erste und älteste knüpft – das zeigt schon die Benennung – an das traditionelle Konzept der Belegsammlung an: das *sample corpus*.

2.3.1. Sample-Korpora

Ein *sample corpus* enthält eine abgeschlossene Reihe strikt gleichlanger Textauszüge eines möglichst breiten Gattungs- oder Sachgebietsspektrums; so umfaßt z.B. das LIMAS-Korpus ca. 1 Mio. Textwörter (*tokens*) aus 500 Textstücken (zu je 2000 Textwörtern) aus allen Bereichen der deutschen Gegenwartssprache (33 Fachgebiete), die nach der Nationalbibliographie 1971/72 ausgewählt und gewichtet wurden.³⁷ Die Wahl der Textstücke kann prinzipiell auch durch Zufallsprozeduren erfolgen (vgl. Sinclair 1982, 1). Sample-Korpora sind damit strikt synchron und enthalten für jedes einzelne Textsegment einen relativ begrenzten und oft fragmentarischen Kotext.³⁸

2.3.2. Projektspezifische Korpora

Alternativ zu diesem Konzept eines möglichst allgemeinen, breiten Korpus für gemein- bzw. standardsprachliche Untersuchungen entstand zu Anfang der Computerisierung die Idee, für jedes einzelne linguistische oder lexikographische Projekt ein ganz spezifisches eigenes Korpus zusammenzustellen.³⁹ Die Realisierung einer solchen Vorstellung – das zeigte sich schnell – scheitert jedoch an dem immensen technischen und daher finanziellen Aufwand, der eine nur einmalige projektspezifische Verwendbarkeit nicht rechtfertigt bzw. verbietet (vgl. Francis 1982, 14f). Maschinell verfügbare

³⁵ Z.B. mit Bezug auf das Brown Corpus of American English und das Lancaster-Oslo/Bergen-Corpus of British English; siehe Johansson 1980.

³⁶ Vgl. z.B. Rey-Debove 1971, Landau 1984.

³⁷ Das LIMAS-Korpus ist Teil der IDS-Korpora; siehe Abschnitt 3.1.

³⁸ Vgl. Clear 1988, 383f über *The characteristics of sample corpora*.

³⁹ Z.B. das „Bonner Zeitungskorpus“ für das „Maschinelle Korpuswörterbuch des BRD- und des DDR-spezifischen Zeitungsvokabulars, siehe Schaefer 1981, 110ff und Hellmann 1984.

Textkorpora müssen für unterschiedliche linguistische Aufgaben herangezogen werden können.⁴⁰

Darüberhinaus gibt es weitere technische Bedingungen wie die Verfügbarkeit von Setzbändern⁴¹ ganz bestimmter, nicht nach linguistischen Kriterien ausgewählter Texte oder Textarten, die allein schon zu einer gewissen Flexibilität in der Festlegung der Korpusgrenzen zwingen. Da die Forderung nach Repräsentativität linguistischer Korpora inzwischen als obsolet gilt (siehe 2.1.3; vgl. Schaefer 1981, 74ff), ist diese Flexibilität linguistisch durchaus zu vertreten, sofern sie begründet und nicht bloß zufällig ist.

2.3.3. Monitor-Korpora

Diese Flexibilität macht sich das dritte Computerkorpuskonzept, das sogenannte *monitor corpus* als ein aus qualifizierten Teilkorpora zusammengesetztes Gesamtkorpus, zunutze. Ausdruck und Idee eines *monitor corpus* in Abgrenzung zu mehreren einzelnen *sample corpora* stammen von Sinclair:

„The difference is one of quality, not quantity. Looking through the computer, the whole state of language can be passed before one's eyes. [...] This new type of corpus I want to call a *monitor corpus*, because of its capacity to hold a 'state of the language' for research purposes. The information which a sample corpus cannot provide can be retrieved by manipulation of a monitor corpus.” (Sinclair 1982, 4)

Die Rezeption dieser Idee,

„a monitor corpus [...] can be manipulated to reveal insights into the state of the language at a given time. The process will involve the continual replacement of old data by new, so that the changing store of text can always reflect current linguistic behaviour.” (Renouf 1984, 4)

macht deutlich, daß bisher vor allem der Bezug auf eine bzw. auf die jeweils gegenwärtige Zeitstufe der zu beschreibenden Sprache im Vordergrund stand.⁴² Die Vorteile des Monitor-Konzepts: Offenheit, Flexibilität und Formbarkeit der Datenbasis scheinen mir jedoch weit mehr und andere

⁴⁰ Vgl. Dokumentation Textkorpora des heutigen Deutsch.

⁴¹ Das sind Magnetbänder, auf denen die Ergebnisse des modernen Computersatzverfahrens gespeichert sind und die von manchen Druckereien bzw. Verlagen für wissenschaftliche Zwecke zur Verfügung gestellt werden. Vgl. Brückner 1989, 91.

⁴² Vgl. Clear 1988, 384f über *The characteristics of monitor corpora*.

Möglichkeiten zu enthalten als die beliebiger synchroner Schnitte.

Gegenüber Sample-Korpora und projektspezifischen Korpora zeichnet sich ein Monitor-Korpus dadurch aus, daß es sukzessive erweiterbar, dadurch sowohl für synchrone als auch für diachrone Untersuchungen nutzbar ist, und Kotexte von maximaler wünschbarer Länge enthält; durch Ausgliederung einer Teilmenge kann je nach Forschungsziel eine spezifische Datenbasis definiert werden. Damit ist es für viele, nicht nur lexikographische Zwecke nutzbar (Ökonomie-Argument); dennoch ist der jeweilige Benutzer (Lexikograph) frei in der Entscheidung, welche Korpusbasis für ein bestimmtes Wörterbuch oder zur Bearbeitung einer bestimmten (lexikologischen, semantischen usw.) Frage zugrundegelegt werden soll.

Auf diese Weise können auch die von der Technik gesetzten Bedingungen einigermaßen „abgefedert“ werden. Wenn z.B. die leichte Verfügbarkeit von Setzbändern zu einer Überproportionalität bestimmter Textarten (etwa einer ganz bestimmten Tageszeitung) geführt hat, kann man dieses Teilkorpus in der Auswertung quantitativ entsprechend schwächer gewichten. Ein derartig offenes Korpuskonzept erlaubt es auch, weiteres, nicht maschinell verfügbares Material (z.B. alle historischen, nicht computerverfügbaren Quellen, alle sehr projektspezifisch und in traditioneller Zettelkastenmanier gesammelten Textbelege) als methodisch gleichrangig mit einzubeziehen (siehe Abschnitt 4.2). Die Gleichartigkeit des maschinell und des traditionell verfügbaren Materials ist gesichert, wenn beides in Form derselben elementaren Einheiten, vornehmlich in Form abgeschlossener Texte, und inbezug auf die gleiche Grundgesamtheit ausgewählt wird.

Offenheit und Flexibilität eines Monitor-Korpus haben zur Folge – und das könnte möglicherweise als nachteilig empfunden werden –, daß der Lexikograph Aufbau und Inhalt der Teilkorpora und sogar die für ihn relevanten Eigenschaften der Einzeltexte kennen muß, um die jeder Frage angemessene Datenbasis auswählen und die Ergebnisse der einzelnen Korpus-Recherchen angemessen einordnen zu können. Diese intellektuelle Arbeit können Computer vor allem deshalb nicht übernehmen, weil eine umfassende Klassifizierung und Markierung aller Korpus-texte bzw. Teilkorpora nach internen und externen Textmerkmalen nur teilweise möglich und für die meisten existierenden Computerkorpora auch gar nicht durchgeführt ist.

Eine Klassifizierung der enthaltenen Teilkorpora und Einzeltexte wäre aus arbeitsökonomischen Gründen nur sinnvoll in Form von mnemotechnisch günstigen Kodierungen (Markierung mittels möglichst weniger alphanumerischer Zeichen), die jedem einzelnen Korpusbeleg, im Prinzip jedem einzelnen token des Gesamtkorpus, beigelegt werden (können). Bei den IDS-Korpora existiert eine solche Kodierung, die allerdings nur rudimentäre und kaum systematische Informationen über Text-„Gattung“, Sachbereich,

Datierung und/oder Autor u.ä. geben (siehe Abschnitte 3.1 und 3.2). Die im Hinblick auf den Typ des Kulturwortschatzwörterbuchs besonders wichtigen kommunikativ-pragmatischen Textmerkmale sind aus solchen Markierungen allenfalls sekundär abzuleiten. Vorausgesetzt, es wäre technisch möglich, alle oder eine große Zahl der relevanten Textmerkmale in einem selbsterklärenden oder mnemotechnisch günstigen Code (von dann wahrscheinlich mindestens zehn Zeichen) abzubilden, so bliebe immer noch das Problem der unerfüllbaren Objektivität oder des wissenschaftlichen intersubjektiven Konsenses einer umfassenden Textklassifikation. Schaefer gibt mit Bezug auf das Bonner Zeitungskorpus (BZK) die „Aussperrung der Pragmatik“ (Schaefer 1979, 235) bei der Textklassifikation zwar zu, fordert sie aber dennoch, ohne auf das Objektivierbarkeitsproblem dabei einzugehen.⁴³

Eine detaillierte, systematische und linguistisch begründete Textklassifizierung ist für ein offenes, vielfach nutzbares Korpus auch deshalb undenkbar, weil sie und die Klassifizierungskriterien nur relativ zum jeweiligen wissenschaftlichen Handlungsziel, d.h. zu einem projektierten Wörterbuch usw. gültig oder von hinreichender Trennschärfe sein können. Beispielsweise ist die bloße Angabe des Sachgebiets unzureichend für die Untersuchung fachexterner gegenüber fachinterner Verwendung von (Fach-)Ausdrücken. Klassifizierung und Beurteilung der jeweils im Korpus enthaltenen Texte im Hinblick auf ein bestimmtes Forschungsziel ist somit Teil der projektspezifischen analytischen Arbeit mit dem Korpus (siehe Abschnitt 2.5.2), die dem einzelnen Lexikographen obliegt. Seine Textkenntnis als wesentlicher Bestandteil seiner linguistischen Kompetenz schließt Textsortenkenntnis und Kenntnis intertextueller Zusammenhänge mit ein.

Nachfolgend eine Liste von Markierungsdimensionen für die Korpus Texte eines Kulturwortschatzwörterbuchs; die Markierungen sind, solange nicht maschinell handhabbar, an die Kompetenz des einzelnen Lexikographen gebunden:

- Ort (nationale, regionale Varietät)
- Zeit (eventuell in Jahrzehnten)
- Medium (schriftlich/ mündlich/ transkribiert, mit/ ohne Bildbezug)
- berufsbezogener Kommunikationsbereich (fachintern/fachextern/ gemeinsprachlich; mit Markierung des Themenbereichs)

⁴³ „Zu berücksichtigen wären wenigstens für die folgenden Arten von Zusammenhängen: (1) Tageszeitung als Massenkommunikationsmittel in einer bestimmten Gesellschaft, (2) die einzelne Tageszeitung als Massenkommunikationsmittel einer oder mehrerer Gruppen in dieser Gesellschaft, (3) der einzelne Zeitungsartikel im Ko- und Kontext der einzelnen Zeitung.“ (Schaefer 1979, 234 Anm. 39).

- gesellschaftsbezogener Kommunikationsbereich (Politik, Jugend, Sport, Verwaltung, Literatur mit Markierung der Gattung, usw.)
- privater oder öffentlicher Kommunikationsbereich (mit Markierung von Einfach-/Mehrfachadressiertheit)
- Sprecher(gruppe) (identifiziert/ nicht identifiziert; Markierung von Äußerungseinbettungen)
- Textintention (anweisend, werbend, informierend, erzählend, usw.)
- originaler/ übersetzter Text
- Textsegment, dem ein Korpusbeleg angehört (Titel/Überschrift, Zwischentitel, Lead, Bildbeischrift usw.)

Für eine Lexikographie auf der Basis der strukturellen lexikalischen Semantik mögen Klassifikation und Markierung der Korpustexte, erst recht in dieser relativ starken Differenzierung, weniger wichtig sein, aber die lexikographische Umsetzung einer pragmatischen, handlungsorientierten Semantikauffassung ist nur dann möglich, wenn die linguistische Interpretation jedes einzelnen Korpusbelegs auf der Basis größtmöglichen Wissens über Ko- und Kontext vorgenommen wird. Je größer die Textkompetenz, desto ergiebiger die Analyse und desto dezidierter die Aussagen über Bedeutung und Verwendung eines Ausdrucks. Welche erkenntniszielspezifischen Klassifikationen von Korpustexten auf der Basis der o.g. Liste von Markierungsdimensionen vorgenommen und als besonders relevant für wortsemantische Fragestellungen eingestuft werden, wird in Abschnitt 2.5.3 an einem Beispiel vorgeführt.

2.4. Die Funktionen lexikographischer Sekundärkorpora

Es gilt, die vorthoretische Bestimmung von Sekundärliteratur als denjenigen Texten, die der Lexikograph für Hintergrund- und Zusammenhangsinformation heranzieht, aus denen aber keine Beispielbelege entnommen werden, linguistisch zu begründen.

Genau wie das Primärkorpus lassen sich Status und Funktion des Sekundärkorpus anhand der vier Parameter: (i) Grundgesamtheit, (ii) elementare Einheiten, (iii) Relation zwischen Korpus und Grundgesamtheit, (iv) wissenschaftliches Handlungsziel bestimmen. Während (i) und (iv) mit den Parametern des entsprechenden Primärkorpus übereinstimmen müssen und für (ii) selbstredend nur Texte infrage kommen, erweist sich die verschiedene Bezogenheit (iii) von Primärkorpus und Sekundärkorpus auf den Kommunikationsbereich, dessen Wortschatz beschrieben werden soll, als distinktiv.

Das Sekundärkorpus bezieht sich auf die Grundgesamtheit in deskriptiv-analytischer Weise und ist nicht exemplarischer Teil der Grundgesamtheit.

Im Hinblick auf das Handlungsziel Kulturwortschatzwörterbuch macht es einen Unterschied, ob in einem Sekundärkorpus sprachkommentierende und -reflektierende Äußerungen von Sprechern eben des Kommunikationsbereichs gesammelt werden, der als die Grundgesamtheit des Primärkorpus definiert ist, oder ob es sich ausschließlich um solche Äußerungen handelt, die außerhalb dieses Kommunikationsbereichs selbst anzusiedeln sind, wie es bei allen sprachwissenschaftlichen, philologischen, sach- und begriffsgeschichtlichen usw. deskriptiven Arbeiten der Fall ist. Insofern die Kulturwortschatzbereiche des 20. Jahrhunderts generell von Interferenzen fach- und wissenschaftssprachlicher Varietäten geprägt sind, gilt es zu entscheiden, ob enzyklopädische und fachinterne Informationsliteratur ausschließlich dem Sekundärkorpus zugeordnet und damit von der späteren Belegzitierung ausgeschlossen werden, oder ob man sie zwar nicht als zentralen aber doch als wesentlichen Faktor der Kommunikation einstuft, weil in ihnen jene terminologischen Festlegungen getroffen werden, die die Ursache für lexikalisch bedingte Probleme und Konflikte im zu beschreibenden Kommunikationsbereich darstellen. Wenn man in der Wörterbuchdarstellung nicht darauf verzichten will, z.B. die fachinterne, abgrenzende Definition von *Ver- und Entsorgung*, von *Störfall* versus *Unfall* oder die Definition von *Aggression* bei S. Freud als Beispielbelege anzuführen, müssen solche Texte einer besonderen Textklasse des Primärkorpus zugeordnet werden können.

Es sind also vor allem solche Fachtexte korpusrelevant, die aufgrund ihrer dominanten Textfunktionen (Normieren, über Normierungen Informieren) terminologische Festlegungen enthalten. Hingegen sind für alle anderen Fachtexte, das heißt diejenigen, in denen Fachausdrücke und -termini verwendet werden, ohne daß dabei auf eine zugrundeliegende Festlegung, Norm oder dergleichen verwiesen würde, als empirische Datenbasis weniger geeignet: Anders als im Falle gemeinsprachlicher Ausdrücke in gemeinsprachlichen Texten kann die Bedeutung eines Fachausdrucks nicht vollständig und korrekt aus Fachtextaktualisierungen heraus analysiert werden.

Information und Repräsentation eines Ausschnitts von sprachlich relevantem Weltwissen sind nicht ausschließlich Leistungen der Texte des Sekundärkorpus, sondern auch der des Primärkorpus. Die in einer eventuellen Artikelposition 'Sachinformation' gemachten Angaben basieren immer auch auf den Informationsinhalten des Primärkorpus, sofern dessen elementare Einheiten Texte und nicht bloß Sätze oder Syntagmen sind. Die spezifische Informativität des Sekundärkorpus besteht darin, daß es vor allem wissenschaftliche Äußerungen über die sprachlichen und kommunikativen Aspekte des Objektbereichs enthält.

Die Abgrenzung von Primär- und Sekundärkorpus wird schwierig, wenn z.B. der nichtfachgebundene bildungssprachliche Wortschatz beschrieben

werden soll, der von Sprach- und Kommunikationswissenschaften geprägt ist, z.B. im Falle der Lemmata *Sprachspiel*, *Metapher*, *Euphemismus*, die z.T. gleichzeitig zu den beschreibungssprachlichen Ausdrücken desselben Wörterbuchs gehören. Die methodologische Trennung von Primär- und Sekundärkorpus ist für den beschreibungssprachlichen Wortschatz nicht strikt aufrechtzuerhalten.

Zusammengefaßt: Ins Sekundärkorpus gehören Texte, die außerhalb des zu beschreibenden Kommunikationsbereichs, d.h. zu ihm nicht in einer Relation der Exemplarizität, Dokumentativität und Illustrativität stehen.

2.5. Informationsprogramm und Korpuskonzept

Ein die konkrete Korpusorganisation vorbereitendes Korpuskonzept für ein Exemplar des Typs Kulturwortschatzwörterbuch⁴⁴ wird ausgehend vom Informationsprogramm, d.h. von der Gesamtheit der je einer Artikelposition zugeordneten Informationsarten bestimmt. Dieses Informationsprogramm liefert diejenigen Fragen, die die Analyse des Korpus leiten. Deshalb soll zunächst als ein Beispiel der Katalog der Informationsarten resümiert werden, die dem Lexikon „Brisante Wörter“ zugrundeliegen, um anschließend die Spezifizierungen des zugehörigen Korpuskonzepts zu erläutern.

2.5.1. Ein Informationsprogramm für den Typ Kulturschatzwörterbuch⁴⁵

Die diskursive Darstellung im Lexikon „Brisante Wörter“ orientiert sich an einer Liste von Informationsarten, die mit bestimmten Artikelpositionen (Textbausteinen) korreliert sind, wobei mehrere Informationsarten in einer Artikelposition zusammengefaßt sein können. Der Leser kann verschiedene Informationsarten vor allem an der Thematik und an bestimmten beschreibungssprachlichen Ausdrücken, weniger an der Typographie erkennen und unterscheiden. Die in bezug auf jede Einzelbedeutung maximale Liste der Informationsarten ist die folgende:

- (1) Lemma (je nach Artikelsorte können ein oder mehrere Hauptlemmata angesetzt werden)
- (2) Sublemmata (sie können zum Hauptlemma entweder in wortbildungsbezogenem oder in semantisch-paradigmatischem Verhältnis stehen)

⁴⁴ Beispiele im folgenden aus Strauß/ Haß/ Harras 1989, besonders aus dem Kommunikationsbereich 'öffentliche Diskussion zum Thema Umwelt'.

⁴⁵ Ich beschränke mich hier auf eine Skizze; die Informationsarten werden eingehender erläutert und begründet in Strauß 1988 und in Strauß/ Haß/ Harras 1989, 2-5 und 14-17.

- (3) Vorspann mit grammatischen, formbezogenen, wortfamiliären und/oder etymologischen Angaben
- (4) Exposition mit Angaben zum Verwendungsbereich (z.B. *fachbezogen*), zur semantischen Funktionsklasse (z.B. *Schlagwort*) und ggf. mit Überblick über die Artikelgliederung nach Einzelbedeutungen
- (5) Angaben zu den Klassen möglicher Bezugsobjekte
- (6) Bedeutungsangabe im engeren Sinn (zur Frage: Wie oder in welcher Perspektive werden die in Position (5) genannten Bezugsobjekte charakterisiert?)
- (7) Angaben zu speziellen, z.B. gruppenspezifischen Verwendungsweisen einschließlich ihrer kommunikationsgeschichtlichen, argumentativ-rhetorischen und sachbezüglichen Hintergründe
- (8) Angaben zu semantisch-paradigmatischen Relationen (vor allem zu (partiell) synonymischen, antonymischen, generalisierenden, spezifizierenden Ausdrücken)
- (9) Angaben zu syntagmatischen Relationen (zu Kollokationen, usuellen, formelhaften Syntagmen und zur Valenz)
- (10) Angaben zu Textvorkommen und Textsortenspezifik
- (11) Sprachkritischer Kommentar (auch als Resümee der kritischen öffentlichen, nicht-linguistischen Sprachreflexion)
- (12) computerbezügliches Datenbankfeld für artikelexterne und artikelinterne Verweise (enthält keine unmittelbaren Informationen für die Adressaten, sondern betrifft die abschließende Erstellung eines Registers)
- (13) Angaben zu Wortbildungsbeziehungen; Position zum „Einhängen“ von Subartikeln (enthält computerintern relevante Informationen)
- (14) Angaben zu kommunikationsgeschichtlichen und sachbezüglichen Bedingungen der Wortverwendung
- (15) Belege in Belegblöcken mit Überschriften
- (16) Literaturhinweise

Ziel zunächst der Korpusorganisation, dann der Korpusanalyse ist es, genau solche Informationen zu erhalten, mit denen – sieht man von den Positionen (12), (13), (15) und (16) ab – zwölf von diesen Positionen gefüllt werden können. Hinsichtlich der Beispielbelege in Position (15) ist bei Annahme des Belegungsprinzips die unmittelbare Relation zwischen lexikographischem Konzept und Korpus selbstredend. Position (16) ist in der Regel eng mit dem Sekundärkorpus verknüpft, und selbst die computerrelevanten Positionen (12) und (13) beruhen letztlich ebenfalls auf lexiko-

graphischen Entscheidungen im Zusammenhang mit der Korpusanalyse.

2.5.2. Konzept eines Korpus für den Typ Kulturwortschatzwörterbuch

(i) Als Grundgesamtheit wird zunächst der Kommunikationsbereich 'öffentliche Kommunikation', eventuell beschränkt auf einen/ mehrere Themenbereich(e) wie Politik/Ideologie, Umwelt, Kultur/Bildung bestimmt und hinsichtlich Zeit (etwa: 1945 bis 1988) und Raum (etwa: Bundesrepublik Deutschland) begrenzt. Ferner werden die die Diskussion führenden Sprechergruppen und zumindest die wichtigsten Textsorten und Medien der Diskussion festgestellt.

(ii) Als elementare Einheiten des Primärkorpus werden Texte der Textsortenklassen 1 bis n bestimmt. Textsorte wird hier eher im Sinne von Gattung, nicht im Sinne von Typ, verstanden als Realisierung eines im Sprecherbewußtsein präsenten und intentional einsetzbaren Textkonzeptes; in Textsortenklassen sind Exemplare verschiedener Konzepte aufgrund linguistischer Ähnlichkeitskriterien, die je nach Wörterbuchplan frei wählbar sind, zusammengefaßt. Die Textsortenklassifikation als Ganzes beruht auf der Annahme, daß die öffentliche Diskussion zu einem Themenbereich mittels Klassen von Textsorten geführt wird, die untereinander in einem systematischen Zusammenhang stehen.⁴⁶

Ein wesentlicher, aber nicht der einzige mögliche Aspekt einer solchen intertextuellen Systematik ist z.B. die Vermittlung fachlicher Inhalte – und damit von Fachwortschatz – an eine breitere Öffentlichkeit. In den Bad Homburger Thesen zum (seinerzeit geplanten) großen interdisziplinären deutschen Wörterbuch wurde noch für die bevorzugte Aufnahme fachexterner Texte plädiert (Mentrup 1979, 189). Um aber die systematischen Wortschatzzusammenhänge einerseits zwischen fachexternem und fachinternem Bereich, andererseits zwischen fachexternem und gemeinsprachlich-öffentlichem Bereich auf der Ebene der Datenbasis empirisch zugänglich machen und nachweisen zu können, ist es erstens notwendig, neben der fachexternen auch die fachinterne und die gemeinsprachliche Kommunikation jeweils zum selben Themenbereich im Korpus zu repräsentieren und zweitens, den fachexternen Kommunikationsbereich intern stärker nach dessen Textsorten(klassen) zu differenzieren.

Ein Beispiel für ein solches System von Textsortenklassen ist das folgende, dessen Klassifikation sich nach der abgestuften fachexternen Kommunikation von der Informationsvermittlung bis hin zur Meinungsbildung richtet; je nach Wörterbuchtyp und Thematik des zu beschreibenden Wortschatzausschnittes sind Klassifikationen nach anderen Kriterien möglich:

⁴⁶ Zur nachweisbaren Interrelation von Textsortenzusammenhängen und Wortschatz vgl. Haß 1986.

1. Textsortenklasse (P1): fachinterne Publikationen und Forschungsberichte; interfachliche Berichte, Verschriftlichungen von Experten-Hearings, Symposien, Gutachten, Stellungnahmen usw.,
2. Textsortenklasse (P2): Gesetze, Verordnungen, andere Texte aus der juristischen Kommunikation zum Themenbereich,
3. Textsortenklasse (P3): Texte fachinterner Wissensvermittlung wie Handbücher, Nachschlagewerke, Einführungen z.B. für Studenten,
4. Textsortenklasse (P4): Texte popularisierender Information und Meinungsbildung, thematische Sachliteratur (meist in Taschenbuchform), Themenmagazine, ausführliche Hintergrundartikel in Tages- und Wochenzeitungen der Rubriken „Aus Forschung und Technik“ o.ä.; Flugblätter und Broschüren aller beteiligten Interessen- und Sprechergruppen,
5. Textsortenklasse (P5): Nachrichtentexte, Medienberichterstattung auf den ersten Seiten der Tages- und Wochenzeitungen,
- als 6. Textsortenklasse (P6) eventuell: sprachreflektierende und sprachkritische Texte wie Glossen, Kommentare, Leserbriefe zum Wortschatz des Themenbereichs.

(iii) Aufgrund eines solchen Modells, das die spezifische Intertextualität der den Kommunikationsbereich konstituierenden Texte in ihren charakteristischen Merkmalen (im Sinne der in 2.1.3 erläuterten exemplarischen Repräsentation) abbildet, läßt sich die Auswahl einzelner, ins Korpus aufzunehmender Texte z.B. wie folgt steuern:

Im Zentrum der lexikographischen Beschreibung steht der Wortschatz, wie er in den Textsortenklassen (P 3), (P 4), (P 5) und ggf. (P 6) verwendet wird; deshalb bilden Textexemplare dieser drei (bzw. vier) Klassen den Grundstock des Primärkorpus. Sie stehen dem Lexikographen vorrangig über Computerabfragen zur Verfügung. Ob Textexemplare der Klasse (P 6) in ausreichendem Maße maschinell miterfaßt worden sind, hängt davon ab, ob die kommunikationsbereichsinterne, d.h. nicht-linguistische Sprachreflexion bzw. Sprachkritik ebenfalls Gegenstand der Wortschatzdarstellung sein soll.⁴⁷

Für das zugehörige Sekundärkorpus können ein oder zwei weitere Textsortenklassen angesetzt werden:

- (S1): allgemeine, historische, fremdsprachige Wörterbücher und Enzyklopädien, Grammatiken,
- (S2): linguistische, medienwissenschaftliche, soziologische, politologische, historische usw. Untersuchungen zum Kommunikations- oder Wortschatz-

⁴⁷ Das war der Fall beim Bereich 'Umwelt' in Strauß/ Haß/ Harras 1989.

bereich.

Für die eventuell noch verbleibenden Texte wie Essays in allgemeinen Kulturzeitschriften (z.B. Kursbuch, Merkur, u.ä.) kann aufgrund der vorliegenden Klassifikation eine begründete Entscheidung für Aufnahme in Klasse (P 6) des Primärkorpus oder in Klasse (S 2) des Sekundärkorpus getroffen werden.

(iv) Die wissenschaftlichen Ziele, die mithilfe dieses intertextuell transparenten Korpuskonzepts erreicht werden können, gehen über das traditionelle Ziel 'Analyse und Beschreibung von Lexembedeutungen und lexikalisch-semantischen Relationen' hinaus; Differenzierung und Ausweitung dieses Ziels wurden bereits im oben skizzierten Informationsprogramm sichtbar: Es wird noch deutlich werden, daß dies vor allem Konsequenzen hinsichtlich traditionell sogenannter Polysemie, (partieller) Synonymie, Antonymie, Komplenymie und Hypo-/Hyperonymie hat, die nicht (mehr) als absolute auf „den“ Wortschatz und seine Struktur bezogene Relationen verstanden, sondern in diskursiv-expliziter Weise beschrieben werden als intentionale oder zumindest funktional begründbare Setzungen von Sprechern im Kommunikationsbereich X (mehr in Abschnitt 4.4).

3. Die Korpora des IDS als Monitor-Korpus

Seit Mitte der 60er Jahre wird im Institut für deutsche Sprache in Mannheim ein maschinell verfügbares Textkorpus der deutschen Gegenwartssprache aufgebaut und ständig erweitert. Die Erweiterung geschieht, indem größere, in sich wiederum nach einzelnen Texten bzw. im Falle von Zeitungsartikeln nach Textgruppen gegliederte Teilkorpora fertiggestellt und dem Gesamtkorpus hinzugefügt werden; einmal fertiggestellte Teilkorpora werden nicht mehr verändert. Das Gesamtkorpus ist als Datenbasis für alle Arten linguistisch-empirischer Untersuchungen nutzbar. Für jedes einzelne Projekt, z.B. für das Lexikon „Brisante Wörter“, werden bestimmte Teilkorpora ausgewählt und als projektspezifische Datenbasis festgelegt. Damit entspricht das IDS-Korpus der oben gegebenen Definition des Monitor-Korpus. Während der Bearbeitungszeit des Lexikons „Brisante Wörter“ stand ein Gesamtkorpus von 20,5 Mio. Textwörtern (tokens) zur Verfügung.

Im folgenden soll am Beispiel vorgeführt werden, wie ein Monitor-Korpus organisiert sein kann, und anschließend, in Abschnitt 4, wird gezeigt, wie mit den als Datenbasis festgelegten Teilkorpora lexikographisch gearbeitet werden kann.

3.1. Die Teilkorpora. Kurzcharakterisierung

Die als Datenbasis des Lexikons „Brisante Wörter“ ausgewählten Teilkor-

pora⁴⁸ werden in der zeitlichen Reihenfolge ihrer Entstehung charakterisiert:⁴⁹

Das Teilkorpus „Mannheimer Korpus 1“ (abfragbar mittels der Sigle MK1) wurde in den Jahren nach 1965 mit dem Ziel angelegt, „möglichst viele relevante Bereiche der geschriebenen deutschen Gegenwartssprache“ (LDV-Info 1/ 1981, 5) zu erfassen. Die 2,2 Mio. Textwörter verteilen sich auf *Schöne Literatur* (Autoren: Bergengruen, Böll, Frisch, Grass, Johnson, Th. Mann, Strittmatter), *Trivallliteratur*, wissenschaftliche und populärwissenschaftliche *Literatur* (Autoren: Bollnow, Bamm, Grzimek, Gail, Heimpel, Heisenberg, Jaspers, Pörtner, Staiger), *Memoiren* (Heuss) sowie *Zeitungen und Zeitschriften* (von „Bild der Wissenschaft“ über die „Frankfurter Allgemeine Zeitung“ bis „Bild-Zeitung“), z.T. nur in Form einzelner Seiten der Jahrgänge 1966 und 1967.

Mitte der 70er Jahre wurde das „Mannheimer Korpus 2“ (Sigle: MK2) erstellt, um die Textsorten des MK1 vor allem um Gebrauchstexte zu ergänzen. Das relativ kleine Korpus (ca. 331.000 Textwörter) enthält diverse Erlasse, Gebrauchsanleitungen, Medikamentenbeipackzettel, Werbebroschüren, Trivallliteratur, technische und wissenschaftliche Fachzeitschriften und einzelne Ausgaben regionaler Tageszeitungen.

Das Bonner Zeitungskorpus (Sigle: BZK) enthält eine statistisch vorgenommene Auswahl der Tageszeitungen „Die Welt“ und „Neues Deutschland“ in Form von „Jahrgangsquerschnitten, die in Intervallen von fünf Jahren aufgenommen wurden und von der Gründung der beiden deutschen Staaten, 1949, bis [...] 1974 reichen“ (LDV-Info 4/1984, 40). In die ca. 4,2 Mio. Textwörter sind alle vorkommenden Textsorten bis hin zu Werbung, Kleinanzeigen und Leserbriefen aufgenommen worden.

Im IDS-Korpus enthalten ist auch das sogenannte LIMAS-Korpus (Sigle: LIM), das am Institut für Kommunikation und Phonetik Bonn erstellt wurde (vgl. Dokumentation Textkorpora des heutigen Deutsch 1982). Es handelt sich hierbei um ein typisches Sample-Korpus, das sich vor allem dadurch von allen anderen Teilkorpora unterscheidet, daß seine Einzeltexte alle auf exakt die gleiche Länge gekürzt sind. Die ca. 1 Mio. Textwörter des LIMAS-Korpus verteilen sich auf 500 Textstücke zu je 2000 Textwörtern aus allen Bereichen der deutschen Gegenwartssprache, d.h. aus 33 Fachgebieten, die nach der Nationalbibliographie 1971/72 ausgewählt und ge-

⁴⁸ Die Vorgeschichte des Korpus geht auf die Bad Homburger Colloquien 1975 und den Plan eines großen interdisziplinären Wörterbuchs der deutschen Sprache zurück. Vgl. Mentrup 1979, 182f.

⁴⁹ Die in den Teilkorpora enthaltenen Einzeltexte sind mit Angaben zur Anzahl der Satzlänge und Zahl der Wortformen aufgelistet in Brückner 1989, Anhang E, E-1 bis E-26. Zum MK1 siehe auch LDV-Info 1/1981, 5-11, zum MK2: LDV-Info 2/1982, 27-33, zum BZK: LDV-Info 4/1984, 38-47.

wichtet wurden; alle Texte sind in den Jahren 1970 und 1971 erschienen.

Das mit gegenwärtig 7,4 Mio. Textwörtern umfangreichste Teilkorpus ist das sogenannte Handbuchkorpus (Sigle: HB), das vor allem im Hinblick auf den Gegenstandsbereich des Lexikons „Brisante Wörter“ erarbeitet wurde und bis Ende 1989 jahrgangsweise (Siglen: H85, H86, H87 usw.) erweitert worden ist. Es bestand bis Sommer 1988 ganz überwiegend aus Artikeln der Tageszeitung „Mannheimer Morgen“ und der Wochenzeitung „Die Zeit“, die nach denjenigen Sach- bzw. Themenbereichen (Politik, Umwelt, Kunst/Kultur, Wirtschaft, Verwaltung/Behörde, Medizin/Gesundheit, Pädagogik/Psychologie, Technik und Sonstiges) ausgewählt wurden, die zu Planungsbeginn als zu behandelnde Lexikon- und Wortschatzbereiche in die engere Wahl gekommen waren. Die Jahrgänge 1985 und 1986 standen für die Arbeit am Lexikon vollständig zur Verfügung; inzwischen sind auch die Jahrgänge 1987 und 1988, ab Jahrgang 1989 auch Artikel aus dem Magazin „Stern“ verfügbar.

Daneben sind zu den Bereichen Politik und Umwelt einzelne Texte wie Parteiprogramme, Verbraucherratgeber v.a. der 80er Jahre ins Handbuchkorpus aufgenommen und maschinell verfügbar gemacht worden.

Als grundsätzlich nicht zur Datenbasis des Lexikons „Brisante Wörter“ gehörend wurden die Teilkorpora „Goethe-Korpus“⁵⁰ (Sigle: GOE), „Thomas-Mann-Korpus“ (Sigle: THM)⁵¹ und das „Freiburger Korpus“ (Sigle: FKO) ausgeschlossen, das aus Transkripten gesprochener deutscher Gegenwartssprache, überwiegend aus öffentlichen Medien, besteht.

3.2. Technische Organisation

Jedes Teilkorpus ist in einzelne Texte untergliedert. Als Texte gelten dabei solche Äußerungen, die von Sprechern des Deutschen als gegenüber dem Ko- und Kontext abgegrenzte, bibliographisch erfaßbare Einheiten wahrgenommen werden und denen konventionalisierte Textsortenmuster zugrundeliegen. Jeder Einzeltext ist durch eine bestimmte Sigle, bestehend aus drei alphanumerischen Zeichen, eindeutig identifiziert. In Abweichung von dieser gemeinsprachlichen Bedeutung von *Text* werden in den Teilkorpora, die Zeitungs- bzw. Zeitschriftenartikel enthalten, aus technischen Gründen meist mehrere – der z.T. sehr kurzen – Texte (Artikel) unter einer gemeinsamen Sigle gebündelt; so werden z.B. alle im Bonner Zeitungskorpus (BZK) ausgewählten Artikel des Jahrgangs 1949 der „Welt“ unter der

⁵⁰ Es enthält die Prosa-Texte der Hamburger Goethe-Ausgabe.

⁵¹ Es ist identisch mit den Gesammelten Werken in 12 Bänden mit einem Nachtragsband, Frankfurt, S. Fischer, 1960 und 1974 (Bd. 13). – Im Lexikon „Brisante Wörter“ finden sich sehr vereinzelt dennoch Belege aus diesem Korpus.

Sigle W49 verwaltet. Aus diesem Grund werden die elementaren Einheiten des Computerkorpus im Sprachgebrauch des IDS als *Dokumente*, nicht als *Texte* bezeichnet. Die belegbegleitenden bibliographischen Angaben berührt die Zusammenfassung mehrerer Texte zu einem Dokument jedoch nicht, sie bleiben erhalten.

Die Funktion der Siglen besteht in erster Linie darin, eine Computerabfrage gezielt auf ein oder mehrere Teilkorpora (z.B. durch Eingabe von MK1 oder MK2) und auf ein oder mehrere Einzeldokumente (z. B. durch Eingabe von LJA für „Uwe Johnson, Das dritte Buch über Achim“, oder von GOS für „Gebrauchsanweisung für Olivetti Schreibmaschine“) beschränken zu können. In gleicher Weise werden die Ergebnisse der Abfragen, sofern es sich um die beiden Arten der Wortformenabfrage und der KWIC-Abfrage (siehe Abschnitte 4.1.1 und 4.1.2) handelt, mittels derselben Siglen markiert.⁵² In zweiter Linie haben die Siglen auch eine mnemotechnische Funktion, die aber bei einem aus drei Zeichen bestehenden Code nur begrenzt sein kann.

Die Funktion, Informationen über die Textsortenzugehörigkeit eines Korpusbelegs zu vermitteln, können die Siglen nur in Ansätzen erfüllen.⁵³ Auf das linguistische Objektivierbarkeitsproblem vor allem einer pragmatischen Textklassifikation wurde schon in Abschnitt 2.3.3 hingewiesen; hinzu kommt, daß es für die unterschiedlichen Nutzungsmöglichkeiten eines Monitor-Korpus nicht ein einziges Klassifikationssystem geben kann: Für den einen Benutzer haben Textgattungsinformationen, für den anderen die exakte Datierung Priorität.

Dennoch sind die an den Siglen abzulesenden z.T. unsystematischen und rudimentären Textinformationen beim ersten Überblick über die Belegsituation eines lexikalischen Ausdrucks von einigem Wert. Im MK1 haben z.B. die jeweils ersten Buchstaben der Sigle eine gattungskennzeichnende Funktion: L steht für „Literatur“, W für „Wissenschaft“, T für „Trivallliteratur“, M für „Memoiren“, Z für „Zeitungen/ Zeitschriften“. Der zweite Buchstabe verweist auf den Autor: G für Grass, aber auch für Grzimek, der dritte Buchstabe auf den Buchtitel. Bei Zeitungen und Zeitschriften verweisen das zweite und dritte Zeichen auf den Namen des Periodikums, z.B. ZBW steht für „Zeitschrift“, „Bild der Wissenschaft“.

Dieses System wurde für das MK2 im wesentlichen beibehalten und auf einige andere Textgattungen ausgeweitet: E für „Erlasse“, G für

⁵² Lediglich im Falle der Abfrageart Satzbelege werden nicht nur die Teilkorpus- und die Dokumentensigle, sondern auch die meist vollständigen bibliographischen Informationen vom Computer mit angegeben.

⁵³ Die Kennzeichnung nach Aspekten „etwa temporaler, regionaler, sozialer, fachbereichlicher und textsortenspezifischen Art“ wurde schon in einer der Bad Homburger Thesen gefordert (Mentrup 1979, 193).

„Gebrauchsanweisungen“ usw.

Im fast vollständig aus Zeitungstexten bestehenden Handbuchkorpus (HB) wurden bei der Siglenvergabe andere Markierungsdimensionen eingeführt. Hier wird außer auf den Namen des Periodikums (M für „Mannheimer Morgen“, S für „Stern“, Z für „Die Zeit“) vor allem auf den Themenbereich verwiesen: U steht für „Umwelt“, K für „Kultur“ neben anderen, sowohl sprechenden als auch nicht-sprechenden Abkürzungen für die vereinbarten Themenbereichskennzeichnungen. Das dritte Zeichen verweist mittels einer Zahlziffer auf den Halbjahrgang, in dem das betreffende Dokument (und damit der einzelne Zeitungsartikel) erschienen ist. So ist z.B. die Sigle UZ1 zu lesen als 'aus einem Artikel der „Zeit“ zum Thema Umwelt im ersten Halbjahr 1985'.

Für die neu zu erstellenden Teilkorpora der 90er Jahre wird zur Zeit die Einführung einer 5-Zeichen-Sigle erwogen, um die Menge der Textinformationen zu einem Beleg erhöhen bzw. spezifizieren zu können.

Menge und Ausführlichkeit der einen Korpusbeleg begleitenden Textinformationen sind abhängig von der Art der Recherche: Wortformenbelege (Abschnitt 4.1.1) und KWIC-Belege (Abschnitt 4.1.2) werden zeilenweise ausgegeben; somit stehen für die Textinformationen nur einige Stellen zu Beginn einer jeden Zeile zur Verfügung, die von der Sigle für das Teilkorpus und das Dokument besetzt werden. Einzig bei der Recherche in Form von Satzbelegen (Abschnitt 4.1.3) werden jedem Einzelbeleg meist vollständige bibliographische Angaben beigelegt. Bei Bedarf steht dem Benutzer eine von der Rechercheform unabhängige Dokumentationsdatei (DOKDAT) zur Verfügung, in der alle bibliographischen Angaben verzeichnet sind.

4. Verfahren der Korpusanalyse

In diesem Abschnitt werden in der Reihenfolge der lexikographischen Praxis die wichtigsten linguistischen Verfahren der Korpusanalyse beschrieben. Die Verfahren orientieren sich an drei Arten der Computer-Recherche, lassen sich aber ebenso auf das in traditioneller Form verfügbare Material anwenden, das dem auf der Basis des Monitor-Konzepts aufgebauten Korpus ebenfalls zuzurechnen ist.

Es soll dabei auch deutlich gemacht werden, wie sich die Bildung von Hypothesen über irgendeinen der zu beschreibenden Aspekte eines lexikalischen Ausdrucks zumeist erst aus einem der ersten Schritte der Korpusanalyse ergibt, dem dann weitere Schritte folgen müssen, um die Hypothesen zu verifizieren. In Abschnitt 4 soll daher auch der Nachweis für die in Abschnitt 2 vertretene Position zum methodologischen Primat des Korpus gegenüber der Idiokompetenz des Linguisten erbracht werden.

4.1. Computer-Recherchen

Das maschinell gespeicherte Gesamtkorpus des IDS ist über das Computerprogramm REFER, „ein interaktives Zugriffssystem für Textkorpora“ (Brückner 1989), zugänglich, d.h. hinsichtlich einer ganzen Reihe lexikologischer Detailfragen auswertbar. Dies geschieht mittels einiger Suchkommandos.⁵⁴ In der lexikographischen Praxis erweist sich schnell, daß sinnvollerweise zuerst diejenigen Suchkommandos gegeben werden, die zu generellen und Überblicksangaben des zu untersuchenden lexikalischen Ausdrucks führen, um dann mit anderen Kommandos schrittweise zu spezielleren Ergebnissen zu kommen. Alle Ergebnisse sind sowohl über den Bildschirm als auch – wahlweise zusätzlich – als Papiaerausdruck verfügbar, je nach individueller Arbeitsweise.

4.1.1. Wortformen-/ Wortbildungsparadigmen

Die erste Art der Recherche bezieht sich auf die maximale für das jeweilige Wörterbuch vorgesehene Datenbasis, d.h. auf sämtliche in Abschnitt 3.1 genannten Teilkorpora. Gefragt wird nach Häufigkeit und Vorkommen eines Ausdrucks insgesamt und in jedem der einbezogenen Teilkorpora; dabei hat der Lexikograph die Wahl, ob diese Angaben ausschließlich für ein Simplex (dessen flektierte Formen werden automatisch erzeugt und mitberücksichtigt) oder für ein Lexem als Simplex u n d als – vorzugsweise linksgebundenes⁵⁵ – Kompositionselement gemacht werden sollen. Sofern die Wortbildungsproduktivität im Wörterbuch nicht ganz unberücksichtigt bleiben soll, wird diese erste Häufigkeitsrecherche sich immer auf ein Lexem als Simplex u n d als Kompositionselement beziehen. Damit erhält man zwei Listen, eine nach Teilkorpora aufgeschlüsselte Liste und anschließend eine nach Dokumenten aufgeschlüsselte wie die folgenden:

⁵⁴ Zur technischen Beschreibung des Programms siehe Brückner 1989.

⁵⁵ Recherchen gleichzeitig zu rechts- und linksgebundenen Kompositionselementen sind möglich, benötigen aber unverhältnismäßig viel Rechenzeit.

Abb. 1

<u>Ergebnis der Anfrage zur Zeichenkette Atom/</u>				
Korpus		Zutreffende Zeichenketten	Häufigkeit absolut relativ	
01	Handbuchkorpus 1985	367	584	0.0140 %
02	Handbuchkorpus 1986	464	920	0.0402 %
03	Handbuchkorpus 1987	100	187	0.0263 %
04				
05				
06				
07				
08				
09				
10				
Summe		931	1691	0.0236 %
Suchbereich: Alle Korpora				

Abb. 2

H87 - UM5	Atomenergie	ABS:	12
H87 - UM5	Atomenergie-Behörde	ABS:	1
H87 - IM5	Atomenergie-Minister	ABS:	1
H87 - UM5	Atomenergie-Organisation	ABS:	3
H87 - UM5	Atomenergieanlagen	ABS:	1
H87 - KM5	Atomexplosion	ABS:	1
H87 - UM5	Atomfabrik	ABS:	1
H87 - UM5	Atomfabriken	ABS:	1
H87 - UM5	Atomfirma	ABS:	1
H87 - IM5	Atomforschung	ABS:	2
H87 - UM5	Atomforschung	ABS:	1
H87 - UM5	Atomforum	ABS:	5
H87 - UM5	Atomforums	ABS:	1
H87 - QM5	Atomfragen	ABS:	1
H87 - FS5	Atomgegner	ABS:	1
H87 - UM5	Atomgesetz	ABS:	9
H87 - UM5	Atomgesetzes	ABS:	5
H87 - UM5	Atomindustrie	ABS:	6
H87 - QM5	Atomkatastrophe	ABS:	1
H87 - UM5	Atomkatastrophe	ABS:	2
H87 - IM5	Atomkern	ABS:	1
H87 - IM5	Atomkerne	ABS:	3
H87 - IM5	Atomkernen	ABS:	2
H87 - IM5	Atomkraft	ABS:	1
H87 - KM5	Atomkraft	ABS:	1
H87 - QS5	Atomkraft	ABS:	1
H87 - UM5	Atomkraft	ABS:	15
H87 - UM5	Atomkraft-Geschäft	ABS:	1
H87 - QS5	Atomkraftgegner	ABS:	1
H87 - UM5	Atomkraftgegner	ABS:	1
H87 - UM5	Atomkraftgegnern	ABS:	1
H87 - UM5	Atomkraftnutzung	ABS:	1
H87 - UM5	Atomkraftpleite	ABS:	1
H87 - QS5	Atomkraftwerk	ABS:	1
H87 - UM5	Atomkraftwerk	ABS:	3
H87 - UM5	Atomkraftwerk-Bauplatzes	ABS:	1
H87 - UM5	Atomkraftwerke	ABS:	9
H87 - IM5	Atomkraftwerken	ABS:	1
H87 - UM5	Atomkraftwerken	ABS:	3
H87 - UM5	Atomkraftwerkes	ABS:	1
H87 - QM5	Atomkraftwerks	ABS:	1
H87 - UM5	Atomkraftwerks	ABS:	3
H87 - UM5	Atomkrafwerk	ABS:	1
H87 - KM5	Atomkrieg	ABS:	1

Den Angaben der ersten Liste (Abb. 1) ist, vorausgesetzt man kennt die ungefähre Datierung der einzelnen Teilkorpora, eine erste Information über den Belegungszeitraum zu entnehmen. Beispielsweise ergibt sich aus den Häufigkeiten der Elemente *Atom-* und *Kern-* die Frage: Wie ist die deutlich höhere Frequenz von *Kern-* in den 80er Jahren und die in älteren Teilkorpora höhere Frequenz von *Atom-* zu erklären und wann hat der Wechsel stattgefunden?

Die Angaben der zweiten Liste (Abb. 2) bieten wichtige Aufschlüsse über das Wortbildungsverhalten und die Wortbildungsproduktivität, und zwar jeweils mit aus den Teilkorpora erschließbaren Belegungszeiträumen korreliert. Außerdem ist durch Vergleich solcher Listen festzustellen, ob es mehrere Schreibvarianten eines Ausdrucks gibt (z.B. *GAU* und *GaU*) und ob eine davon deutlich häufiger, d.h. konventionalisierter ist, z.B. ob als Verbalableitung zu *Recycling* eine der verschiedenen Schreib- bzw. Ableitungsformen (*recyclen*, *recyceln*, *recykeln*, *rezyklieren*) bereits konventionalisiert ist und wie diese dann flektiert wird. Sofern die Listen orthographische und/oder morphologische Unsicherheiten im gegenwärtigen Gebrauch aufzeigen, ist daraus ein in der entsprechenden Artikelposition zu berücksichtigender Erläuterungsbedarf abzuleiten.

Bereits zu diesem frühen Zeitpunkt der Korpusanalyse wird deutlich, daß wesentliche Hypothesen sich erst aufgrund von Korpusdaten überhaupt stellen und daß diese Hypothesen aufgrund der linguistischen Idiokompetenz gar nicht hätten gebildet werden können. Vor allem aber wird aufgrund der Angaben zur absoluten Häufigkeit (Summe der Häufigkeiten in allen Teilkorpora) entschieden, mit welcher Art der Recherche anschließend weiter verfahren wird. Bei einer Häufigkeit von über 500 muß die Datenbasis reduziert werden, da hier eine Sichtung jedes einzelnen Satzbelegs samt Kontext nicht mehr möglich ist.⁵⁶

Innerhalb eines Monitor-Korpus kann die Reduzierung geschehen

- mittels eines in das Programm REFER eingebauten Zufallsgenerators
- durch linguistisch begründete Selektion eines oder weniger Teilkorpora, und weiter durch Selektion bestimmter Dokumente (Einzeltexte)⁵⁷ und anschließender Anwendung der Recherche-Art 'Satzbelege' oder zunächst
- durch Anwendung der Recherche-Art 'KWIC-Belege'.

⁵⁶ Zur Menge der noch bewältigbaren Korpusbelege vgl. auch Blumenthal 1988, 345, der zur Lösung des Quantitätsproblems allerdings „(semi-)automatische Verfahren der Korpusauswertung“ ins Auge faßt.

⁵⁷ Zur Unterscheidung von „Dokument“ und „Text“ siehe oben Abschnitt 3.2.

4.1.2. KWIC-Belege

Key-Word-In-Context-Belege zeigen das Suchwort (auch hier wahlweise als Simplex und/oder Kompositionselement) im unmittelbaren Satzkotext. Der belegte Kotext wird strikt auf eine Zeile begrenzt, an deren Anfang die Siglen des Teilkorpus und des Dokuments stehen und damit minimale Kontextinformationen angeben, und bricht also meistens mitten im Satz ab. (siehe Abb. 3).

Abb. 3

JM5	m ist die keineswegs perfekte Strategie Rita Süsmuths ohne	Alternative	, da die Mehrheit der Experten Zwangsmaßnahmen al
JM5	alten werden) , bieten die Generika gewiß eine verlockende	Alternative	, heißt es weiter .
JM5	-Methoden " gebe es jedoch zum Kondom als Aids-Schutz keine	Alternative	zur Schulmedizin an .
JS5	ngen und Hautauschlägen bietet sich die Naturheilkunde als	Alternative	aufzubieten hat und genau mit jenem geschmäckert
KM5	Opernmuseum , das zehn Jahre nach Chereau immer noch keine	Alternative	zur tristen Eisporthele geschaffen , an die sic
KM5	damit wurde eine	Alternative	zwischen " Demokratie und sozialen Stillstand " g
QM5	gungalos hat Michail Gorbatschow sein Volk noch nie vor die	Alternative	wird die sogenannte " soziale Verteidigung propag
QM5	als	Alternative	gibt .
QM5	es wird viel die Rede davon sein , daß es zu Europa keine	Alternative	für die Hausmüllentsorgung in Südhessen , als die
QM5	nach dem gegenwärtigen Stand sehen CDU und FDP keine	Alternative	Liste (AL) , die ja den Volkszählung-Boykott be
QM5	t " , sagte ein nachdenklicher Delegierter , " wir sind die	Alternative	zu den herrschenden Parteien .
UM5	der grüne Umweltminister betrachtet als einzig mögliche	Alternative	die Stilllegung der Produktion von Brennelementen
UM5	sch Tschernobyl hätte die Kernenergie eine Chance als echte	Alternative	gegenüber anderen Energieträgern .
UM5	zu den Plänen für Sinheim gibt es als	Alternative	die Nutzung der Müllverbrennungsanlage Mannheim i
BM5	r demokratisch verfaßten Staatsgemeinschaft machen könnte ,	Alternativen	erarbeitet .
FM5	um das bewußt zu machen , aber auch um	Alternativen	aufzuzeigen , veranstalteten jetzt die Mannheime
FM5	u+ auf der Suche nach	Alternativen	+ u .
FM5	dem Gesetzgeber jedoch zur Aufgabe , vor jeder Zählung nach	Alternativen	zur Volkszählung sind " noch mit zu großen Fehle
FM5	x+ Punkt + x gibt es	Alternativen	zu suchen .
FM5	bei der Suche nach	Alternativen	? .
JM5	agnachmittagen geht es um natürliche Verhütungsmethoden als	Alternativen	steht die Frage im Vordergrund , ob und durch we
JM5	u+ Suche nach	Alternativen	zur Pille .
JM5	vielmehr sollten	Alternativen	+ u .
JM5	ugendtaucht und Jugend-Sehnsüchte zu wecken , vor allem aber	Alternativen	zur Gentechnologie gründlich erörtert und entwic
JS5	deswegen wurden in letzter Zeit mehrere	Alternativen	aufzuzeigen , solchen Sehn-Süchten aus eigener K
JS5	muß dem Menschen annehmbar , vernünftige und alltagsgemäße	Alternativen	entwickelt .
KM5	mit auf Kreisebene das Scheitern der Zusammenarbeit mit den	Alternativen	anbieten , die die Weiterverbreitung von Aids wi
QM5	den die	Alternativen	in Kauf .
QM5	140 Organisationen , von der grünen	Alternativen	so oft lähmenden Streit um ein rot-grünes Bündni
UM5	um Atomkraft und ihre	Alternativen	Liste über das Kindertheater " Rote Grütze " bis
UM5	CDU) erklärte , eine Politik , die den Ausstieg wolle ohne	Alternativen	zur herkömmlichen Ölheizung und Gasheizung für
UM5	Wirtschaft mit Wachstum und Umweltschutz dürften nicht als	Alternativen	ging es bei der Bürgeraktion Umweltschutz Rhein-
UM5	ionaussage zugunsten der SPD für den 5. April die Wahl der	Alternativen	sah Dorit Wolf etwa im Ausbau des Nah-wärmenetze
UM5	s und Verwendungszweck des Lockes , immer relativ zu vielen	Alternativen	aufzuzeigen , sei " Pfusch " .
UM5		Alternativen	gesehen werden .
UM5		Alternativen	Partei empfehlen .
UM5		Alternativen	betrachten .

Die Zeile auf dem Papierausdruck ist allerdings bedeutend länger als diejenige auf dem Bildschirm (und auch länger als die in Abb. 3); aus diesem und aus weiteren Gründen lassen sich KWIC-Belege besser auf dem Papier auswerten. Dabei sind Aufschlüsse zu gewinnen über:

- Valenz und präpositionale Anschlüsse eines Ausdrucks (z.B. *Alternative zu etwas, etwas nach X* [=Ortsname] oder *in eine Region entsorgen*),
- die Existenz konventionalisierter Kollokationen⁵⁸ und formelhafter Wendungen (z.B. *Versöhnung/ Ausgleich von Ökonomie und Ökologie*),
- die (dominante) Funktion des Ausdrucks innerhalb thematischer/ rhematischer Äußerungsteile und damit über das Vorhandensein einer (dominant) prädzierenden Funktion; das ist wichtig für eine empirisch fundierte Klassifizierung eines Ausdrucks als Schlagwort, Wertungsausdruck o. ä.,
- eine mögliche/ konventionalisierte Kookkurrenz bestimmter Lexemklassen und Textsegmente, z.B. von *Atom*-Kompositen und Zeitungsartikelüberschriften,
- die Verwendung im Rahmen bestimmter Themen- oder Sachbereiche,
- eine eventuell vorhandene metaphorische Produktivität, z.B. von *Entsorgung, entsorgen*, die zu Hypothesen über die Ursachen einer solchen Produktivität anleitet,
- die möglichen Bezugsklassen des Ausdrucks (als Aspekte der Semantisierung und Bedeutungsabgrenzung); damit werden erste und noch vorläufige Hypothesen über die Gliederung des Denotatsbereichs und weiter über den Aufbau des Wortartikels möglich.

Im Verlauf der Analyse von KWIC-Belegen vor allem unter diesen genannten Aspekten kann, falls notwendig, die Datenbasis für die anschließend durchzuführende Recherche in Form von Satzbelegen reduziert und begründbar spezifiziert werden.

4.1.3. Satzbelege

Ein Satzbeleg wird computerbezogen definiert als Zeichenfolge, die von zwei Satzendeckpunkten eingeschlossen ist. Diese Definition setzt eine von der üblichen Interpunktion abweichende Schreibkonvention voraus, bei der zwischen das letzte Wort eines Satzes und das abschließende Satzzeichen ein blank (Leerzeichen) eingefügt und auch die mit Frage- oder Ausrufezeichen endenden Sätze zusätzlich mit der Zeichenfolge 'blank plus Punkt

⁵⁸ Siehe Abschnitt 4.4, vgl. Viehweger 1988, 108f.

plus blank' abgeschlossen werden. Alle ins Korpus aufzunehmenden Texte müssen dieser und noch einigen anderen Anpassungen unterzogen werden, um recherchierbar zu sein. (LDV-Info 5/1985, 42-81 und Brückner 1989a, 24-30).

Ein Satzbeleg zeigt das Suchwort innerhalb des Ko- und Kontextes, der von zwei Satzendepunkten begrenzt wird. Bei der Durchsicht der einzelnen Satzbelege der gewählten Teilkorpora oder Dokumente kann dieser Ko- und Kontext in beide Richtungen um beliebig viele Sätze erweitert werden. Erst dadurch wird die wesentliche Leistung dieser Recherche-Art, die endgültige Semantisierung durch Analyse verschiedener Kontextfaktoren, erreicht.

Allgemeingültige Kriterien zur Bestimmung des für eine semantische Analyse optimalen Belegumfangs, wie sie z.B. von Sinclair gefordert werden,⁵⁹ sind m.E. undenkbar, und zwar deshalb, weil die semantische Analyse unter einer Vielzahl von Aspekten betrieben werden kann, die in Abhängigkeit vom jeweiligen wissenschaftlichen Handlungsziel, z.B. dem Wörterbuchtyp, stehen und die Analyse deshalb nicht objektivierbar, allenfalls intersubjektiv transparent gemacht werden kann. Alle Versuche, sich dem Ziel des optimalen Umfangs eines Korpusbelegs rein mechanisch-technisch zu nähern,⁶⁰ verfehlen den letztlich hermeneutischen Charakter der semantischen Analyse. Die nachfolgenden Fragen, an denen sich die Analyse der Satzbelege orientiert, sollen der Herstellung der genannten intersubjektiven Transparenz dienen:

- Welches ist der Extensionsbereich und welches sind die Extensionsklassen des Ausdrucks? Die Feststellung geschieht durch Auflistung aller den Einzelbelegen zu entnehmenden Bezugsobjekte des Ausdrucks und ihre anschließende Klassifizierung.
- Welche expliziten und impliziten Prädzierungen werden über die Bezugsobjekte des Ausdrucks gemacht? (Feststellung durch Auflistung wie oben) Welche davon sind als usuell/typisch und welche als okkasionell zu klassifizieren?
- In welche Isotopielinien⁶¹ (pronominale Ketten) ist der lexikalische Ausdruck eingefügt, welches sind die anderen lexikalischen Glieder dieses semantischen Paradigmas und welche begriffshierarchischen

⁵⁹ „Criteria for evaluating citations are urgently required, closely followed by criteria for estimating the optimum length of a given citation.“ (Sinclair 1985, 86). Vgl. Tollenaere 1973.

⁶⁰ Vgl.: „to save waste of time reading unnecessary context“ Tollenaere (1973, 27), sowie Blumenthal (1988, 345f).

⁶¹ Textisotopien werden mit Hilfe graphischer Darstellung erläutert in Agricola 1976, 26; vgl. Kallmeyer et al. 1974, 101.

Relationen zwischen den Gliedern werden textimmanent konstituiert? Z.B. gilt die Relation der *differentiae specifica* *Störfall* und *Unfall* gegenüber dem *genus proximum* *Vorkommnis* nur in juristischen Texten; hingegen sind in öffentlichen Texten die Ausdrücke *Störfall*, *Betriebsstörung*, *Beinahe-GAU* nicht begriffshierarchisch, sondern evaluativ unterschieden. Ist eine Menge textimmanenter Relationen konstatiert (durch Auflistung wie oben), können die textsorten- und kommunikationsbereichsspezifisch usuellen/typischen von den okkasionellen unterschieden werden (Ausführlicheres in 4.3 und 4.4 zu Position [8]).

- Wird der Ausdruck selbst unmittelbar zur Prädisierung welcher Bezugsobjekte eingesetzt? Welcher Art (kennzeichnend, charakterisierend, einstuftend, auf- oder abwertend) ist die Prädisierung?
- An welche Themen- bzw. Sachbereiche, Sprechergruppen, Zeitstufen, Textsorten, Äußerungsabsichten usw. sind die jeweiligen Extensionalisierungen und Prädisierungen gebunden?

4.2. Einbeziehung nicht-computerverfügbaren Materials

In Ergänzung zum maschinell gespeicherten Textkorpus ziehen Lexikographen zumeist noch eigene Sammlungen von Belegmaterial heran. Methodologisch gesehen kann dieses Material als in das Korpus des Wörterbuchs integriert gelten, wenn es im wesentlichen die gleichen Bedingungen erfüllt, wie sie für das Computerkorpus festgelegt wurden.

Erstens müssen sich die zusätzlichen Sammlungen auf dieselbe Grundgesamtheit wie die des Computerkorpus beziehen. Zweitens müssen sie aus denselben elementaren Einheiten bestehen; im Falle des für Kulturwortschatzwörterbücher vorgeschlagenen Konzepts heißt das, daß die einzelnen Belege Textstatus haben müssen, nicht bloß Wort- oder Satzstatus. Z.B. genügt es dann nicht, Einheiten in Form von einigen Zeilen Belegtext mit Quellenangabe, Datum und Seitenzahl zu sammeln; es müssen, wenn schon nicht der gesamte Text, z.B. ein Zeitungsartikel gesammelt worden ist, mindestens Angaben über Titel, Untertitel, Autor, eventuell über die Rubrik (Nachrichtenteil, Wirtschaftsteil, Feuilleton, usw.) und über die Textsorte (Meldung, Kommentar, Interview, Buchbesprechung, usw.) mit enthalten sein. Ferner müssen sich diese Belege den Textsortenklassen zuordnen lassen, die auch der Organisation des Computerkorpus zugrundeliegen. Dabei ist zu überlegen, ob das zusätzliche Material eher die zentralen Textsortenklassen (siehe oben 2.5.2) hinsichtlich z.B. ihres Meinungsspektrums erweitern soll, oder ob es eher die peripheren Textsortenklassen, z.B. die eher fachinternen, vertreten soll, die im Computerkorpus tendenziell unterrepräsentiert sind.

Von den drei auf computerverfügbares Material anwendbaren Analysever-

fahren ist nur eines, nämlich das der kontextualisierten Satzbelege ebenso auf nicht maschinell gespeichertes Material anwendbar, was wegen der überschaubareren Belegmenge kein Nachteil ist. Die Zusatzbelege werden also erst im Stadium der Feinanalyse hinzugezogen.

4.3. Analyse-Beispiel

Im folgenden soll am Beispiel des Stichworts *Störfall* veranschaulicht werden, wie zum einen die Unterscheidung verschiedener Verwendungsweisen, hier **Störfall 1a**, **1b** und **2**, in der Arbeit mit dem Textkorpus entsteht, und zum andern, auf welchen Analyseergebnissen die später im Wortartikel gemachten Aussagen z.B. über Bedeutungsähnlichkeit, Sprechergruppenspezifika und über metaphorische Verwendungen beruhen.

Im Artikel **Störfall** (Strauß/ Haß/ Harras 1989, 524-531) wird unterschieden zwischen einerseits der Verwendung des Ausdrucks durch Fachexperten bzw. Industrievertreter und Politiker mit ähnlichen Interessen (**Störfall 1**) und andererseits der Verwendung des Ausdrucks in der allgemeinen öffentlichen Diskussion (z.B. durch Journalisten) (**Störfall 2**); die fachbezogene Verwendung ist wiederum dadurch unterteilt, daß man sich mit *Störfall* auf zwei ganz unterschiedliche Arten von Ereignissen beziehen kann (**1a** und **1b**).

Der folgende Korpusbelegsatz:

Die Eintrittswahrscheinlichkeit eines Störfalls in der Bundesrepublik ist durch technische und organisatorische Maßnahmen sicherlich verringert worden. (Die Zeit 10.5.1985, 35)

läßt erst durch die zugehörige Artikelüberschrift *Wege zur sanften Chemie* erkennen, daß *Störfall* auf Ereignisse in Chemiebetrieben bezogen werden kann. Der innerhalb des Korpuscontextes nachfolgende Satz:

Seit Seveso sind zusätzliche Maßnahmen wie Störfallanalysen hinzugekommen.

weist durch den Ortsnamen auf ein bestimmtes, offensichtlich exemplarisch bedeutsames Ereignis hin, das vielleicht zuerst als *Störfall* bezeichnet wurde. Dies kann durch sachgeschichtliche Nachforschungen innerhalb des Sekundärkorpus geklärt werden.

Der Korpusbelegsatz:

Auch 1984 gab es weder einen Störfall noch einen Unfall. (Mannheimer Morgen 14.8.1985, 85)

und auch die dazugehörige Artikelüberschrift *Sicher und sauber* geben den Bezugsbereich nicht eindeutig zu erkennen. Dies gelingt erst durch

Berücksichtigung der beiden innerhalb des Korpus textes vorhergehenden Sätze:

Der jährliche Bericht des Bundesinnenministers über besondere Vorkommnisse in Kernkraftwerken unterstreicht einmal mehr: Deutsche Reaktoren arbeiten sicher und sauber.

Die beiden letzten Belegsätze bieten auch wesentliche Informationen zur Verwendungsspezifik und zu semantischen Relationen: Erstens handelt es sich bei diesem Zeitungsartikel um die Wiedergabe eines ministeriellen Textes, bei der auch ohne wörtliches Zitat Wahl und Verwendung der Wörter durch den ursprünglichen Sprecher bzw. Autor genau wiederholt werden. Zweitens führt die Textanalyse zur Bildung einer Hypothese, daß nämlich in Verbindung mit dieser Sprecherposition eine bewußte Unterscheidung zwischen *Störfall* und *Unfall* gemacht wird, eine Hypothese, die allein aufgrund der linguistischen Sprachkompetenz nicht hätte gebildet werden können. Zur Verifizierung der vermuteten terminologisierenden Unterscheidung kann die fachspezifische Literatur des Sekundärkorpus herangezogen werden.

Außer zu *Unfall* besteht eine semantische Relation⁶² zu dem vom Minister verwendeten Ausdruck (*besondere*) *Vorkommnisse*, der im weiteren Kontext der hier wiedergegebenen Sätze mit den immer allgemeiner werdenden Ausdrücken *gemeldete Vorkommnisse*, *meldepflichtige Tatbestände*, *in fünf Fällen* eine Koreferenzebene bildet. Auf die Feststellung der Koreferentialität folgt die semantisch-pragmatische Differenzierung der Glieder dieses Paradigmas; z.B. bilden innerhalb des folgenden Korpusbelegs *Störfall* und andere Ausdrücke ein koreferentielles Paradigma:

Dies bestätigte am Mittwoch der Hanauer Oberstaatsanwalt Albert Farwick. Seinen Ermittlungen zufolge ereignete sich der Störfall in Philippsburg am 1. Juli 1983. Der Vorfall ist nach seiner Kenntnis ordnungsgemäß gemeldet worden. Bei dem Zwischenfall sind nach den Ausführungen des Oberstaatsanwaltes sechs Metallteile eines defekten Kernbrennstabes sowie ein Pellet (Brennstofftablette) in ein 14 Meter tiefes Loch gefallen. (Frankfurter Rundschau 25.2.1988, 1)

Als nächstes stellt sich die Frage, wie von einzeltextbezogenen Koreferenzbeziehungen auf irgendeine Art sprach- und varietätenbezogene Bedeutungsverwandtschaft⁶³ geschlossen werden kann. Selbst wenn sich eine Ausdrucksreihe wie *Störfall*, *Vorfall*, *Zwischenfall* in vielen Texten auffal-

⁶² Die Termini Synonymie, Autonymie usw. werden, weil sie traditionell an die strukturelle lexikalische Semantik gebunden sind, hier bewußt nicht verwendet.

⁶³ Die Ausdrücke *bedeutungsverwand* und *sinnverwand* gehören der Beschreibungssprache des Lexikons „Brisante Wörter“ an.

lend wiederholt, kann dies lexikographisch nicht als textunabhängig verstandene Synonymie (vom Typ *Metzger* – *Fleischer*) dargestellt werden. Bei expliziter und diskursiver Darstellungsweise besteht aber die Möglichkeit, die verschiedensten semantischen Relationen in ihrer Bindung an Kommunikationsbereich, Textsorte und/ oder Sprechergruppe zu erläutern und dadurch wesentliche Bedeutungsaspekte eines Ausdrucks sichtbar zu machen. Dies gilt besonders dann, wenn einer Reihe wie *Störfall*, *Vorfall*, *Zwischenfall* noch eine zweite, aus den Koreferenzebenen anderer Texte oder Sprechergruppen ableitbare Reihe gegenübersteht, so wie es bei der aktuellen Berichterstattung über Tschernobyl der Fall war:

Unter der Überschrift *Tschernobyl ist kein Einzelfall* heißt es: *Über das Ausmaß des Unfalls in dem sowjetischen Kernkraftwerk Tschernobyl hat die Nachrichtenagentur TASS nur indirekte Angaben gemacht; sicher scheint, daß es sich nicht um einen kleineren Störfall handelt. Nachfolgend die schwersten, bisher bekannt gewordenen Unfälle in der Geschichte der friedlichen Nutzung der Kernenergie.* (Mannheimer Morgen 29.4.1986, 2) und einen Tag später unter der Überschrift *Fünf Zeilen, die die Welt erschüttern: Von Notstandsmaßnahmen, die auf eine Katastrophe tragischen Ausmaßes hindeuten, war indessen zunächst nichts zu hören. Andererseits konnte es auch kein kleiner Störfall gewesen sein. Dagegen sprach der massive Ausstoß an Radioaktivität.* (Mannheimer Morgen 30.4.1986, 2)

Hier werden die Ausdrücke *Unfall*, *Störfall* und *Katastrophe* als bedeutungsähnliche verwendet und nach dem Gesichtspunkt der abgestuften Schwere des bezeichneten Ereignisses unterschieden; dabei wird *Störfall* am unteren Ende einer Skala eingeordnet, erkennbar an den Epitheta *klein*, *kleiner*, und *Katastrophe* am oberen Ende.

Betrachtet man insgesamt die beträchtliche Menge von Zeitungsartikeln, die das Textkorpus über das Ereignis Tschernobyl enthält, so tritt der Bezug zwischen *Störfall* und *Katastrophe* semantisch als Anfangs- und Endpunkt einer Skala von Einstufungsprädikaten ganz deutlich hervor (um die Koreferenzebene hervorzuheben, ist das folgende Textbeispiel stark gekürzt):

Bei dem Reaktorunglück von Tschernobyl ist einer amerikanischen Untersuchung zufolge [...] In einer weiteren Studie wird Zweifel an der Wirksamkeit des sowjetischen Plans geäußert, den Unglücksreaktor mit Beton zu versiegeln [...] Westliche Kernreaktoren sind [...] „eindeutig unfallanfälliger“ als die sowjetischen Reaktoren. Das geht aus den Unfallanalysen des Reaktorunglücks von Tschernobyl hervor [...] Bei der Reaktorkatastrophe in Tschernobyl sind Webb zufolge [...] geht es um die international festgelegte gegenseitige Verpflichtung zur frühestmöglichen Unterrichtung bei Störfällen [...] Dem SPD-Fraktionsvorsitzenden Hauff dagegen gehen

die geplanten Abmachungen nicht weit genug [...] die Staaten müßten sich gegenseitig schon über ihre Planungen bei Atomanlagen unterrichten, sich zu einer internationalen Haftung bei Unfällen bereit erklären und gemeinsame, grenzüberschreitende Katastrophenpläne erarbeiten. (Süddeutsche Zeitung 24.9.1986, 8)

Die Reihe *Störfall, Unglück, Unfall, Katastrophe* wird in diesem Text von einem Politiker verwendet, Hauff, der sich seinerzeit als Kernkraftgegner zu profilieren suchte. Hingegen wird die Reihe *Störfall, Vorfall, Zwischenfall, Betriebspanne, Vorkommnis* typischerweise – das kann aufgrund der hohen Belegzahl gesagt werden – in Texten von Kernkraftbefürwortern sowie von Vertretern und Verteidigern der Chemieindustrie verwendet. Es lassen sich also beide Reihen in Beziehung setzen zu den unterschiedlichen Problemperspektiven der Sprecher- und Interessengruppen: Was für die einen als „allgemeine Bedrohung, Gefahr“ gilt, ist für die anderen „Störung eines technischen Ablaufs“. Diese kontroversen Problemperspektiven werden auch in anderen Texten sowohl des Primär- als des Sekundärkorpus z.T. sehr ausdrücklich artikuliert; die textanalytisch gewonnene semantische Hypothese läßt sich dadurch weiter stützen.

Für die Analyse von Wertungsfunktionen sind vor allem solche Texte hervorzuheben, mittels derer in besonders expliziter Weise Bewertungen oder Assoziationen vermittelt werden; zu ihnen gehören Sprachglossen: *Eine Riesenkatastrophe ist ein GAU, eine weniger große heißt „Störfall“*. (Badische Zeitung 11.9.1980). Die Bedeutung solcher Texte für die Interpretation von Charakterisierungsfunktionen eines Ausdrucks besteht nicht nur darin, daß sie Bewertungen explizit und tendenziell zu Topoi machen, sondern auch darin, daß sie sie in der öffentlichen Diskussion multiplizieren.

Ähnlich, nur weniger explizit, vermitteln und etablieren metaphorische Verwendungen, welche Art von Charakterisierung oder Bewertung mit einem Ausdruck verbunden sein kann.

Der Anstoß zu metaphorischen Verwendungen geht offensichtlich gerade von der Umstrittenheit eines Ausdrucks als „richtige“ oder „falsche Bezeichnung“ für etwas aus. Unter der Überschrift *Bundesrat ändert Störfallverordnung* meldet Die Welt am 27.2.1988 (S.4): *Der Bundesrat hat am Freitag an der neuen Störfallverordnung Änderungen vorgenommen, bevor er zustimmte. Über dieselbe Angelegenheit berichtet auch die „tageszeitung“; sie bedient sich dabei allerdings einer zweideutigen Überschrift: Störfälle im Bundesrat; dann heißt es: Die gesetzlichen Konsequenzen aus der Brandkatastrophe beim Schweizer Chemiekonzern Sandoz am 1. November 1986 sind in der Bundesrepublik differenzierter ausgefallen als von der Regierung vorgesehen. Der Bundesrat [...] nahm am Freitag vor seiner Zustimmung zur Neufassung der Störfallverordnung rund 30 Änderungen vor.* (taz 27.2.1988, 4). Hier wird *Störfall* in einer noch sehr

versteckten Weise metaphorisch verwendet, um aus der Sicht der Bundesregierung das Handeln des Bundesrats als Betriebsstörung des Regierungsbetriebs zu charakterisieren; d.h. die "tageszeitung" vergleicht das Demokratieverständnis der Regierung mit der Dienstauffassung eines Technikers, dessen oberstes Ziel der reibungslose Betrieb seiner Maschine ist. Mit der metaphorischen Verwendung von *Störfall* wird also der Gesichtspunkt 'Unterbrechung einer ansonsten reibungslosen technischen Routine' hervorgehoben.

Nicht mehr verdeckt, sondern offen ist die Metapher *Störfall* in: *Die heimlichen Basler Herren haben Schwierigkeiten mit dem, was sie in ihren Zirkeln womöglich Störfälle im örtlichen Bewußtsein nennen. Plötzlich hagelte es 500 Einsprüche gegen die geplante Sondermüllverbrennung.* (Natur 9/1987, 25). Hier wird mit *Störfall* das Handeln von scheinbar Unterlegenen aus der Sicht von Machthabern charakterisiert, durch das diese in der bisher reibungslosen Ausübung ihrer Macht gestört und deshalb in ihrem Selbstverständnis verstört werden. Durch die Metapher wird die technische Routine betont, mit der Konzern-„Herren“ ihre Beziehungen zur Öffentlichkeit abwickeln wollten.

In einem weiteren Korpusbeleg wird mit der metaphorischen Verwendung des Ausdrucks *Störfall* im Sinne von 'Durchbrechen einer Routine' die Meinung als falsch entlarvt, etwas wie z.B. die Demokratie funktioniere wie eine Routine und man könne ihrer sicher sein: [...] *ein Blick nach Schleswig-Holstein auf die sogenannte „Kieler Affäre“ beweist überdeutlich, wie zerstörerisch Macht zu wirken vermag: auf die Ausübenden und auf die sie Ertragenden. Freilich: der Fall Barschel ist kein Super-Gau, nur ein kleiner „Störfall“, doch macht er klar sichtbar, mit welcher latenten Bedrohung man auch in Demokratien zu rechnen hat.* (Günter Kunert in Die Zeit vom 12.2.1988, 42 unter der Überschrift: *J'accuse oder über den Sinn des Protestierens.*)

Metaphorische Verwendungen bewirken die stärkere Betonung eines der zentralen Bedeutungsaspekte eines Ausdrucks, die mit einer bestimmten kritischen Einstellung der Sprecher diesem Aspekt gegenüber verbunden sein kann. Dies kann auch die nicht-metaphorischen Verwendungen beeinflussen, wie eine die Chronologie der Korpus Texte berücksichtigende Analyse zeigt. In späteren Texten treten zuvor eher nebensächliche Bedeutungsaspekte (wie der Bezug auf eine technische Routine im Falle von *Störfall*) stärker hervor.

Die Analysebeispiele dieses Abschnitts sollten deutlich gemacht haben, daß auf der Basis eines Korpus, das umfangreich genug ist und dessen Textsorten ein definiertes Spektrum abdecken, Hypothesen über alle im Hinblick auf das anvisierte Wörterbuch relevanten Verwendungen samt ihres semantischen Zusammenhangs gebildet und auch hinreichend verifiziert werden

können. Es sollte ferner deutlich geworden sein, daß handlungssemantisch ausgerichtete Lexikographie (und Lexikologie) in ganz besonderer Weise korpusfundiert sein muß.

4.4. Analyse-Ergebnisse im Hinblick auf die einzelnen Informationsarten

Gegenüber Abschnitt 4.3 wechselt nun die methodologische Perspektive, denn es sollen die Verfahren der Korpusanalyse von den Informationsarten her erläutert werden, die das Informationsprogramm eines kulturwortschatzbezogenen Wörterbuchs bilden. Es soll am Beispiel des Informationsprogramms des Lexikons „Brisante Wörter“ gezeigt werden, wo und in welcher Weise der Lexikograph das Korpus heranzieht, wenn er die für einen Wortartikel vorgesehenen Positionen mit den notwendigen Informationen füllen will.

Die lexikographisch-diskursive Darstellung des Lexikons „Brisante Wörter“ orientiert sich an der folgenden Liste aus fünfzehn teils obligatorischen, teils fakultativen Informationsarten, die mit bestimmten Artikelpositionen (oder Textbausteinen) korreliert sind, wobei mehrere Informationsarten in einer Artikelposition zusammengefaßt sein können (Strauß 1988):

- [1] Lemma
- [2] grammatische und orthographische Angaben
- [3] Vorspann zur eventuellen Verdeutlichung wortfamiliärer oder etymologischer Zusammenhänge
- [4] Angaben zur semantischen Funktionsklasse
- [5] Angaben zum Bezugsbereich (Extension) und ggf. zur Unterscheidung der Bedeutungspositionen (Artikelgliederung)
- [6] Angabe der Bedeutung im engeren Sinn (Intension)
- [7] Angaben zu speziellen, z.B. sprechergruppengebundenen Verwendungsweisen einschließlich ihrer kommunikationsgeschichtlichen, argumentativ-rhetorischen und sachbezüglichen Hintergründe
- [8] Angaben zu paradigmatischen semantischen Relationen
- [9] Angaben zu syntagmatischen Relationen und zur Valenz
- [10] Angaben zu Textvorkommen und Textsortenspezifik

- [11] Sprachkritischer Kommentar und Angaben zur öffentlichen Sprachreflexion
- [12] Angaben zu Wortbildung und ggf. zu Sublemmata
- [13] Enzyklopädische und wortgeschichtliche Informationen
- [14] Beispielbelege
- [15] Literaturhinweise

Zu Position [1] Lemma: Die Auswahl der Lemmata kann für ein Kulturwortschatzwörterbuch nicht unmittelbar auf der Basis des Korpus, etwa durch Frequenzlisten, vorgenommen werden; allerdings eignen sich Korpus-texte wie Sprachglossen oder Medienkritiken, das heißt also Texte, in denen Kommunikations- und sprachliche Ausdrucksweisen der öffentlichen Diskussion über ein Thema wie z.B. Umwelt thematisiert werden, sehr gut als Gradmesser für die „Brisanz“ bestimmter Ausdrücke und lassen sich daher zur Überprüfung der auf andere Weise gewonnenen Lemmaliste heranziehen.

Für die Festlegung der zu lemmatisierenden Form bei Vorliegen von Schreibvarianten ist das Korpus unmittelbar nutzbar, weil es Auskunft über den Konventionalisierungsgrad, über die mehr oder weniger weit fortgeschrittene Integration eines Fremdworts usw. geben kann. Solche Ergebnisse können als explizite differenziertere Information in der Artikelposition [2] gegeben werden.

Zu Position [2] grammatische und orthographische Angaben: Ein Korpus ist insbesondere im Hinblick auf eine deskriptiv verfahrenende Fremdwort- und Neologismenlexikographie wertvoll, weil man ihm Informationen sowohl über die noch unabgeschlossene Integration eines Lexems als auch über die daraus resultierenden Probleme und Unsicherheiten der Sprachteilhaber entnehmen kann. Voraussetzung dafür ist allerdings, daß das Korpus einen gewissen Umfang hat.

Zu Position [3] Vorspann: Dazu wird eher das Sekundärkorpus herangezogen. Allerdings können die mit dem Computer erstellten Wortbildungspatrigmen (siehe oben 4.1.1) sicherstellen, daß bei den Angaben in dieser Position keine wichtige Wortfamilienbeziehung übersehen wird.

Zu Position [4] Angaben zur Funktionsklasse: Satzbelege (siehe oben 4.1.3) und zum Teil bereits KWIC-Belege (siehe oben 4.1.2) lassen aufgrund der Position des Ausdrucks innerhalb von Thema-Rhema-Strukturen, seiner Position innerhalb von Koreferenzebenen, aufgrund des möglichen Vor-

kommens in Schlagzeilen und Überschriften und aufgrund des gesamten kommunikativen Rahmens erkennen, ob der Ausdruck dominant zum Bezeichnen, Kennzeichnen, Charakterisieren, Bewerten (Aufwerten/ Abwerten), oder Einstufen verwendet wird.

Zu Position [5] Angaben zum Bezugsbereich (Extension): Die Bezugsklassen sind empirisch ermittelbar, indem aufgrund von Satzbelegen eine Liste aller oder doch einer größeren Zahl von Bezugsobjekten erstellt und anschließend klassifiziert wird. Die Unterscheidung verschiedener Bezugsklassen kann dann die Basis für die Gliederung des Wortartikels nach Bedeutungspositionen bilden, sofern das lexikographische Konzept nicht andere Gliederungsprinzipien vorsieht.

Zu Position [6] Angabe der Bedeutung im engeren Sinn (Intension): Die Korpusanalyse besteht hier im wesentlichen in Verfahren der semantischen Bestimmung von Lexikoneinheiten, die das Korpus in ihrer ursprünglichen Vertextung repräsentiert. In einem ersten Schritt der Analyse werden die textuellen Bedeutungen sortiert und als Vertextungstypen klassifiziert. Auf der Basis dieser Klassen werden in einem weiteren Schritt die lexikalischen Bedeutungen als Abstraktionsprodukt formuliert.⁶⁴

Diese strikte Methode der Semantisierung mittels des Korpus wird in der lexikographischen Praxis zumeist flankiert durch das vorbereitende und vergleichende Heranziehen der Bedeutungsangaben in existierenden (Gebrauchs-)Wörterbüchern (Duden, Wahrig u.ä.). Um beide Verfahrensweisen miteinander zu verbinden, ist es m. E. sinnvoll, kontrastiv zu verfahren, das heißt, bei der Korpusanalyse darauf zu achten, inwieweit sich Abweichungen zu den Bedeutungsangaben der herangezogenen Wörterbücher erkennen lassen. Es ist sicherzustellen, daß die nach Vertextungstypen organisierte Klassifizierung der Korpusbelege nicht von vorneherein auf ein Zur-Deckung-Bringen mit den Klassifikationen der Bedeutungsangaben vorhandener Wörterbücher hin angelegt ist.

Zu Position [7] Angaben zur Verwendungsspezifik: Zur empirischen Ermittlung von gruppen- und situationsspezifischen Verwendungsweisen, von der Verwendung des Lexems innerhalb typischer Argumentationen oder „Sprachspiele“ müssen vor allem der textuelle und kommunikative Rahmen der einzelnen Satzbelege herangezogen werden; das heißt, bei dieser Informationsart kommt die Textualität des Primärkorpus besonders zum Zuge. Durch reine Belegkorpora könnte eine Position wie [7] nicht empirisch fundiert werden.

Zu Position [8] Angaben zu paradigmatischen semantischen Relationen und [9] Angaben zu syntagmatischen Relationen: Beide Positionen, beson-

⁶⁴ Darstellung des dort Monosemierung genannten lexikologischen Verfahrens siehe Kallmeyer et al. 1974, 132f.

ders Position [8], wird hier ausführlicher als die übrigen Positionen behandelt, weil sich das Ziel einer handlungssemantisch orientierten Kulturwortschatzlexikographie in besonderer Weise auf die analytische Erschließung semantischer Relationen aus einem Textkorpus auswirkt.

Die Grundannahme, „daß der Wortschatz Struktur hat“ (vgl. Lutzeier 1985, 73), daß es also semantische Relationen zwischen lexikalischen Ausdrücken über eine kotextuelle Verknüpfung hinaus gibt, „kann unter Umständen [...] ohne Klarstellung des [von Seiten des Sprechers, U.H.] beabsichtigten Hintergrundes nicht behauptet werden.“ (Lutzeier 1985, 76).⁶⁵ Die Aussage, zwei lexikalische Ausdrücke stünden in einer bestimmten semantischen Relation zueinander, ist mit Lutzeier (1985, 77) abhängig vom Vorhandensein eines gemeinsamen „Aspektes“ oder „Bedeutungsrahmens“ und von der Kompatibilität jedes der lexikalischen Ausdrücke mit diesem Aspekt, der zumindest teilweise außersprachlich, mit der kommunikativen Situation gegeben ist. Wolski fordert für die Lexikographie einen ko- und kontextbezogenen Synonymitätsbegriff: „Es muß für entsprechende monosemierte Sprachzeichenkandidaten ein Geltungsbereich bzw. Bezugsbereich angegeben werden, auf den Urteile der Art „X ist bedeutungsgleich mit Y“, „X ist bedeutungsähnlich mit Y“ bezogen und damit auch relativiert sind.“ (Wolski 1989, 616). Es ist beobachtbar, daß, sich der „Bedeutungsrahmen“ (Lutzeier) bereits mit dem Übergang von einer Textsorte zur anderen wandeln kann (Lutzeier 1985, 77) und daß sich demzufolge auch die an diesen Rahmen gebundenen semantischen Relationen zwischen lexikalischen Ausdrücken ändern. Damit soll die lexikographische Beschreibung von semantischen Relationen auf der Basis textanalytischer Verfahren als theoretisch legitimiert gelten, und es geht um die methodologische Frage: Was entspricht den systematischen paradigmatischen und syntagmatischen Relationen, in denen ein lexikalischer Ausdruck steht, auf der Textebene, d.h. im Korpus?

Es gibt mehrere Arten korpusanalytisch erfaßbarer Kontextrelationen: (a) paradigmatische, dazu gehören vor allem Koreferentialisierungen mittels semantischer Äquivalente (vgl. Agricola 1976, 26) und (b) syntagmatische, d.h. Kollokabilität.

Das handlungssemantische Konzept semantischer Äquivalenz oder von „Bedeutungsverwandtschaft“ bezieht sich auf das in der Textlinguistik entwickelte Modell der Isotopie. Isotopieebenen schließen Koreferenzebenen ein und sind nach Kallmeyer et al. (1974, 101) allgemein bestimmbar als „den Texten zugrunde liegende semantische Orientierungsebenen“, die die textuellen Bedeutungen der lexikalischen Ausdrücke bestimmen

⁶⁵ Kritisches zum strukturalistischen Konzept der Antonymie mit einem Hinweis auf die außersprachlichen Bedingungen paradigmatischer semantischer Relationen schon bei Agricola 1972, 81f.

oder zumindest weitgehend beeinflussen. Isotopien werden durch Pro-Relationen (Terminus Braunmüller 1977) sowie durch Abgrenzungs-, Gegenüberstellungs-, Vergleichs- und Paraphrasierungsrelationen u.ä. konstituiert und bilden lexikologisch analysierbare Vertextungsketten lexikalischer Ausdrücke.⁶⁶

Ausgehend von der Wortartikelposition [8] stellt sich die korpusanalytische Aufgabe, alle Glieder eines Koreferenzparadigmas zu ermitteln und anschließend die für den zu beschreibenden Kommunikationsbereich charakteristischen Paradigmenglieder zu vergleichen und semantisch-pragmatisch gegeneinander abzugrenzen. Dabei kann man sich jene Unterscheidung zwischen syntaktischen, semantischen und pragmatischen Pro-Relationen (Braunmüller 1977, 146ff) zunutze machen. Pragmatische Pro-Relationen beruhen auf situativem und enzyklopädischem Wissen und haben vor allem eine prädikative Funktion (ebenda 181),⁶⁷ die z.B. sichtbar wird, wenn der Ausdruck *Störfall* in Texten der Sprechergruppe A mittels *Vorkommnis*, in Texten der Sprechergruppe B mittels *Beinahe-Katastrophe* wiederaufgenommen und so zwei vor allem pragmatisch zu unterscheidende Koreferenzparadigmen konstituiert werden.

Solche Paradigmen können, vor allem wenn sie konventionalisiert werden, kulturell relevante Konzepte repräsentieren, wie dies für die Kette *Natur, Umwelt, natürliche Lebensgrundlagen, Wasser/Boden/Luft* gilt. Die Setzung von Koreferentialität als Polarität („Antonymie“), Ergänzung („Komplenymie“), Skalierung u.a. ist Teil der kommunikativen Strategie von Sprechern und trägt zur Konzeptualisierung der Wahrnehmung bei.

Außer der Isotopie gibt es im Bereich der Textlinguistik weitere, für korpusanalytische Zwecke nutzbare Modelle. Genannt sei lediglich einmal van Dijks Modell sogenannter Makrostrukturen, mit dem ein systematischer Zusammenhang hergestellt wird bzw. herstellbar ist zwischen den übergeordneten Makropositionen eines Textes, dem Textthema, einzelnen Themasätzen und deren Aktualisierungen in Form von Textüberschriften (van Dijk 1980, 47ff). Zum zweiten Viehwegers Untersuchung von Strukturregeln nominativer Ketten, die unter anderem den besonderen Status der Ersterwähnung in einer nominativen Kette erklärt (Viehweger 1976).

Für textlinguistische Modelle gilt jedoch generell, daß sie sich nicht unmittelbar in lexikologische Methoden überführen lassen. Während der Textlinguistik die differenzierte und umfassende Textanalyse als oberstes Ziel

⁶⁶ Als lexikographisches Verfahren erstmals angewandt von Reichmann 1986, 126f.

⁶⁷ Ähnliches meint Viehweger, wenn er von funktionaler semantischer Äquivalenz spricht, die auf einer Einordnungsinstanz in Abhängigkeit von Weltwissen beruht (Viehweger 1976, 103).

gilt, ist die Textanalyse in der praktischen Lexikologie und Lexikographie ein heuristisches Instrument zur Erreichung des Ziels 'Wortschatzbeschreibung'.

Im Hinblick auf [9] (syntagmatische Relationen) wird auf der Ebene des Textkorpus die Kollokabilität, d.h. die kotextuelle Relation zwischen lexikalischen Ausdrücken untersucht. Im Lexikon „Brisante Wörter“ wird Kollokabilität ausschließlich als usueller Anschließbarkeit verstanden, die sich mit Hilfe des Korpus von okkasioneller Anschließbarkeit unterscheiden läßt.⁶⁸ Position [9] enthält beispielhafte und formelhafte Verbindungen, die z.T. mehrgliedrig sind und/oder die Form ganzer Sätze haben. Die formelhaften Syntagmen veranschaulichen vor allem pragmatische Bedeutungsaspekte, die beispielhaften vor allem die semantisch-syntaktische Valenz.

Für die Positionen [8] und [9] ist charakteristisch, daß sie objektsprachliche Angaben enthalten. Darunter sind (einzelne) Ausdrücke, mehrgliedrige Syntagmen bis hin zu ganzen Sätzen zu verstehen, die dem Objektbereich des Wörterbuchs und nicht der Sprachebene des lexikographischen Kommentars angehören, in den sie aber im Unterschied zu den Beispielbelegen eingebettet sind. Diese Einbettung wird üblicherweise dadurch kenntlich gemacht, daß man alle objektsprachlichen Ausdrücke durch Kursivdruck vom umgebenden Text abhebt.

Aus der Anwendung des Korpusprinzips folgt konsequenterweise, daß alle objektsprachlichen Angaben, nicht nur die Beispielbelege (Position [14]), sondern auch diejenigen der Positionen [8] und [9] unmittelbar dem Korpus entnommen werden.

Wenn also z.B. im Artikel **Alternative** unter **a** als eine der typischen Verbindungen *jmd vor die Alternative stellen* und unter **b** u.a. *die Alternative zu etwas, keine Alternative haben* angeführt wird, so sind diese Syntagmen im Textkorpus tatsächlich vorhanden und bei der Analyse als besonders typisch aufgefallen (KWIC-Belege eignen sich dazu besonders gut). Die einzige Veränderung, die der Lexikograph an den Korpus syntagmen vornimmt, ist die Ersetzung konkreter Personen- und Gegenstandsbezeichnungen durch *jmd* und *etwas*, die Ersetzung der Kasus- und Numerusform durch den Nominativ Singular und die Ersetzung flektierter Verbformen durch den Infinitiv aktiv oder passiv, sofern das Charakteristische des Ausdrucks oder der Verbindung durch diese Normalisierungen erhalten bleibt.

Zu Position [10] (Angaben zu Textvorkommen und Textsortenspezifik): Sie ergeben sich unmittelbar aus einem Textkorpus, über dessen Einheiten

⁶⁸ Zu Anschließbarkeit siehe Kallmeyer et al. 1974, 124; zum Problem der Bestimmung von Kollokabilität vgl. Viehweger 1988.

nicht nur bibliographische, sondern auch alle wesentlichen kommunikationsbezogenen Informationen während der Analyse verfügbar sind.

Zu Position [11] Sprachkritischer Kommentar: Aus der kontextbezogenen Analyse von Satzbelegen läßt sich erstens ermitteln, ob und inwiefern innerhalb des anvisierten Kommunikationsbereichs und überindividuell ein sprachkritisches Bewußtsein in bezug auf einen bestimmten Ausdruck bzw. eine bestimmte Art der Verwendung (Paradebeispiel *Entsorgung*, *Entsorgungspark*) vorliegt; zweitens lassen sich in dieser Position – durch eine eventuelle, vorhergegangene Feststellung manipulativer Tendenzen bei der Wortverwendung motiviert – kritische oder warnende Hinweise von lexikographischer Seite aus formulieren.

Zu Position [12] Angaben zu Wortbildung und ggf. zu Sublemmata: Diesbezügliche Angaben ergeben sich vor allem aus der Analyse der Wortformen- und Wortbildungsparadigmen (siehe 4.1.1), anhand derer sich diejenigen Wortbildungselemente feststellen lassen, die besonders produktiv sind, und die die Unterscheidung nicht-lexikalisierten von lexikalisierten und eventuell als Sublemmata anzusetzenden Kompositen erlauben.

Zu Position [13] Enzyklopädische und wortgeschichtliche Angaben: Relativ zum jeweiligen Wörterbuchkonzept und in Abhängigkeit von dem mehr oder weniger großen Schwergewicht, das außersprachlichen und historischen Einflußfaktoren auf einen synchron betrachteten Wortschatz zugemessen wird, besteht die Aufgabe des Lexikographen bei dieser Wortartikelposition darin, eine geeignete Auswahl aus der Fülle der ihm vor allem, aber nicht ausschließlich durch das Sekundärkorpus zugänglichen Informationen zu treffen. Je nach den wörterbuchtypspezifischen Kriterien dieser Auswahl ergeben sich vor allem die typischen, weil rekurrenten Sach- und Argumentationszusammenhänge, in denen ein lexikalischer Ausdruck steht, eher aus der kontextbezogenen Analyse von Satzbelegen des Primärkorpus.

Beispielsweise lauteten die Auswahlkriterien im Lexikon „Brisante Wörter“ dahingehend, daß die im Hinblick auf die Gegenwart entscheidenden Wendepunkte in der kommunikativen Geschichte eines lexikalischen Ausdrucks, eventuell mit ihrer möglicherweise kausalen Verkettung, angegeben werden. Die Auswahlkriterien für die enzyklopädischen Informationen orientierten sich dort ebenfalls an ihrem Erklärungspotential im Hinblick auf die gegenwärtige Verwendung eines Ausdrucks.

Dabei stehen dem Lexikographen als heuristische Instrumente einmal die diversen historischen Wörterbücher und Enzyklopädien, lexikologische Spezialliteratur und fachinterne und vor allem fachdidaktische Literatur des infragestehenden Fach- bzw. Sachbereichs zur Verfügung. Zum andern kann eine auf diese Wortartikelposition bezogene Analyse des Primärkorpus durchgeführt werden, die sich die Perspektive sprachwandeltheoreti-

scher Überlegungen zunutze macht. Voraussetzung ist natürlich, daß das Korpus so organisiert ist, daß Abfrageergebnisse in Relation zu Zeitstufen gesetzt werden können, etwa indem selektierbare Teilkorpora jeweils einen bestimmten Zeitraum repräsentieren.

Die bei dieser Art Korpusanalyse einzunehmende Perspektive hängt davon ab, was als entscheidender Faktor des Bedeutungswandels, d.h. der Herausbildung der einzelnen Bedeutungen oder Verwendungsweisen und ihres Zusammenhangs angesehen wird. Mattheier (1981) nennt generell als auslösende und steuernde Faktoren des Sprachwandels Sprachprestige und Sprachbewertung, die in direktem Zusammenhang mit außersprachlich-gesellschaftlichen Faktoren stehen und als sprachbezogene Wertsysteme innerhalb des Wissens der Sprachteilhaber wirksam sind. Gegenüber diesen beiden Hauptfaktoren ist die Beteiligung innersystematischer Faktoren am Sprachwandel eher sekundär.⁶⁹

Aus diesem, hier nur sehr knapp skizzierten, sprachwandeltheoretischen Modell lassen sich vor allem für eine auf das 19. und/oder 20. Jahrhundert bezogene Kulturwortschatzlexikographie jene o.g. Auswahlkriterien diachroner Informationen ableiten. Die zentrale Bedeutung der funktionalen Varietäten, d.h. der technisch-wissenschaftlichen Sprache und ihres Wortschatzes, steht für diesen Zeitraum außer Frage; ihr wird von allen gesellschaftlichen Gruppen ein außerordentlich hohes Prestige und besondere Objektivität zugesprochen. Fachsprachlich gebildete Bezeichnungen gelten im allgemeinen Sprachbewußtsein, oft im ausdrücklichen Gegensatz zu denen der Gemeinsprache, als besonders realitätsangemessen.

Vor dem Hintergrund eines solchen Sprachwertsystems lassen sich die Befunde eines Textkorpus auch kommunikationsgeschichtlich interpretieren und in eine notwendig stark selektive, aber in sich konsistente und plausible Darstellung der diachronen Herausbildung von Bedeutungssträngen eines Ausdrucks einbeziehen. Die synchron beschreibende Gliederung z. B. des Wortartikels *Umwelt*⁷⁰ in drei Einzelbedeutungen: (a) eine gemeinsprachliche für die '(meist soziale) Umgebung des Menschen', (b) eine biologisch-fachsprachliche für die 'umgebenden Lebensbedingungen eines Lebewesens' und (c) eine gemeinsprachliche für ein 'Thema der öffentlichen Diskussion' läßt sich dann mithilfe der für Position [13] ermittelten diachronen Informationen als ein Kontinuum beschreiben, in dem Sprach- und Gesellschaftsgeschichte aufeinander bezogen sind.

Zu Position [14] Beispielbelege: Siehe Abschnitt 5.

⁶⁹ Sie bilden einen „Bedingungsrahmen“ oder jene „Systemstrukturen“, die eine „Adaption erleichtern oder sie hindern“ (Mattheier 1981, 282f).

⁷⁰ Im Lexikon „Brisante Wörter“ (Strauß/ Haß/ Harras 1989, 531-536).

Zu Position [15] Literaturhinweise: Diese sind zu verstehen als wörterbuch-adressatenbezogene Auswahl aus Titeln des Sekundärkorpus.

5. Beispielbelege

In diesem Abschnitt wird einer von den auf die Korpusanalyse folgenden lexikographischen Arbeitsschritten herausgegriffen, der die Herstellung des Textbausteins 'Beispielbelege' im Wortartikel zum Ziel hat. Wie die vorangehenden Abschnitte auch beziehen sich die Überlegungen zu den Beispielbelegen auf den Typ des korpusfundierte und textdokumentativen Kulturwortschatzwörterbuchs.

Aus methodischen Gründen scheint es mir notwendig, zwischen Belegen mit Datenbasisstatus und Belegen mit Beispiel- und Nachweisstatus zu unterscheiden. Belege mit Datenbasisstatus sind das Ergebnis der Segmentierung eines Korpus textes, unabhängig von der Art der Segmentierung z.B. in Form eines KWIC-Index oder in Form eines Satzbelegs mit oder ohne weiteren Kontext; sie werden hier auch Textsegmente genannt. Belege mit Beispiel- und Nachweisstatus sind das Ergebnis von Verfahren der Wortartikelproduktion, denen eine Teilmenge der Textsegmente nach bestimmten adressatenbezogenen Kriterien unterzogen wurde. Sie werden hier Beispielbelege, in eindeutigen Kontexten auch abkürzend Belege genannt.⁷¹

5.1. Zu Funktion und Definition von Beispielen und Belegen

In der metalexikographischen Diskussion sind Beispiele und Belege erst vor relativ kurzer Zeit zum Gegenstand theoretisch-methodischer Überlegungen⁷² geworden; in ihnen stehen vor allem die kommunikativen Funktionen der Beispiele innerhalb des Wörterbuchs bzw. eines Wortartikels im Mittelpunkt.

Unabhängig vom anvisierten Wörterbuchtyp und unabhängig von der Zugrundelegung des Korpusprinzips gilt als zentrale Funktion aller möglichen Arten von Beispielen die (implizit) demonstrative und illustrative Darstellung der im lexikographischen Kommentar vorher oder nachher explizit formulierten Regeln. Daneben spielt eine weitere Funktion eine je nach Wörterbuchtyp kleinere, für Kulturwortschatzwörterbücher aber mindestens ebenso große Rolle: die Funktion des Dokumentierens und Begründens der wissenschaftlichen Behauptung, die mit der Formulierung

⁷¹ Vgl. Reichmann (1988, 413): „Beispielbelege sind [...] Ausschnitte aus originalen Texten der Sprache oder der Sprachvarietät, die lexikographisch zu beschreiben ist.“

⁷² Stellvertretend nenne ich die Arbeiten von Harras (1990), Hermanns (1988), Reichmann (1988), Rey-Debove (1971).

einer sprachlichen Regel zugleich auch aufgestellt wird.

Je nachdem, welche der beiden genannten Funktionen als die vorrangige angesehen wird, hat sich die Redeweise vom Beispiel (zur Demonstration/ Illustration) und vom Beleg (zur Dokumentation/ Begründung) eingebürgert.

Speziell für kulturwortschatzbezogene Wörterbücher ist noch von einer dritten Funktion⁷³ auszugehen, die die Beispielbelege eines Wörterbuchs insgesamt erfüllen. Ausgewählte, mehr noch qua Kompetenz gebildete Beispiele sind innerhalb einer Semiotik des Wörterbuchs sowohl Zeichen für die kulturelle, historische, soziale, politische usw. Situation des bzw. der Lexikographen, als auch Zeichen für diejenige soziokulturelle Wirklichkeit, von deren Kommunikation das Korpus einen Ausschnitt darstellt. Zeichencharakter in diesem Sinne haben auch Stichwortauswahl, Erläuterungstypen, Kommentarsprache, das Wörterbuchkonzept insgesamt; aber die „kulturelle Semantik“ der Beispielbelege ist vergleichsweise unmittelbarer, d.h. auch von Nicht-Lexikographen verstehbar und als Faktor bei der Wörterbuchbenutzung existent.⁷⁴

Im Rahmen von Überlegungen zum lexikographischen Beispiel, die sich mehr oder weniger ausdrücklich auf gegenwartssprachige Gebrauchs- und Lernerwörterbücher beziehen, werden Syntagmen gleich welchen Umfangs und gleich welcher Herkunft unter die Bezeichnung *Beispiel* subsumiert. Als Prototyp oder Normalfall des Beispiels gilt dabei das nach bestimmten Maximen vom Lexikographen selbst gebildete Kompetenzbeispiel, das so ideal ist oder sein soll, daß ein einziges je Einzelbedeutung ausreicht.

Hingegen sind Kompetenzbeispiele undenkbar in Wörterbüchern solcher Sprachvarietäten und Kommunikationsbereiche, in denen der Lexikograph selbst als Sprecher nicht hinreichend kompetent sein kann, d.h. in historischen, dialektalen, fach- und gruppensprachlichen Wörterbüchern, die dem Typ Kulturwortschatzwörterbuch zuzuordnen sind. Hier können Beispiele nur in Form von dem Textkorpus entnommenen und bibliographisch belegten Zitaten eingesetzt werden.

Beispielbelege bestehen aus einem Syntagma, das gewöhnlich mindestens ein Satzglied umfaßt und mehrere Sätze umfassen kann, und aus einer eventuell abgekürzten Quellenangabe. Das Syntagma ist am Anfang und

⁷³ Die Anregung dazu verdanke ich G. Hoppe, Belege und Beispielsätze, unveröff. Ms. IDS Mannheim, 1978, 24.

⁷⁴ Vgl. Rey-Debove (1971, 272): „La macrostructure déjà, en tant que lexicque, reflète l'état social; mais les exemples, en tant que discours sur le monde, nous restituent les centres d'intérêt les jugements communs de la société dont la langue est décrite. C'est pourquoi nous prenons tant de plaisir à la lecture des dictionnaires anciens”.

am Ende dekontextualisiert und es kann sowohl in der Mitte gekürzt sein als auch vom Lexikographen vorgenommene und als solche kenntlich gemachte Ergänzungen enthalten, mit denen die nötige Verständlichkeit gesichert oder auch die Aufmerksamkeit des Lesers auf bestimmte Aspekte der Wortverwendung gelenkt werden soll.

Die Quellenangabe soll nicht nur die eindeutige bibliographische Identifizierbarkeit sichern, sondern sie kann auch über die Zeit, die Textsorte und die kommunikative Funktion des Textemplars informieren, wenn sie in diesen Hinsichten sprechend gewählt wird.

Für einen Beispielbeleg gilt außerdem die Bedingung, daß er aus demselben Text- oder Belegkorpus genommen ist, auf das sich die semantische Analyse bezog, und zwar ganz besonders dann, wenn die Beispielbelege die Funktion des wissenschaftlichen Begründens und Dokumentierens erfüllen sollen.

5.2. Auswahl der Beispielbelege

Das Problem der Auswahl von Beispielbelegen aus einer vielfach größeren Menge von Korpusbelegen stellt sich umso eher, je größer das zugrundeliegende Text- oder Belegkorpus ist. Beispielsweise gibt die Göttinger Arbeitsstelle des Deutschen Wörterbuchs (Neuaufgabe) als Auswahlrate „etwa 10 % besonders anschauliche Belege“ (Schlaefter 1986, 16) an, wobei zum Kriterium der Anschaulichkeit weiter keine Angaben gemacht werden.

Demgegenüber wird im folgenden davon ausgegangen, daß die Auswahl der Beispielbelege nicht nach quantitativen, von der Größe des Korpus abhängigen, sondern nach qualitativen und vom jeweiligen lexikographischen Konzept abhängigen Kriterien vorgenommen werden sollte. Vor einer Entscheidung für bestimmte Auswahlverfahren ist zunächst zu klären, ob die Beispielbelege vor allem das Allgemeine und Typische oder vor allem das Besondere (sofern es nicht nur individuell ist) oder beides in gleichem Maße veranschaulichen sollen. Für den hier anvisierten Wörterbuchtyp, der die – im weitesten Sinne verstanden – kulturellen Aspekte des Wortschatzes herausstellen will und dies in Form diskursiver Erläuterungen tut, gilt es, das Typische ebenso zu belegen wie das (für Sprechergruppen, Textsorten, Situationen usw.) Spezifische.

Die Kriterien für die Auswahl von Korpusbelegen für die Übernahme in einen Wortartikel lassen sich ableiten aus dem Informationsprogramm des Wörterbuch(typ)s, genauer aus den in einem Wortartikel je Einzelbedeutung realisierten, maximal 15 (vgl. Abschnitt 2.5.1) Informationsarten. Im Zusammenhang mit diesen Kriterien steht die Frage nach einer möglichen Klassifizierbarkeit der Korpusbelege. Die Ableitbarkeit von Auswahlkriterien aus den Informationsarten des Wörterbuchs läßt sich wie folgt begründen:

Lexikographische Informationen sind generalisierende und von konkreten Äußerungen abstrahierende Formulierungen sprachlicher Regeln; denen gegenüber erfüllt der Beispielbeleg seine demonstrativ-illustrative Funktion und zeigt wieder die oder eine konkrete Äußerung. Über die Erfüllung dieser Funktion hinaus kompensiert der Beispielbeleg auch eine z.T. unzureichende oder vom Leser nicht voll verstandene Formulierung der betreffenden Regel. Wenn nun die Informationen vom Lexikographen systematisch unterschieden und dem Leser mittels der Strukturierung des Wortartikels unterscheidbar dargestellt werden, dann müssen die die Informationen veranschaulichenden bzw. ergänzenden Beispielbelege nach den gleichen Kriterien unterscheidbar sein.

Außerdem ist das Informationsprogramm, d.h. die Reihe der Informationsarten auch das Hauptinstrument bei der Analyse der Korpusbelege (siehe Abschnitt 4.4); zur Regelformulierung können also nur diejenigen Ergebnisse werden, die sich aus der Analyse der Korpusbelege ergeben haben. Es muß sich folglich für jede lexikographische Information eine Teilmenge der Korpusbelege anführen lassen, die die implizite Behauptung „es gibt eine sprachliche Regel X“ sowohl anderen Linguisten als auch dem nicht-wissenschaftlichen Wörterbuchadressaten gegenüber zu stützen vermag.

Es geht beim ersten Schritt der Auswahl der Beispielbelege darum, jeder in einem Wortartikel gegebenen Information den- oder diejenigen Korpusbelege zuzuordnen, die die dort formulierte Regel in deutlich herausragender Weise veranschaulichen bzw. die lexikographische Behauptung einleuchtend stützen. Man kann z.B. schon während der Analyse einige Korpusbelege mit Markierungen für eine oder zwei bestimmte Informationsarten versehen. Zwar ist theoretisch jeder Korpusbeleg eine Beispiel für alle oder fast alle zur Verwendung eines lexikalischen Ausdrucks gegebenen Informationen gleichzeitig, in der Praxis aber zeigen sich Dominanzen, stärkere oder schwächere Illustrativität und Plausibilität, die man bei der Zuordnung ausnutzen kann. Es beeinträchtigte sogar die demonstrativ-illustrative Funktion eines Beispielbelegs, wenn er sie für mehr als zwei bis drei Regelinformationen gleichzeitig erfüllen müßte. Die Verknüpfung zwischen den einzelnen lexikographischen Informationen und den zuzuordnenden Belegen geschieht mittels expliziter Verweise im Text des Wortartikels (z.B. *siehe Bel. 7*). Sinn der Verknüpfung ist einmal größtmögliche Transparenz der Belegauswahl, zum zweiten kann der Wörterbuchadressat die Verknüpfung interpretieren als: „Jener Beleg illustriert in besonderer Weise die allgemein gehaltene Aussage dieses Textabschnitts.“ oder als: „Jener Beleg begründet eine eventuell als strittig empfundenen Aussage eines Textabschnitts über die generelle Begründung durch Literaturangaben hinaus.“

Nachfolgend seien am Beispiel des Informationsprogramms des Lexikons „Brisante Wörter“ (siehe Abschnitt 2.5.1) Verfahren der Belegauswahl

erläutert und in Form von Grundsätzen formuliert.⁷⁵

5.3. Relation zwischen Beispielbelegen und Informationsarten

Mit Bezug auf diejenigen Artikelpositionen, die die semantischen Kerninformationen enthalten, das sind die Positionen [4] 'Angaben zur (semantischen) Funktionsklasse', [5] 'Angaben zum Bezugsbereich', [6] 'Bedeutungsangabe im engeren Sinn' und ggf. [7] 'Angaben zur Verwendungsspezifik',⁷⁶ sind solche Korpusbelege auszuwählen, die zum Teil gleichzeitig die folgenden Bedingungen erfüllen:

- Die Belege decken insgesamt das extensionale Spektrum ab und lassen die einzelnen Bezugsgegenstände deutlich erkennen;⁷⁷ für prädiszierende Verben wie z.B. *entsorgen* und Adjektive wie z.B. *ökologisch* machen die Belege das Spektrum der Gegenstände deutlich, auf die mit dem Verb oder Adjektiv prädiszierend Bezug genommen wird.
- Falls im Korpus vorhanden, wird eine Äußerung als Beleg ausgewählt, in der ein (nicht-linguistischer) Sprecher den Lemmaausdruck mehr oder weniger explizit einer Funktionsklasse wie Schlagwort oder Euphemismus zuordnet; ein solcher Beleg trägt dazu bei, den Stand des allgemeinen öffentlichen Sprachbewußtseins zu dokumentieren.⁷⁸
- Bezüglich der Bedeutungsangabe in Position [6] können zwei Klassen von Belegen unterschieden werden, die beide bei der Auswahl berücksichtigt werden sollten.

Die eine Klasse wird von explizit definitionsartigen Äußerungen gebildet⁷⁹ (*Definition* wird hier im allgemeinsprachlichen Sinn verwendet); dabei handelt es sich um beispielhafte, für einen bestimmten (politischen, weltanschaulichen) Standpunkt repräsentative, interessante oder auch streitbare Erklärungsversuche des Lemmaausdrucks, die von Sprechern meist mit argumentativer oder autoritativer Absicht vorzugsweise mittels be-

⁷⁵ Zur Erläuterung der Belegauswahl für Nicht-Linguisten siehe Strauß/ Haß/ Harras 1989, 21f.

⁷⁶ Zur ausführlichen Erläuterung der Informationsklassen siehe Strauß 1988.

⁷⁷ Vgl. z.B. s.v. **Gau** die Belege 2 und 3 zum Bezugsbereich 'Atomreaktor', Beleg 5 zum Bezugsbereich 'Anlage der Industrie, besonders Chemieindustrie, Beleg 7 zum Bezugsbereich 'Aids-Virus in der Biologie/Gentechnologie' und Beleg 8 zum Bezugsbereich 'anthropogene Klimaveränderungen' (Strauß/ Haß/ Harras 1989, 473f).

⁷⁸ Vgl. z.B. s.v. **Störfall** den Beleg 19 (Strauß/ Haß/ Harras 1989, 530).

⁷⁹ Vgl. „Typ 1 des Beispielbelegs: die textinterne definitionsartige Erläuterung“ in Reichmann 1988, 421, mit Beispielen in Reichmann 1983, 120f.

stimmter Textsorten⁸⁰ in die öffentliche Diskussion eingebracht werden.⁸¹ Darunter fallen natürlich nicht Erklärungen bzw. Definitionen aus begrifflichen Nachschlagewerken, die dem Sekundärkorpus zugeordnet sind. Auf diesen Belegtyp wird meist schon bei der Korpuszusammenstellung besonders geachtet⁸² und ihm wird auch eine herausgehobene Position innerhalb des Belegblocks zugeteilt (siehe unten). Mithilfe solch sprachthematisierender Äußerungen kann der Stand der öffentlichen Sprachreflexion dokumentiert und resümiert werden. Sie können sichtbar machen, welche Sprecher(gruppen) sich wie und mit welchem Erfolg am kollektiven „Aushandeln“ von Wortbedeutungen beteiligen und von welchen Bedingungen dieses „Aushandeln“ abhängt.

Falls sich zwei, politisch-weltanschaulich konträre Definitionsversuche eines Ausdrucks im Primärkorpus finden, sollten diese dann im Belegblock unmittelbar nebeneinander stehen. Die explizit definitionsartigen Belege können ihrer besonderen, z.T. kommunikationssteuernden Bedeutung wegen im Belegblock an herausragender Stelle, etwa am Anfang platziert werden.⁸³

Die andere Klasse von Belegen, die sich der Position [6] zuordnen lassen, kann implizit definitionsartig oder prädikativ genannt werden, weil ein Beleg oder wenige Belege gemeinsam alle wesentlichen Aspekt der (Intensions-)Bedeutung deutlich werden läßt bzw. lassen,⁸⁴ d.h. sehr anschaulich sind, ohne dabei wie die zuerst genannte Klasse begriffs- oder sprachreflektierend zu sein. Als Auswahlgrundsatz formuliert: der Lexikograph sollte darauf achten, daß die Belege je Einzelbedeutung insgesamt auch sämtliche Elemente der Bedeutungsangabe im engeren Sinne spiegeln. Möglicherweise kann man die Klasse der prädikativen Belege, d.h. derer, die Prädikationen über den Bezugsbereich des Lemmaausdrucks enthalten, noch weiter differenzieren, z.B. um zufällige von typischen oder konventionalisierten Prädikationen zu unterscheiden.⁸⁵

⁸⁰ Zur Affinität zwischen diesem Belegtyp und bestimmten Textsorten siehe auch Reichmann 1988, 422.

⁸¹ Vgl. z.B. s.v. **Gau** Beleg 14 (Strauß/ Haß/ Harras 1989, 474).

⁸² Sie bilden einen großen Teil des zusätzlich gesammelten, nicht-computerverfügbaren Materials.

⁸³ Diese Anregung verdanke ich Sabine Plum, Aachen.

⁸⁴ Vgl. „Typ III des Beispielbelegs: Charakterisierungen“ in Reichmann 1988, 428-436. Vgl. s.v. **alternativ** die Belege 7 und 8, in denen jene Aspekte aufgezählt werden, die den Inhalt des Prädikats **alternativ** ausmachen (Strauß/ Haß/ Harras 1989, 419).

⁸⁵ Solche und andere Differenzierungen dieses Typs bei Reichmann 1988, 423f.

- Den Informationen in Position [7] entsprechen z.T. schon die explizit definitionsartigen Belege.⁸⁶ Die Auswahl läßt sich durch Kontrollfragen etwa der folgenden Art steuern: Sind Belege vorhanden, die je einen der politischen oder weltanschaulichen Standpunkte wiedergeben, die mit dem Gebrauch des Lemmaausdrucks typischerweise, häufig oder eng verknüpft sind? Sind die verschiedenen Sprecher und Interessengruppen mit je einem Beleg vertreten und spiegeln die Belege somit das Meinungsspektrum innerhalb des Kommunikationsbereichs wieder, der durch das Textkorpus repräsentiert wird?⁸⁷ Sind Belege darunter, die besonders charakteristische Argumentationen im Zusammenhang mit dem Lemmaausdruck wiedergeben?

Solche Kontrollfragen enthalten auch Kriterien, nach denen der Lexikograph entscheiden kann, ob und welche Ergänzungen in ausgewählte Belege einzufügen sind. Ergänzungen sorgen dafür, daß der Leser eine Äußerung auch eindeutig einem bekannten Sprecher, einer Partei- oder Interessengruppe oder auch einer bestimmten Diskussionssituation (z.B. vor oder nach Tschernobyl) zuordnen kann. Die Quellenangabe allein reicht dazu oft nicht aus.

Mit Bezug auf Position [8] 'Angaben zu paradigmatischen semantischen Relationen' gelangt man zur Auswahl solcher Belege, die veranschaulichen, inwiefern der Lemmaausdruck mit anderen Ausdrücken semantisch parallelisiert, identifiziert, konterkariert, von anderen Ausdrücken abgegrenzt oder in ein genus-differentia-Verhältnis gesetzt wird.⁸⁸

Um solche Koreferenz- und Isotopieketten besser hervortreten zu lassen, wird zunächst ein relativ umfangreicher Textausschnitt gewählt, der dann um alles dasjenige gekürzt wird, was nicht zur Koreferenz- bzw. Isotopieebene gehört. Da das Ergebnis der Kürzung aber noch lesbar sein muß, kommt nur ein kleiner Teil der in dieser Hinsicht lexikologisch aussagekräftigen Korpusbelege für die lexikographische Darstellung infrage.

Mit Bezug auf die Informationen in Position [9] 'syntagmatische Relationen', die bereits weitgehend objektsprachlich formuliert sind, werden in der Regel die bereits aus anderen Gründen ausgewählten Belege die

⁸⁶ Vgl. s.v. **Super-Gau** Beleg 17, der Versuche von Begriffsbesetzung in Abhängigkeit von bestimmten Meinungspositionen deutlich macht.

⁸⁷ Dies trifft s.v. **Gau** auf die Belege 9 und 10 für die eine, auf Beleg 11 für die andere Gruppe zu (Strauß/Haß/Harras 1989, 474).

⁸⁸ Vgl. „Typ II des Beispielbelegs: Identifizierungen“ in Reichmann 1988, 426-428. Vgl. z.B. s.v. **Gau** die Belege 13 und 15, in denen eine Identifikation in Abhängigkeit von Sprechergruppen vorgenommen wird (Strauß/ Haß/ Harras 1989, 474), sowie s.v. **alternativ** Beleg 18, in dem bedeutungsähnliche Ausdrücke distinktiv voneinander unterschieden werden (Strauß/ Haß/ Harras 1989, 420).

Funktionen des Veranschaulichens und Dokumentierens erfüllen, indem mindestens einer von ihnen die typischen Satzverknüpfungen des Lemmaausdrucks oder deren leichte Variation zeigt. Beim Vorliegen typischer formelhafter Syntagmen (z.B. *Versöhnung von Ökologie und Ökonomie*) sollte ein Beleg ausgewählt werden, der die Verwendung der Formel in ihrem argumentativen Kontext, in ihrer Sprechergruppenbindung u.ä. demonstriert.

Das Veranschaulichen der Informationen in Position [10] 'Textvorkommen, Textsortenspezifik' leisten vor allem sprechende Quellenangaben und diesbezügliche *Ergänzungsangaben* in den Belegen⁸⁹ (siehe Abschnitt 5.4.3).

Die sprachkritischen Informationen in Position [11] tragen mehr als alle anderen Informationen den Charakter der wissenschaftlichen Behauptung und machen in besonderem Maße die Auswahl von Belegen erforderlich, die geeignet sind, diese Behauptungen zu stützen, d.h. die vor allem ihrer wissenschaftskommunikativen Funktion wegen ausgewählt werden.

Insofern Position [11] auch dazu benutzt werden kann, die nicht-linguistische öffentliche Sprachkritik zu einem bestimmten Ausdruck zu resümieren (siehe Abschnitt 4.4), haben diesbezüglich ausgewählte Belege aber auch eine demonstrative Funktion. Im Wortschatz der Umweltdiskussion ziehen nicht wenige Ausdrücke Bewertungen wie 'euphemistisch' auf sich. Im Fall von *Entsorgungspark* hat die öffentliche Sprachkritik bewirkt, daß der Ausdruck sehr schnell nicht mehr denotativ mit Bezug auf einen konkreten Gegenstand verwendet wurde (nur 1 Beleg!).⁹⁰ Deshalb erhalten Belege, in denen sich Sprecher reflektierend oder bewertend über Ausdrücke und ihre Verwendung äußern, im Lexikon „Brisante Wörter" bei der Auswahl einen besonders hohen Stellenwert, und zwar auch dann, wenn die linguistische Analyse und Beschreibung einer solchen Bewertung widerspricht.⁹¹ Sprachreflektierende Belege lassen sich außer zu Position [11] (Sprachkritik) häufig auch zu Position [7] 'Verwendungsspezifik' oder [8] 'paradigmatische semantische Relationen' in Beziehung setzen. Bei diesem Belegtyp ist besonders wichtig, daß Quellenangaben und Ergänzungen so eingerichtet sind, daß die kommunikative Funktion einer Äußerung erkennbar wird.

Belege mit Bezug zu [12] 'Wortbildungsbeziehungen' auszuwählen, kann dann sinnvoll sein, wenn der Lemmaausdruck als Wortbildungselement besonders produktiv ist. Dies gilt vor allem für meist unselbständige Wort-

⁸⁹ Vgl. s.v. **Alternative(r)** Beleg 3 (Strauß/ Haß/ Harras 1989, 423).

⁹⁰ S.v. **Entsorgungspark** Beleg 1 (Strauß/ Haß/ Harras 1989, 468).

⁹¹ Vgl. s.v. **Entsorgung** Beleg 21 (Strauß/ Haß/ Harras 1989, 465f), in dem der Euphemismus-Vorwurf sich auf eine linguistisch falsche Benennungsmotivierung stützt.

bildungselemente wie *bio-*, *öko-*, aber auch für selbständige wie *Umwelt-*, *Recycling-*, *Entsorgungs-*. Die Belegauswahl hängt dann davon ab, inwieweit man Sublemmata einzeln oder in Gruppen zusammengefaßt behandelt hat und inwieweit die Produktivität bereits kommentarsprachlich erläutert worden ist.

Die in Position [13] gegebenen enzyklopädischen und/oder wortgeschichtlichen Informationen können z.T. ebenfalls durch eigene Belege veranschaulicht werden, je nachdem welches Gewicht man diesen Informationen im Artikel beimißt. Im Lexikon „Brisante Wörter“ folgen im Anschluß an die explizit definitionsartigen Belege der früheste oder höchstens drei frühe Belege, wenn dadurch wort- und kommunikationsgeschichtlich wichtige Markierungen veranschaulicht werden, wie z.B. die fachlexikalische Bedeutungsveränderung von *Umwelt* in der Biologie in den 20er Jahren unseres Jahrhunderts. Der „Erstbeleg“ besitzt keinen Wert an sich; er spielt nur dann eine Rolle, wenn sich durch ihn ein Bedeutungswandel in Abhängigkeit vom gesellschaftlichen Wandel veranschaulichen läßt.

Besondere enzyklopädische Informationen machen kaum eigens deshalb ausgewählte Belege erforderlich, da die sachlichen Bezüge eines Bezugsgegenstands und das Reden von Zielen und Zwecken im Umgang mit einem Gegenstand in den Belegen insgesamt in der Regel hinreichend sichtbar werden, vorausgesetzt das lexikographische Konzept läßt einigermaßen umfangreiche Textbelege zu. Da bei diesem Belegtyp der Inhalt der Äußerung eine besondere Rolle spielt, sollten keine Belege mit nachweislich falschen Sachaussagen (z.B. Zahlen) angeführt werden.⁹² Als Maßstab hierfür reicht in der Regel die Literatur des Sekundärkorpus aus.

Auch zu den Positionen [1] 'Lemma', [2] 'grammatische Angaben' und [3] 'Vorspann zu Etymologie und Wortfamilie' ist eine spezifische Belegauswahl und -zuordnung unter Umständen denkbar, etwa wenn unterschiedliche Schreibweisen bei *GAU* (*Gau*, *GaU*), ungewöhnliche Steigerbarkeit (bei *ökologisch*) oder (volks)etymologische Assoziationen zu dokumentieren sind.

Je nachdem, welche Umfangsbegrenzung für den Belegteil der Wortartikel besteht, wird es nötig sein, die genannten Auswahlkriterien so anzuwenden, daß ein Beleg seine Funktion(en) mit Bezug auf mehrere Informationsarten erfüllt. Nachdem ein wie hier erläutertes Auswahlverfahren durchgeführt ist, muß der Lexikograph in einem eventuell zur Kontrolle eingesetzten Umkehrschlußverfahren in der Lage sein, explizit die spezifische Funktion eines jeden Beispielbelegs anzugeben, d.h. dessen Unverzichtbarkeit zu begründen.

⁹² Vgl. „Typ IV des Beispielbelegs: Rahmenkennzeichnende Belege“ in Reichmann 1988, 436-439.

5.4. Vom Korpusbeleg zum Beispielbeleg: Bearbeitungsverfahren

Die zuvor erläuterten Auswahlverfahren verfolgten das Ziel, die der Informationsart 'Beispielbelege' wörterbuchtypspezifisch zugewiesene Funktionalität herzustellen bzw. zu sichern. Zur Erreichung dieses Ziels gehören noch einige weitere Verfahren, die die Gestaltung des Belegblocks, d.h. die Anzahl der Belege je Wortartikel bzw. je Einzelbedeutung, ihre Reihenfolge, den möglichen Status eines durch das Belegsortiment neu entstandenen (Wortartikelteil-)Textes sowie die Gestaltung eines Korpussegmentes, das Beispielbeleg werden soll, betreffen.

5.4.1. Der Belegblock

Die bei der Gestaltung des Belegblocks möglichen Verfahren werden nachfolgend am Beispiel des Wortschatzbereichs 'Umwelt' aus dem Lexikon „Brisante Wörter“ erläutert.

Belegungszeitraum: Das zugrundegelegte Textkorpus ist gegenwartsbezogen synchron; in der Praxis heißt das, daß vor allem Korpusbelege von 1949 bis 1988 mit Schwerpunkt auf den Jahren 1985 und 1986 verarbeitet wurden. Daraus wurde der Grundsatz abgeleitet, daß die Wörterbuchbelege diesen Zeitraum unter Berücksichtigung des Schwerpunkts etwa gleichmäßig abdecken sollten, es sei denn ein Lemmaausdruck ist schwerpunktmäßig nur in bestimmten Jahren belegt und dieses läßt sich kommunikationsgeschichtlich begründen. Wenn z.B. *Recycling* schon Anfang der 70er Jahre belegt ist, dann wurde darauf geachtet, daß die Beispielbelege nicht nur von 1986 stammen.

Beleganzahl: Bezüglich der Anzahl der Belege je Einzelbedeutung sind strikte Grundsätze kaum zu formulieren, aber sie hängt in gleicher Weise wie der Belegtyp von den lexikographischen Informationen ab, aus denen sich in jedem Einzelfall ein mehr oder weniger großer Illustrations- und/oder Dokumentationsbedarf ergibt. Wo viele und differenzierte Informationen zu einem Lemmaausdruck gegeben werden, müssen auch mehr Belege ausgewählt werden, die diese Informationen veranschaulichen oder stützen. Der Leser wird (in der Einleitung) darauf hingewiesen, daß aus der Menge der Belege keine Rückschlüsse auf die Häufigkeit eines Ausdrucks zu ziehen sind. Es ist unvereinbar mit den genannten Funktionen von Beispielen und Belegen und mit dem exemplarischen, nicht repräsentativen Charakter des Korpus (vgl. oben Abschnitt 2.1.3), die Anzahl der Wörterbuchbelege direkt abhängig zu machen von der Anzahl der Korpusbelege und damit eine Aussage über die Frequenz eines Ausdrucks zu verbinden.

Belegreihenfolge: Über sie kann auf der Basis der Unterscheidung verschiedener Belegtypen entschieden werden. Hinsichtlich der explizit definitionsartigen Belege wurde oben (siehe Abschnitt 5.2) schon gesagt, daß sie herausgehoben sind durch ihre Anfangsstellung im Belegblock. Ihnen an-

schließlich folgt der oder folgen die frühesten Belege, wenn sie die genannten Bedingungen (Markierung von wort- bzw. bedeutungsgeschichtlichen Wendepunkten) erfüllen. Den Schluß eines Belegblocks bilden meist, sofern vorhanden, Belege mit metaphorischen Verwendungen, und zwar deshalb, weil sie den beschriebenen Wortschatzbereich 'Umwelt' nicht selten verlassen.⁹³ Für sie gilt natürlich auch, daß sie sich in Beziehung setzen lassen müssen zu Informationen meist der Positionen [7] und/oder [11]. Ansonsten bietet es sich an, die Belege beieinander zu lassen, die sich auf dieselbe Position oder auch auf denselben Bezugsbereich (Sachverhalt, Diskussionsthema) beziehen.⁹⁴

Der Belegblock als Text: Der Lexikograph kann bei der Festlegung der Belegreihenfolge in Versuchung kommen, einen Belegblock als Text sui generis aufzufassen und unausgesprochene Zusammenhänge zwischen Belegen als Sätzen eines kohärenten Textes herzustellen, etwa indem neben die Aussage eines Verwaltungsgerichts, das von einem bestimmten Kernkraftwerk ausgehende Restrisiko sei unzumutbar, die Aussage des Bundesgerichtshofs gestellt wird, dasselbe Restrisiko sei tragbar. Das Nebeneinandersetzen konträrer oder widersprüchlicher Äußerungen, die innerhalb des beschriebenen Kommunikationsbereichs meist eher disparat vorkommen und deren Kontrarität deshalb verdeckt ist, hat innerhalb der Semiotik des Wörterbuchs unzweifelbar Bedeutung.⁹⁵

Es gilt also, methodisch zu reflektieren, daß die Textualität eines Wortartikels und des Belegblocks als eines seiner Textelemente eine mehrfache Lesbarkeit impliziert. Beispielbelege werden einmal gelesen im Hinblick auf die Veranschaulichung und die Plausibilität von Regelformulierungen über die Verwendung eines Ausdrucks. Sie sind aber zum zweiten lesbar als Ausschnitte aus Texten, in denen von der Welt gehandelt wird.⁹⁶ Beispielbelege stellen dekontextualisierte Äußerungen mit Sinnpotentialen dar, die nicht endgültig bestimmbar sind; sie werden durch lexikographische Manipulationen in neuer Weise und mehrfach rekontextualisiert. Die Art der Rekontextualisierung und die teilweise Aktualisierung der Sinnpotentialie

⁹³ Ein Beispiel ist die inzwischen berühmt gewordene *Entsorgung der Vergangenheit*.

⁹⁴ Zur Anordnung der Belege vgl. auch Vorwort zum Bereich 'Umwelt' in Strauß/Haß/Harras 1989, 408.

⁹⁵ Rey-Debove (1971, 269) bestreitet den textuellen Zusammenhang der Beispiele untereinander ("Les exemples ne se lisent pas „bout à bout“ ... comme les phrases d'un roman."), muß aber Einschränkungen dieses Prinzips in der Praxis zugeben: „certains lexicographes sont soucieux de la lecture totale, où la suite des mentions lues bout à bout garde un semblant de sémantacité au niveau transphrastique, ménageant des transitions et évitant les rencontres ridicules. Le GR en donne de nombreux échantillons." (ebenda)

⁹⁶ Vgl. Hermanns 1988, 166ff und Rey-Debove 1971, 263.

hängt dabei nicht nur vom Lexikographen, sondern auch vom Rezipienten, d.h. vom Wörterbuchbenutzer, dem Benutzungsanlaß, der Situation usw. ab. Nicht zuletzt aufgrund der unabschließbaren Rekontextualisierung von Beispielbelegen veralten Kulturwortschatzwörterbücher, anders als Gebrauchswörterbücher, nicht, sie werden mit verändertem Zeitabstand nur anders gelesen. Diese hier nur angedeuteten wörterbuchtypspezifischen Rezeptionsbedingungen können nicht ohne Einfluß auf die lexikographische Methodik sein.

5.4.2. Belegschnitt

Mit Belegschnitt werden alle Verfahren der Dekontextualisierung eines Textsegments im Hinblick auf dessen Funktion als Beispielbeleg zusammengefaßt. Dazu gehören erstens die Festlegung des Beleganfangs, zweitens die Festlegung des Belegschlusses und drittens beleginterne Kürzungen.

Als übergeordnete Orientierung galt im Lexikon „Brisante Wörter“, daß ein Beispielbeleg möglichst eine kommunikativ vollständige Äußerung darstellen sollte, d.h. daß der Belegschnitt so gelegt wird, daß die der Äußerung übergeordnete textuelle Makroposition im Sinne van Dijks (1980 und 1985) erkennbar bleibt und der Gebrauch eines Worts innerhalb dieses Rahmens gezeigt bzw. verstanden wird. Nur auf diese Weise können Informationen über äußerungsfunktionale, kontextbezogene Aspekte des Wortgebrauchs hinreichend veranschaulicht und dokumentiert werden. Deshalb besteht der Prototyp des Beispielbelegs in der Regel aus einem Satzgefüge oder aus mehreren konjunkional verknüpften Hauptsätzen, das/die intern gekürzt ist/sind; Belege in Form eines abhängigen Gliedsatzes oder einer Satzgliedphrase kommen hingegen nur ausnahmsweise vor.

Z.B. hat ein Beleg s.v. **umweltfreundlich**, **umweltverträglich** nicht die Form: *Sondermüll umweltverträglich [...] verbrennen*, sondern die Form: *Töpfer glaubt, daß durch „hohen Einsatz an Technik und viel Geld Sondermüll umweltverträglich“ zu verbrennen ist*.

Ein weiterer Beispielbeleg lautet nicht: *Der Hinweis „phosphatfrei“ werde so verwendet, als könne er mit „unbedenklich“ oder „umweltfreundlich“ gleichgesetzt werden*, sondern: *[...] der Umweltexperte [...] hat die Werbung für phosphatfreie Waschmittel als „irreführend“ bezeichnet. Der Hinweis „phosphatfrei“ werde so verwendet, als könne er mit „unbedenklich“ oder „umweltfreundlich“ gleichgesetzt werden.*⁹⁷

Bedingt durch den Kommunikationsbereich öffentliche Diskussion spielen nicht selten makropositionale Zusammenhänge zwischen kotextuell nicht unmittelbar benachbarten Sätzen, z.B. zwischen einem Zeitungsartikelsatz

⁹⁷ Strauß/ Haß/ Harras 1989, 541.

und der zugehörigen Überschrift (Schlagzeile) eine Rolle; solche Zusammenhänge werden beim Belegschnitt ebenfalls berücksichtigt.

Ein Vorteil unter anderen dieses Verfahrens ist, daß die referentielle Bedeutung pronominaler Formen und damit die Lesbarkeit weitgehend erhalten bleibt.

Je weiter der kotextuelle Rahmen zur Belegung des Lemmaausdrucks gesteckt wird, desto notwendiger ist die beleginterne Kürzung. Gekürzt wird alles dasjenige, was nicht zu den relevanten Ko- und Kontextfaktoren gehört, d.h. was nicht zu der dem einzelnen Beleg speziell zugewiesenen Funktion in Relation zu einer bestimmten Informationsart (siehe Abschnitt 5.3) gehört. Z.B. werden in einem Beleg, der primär paradigmatische semantische Relationen des Lemmaausdrucks veranschaulichen soll, jene Satzglieder weggekürzt, die den Bezugsbereich veranschaulichen, weil oder insofern diese Aufgabe von anderen Belegen erfüllt wird.

Allerdings kann auch der Zusammenhang verschiedener, pragmatisch relevanter Kontextfaktoren zu denjenigen Besonderheiten der Wortverwendung gehören, die mittels eines Beispielbelegs vorgeführt werden sollen, z.B. der typische Zusammenhang zwischen einem bestimmten Sprecherstandpunkt/ Weltbild und einer formelhaften Verbindung oder mit der Bevorzugung einer von zwei Bezeichnungsalternativen. Ein solcher Zusammenhang wird nicht durch Belegkürzung zerstört, sondern eher noch durch einen Belegverweis mit seiner abstrahierend-generalisierenden Erläuterung im Wortartikel verknüpft. Für Belegkürzungen gibt es keine generellen Regeln, sondern nur solche in Abhängigkeit von den im jeweiligen Wörterbuchkonzept und im einzelnen Wortartikel unterschiedenen Belegtypen.

5.4.3. Belegergänzung und -kommentierung

Hierbei handelt es sich um eindeutig (im Lexikon „Brisante Wörter“ durch eckige Klammern) gekennzeichnete, kommentarsprachliche Zusätze in einem oder im Anschluß an einen Beispielbeleg. Ihre Funktion ist eine noch weitergehende Rekontextualisierung als sie mittels Belegschnitt erreichbar ist. Zunächst dienen sie der (stellvertretenden) Ergänzung von durch Kürzung weggefallenen Satzgliedphrasen und der Referenzierbarkeit pronominaler Formen, und zwar besonders in den Fällen, in denen sie Personen und Institutionen als Sprecher bzw. Handelnde der öffentlichen Diskussion (z.B. Politiker, Ministerien) vertreten. Bei der Wortwahl dieser Ergänzungen wird wenn irgend möglich der originale Wortlaut aus dem weiteren Kontext wiederholt und nicht frei vom Lexikographen paraphrasiert, z.B.⁹⁸

⁹⁸ Die Wörterbuchbelege werden im Folgenden gekürzt.

Hier [im tropischen Regenwald] finden viele Tierarten ... (Strauß/ Haß/ Harras 1989, 497)

Solche Systeme [natürliche Ökosysteme] ... (Strauß/ Haß/ Harras 1989, 503)

„Betrug mit der Sprache“ warf Wallmann [Bundesumweltminister] allen Kernenergie-Gegnern vor, die ... (Strauß/ Haß/ Harras 1989, 474)

Damit die ausgedienten Autos künftig recycelt werden, verlangt Baum [FDP-Politiker] verwertungsfreundliche Konstruktionen und ... (Strauß/ Haß/ Harras 1989, 505)

hat v. Bennigsen [von der Kernkraftwerksfirma KWU] betont ... (Strauß/ Haß/ Harras 1989, 473)

Traute und Michael [das „grüne“ Lehrerehepaar] ... (Strauß/ Haß/ Harras 1989, 419).

Belege, die als Ganzes Äußerungen herausragender Sprecher der öffentlichen Diskussion sind, werden mit dem Kommentar „Zitat“ plus Angabe des Namens und eventuell der Funktion versehen, z.B.

[Zitat Winnacker von der Enquete-Kommission Gentechnologie] (Strauß/ Haß/ Harras 1989, 473)

[Aus den 13 Thesen der Gewerkschaft NGG] (Strauß/ Haß/ Harras 1989, 503)

[aus Regierungserklärung des Bundeskanzlers] (Strauß/ Haß/ Harras 1989, 437)

[Leserbrief] (Strauß/ Haß/ Harras 1989, passim).

Der Kommentar kann auch die Textsorte oder die Kommunikationssituation angeben, der ein Beispielbeleg angehört, wenn dies besonders symptomatisch ist, z.B.

[aus e. Anzeigenwerbung]

[Artikelüberschrift]

[Buchtitel]

[nach Tschernobyl]

Belegergänzungen und -kommentare tragen gemeinsam mit den aus den grundsätzlich datierten Quellenangaben erschließbaren Informationen wesentlich dazu bei, daß die je spezifische soziale, historische, politische und kulturelle Gebundenheit des Wortgebrauchs, so wie er in den Beispielbelegen gezeigt wird, nicht nur implizit deutlich, sondern explizit angebbar wird.

Sämtliche am „Rohmaterial“ des Korpus vorgenommenen Bearbeitungsverfahren tragen zur Lesbarkeit der Beispielbelege und damit des Wortartikels bei, sind also umso notwendiger, je diskursiver ein Wörterbuch konzipiert ist.⁹⁹

⁹⁹ Von Benutzern des Lexikons „Brisante Wörter“ wissen wir, daß vor allem die Belege als sprachdidaktisches Material im Schulunterricht eingesetzt, hingegen die erläuternden Artikelteile eher zur Lehrervorbereitung herangezogen werden.

6. Literatur:

- (Agricola 1972) = Agricola, Erhard: *Semantische Relationen im Text und im System*, The Hague 1972.
- (Agricola 1976) = Agricola, Erhard: *Vom Text zum Thema*, in: F. Danes/ D. Viehweger (Hrsgg.): *Probleme der Textgrammatik*, Berlin (DDR) 1976, 13-27.
- (Aitken/ Bailey/ Hamilton-Smith 1973) = Aitken, A. J./ Bailey, R.W./ Hamilton-Smith, N. (Hrsgg.): *The Computer and Literary Studies*, Edinburgh 1973.
- (Ballweg-Schramm 1978) = Ballweg-Schramm, Angelika: *Zur prinzipienlehre der lexikographie*, in: *Zeitschrift für germanistische Linguistik* 6/1978, 1-17.
- (Bausch 1975) = Bausch, Karl-Heinz: *Zur Problematik der empirischen Basis der Linguistik*, in: *Zeitschrift für germanistische Linguistik* 3/1975, 123-148.
- (Bausch 1979) = Bausch, Karl-Heinz: *Intuition und Datenerhebung in der Linguistik*. In: *Bergenholtz/Schaeder 1979*, 71-88.
- (Bergenholtz/ Schaeder 1979) = Bergenholtz, Henning / Schaeder, Burkhard (Hrsgg.): *Empirische Textwissenschaft*, Königstein 1979.
- (Bergenholtz/ Schaeder 1985) = Bergenholtz, Henning / Schaeder, Burkhard: *Deskriptive Lexikographie*, in: L. Zgusta (Hrsg.): *Probleme des Wörterbuchs*, Darmstadt 1985, 277-319.
- (Blumenthal 1988) = Blumenthal, Andreas: *Anmerkungen zur Konzeption eines Lexikographie-Systems*, in: *Snell-Hornby 1988*, 341-350.
- (Braunmüller 1977) = Braunmüller, Kurt: *Referenz und Pronominalisierung*, Tübingen 1977.
- („Brisante Wörter“) = Strauß/ Haß/ Harras 1989.
- (Brückner 1989) = Brückner, Tobias: *REFER – ein interaktives Zugriffssystem für Textkorpora*. Benutzerhandbuch, hrsgg. vom Institut für deutsche Sprache, Mannheim 1989.
- (Brückner 1989a) = Brückner, Tobias: *REFER. Konventionen für die Korpusaufnahme*, unveröff. Ms., Mannheim 1989.
- (Buck 1971) = Buck, G.: *Beispiel, Exempel, exemplarisch*. In: J. Ritter (Hrsg.): *Historisches Wörterbuch der Philosophie*. Basel 1971, Bd. 1, 818-823.

- (Bungarten 1979) = Bungarten, Theo: Das Korpus als empirische Grundlage in der Linguistik und Literaturwissenschaft, in: Bergenholtz/Schaeder 1979, 28-51.
- (Clear 1988) = Clear, Jeremy: Trawling the Language: Monitor Corpora, in: Snell-Hornby 1988, 383-389.
- (van Dijk 1980) = van Dijk, Teun A: Textwissenschaft, Tübingen 1980.
- (van Dijk 1985) = van Dijk Teun A.: Semantic Discourse Analysis, in: ders. (Hrsg.): Handbook of Discourse Analysis, London 1985, Vol. 2, 103-136.
- (Dokumentation Textkorpora des heutigen Deutsch 1982) = Dokumentation Textkorpora des heutigen Deutsch, hrsgg. vom Institut für deutsche Sprache, Mannheim 1982.
- (Francis 1982) = Francis, W. Nelson: Problems of Assembling and Computerizing Large Corpora, in: Johansson 1982, 7-24.
- (Gadamer 1975) = Gadamer, Hans-Georg: Wahrheit und Methode, Tübingen 4. Aufl. 1975.
- (Greimas 1971) = Greimas, Algirdas J.: Strukturelle Semantik. Methodologische Untersuchungen, Braunschweig 1971.
- (Harras 1989) = Harras, Gisela: Zu einer Theorie des lexikographischen Beispiels, in: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta (Hrsgg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie (= Handbücher zur Sprach und Kommunikationswissenschaft; 5.1), 1. Teilband, Berlin 1989, 607-614.
- (Hartmann 1983) = Hartmann, R.R.K. (Hrsg.): Lexicography: Principles and Practice, London 1983.
- (Haß 1986) = Haß, Ulrike: Textsorten als Wirkungssystem. Zum popularisierenden Transfer 'gelehrter' Inhalte in schriftlichen Texten an einem Beispiel aus dem 15./16. Jahrhundert, in: Deutsche Sprache 3/1986, 224-234.
- (Haß 1990) = Haß, Ulrike: Zu Bedeutung und Funktion von Beleg- und Kompetenzbeispielen im Deutschen Wörterbuch in: A. Kirkness/ P. Kühn/ H.E. Wiegand (Hrsgg.): Studien zum Deutschen Wörterbuch von Jacob Grimm und Wilhelm Grimm, Tübingen (ersch.).
- (Hausmann 1985) = Hausmann, Franz Josef: Lexikographie, in: Chr. Schwarze/ D. Wunderlich (Hrsgg.): Handbuch der Lexikologie, Königstein 1985, 367-411.
- (Hausmann 1989) = Hausmann, Franz Josef: Wörterbuchtypologie, in: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta (Hrsgg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie (= Handbücher

- zur Sprach- und Kommunikationswissenschaft; 5.1), 1. Teilband, Berlin 1989, 968-981.
- (Hellmann 1984) = Hellmann, Manfred W. (Hrsg.): Ost-West-Wortschatzvergleiche. Maschinell gestützte Untersuchungen zum Vokabular von Zeitungstexten aus der BRD und der DDR, Tübingen 1984.
- (Hermanns 1988) = Hermanns, Fritz: Das lexikographische Beispiel. Ein Beitrag zu seiner Theorie, in: G. Harras (Hrsg.): Das Wörterbuch: Artikel und Verweisstrukturen, Düsseldorf 1984, 161-195.
- (Hildebrandt 1977) = Hildebrandt, Rudolf: Die Diskussion um den linguistischen Begriff der Intuition, in: Sprachwissenschaft 2/1977, 134-150.
- (Hoffmann 1984) = Hoffmann, Walter: Probleme der Korpusbildung in der Sprachgeschichtsschreibung und Dokumentation vorhandener Korpora, in: W. Besch /O. Reichmann /S. Sonderegger (Hrsgg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung (= Handbücher zur Sprach- und Kommunikationswissenschaft; 2.1), 1. Teilband, Berlin 1984, 670-682.
- (Ilson 1985) = Ilson, Robert (Hrsg.): Dictionaries, Lexicography, and Language Learning, Oxford 1985.
- (Johansson 1980) = Johansson, Stig: The LOB Corpus of British English Texts: Presentation and Comments, in: ALLC Journal 1/1980, 25-36. 1980.
- (Johansson 1982) = Johansson, Stig (Hrsg.): Computer Corpora in English Language Research, Oslo 1982.
- (Kallmeyer 1974) = Kallmeyer, Werner et al.: Lektürekolleg zur Textlinguistik. Band 1, Frankfurt/M. 1974.
- (Koselleck 1989) = Koselleck, Reinhart: Sprachwandel und Ereignisgeschichte, in: Merkur 8/1989, 657-673.
- (Kühn 1978) = Kühn, Peter: Deutsche Wörterbücher. Eine systematische Bibliographie. Tübingen 1978.
- (Landau 1984) = Landau, Sidney I.: Dictionaries. The Art and Craft of Lexicography. New York 1984.
- (LDV-Info) = LDV-Info: Informationsschrift der Arbeitsstelle für linguistische Datenverarbeitung, hrsgg. vom Institut für deutsche Sprache, Mannheim.
- (Lexikon „Brisante Wörter“) = Strauß/ Haß/ Harras 1989.
- (Lutzeier 1985) = Lutzeier, Peter Rolf: Linguistische Semantik. Stuttgart 1985.

- (Martin 1989) = Martin, Robert: L'exemple lexicographique dans le dictionnaire monolingue, in: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta (Hrsgg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie (= Handbücher zur Sprach und Kommunikationswissenschaft; 5.1), 1. Teilband, Berlin 1989, 559-607.
- (Mattheier 1981) = Mattheier, Klaus J.: Wege und Umwege zur neuhochdeutschen Schriftsprache, in: Zeitschrift für germanistische Linguistik 9/1981, 274-307.
- (Mentrup 1979) = Mentrup, Wolfgang: Überlegungen zur Zusammenstellung und Verwendung eines Korpus für ein großes interdisziplinäres Wörterbuch der deutschen Sprache, in: Bergenholtz/ Schaefer 1979, 182-203.
- (Mittelstraß 1984) = Mittelstraß, Jürgen (Hrsg.): Enzyklopädie Philosophie und Wissenschaftstheorie, 2 Bde. Mannheim 1984.
- (Neubauer/ Petöfi 1981) = Neubauer, Fritz/ Petöfi, János S.: Wortsemantik, Lexikonsysteme und Textinterpretation, in: W. Heydrich (Hrsg.): Lexikoneinträge. Hamburg 1981, 149-186.
- (Pfeffer 1979) = Pfeffer, J. Allen: Drei deutsche Korpora am Institut für Grunddeutsch. Ziel und Stand ihrer Auswertung, in: Bergenholtz/ Schaefer 1979, 307-323.
- (Quirk 1986) = Quirk, Sir Randolph: Opening Remarks, in: R. Ilson (Hrsg.): Lexikography. An Emerging International Profession (Proceedings of colloquia sponsored by the ... Fulbright commission vol.I) Manchester 1986, 1-6.
- (Reichmann 1983) = Reichmann, Oskar: Möglichkeiten der Erschließung historischer Wortbedeutungen, in: M.A. van den Broek/ G.L. Jaspers (Hrsgg.): Festschrift für Anthony van der Lee, Amsterdam 1983, 111-140.
- (Reichmann 1986) = Reichmann, Oskar: Lexikographische Einleitung, in: R.R. Anderson/ U. Goebel/ O. Reichmann (Hrsgg.): Frühneuhochdeutsches Wörterbuch, Berlin 1986, Bd. 1, 10-164.
- (Reichmann 1988) = Reichmann, Oskar: Zur Funktion, zu einigen Typen und zur Auswahl von Beispielbelegen im historischen Bedeutungswörterbuch, in: K. Hyldgaard-Jensen/ A. Zettersten (Hrsgg.): Symposium on Lexicography III, Tübingen 1988, 413- 444.
- (Renouf 1984) = Renouf, Antoinette: Corpus Development at Birmingham University, in: J. Aarts/ W. Meijs (Hrsgg.): Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research. Amsterdam 1984, 3-39.
- (Rey 1987) = Rey, Alain: Le dictionnaire culturel, in: Lexicographica 3/1987, 3-50.

- (Rey-Debove 1971) = Rey-Debove, Josette (1971): *Étude linguistique et sémiotique des dictionnaires français contemporains*, The Hague, Paris 1971.
- (Schaeder 1979) = Schaeder, Burkhard: *Zur Methodik der Auswertung von Textkorpora für die Zwecke der Lexikographie*, in: Bergenholtz/Schaeder 1979, 220-267.
- (Schaeder 1981) = Schaeder, Burkhard: *Lexikographie als Praxis und Theorie*, Tübingen 1981.
- (Scheuerl 1969) = Scheuerl, Hans: *Die exemplarische Lehre. Sinn und Grenzen eines didaktischen Prinzips*, 3. Aufl. Tübingen 1969.
- (Schlaefler 1986) = Schlaefler, Michael: *Das Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm*. Informationsschrift hrsgg. von der Arbeitsstelle Göttingen des Deutschen Wörterbuchs (unveröff. vervielfältigtes Ms.) 1986.
- (Schmidt 1988) = Schmidt, Hartmut: *Sprachkompetenz und Sprachreflexion als Probleme der historischen Semantik*, in: *Sprachwissenschaft* 13/1988, 334-355.
- (Sherman 1979) = Sherman, D.: *Retrieving Lexicographic Citations from a Computer Archive of Language Materials*, in: R. R. K. Hartmann (Hrsg.), *Dictionaries and their Users*. Exeter 1979, 136-142.
- (Sinclair 1982) = Sinclair, John McH.: *Reflections on Computer Corpora in English Language Research*, in: Johansson 1982, 1-6.
- (Sinclair 1985) = Sinclair, John: *Lexicographic Evidence*, in: Ilson 1985, 81-92.
- (Snell-Hornby 1988) = Snell-Hornby, Mary (Hrsg.): *Züri-LEX '86 Proceedings*, Tübingen 1988.
- (Steger 1980) = Steger, Hugo: *Soziolinguistik*, in: H. P. Althaus/ H. Henne/ H.E. Wiegand (Hrsgg.): *Lexikon der germanistischen Linguistik*, 2. Aufl. Tübingen 1980, Bd. II, 347-358.
- (Strauß 1988) = Strauß, Gerhard: *Artikelsorten und Artikelstrukturen im „Lexikon schwerer Wörter im Deutschen“*. In: G. Harras (Hrsg.), *Das Wörterbuch: Artikel und Verweisstrukturen*, Düsseldorf 1988, 196- 222.
- (Strauß 1989) = Strauß, Gerhard: *Angabe traditioneller Wortarten oder Beschreibung nach funktionalen Wortklassen*, in: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta (Hrsgg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie (= Handbücher zur Sprach- und Kommunikationswissenschaft; 5.1)*, 1. Teilband, Berlin 1989, 788-796.

- (Strauß/ Haß/ Harras 1989) = Strauß, Gerhard/ Haß, Ulrike/ Harras, Gisela: *Brisante Wörter von Agitation bis Zeitgeist. Ein Lexikon zum öffentlichen Sprachgebrauch*, Berlin 1989.
- (Strauß/ Zifonun 1985) = Strauß, Gerhard/ Zifonun, Gisela (1985): *Die Semantik schwerer Wörter im Deutschen*, 2 Bde., Tübingen. 1985.
- (Tollenaere 1973) = Tollenaere, F. de: *The Problem of the Context in Computer-aided Lexicography*, in: Aitken/ Bailey/ Hamilton-Smith 1973, 25-35.
- (Viehweger 1976) = Viehweger, Dieter: *Semantische Merkmale und Textstruktur*, in: F. Danes/ D. Viehweger (Hrsgg.): *Probleme der Textgrammatik*, Berlin (DDR) 1976, 195-206.
- (Viehweger 1988) = Viehweger, Dieter: *Kollokationen im Lexikon und deren Darstellung im Wörterbuch*, in: K. Hyldgaard-Jensen/ A. Zettersten (Hrsgg.): *Symposium on Lexicography III*, Tübingen 1988, 107-135.
- (Wiegand 1983) = Wiegand, Herbert Ernst: *Ansätze zu einer allgemeinen Theorie der Lexikographie*, in: J.Schildt/ D. Viehweger (Hrsgg.): *Die Lexikographie von heute und das Wörterbuch von morgen*, Berlin (DDR) 1983, 92-117.
- (Wiegand 1985) = Wiegand, Herbert Ernst: *Einige grundlegende semantisch-pragmatische Aspekte von Wörterbucheinträgen*, in: L. Zgusta (Hrsg.): *Probleme des Wörterbuchs*, Darmstadt 1985, 342-377.
- (Wolski 1989) = Wolski, Werner: *Die Synonymie im allgemeinen einsprachigen Wörterbuch*, in: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta (Hrsgg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie* (= *Handbücher zur Sprach- und Kommunikationswissenschaft*; 5.1), 1. Teilband, Berlin 1989, 614-628.