

## Zur Mikrostruktur im Hypertext-Wörterbuch

- |  |                                    |
|--|------------------------------------|
| 1 Mediale Bedingungen  | 4 Vom Verweis zur Linktypologie    |
| 2 Mikrostrukturelle Grundsatzentscheidungen                        | 5 Zwang zu theoretischer Reflexion |
| 3 Interne Differenzierung und<br>Modularisierung der Mikrostruktur | 6 Literatur                        |

In diesem Beitrag werden erste Erfahrungen mit und Überlegungen zu der Aufgabe dargelegt, ein Mikrostrukturenprogramm für ein Hypertext-Wörterbuch zu entwerfen.<sup>1</sup> Zur Hypertextualisierung gedruckter Wörterbücher gibt es inzwischen erste Veröffentlichungen;<sup>2</sup> meist bleibt hier die Bindung an eine gedruckte Vorlage, und sei die Hypertextualisierung noch so konsequent, bestehen.<sup>3</sup> Im Unterschied zu solchen Hypertext-Wörterbüchern gehen nachfolgende Überlegungen von einem vorlagenunabhängigen Hypertext aus, dessen allgemeines Ziel es ist, Informationen zum deutschen Wortschatz zu vermitteln. Die hier vorgestellten Erfahrungen und Überlegungen sind an ein konkretes Projekt gebunden: LEKSIS – das lexikalisch-lexikologische Informationssystem des Instituts für Deutsche Sprache, Mannheim<sup>4</sup>. Auf eine (weitere) Projektbeschreibung wird hier aber verzichtet; sie findet sich in Fraas/Haß-Zumkehr (1999), ferner auf der Homepage unter <http://www.ids-mannheim.de/wiw>. Vor dem Hintergrund dieses Projektes stehen die Bedingungen bzw. lexikografischen Konsequenzen des Mediums Hypertext im Unterschied zum Druck zur Diskussion.

### 1 Mediale Bedingungen

LEKSIS versteht sich als Informationssystem zum Wortschatz der deutschen Gemeinsprache bzw. zu seinen Elementen; es soll online verfügbar sein und stets aktuell gehalten werden. Die Informationsmenge<sup>5</sup> wird in einer (objektorientierten) Datenbank abgelegt und verwaltet, an deren ‚Füllung‘ sich mehrere Forschergruppen zu unterschiedlichen (lexikologischen) Gebieten (zunächst des IDS) beteiligen sollen; die Beiträge dieser Gruppen heißen in unserem Zusammenhang (Projekt-)Module.

Im Unterschied zur Planung eines – großen – Wörterbuchs bedeuten zeitliche Nicht-Befristung, ständige Aktualität und frühzeitige Verfügbarkeit, dass es unmöglich ist, fest-

---

<sup>1</sup> Für die sehr fruchtbaren Diskussionen, die den Hintergrund dieses Beitrags bilden, danke ich den Mitgliedern meines Teams (in alphabetischer Reihe) Cyril Belica, Claudia Fraas, Carolin Müller, Sonja Müller-Landmann und Kathrin Steyer.

<sup>2</sup> Da dies kein Forschungsbericht zum Thema Hypertext-(Lexikografie) ist, sei exemplarisch und wegen der dort angegebenen weiterführenden Literatur verwiesen auf Lemberg/Petzold/Speer (1998) und Storrer/Harriehausen (1998). Zu Hypertext vgl. Storrer (2000).

<sup>3</sup> Ein Negativ-Beispiel erläutert Wiegand (1998[99]).

<sup>4</sup> Seit Frühjahr 2000 mit dem Namen *Wissen über Wörter / WiW*.

zulegen, in wie vielen Jahren mit wie vielen Mitarbeiterinnen und Mitarbeitern ein ‚fertiges‘ Produkt vorliegen kann. Da das System für diverse Projektmodule offen bleiben muss, auch für solche, die es heute noch gar nicht gibt, kann und muss die Mikrostruktur so umfassend und differenziert wie möglich sein.

Die lexikografischen AutorInnen des Systems werden auf die Datenbank zu eigenen und verwandten wissenschaftlichen Zwecken (auch) unmittelbar zugreifen; die eigentlichen Nutzer allerdings werden ausschließlich über ein Hypertextsystem je nach ihren speziellen Interessenlagen und unter Berücksichtigung auch nicht-linguistischer Vorbildung recherchieren können. Das Hypertextsystem soll aus der Datenbank die für eine Rechercheanfrage nötigen Informationen herausholen und mithilfe bestimmter Seitenlayouts im Internet ansprechend präsentieren. So kann der eine Nutzer beispielsweise die Bedeutungsparaphrasen aller Glieder eines onomasiologischen Paradigmas nebeneinander auf dem Bildschirm sehen wollen; die andere Nutzerin interessiert sich hingegen für die Vernetzung des semasiologischen Feldes, der dritte vielleicht für den Vergleich von Kollokations- und Valenzangaben aller Einzelbedeutungen eines Lemmas; und die vierte sucht ausschließlich wortgeschichtliche Informationen zu einem bestimmten Lemma samt Belegen. Daneben gibt es diejenigen Nutzer, die überhaupt ‚erst einmal schauen‘ wollen, d.h. die einen ersten Überblick über die je Lemma gebotenen Informationen gewinnen und nicht gleich ins Detail gehen wollen. Für sie müssen in geeigneter Weise zusammenfassende und allgemein verständliche, nicht-linguistisch formulierte Angaben sowie Überblicks-‚karten‘ je Wortartikel vorgehalten werden

Beim gedruckten Wörterbuch hängt die Festlegung des Mikrostrukturenprogramms normalerweise von der Einschätzung des Adressatenkreises ab (vgl. Wiegand 1998, 249) – bei einem Hypertextsystem wie LEKSIS eher nicht mehr: Hier wird die Strukturierung der Daten von ihrer Präsentation getrennt vorgenommen, denn die Ausarbeitung der Präsentationsform(en) wird als eigene lexikografische Handlung geplant, die auf die Erarbeitung der Mikrostrukturen nur zurückgreift. Das Mikrostrukturenprogramm steht also der Datenbank und ihrer Struktur näher als der nutzerzugewandten Hypertextoberfläche. Von allen möglichen Nutzungsprofilen muss das Mikrostrukturenprogramm zunächst insoweit unabhängig bleiben, als in diesem Informationspotential lediglich diejenigen Elemente definiert und geordnet werden, die bei irgend einer Recherche relevant werden *könnten*. Es handelt sich also um eine Art lexikografisches Maximalprogramm,

- das (1.) viel mehr Wörterbuchtypen in sich vereint als im Druckmedium möglich,
- das (2.) ein viel umfassenderes Mikrostrukturenprogramm erfordert als im Druckmedium nötig,
- das (3.) ständig ergänzt und korrigiert werden können muss und
- (4.) dessen interne Struktur nicht die eines primär linear aufgebauten Textes ist, sondern die eines Hypertextes.

Aus Letzterem folgt vor allem, dass die Textsegmente bzw. die auf irgend eine Weise definierten atomaren Elemente der Mikrostruktur unabhängig voneinander rezipierbar, dennoch kombinierbar sein und entsprechend kontextfrei formuliert werden müssen, denn jeder Nutzer, jede Nutzerin rezipiert sie in einer potenziell anderen Reihenfolge oder in einer potenziell anderen typografischen Nähe.

<sup>5</sup> Der Ausdruck *Information* wird hier – angesichts der interdisziplinär uneinheitlichen Bestimmung und Abgrenzung von *Information*, *Wissen* und *Daten* – absichtlich völlig unterterminologisch verwendet.

Das Maximalprogramm bedeutet aber nicht, dass zu jedem Lemma die selbe informatorische Dichte vorhanden sein wird – im Gegenteil: der im Prinzip unendliche Prozess des Auf- und Ausbaus eines online-Informationssystems wie LEKSIS wird Nutzer und Nutzerinnen daran gewöhnen, dass nicht zu jedem Lemma die gleiche Menge, Dichte und Tiefe an Information zur Verfügung steht. Dies ist aber kein Nachteil gegenüber umfangreichen gedruckten Wörterbuchprojekten, bei denen jahre- und jahrzehntelang nur ein Teil der Alphabetstrecke benutzbar ist, bei denen Verweise bis zum Schluss vielfach ins Leere gehen und bei denen die ersten Lieferungen veralten, bevor die letzten erscheinen. In einem Informationssystem muss den Nutzern selbstverständlich zu einem sehr frühen Zeitpunkt ihrer Recherche ein Überblick über die vorhandenen und die nicht-vorhandenen Informationen gegeben werden; Hypertextsysteme können dies mithilfe sog. fish-eye-views, bei denen ähnlich dem gleichnamigen Kamera-Objektiv die unmittelbare Informationsumgebung größer oder detaillierter gezeigt wird als die weiter entfernt liegende, und dynamischer site maps verwalten: letztere sind Landkarten ähnliche Darstellungen des Informationsgehalts, die sich je nach Benutzerinteresse verwandeln (lassen). Gedruckte Wörterbücher wie etwa das *Deutsche Wörterbuch* können dies nicht oder erst bei ihrer elektronischen Konversion und entsprechenden Zusätzen.

Der Entwurf des Mikrostrukturenprogramms für ein online verfügbares Hypertext-Wörterbuch hat also von einem maximalen Informationsangebot auszugehen und *alles* zu berücksichtigen, was in irgendeiner Weise gesucht werden oder bei einer Suche einbezogen werden könnte. Damit einher geht die Notwendigkeit, das Informationsangebot weitestgehend zu atomisieren, weitergehend jedenfalls als es bei gedruckten Wörterbüchern üblich und angemessen ist und auch weitergehend als bei Wörterbüchern, die von einer Druck- in eine elektronische Fassung konvertiert werden. Das auf eine Datenbank hin angelegte Mikrostrukturenprogramm umfasst somit textuelle Elemente (z.B. narrativ formulierte Bedeutungsgeschichten, Zusammenfassungen zu den einzelnen Informationsdimensionen wie Semantik, Grammatik, siehe unten), atomare Elemente wie alphanumerische Angaben (z.B. Frequenzangaben, Datierungen) und Attribute wie Genus- oder Valenzangaben, die auf der Nutzeroberfläche erst zu einem lesbaren Teiltexst zusammengefügt bzw. ergänzt werden müssen, vorstellbar etwa in Form eines Lückentextes wie

Ungeachtet der Unterscheidung der [...] Lesarten kommt das Stichwort im Gesamtkorpus [...] mal, im presssprachlichen Subkorpus [...] mal und im Subkorpus der sprechsprachnahen Texte [...] mal vor.

Etwa um [...] wurde das Stichwort aus dem [...] ins Deutsche entlehnt.

Oder: Entlehnung ins Deutsche aus: [...] um [...]

Was hier durch eckige Auslassungsklammern gekennzeichnet ist, muss aus der Datenbank geholt und eingesetzt werden. Voraussetzung ist, dass die jeweiligen Informationselemente dort eindeutig identifiziert sind und eine Form haben, die in solch einen Lückentext „passt“.

Aus dem Vorangehenden wird deutlicher, worin die Aufgabe des Entwurfs des Mikrostrukturenprogramms besteht:

1. aus der eindeutig identifizierenden Bestimmung der Elemente oder ‚informationellen Atome‘,
2. aus der Festlegung ihrer jeweiligen Datenart und Eigenschaften (numerisch, alphanumerisch, standardisierter Text, freier Text, beschreibungssprachlich, objektsprachlich, u.a.),

3. aus der Erarbeitung einer (teilhierarchischen) Elementenstruktur, die relativ unabhängig von Benutzungssituationen ist, und
4. aus der Festlegung möglicher Interrelationen zwischen den Elementen (Verlinkungssystem).

Die Lösung der genannten Aufgaben ist auf computerlinguistische und texttechnologische Kenntnisse ebenso angewiesen wie auf lexikografisch-linguistische – dies alles kann wohl nur in einem Team zusammengeführt werden. Betonen möchte ich an dieser Stelle, dass das Festlegen einer einheitlichen Datenstruktur und das Festlegen eines Datenbankmodells – etwas anders als die Einschätzung in Wiegand (1998, 208) – sich als eine durch und durch lexikografische Tätigkeit erweisen, die allerdings durch die medialen Bedingungen und durch die Teammitglieder mit informatischen Kompetenzen viel stärker in Richtung Entscheidungseindeutigkeit, logische Konsistenz und Redundanzfreiheit verändert wird – eine Veränderung, die man als Verbesserung bezeichnen muss. Von ‚Standardisierung‘ lexikografischer Tätigkeit spreche ich hier absichtlich nicht, denn damit wird oft implizit auf eine Beschränkung lexikografischer Freiheit und Behinderung von Kreativität Bezug genommen, die im Fall eines dynamischen Hypertextsystems keineswegs gegeben sind. Im Gegenteil, der Zwang, ein umfassendes Mikrostrukturenprogramm zu entwerfen, das in eine SGML/XML-DTD und weiter in eine Datenbankstruktur übersetzt werden soll, hat – wie nachfolgend gezeigt wird – Kreativität freigesetzt.

Es stellt sich die Frage, warum man erst eine Document Type Definition (DTD) für SGML/XML entwerfen sollte, wenn man eigentlich ein Datenbankmodell braucht. Die Antwort lautet: weil eine SGML/XML-Datenstruktur datenbankunabhängig ist. Dies fällt allerdings ins Gewicht, wenn man zu Beginn noch nicht weiß, welche Datenbanksoftware man sich in Zukunft leisten kann, wie zufrieden man mit ihr sein wird, ob nicht gewisse Systemteile einmal als gedrucktes Wörterbuch herausgezogen werden sollen, und: ob nicht einmal externe Kooperationspartner mit anderer Software sich am Ausbau des Informationssystems werden beteiligen wollen. Alles dies ist unproblematisch, wenn die komplette Datenstruktur in einem datenbankunabhängigen Format wie SGML/XML vorliegt.

## 2 Mikrostrukturelle Grundsatzentscheidungen

In der Metalexikografie wird davon ausgegangen, dass zu jedem Lemmazeichentyp je eine spezifische abstrakte hierarchische Mikrostruktur gehört (Wiegand 1998, 215). Werden die Mikrostrukturen aber in Form einer DTD dargestellt, so ist die Existenz mehrerer paralleler Strukturen problematisch: Die Redundanzen partiell identischer Mikrostrukturen sind Fallen für widersprechende Angaben, so dass ein informatischer Rat zu beherzigen ist: Keine Angabe darf zweimal in den gleichen Wortartikel eingetragen werden (müssen). Und: Es darf keine doppelten Strukturen geben. Die (einzige) Alternative zum Ansatz mehrerer, lemmatypspezifischer Mikrostrukturen besteht im Ansatz einer einzigen, „mehrzweckgeeigneten“, in sich modularen Mikrostruktur, die alle lemmatyp-spezifischen Artikelpositionen/Informationsarten enthält und aus der auf der Autorenoberfläche spezifische Untermengen je Lemmatyp ausgegliedert werden. Die Unterscheidung lemmatypspezifischer Informationsprogramme ist dann Aufgabe der Projektmodule bzw. Forschergruppen, die zum Gesamtsystem beitragen. Sie müssen vor allem den semantisch-pragmatischen Lemmatypen gerecht werden, indem sie die obligatorischen Angaben und not-

wendige Eigenschaften der Nutzeroberfläche festlegen. Hier sorgt das Hypertextsystem dafür, dass die nicht relevanten und im System unausgefüllt gebliebenen Angabearten ausgeblendet werden.

Der modular-hierarchische Charakter der Mikrostruktur zeigt sich z.B. schon darin, dass sich die lexikografischen Autoren z.B. *erst* über die Zuordnung zu einer Wortart entscheiden und *danach* die Flexionsangaben machen müssen, für die das System die passenden Kategorien vorschlägt

Die vorläufig komplexeste Klammer der Mikrostruktur ist der Wortartikel, wobei unter „Wort“ alles dasjenige fällt, was als Lemmazeichen infrage kommt, unter Umständen also auch unselbstständige Wortbildungselemente und Mehrworteinheiten. Auch außerhalb der „Wortartikel“ wird wohl jedes lexikalische Informationssystem noch weitere, den sog. Umtexten analoge Teile enthalten: Benutzungshinweise, Informationen über das Projekt, die Konzeption des Ganzen und diejenige der realisierten Module, bibliografische Daten, ein terminologisches Glossar, ggf. eine Grammatik und ähnliches. Auf diese wird nachfolgend nicht weiter eingegangen.

Wie ist die Mikrostruktur intern gegliedert, wenn nicht (mehr) von einem linearen Rezeptionsprozess beginnend bei der Lemmazeichengestaltangabe und endend etwa bei den Belegangaben ausgegangen werden kann? Eine Trennung von Datenmodellierung und Nutzeroberfläche bedeutet nicht, dass die einzelnen Angaben in gar keinen strukturbedingten Interrelationen mehr zu sehen sind. Sie alphabetisch, nach Datentyp oder nach Umfang zu sortieren, hieße Zusammengehöriges auseinanderzureißen und wortartikelinterne Klammerungen auszuschließen. Es wäre absurd, wollte man die lexikografischen AutorInnen je Lemma der alphabetischen Reihe nach Abkürzungsangabe, Ableitungsangabe... Wortbildungsproduktivitätsangabe, Worttrennungsangabe bis zuletzt die Zitatangabe eintragen lassen. Es sind lediglich partiell *andere* Inhaltsaspekte, die die Datenmodellierung einerseits und die Gestaltung der Nutzeroberfläche andererseits leiten.

Hinzu kommt die Arbeitsteilung, die bei größeren und längerfristigen lexikografischen Projekten unausweichlich und notwendig ist. Zwar wurde oben gesagt, dass ein Informationssystem wie LEKSIS wörterbuchtypenspezifisch ist, das heißt aber nicht, dass die AutorInnen bei ihrer Wortschatzbeschreibung keine Schwerpunkte analog zu Wörterbuchtypen setzen. Die partiell gleiche Lemmastrecke wird beispielsweise überlappend von einer Forschergruppe Synonymik, von einer Forschergruppe Fremdwortschatz, einer Forschergruppe Neologie und einer Forschergruppe Kollokationen bearbeitet. Jede dieser Gruppen nimmt je spezifische Angaben etwa zu einem Lemma wie *realisieren* vor und muss diese Angaben mit anderen, inhaltlich „benachbarten“ kompatibel halten. Die mikrostrukturellen Vorgaben für die Autoren dürfen also in linguistisch-lexikologischer Hinsicht nicht kontextblind sein.

Außerdem ist trotz der erwähnten planerischen Trennung von Datenmodell und Nutzeroberfläche die frühzeitige Beschäftigung mit nutzerseitigen Aspekten nicht obsolet. Die Bestimmung der atomaren Informationseinheiten darf nicht nur aus linguistischer Sicht im Hinblick auf linguistische Nutzer geschehen, sondern muss bestimmte Informationseinheiten auch aus sprachinteressierter Sicht und damit gröber, zusammenfassender definieren. Bei Letzterem kommen die tradierten Erwartungen an die Textsorte Wörterbuch stärker ins Spiel. Sie sollten auf der neuen Entwicklungsstufe der Textsorte, die bei einem Hypertext-Wörterbuch ansteht, aber nicht so dominant werden, dass die Erwartungen, die an den Papierdruck gebunden sind, lediglich elektronisch kopiert werden.

### 3 Interne Differenzierung und Modularisierung der Mikrostruktur

Auf der Basis der in den vorhergegangenen Abschnitten erläuterten Aufgaben und Grundsatzentscheidungen ließe sich sicherlich mehr als eine Mikrostruktur konstruieren. Entscheidend ist daher, dass sich diejenige, für die sich ein Projekt wie LEKSIS (möglichst zügig) entscheiden muss, gut begründen lässt.

Interne Differenzierung und Ordnung ergibt sich erstens aus der Klassifizierung der elementaren Informationseinheiten, zweitens aus der hierarchischen Setzung weiterer Klammerungen innerhalb der Klammer Wortartikel.

In der Mikrostruktur wird zwischen Klassen von *Angaben*, *Kommentaren* und *Erläuterungen* unterschieden; alle gemeinsam werden hier als *Angaben* zusammengefasst. Nach Reichmann (1986, 152ff.) unterscheiden sich Angaben und Kommentare dadurch, dass Angaben „das als faktisch Hingestellte“ (ebd. 152) und Kommentare darauf bezügliche Gewichtungen, Interpretationsvorschläge und Verständnishilfen enthalten. Diese Unterscheidung wird in LEKSIS übernommen, aber noch strikter gefasst, insofern Kommentare keinerlei ‚Zusatzinformationen‘ enthalten dürfen, für die in einem informatorischen Maximalprogramm ja anderswo Platz vorgesehen ist.

Kommentare setzen mindestens eine Angabe voraus, auf die sie sich beziehen; sie dürfen keine neuen Informationen bringen und keine Daten enthalten, die für die Datenbank identifizierungsrelevant sind. Kommentare können zu *jeder* Angabe vorgenommen werden; eine Beschränkung auf einige wenige ist bei Wegfall des im Druck gegebenen Platzproblems nicht mehr zu rechtfertigen.

Der Angabetyp *Erläuterung* dient der Strukturierung der Umtexte.

*Angabe(klasse)n* im engeren Sinne unterscheiden sich in verschiedenen Hinsichten: Aus lexikografischer Sicht gibt es obligatorische und fakultative relativ zum Gesamtsystem und relativ zu einem bestimmten Bearbeitungsmodul. Beispiel: Die Zahl der obligatorischen Angaben wird in einem umfassenden, bündelnden Informationssystem wie LEKSIS eher gering sein und neben der Lemmazeichengestaltangabe vor allem die mittels korpuslinguistischer Verfahren (teil)automatisch zu gewinnenden Angaben enthalten. Für den Teilwortschatz „Neologismen“ oder den Teilwortschatz „sprachhandlungsbezeichnende Verben“ werden darüber hinaus eigene Obligatoriken entworfen, hier die Angaben zur Entlehnung oder Wortbildung, dort die Angaben zu Synonymie, Antonymie und Pragmatik.

*Angabe(klasse)n* unterscheiden sich ferner nach ihrem Status als (a) autonomer Text, (b) funktionales Textsegment, (c) tabellarisches Element. Für gedruckte Wörterbücher definiert Wiegand (1989, 464) Angaben als funktionales Textsegment – in einem datenbankbasiertem Hypertextsystem kommen zwei weitere Möglichkeiten hinzu: Als autonome Textsegmente fungieren solche Angabeklassen, die auf der Benutzeroberfläche in genau derselben Form und ohne Zusätze präsentiert werden, wie dies z.B. für Bedeutungs- ‚Geschichten‘ oder für die Beschreibung pragmatisch komplexer Verwendungszusammenhänge erforderlich ist. Falls in solchen ‚Geschichten‘ identifizierbare Angaben zu Zeit, Raum oder zu Kategorien relativ begrenzter Kategorieninventare (z.B. abgekürzte Bezeichnungen für Herkunftssprachen) enthalten sind, sollten diese im Fließtext markiert, d.h. getaggt werden, damit sie der Datenbank nicht verloren gehen. Als (b) funktionales Textsegment werden solche Angaben klassifiziert, die sich grob gesprochen in einen standardisierten Lückentext (s.o.) einfügen lassen müssen. Solche Lückentexte dienen auf der Benutzerseite dazu, Angaben zu erläutern, in Zusammenhänge einzuordnen und so ‚verdaulicher‘ zu machen. Geeignet hierfür sind z.B. formgeschichtliche Angaben, Angaben zum

Zusammenhang eines semasiologischen Feldes oder Angaben zur internen Struktur von Kollokationen. Als (c) tabellarische Elemente werden solche Angaben klassifiziert, die keine kotextuellen Erläuterungen erfordern, wie dies exemplarisch bei Genusangaben, Flexionsparadigmen u.ä. der Fall ist. Auf Benutzerseite können hier bei Bedarf minimale Überschriften oder Etikettierungen wie „Flexion: ...“ hinzutreten.

Für jede Angabeklasse muss ferner eine Reihe von Eigenschaften oder Attributen festgelegt werden, die weitere Klassifizierungen zuließen. Diese Festlegungen antworten vereinfacht gesagt auf die Frage: Woraus besteht diese Angabe? Aus objektsprachlichen Ausdrücken (wie bei einer Synonymenangabe), aus beschreibungssprachlichem Text (wie bei einer Bedeutungsparaphrasenangabe), aus beidem? Wenn beschreibungssprachlicher Text zu produzieren ist, müssen dabei standardisierte Kategorialexpressionen (für Genus, für Varietätzugehörigkeit, für semantische Verschiebungen, u.ä.) verwendet werden oder nicht?

Nur eine Kategorie spielt bei der Klassifizierung von Angaben im Hypertext-Wörterbuch *keine* Rolle: die Position im (gedruckten) Wortartikel.

Soweit die Überlegungen zu den Elementen der Mikrostruktur. Nun zum System, das die Angabeklassen bilden.

Der LEKSIS-Wortartikel wird in sechs informationelle Dimensionen bzw. – auf der Ebene der Datenmodellierung: – komplexe Objekte geteilt, die untereinander gleichrangig und nicht-sequenziell sind. Für deren Ansatz ist zunächst eine auf die Nutzerseite bezogene Entscheidung verantwortlich: Nach Wahl eines Lemmas werden den Nutzern kreisförmig angeordnete Schaltflächen zu folgenden Informationsdimensionen geboten, hinter denen sich die Suchwege in Details verzweigen. Die linguistisch systematische Motiviertheit dieser Dimensionen ist unschwer zu erkennen:

- Schreibung & Aussprache/auf die Ausdrucksseite bezogen
- Bedeutung & Verwendung/auf die Inhaltsseite, semantisch und pragmatisch, bezogen
- Grammatik
- Historisches & Sachliches/Formgeschichte, Bedeutungsgeschichte, Enzyklopädisches
- Dokumentation/Belege, Zitate, korpusanalytische Daten, fremdsprachige Äquivalente, Vergleichsdaten u.a.
- Sprachkritisches & Normatives / Sprachreflexives; hier soll die normative Wirkung deskriptiver Angaben kommentierend transparent gemacht, zur Diskussion gestellt und relativiert werden.

Ein Problem, das sich noch nicht bei der Erarbeitung der Datenmodellierung, aber später stellt, ist die Benennung dieser Schaltflächen.

Die klassische Lexikografie braucht wegen des Zwangs zu Knappheit und hoher Textverdichtung kaum jemals explizite Benennungen der Angabe- oder Informationsarten je Artikelposition zu geben; die Nutzer lernen sie durch Übung und anhand der Typografie zu identifizieren und auseinanderzuhalten, seltener auch durch Benutzungshinweise; sie haben in der Regel aber keine Begriffe dafür.

In der Mikrostruktur klassischer Wörterbücher wird der Terminus Bedeutungsidentifizierungskennzeichnung verwendet (Wiegand 1998, 226), um Bedeutungsstellennummern einer strukturellen Kategorie zuzuordnen. Gemeint sind Angaben der Art „zu Absatz (3) gehört das Kompositum *Absatzzahlen*“. Solche Bedeutungsstellennummern sind in einem Hypertext-Wörterbuch problematisch, wenn nicht gänzlich unbrauchbar, weil ein Wortartikel über mehrere Bildschirmseiten verteilt ist und die Nutzer nicht mehr durch einen Blick in den oberen Druckraum ansehen können, welche Einzelbedeutung die Nummer „(3)“ hat. Die semantisch „leeren“ Zahlziffern müssen in einem hypertextuellen Informa-

tionssystem durch Identifikatoren ersetzt werden, die die Zuordnung zu einer der Einzelbedeutungen ohne Umwege gestatten. Wir haben dafür eine sehr kurze Form der semantischen Paraphrase gewählt, die als Bedeutungssettikettierungsangabe bezeichnet und immer dann eingesetzt wird, wenn einzelbedeutungsbezogene Angaben gemacht werden müssen. Etwa: „zu *Absatz* in der Bedeutung ‚Verkauf‘ gehört das Kompositum *Absatzzahlen*“; hier wird *Verkauf* als Bedeutungssettikettierungsangabe getaggt. Selbstverständlich wird von solch einem semantischen Ettikett immer per Link auf die vollständige semantische Paraphrase verwiesen.

Elektronische Informationssysteme zwingen die Nutzer immer wieder zu Informationsselektionen, was von den Autoren des Systems verlangt, die selegierbaren Teile ihres Informationsangebots in den vorgeschalteten Menüs so zu benennen, dass Nutzer die ihren Bedürfnissen entsprechende Wahl treffen können. Es müssen beschreibungssprachliche Etikettierungen oder – was noch schwieriger sein dürfte: non-verbale Icons – gefunden werden, wobei die terminologische Kluft zwischen Linguisten und Sprachinteressierten scharf zutage tritt. Was LinguistInnen hinter „Angaben zur Semantik“ oder „Angaben zur Pragmatik“ oder „zum Gebrauch“ vermuten, ist – wie ein Gespräch mit Studierenden zeigte – nicht selten partiell identisch mit dem, was unter Etymologie („was das Wort ursprünglich heißt“) fällt; „Gebrauch“ wird leicht mit morphologisch-syntaktischen Normen assoziiert („ob man *wegen* mit Dativ oder Genitiv gebraucht“) oder auch mit semantischer Variation („ob das Wort übertragen gebraucht wird“); Belege bzw. Zitate werden als „Beispiele“ klassifiziert und entsprechend, d.h. tendenziell als vorbildlich verstanden (!), aber kaum als Nachweis der in den Angaben formulierten Regeln und als Verifikationsinstanz der lexikologischen Methode.

Für die Wörterbuchbenutzungsforschung öffnen sich hier neue Horizonte. Und: lexikalische Informationssysteme können das begriffliche Inventar sprachreflexiven Wissens der Nutzer und damit vielleicht tendenziell der Sprachgemeinschaft verändern. Das Benennungsproblem ist also kein nebensächlicher Aspekt.

Für die Definition von Elementen und damit der Tags für die DTD genügen Abkürzungen von Benennungen, die die AutorInnen verstehen, selbst wenn sie einem projektspezifischen Jargon entstammen. Allerdings sollte man bei längerfristigen elektronischen Wörterbüchern (und welche bezeichnen sich freiwillig als kurzfristig?!) auf die „Nachhaltigkeit“ und transparente Dokumentation des Benennungssystems achten, damit nachfolgende Wissenschaftlerinnen und Wissenschaftler Inhalt und Funktion einer Angabeklasse eindeutig identifizieren können. Wichtig ist aber vor allem die begründete Bestimmung der Angabeart selbst, weshalb wir parallel zum Aufbau der Mikrostruktur ein Redaktions- oder Instruktionshandbuch „mitschreiben“, in dem die Überlegungen zur Ansetzung und zur Charakteristik jeder Angabeart festgehalten werden können. Dieses wird den AutorInnen später während der Abfassung der Wortartikel online zur Verfügung stehen.

Die o.g. sechs Informationsdimensionen werden formal als komplexe Objekte oder Klammern aufgefasst, die bestimmte Angaben enthalten, die ihrerseits weitere Angaben einklammern, und/oder eine Reihe atomarer, d.h. nicht weiter zerlegbarer Elemente enthalten. Beispielweise enthält die Bedeutungsdimension u.a. ein komplexes Objekt/ eine Klammer „Paradigmatik“, in der die Angaben zu Synonymen, Antonymen usw. zusammengefasst sind, wobei man die Antonymenangabe bei Bedarf typologisch weiter differenzieren könnte. Ausschlaggebend für den Differenzierungsgrad ist das antizipierte lexikologische Interesse an den Datenbankinformationen und ihrer Verknüpfung. Auch zukünftige Forschungen sollten, so schwierig dies erscheint, weitmöglichst berücksichtigt werden, etwa indem ein noch undifferenziertes Datenobjekt vorsorglich eingebaut wird. Atomare



Objekte sind „objektsprachlicher Text“ und „beschreibungssprachlicher Text“, obwohl auch sie genau genommen aus einer Menge Text zuzüglich obligatorischer Autoren- und Datumsangabe bestehen. Letzteres ist für die langfristige Verwaltung und Aktualisierung der Daten unerlässlich.

Einerseits enthalten die übergeordneten sechs informatorischen Komplexe genau die Menge an Angabearten, die linguistisch wünschbar ist, andererseits müssen in ihnen aber auch solche Angabearten enthalten sein, die nicht-linguistische Nutzer mit Bedarf an Überblicks- und zusammenfassender Information nachfragen könnten. Man kann davon ausgehen, dass solche Zusammenfassungen sich gerade nicht schematisch oder automatisch aus den linguistisch detaillierten Angaben ableiten lassen. Diese zu komprimieren, erhöht die Abstraktion noch. Vielmehr werden sie partiell gesondert zu formulieren und an entsprechender Stelle in der Datenbank abzulegen sein. An diesem Punkt zeigt sich ein weiterer Vorteil des Ansatzes der sechs komplexen Informationsobjekte. Diese können neben linguistisch motivierten Angabarten immer auch eine oder mehrere Angaben vom Typ „zusammenfassender Kommentar für sprachinteressierte Nicht-Linguisten“ enthalten, auf die das Hypertext generierende System immer dann zurückgreift, wenn zuvor eine entsprechende Benutzungssituation oder ein Benutzertyp ausgewählt worden ist. Es ist sogar denkbar, dass diese Arbeit nachträglich oder gesondert von hierzu besonders befähigten Autoren vorgenommen wird.

#### 4 Vom Verweis zur Linktypologie

Was im gedruckten Wörterbuch der Verweis ist, ist im Hypertext der Link. Vor dem Hintergrund möglicher Links, die in eine Mikrostruktur einzubauen sind, wird allerdings sehr schnell klar, dass es viele, sehr unterschiedliche Arten von Verweisen gibt, die im gedruckten Wörterbuch meist in einen Topf geworfen werden. Verweise sind im gedruckten Wörterbuch in der Regel – von den diese bestätigenden Ausnahmen abgesehen – untypisiert und haben einen weiten bzw. undifferenzierten Skopus.

Der entscheidende Unterschied zwischen Verweis und Link (=Verweis im Hypertext) besteht darin, dass Verweise entweder auf einen anderen Wortartikel als Ganzes (externer Verweis des Typs „**Absatz** [...] → Abschnitt“) oder auf eine Bedeutungsstellennummer im selben Artikel zielen (interner Verweis des Typs „**Absatzflaute** [...] Flaute im Absatz (3) [...] **absatzweise** [...] in Absätzen (2b)“, während ein Hyperlink sowohl auf einen ganzen Wortartikel (extern verweisend) als auch auf einen Wortartikelteil, etwa alle auf eine bestimmte Einzelbedeutung bezogenen Angaben, und auf eine einzelne Angabe in irgendeinem Wortartikel (extern und intern verweisend) zielen kann. Auch der Ort im Datenmodell, wo sich ein Link befindet (Verweisanker), ist klassifizierungsfähig. Außerdem ist ein elektronischer Verweis im Prinzip bidirektional, ein gedruckter nicht oder nur im Ideal- und Ausnahmefall.

Das neue Medium fordert also eine Linktypologie, die außer den Kriterien Anker und Ziel eine Reihe weiterer klassifizierender Kriterien berücksichtigt.

Denkbar ist eine Unterscheidung nach der Funktion eines Links: der eine Link führt von einer Angabe zum zugehörigen Kommentar (nennen wir ihn *Kommentarlink*), ein anderer von einem Fachausdruck zu dessen Erläuterung ins Glossar (nennen wir ihn *glossierender Link*), ein dritter von einem objektsprachlichen Ausdruck zum entsprechenden Lemma

(*Lemma-Link*) oder zur entsprechenden Einzelbedeutung (*Lesarten-Link*), ein vierter, wichtiger führt innerhalb der mikrostrukturellen Hierarchie nach „oben“ zu allgemeineren oder verwandten Angaben oder nach „unten“ in weitere Details (*Strukturlink*), ein fünfter verknüpft die Elemente eines der semantischen bzw. lexikologischen Netze, die anzusetzen sind, z.B. Handlungsnetze, diverse paradigmatische und syntagmatische Netze, Wortbildungsnetze, u.a. (*Vernetzungslink*), ein sechster Linktyp führt in eine bestimmte Stelle der Umtexte, z.B. auf die Erläuterungen zu einem linguistischen Projekt, dessen Mitglieder für eine Angabe verantwortlich zeichnen (*Projektlink*, *Autorenlink*), oder auf die Erläuterungen zu einem (Teil-)Korpus (*Korpuslink*). Zuletzt gibt es noch die rein assoziativen Links, die sich (zu einem gegebenen Zeitpunkt noch) keiner der anderen Funktionen zuordnen lassen (*Hotlink*).

Welche Linktypen wie identifizierbar gemacht werden und wie in eine Mikrostruktur einzubauen sind, hängt natürlich stark von der Recherche-Software und den Möglichkeiten ihrer Benutzeroberfläche ab, die zur Verfügung steht.

Bei einem potenziellen Maximum an Information, wie LEKSIS es vorsieht, und bei einer primär linguistisch motivierten Mikrostruktur kann es leicht passieren, dass sich als Prinzip einer Verlinkungsstruktur das „Alles-hängt-mit-allem-zusammen“ aufdrängt. Nutzer und Autoren wären damit aber sicherlich überfordert; außerdem sind Redundanzen zu erwarten (vgl. Wiegand 1998[99], 247f.). Es gibt vermutlich eine quantitativ bestimmbare Obergrenze der Anzahl von Links, die Nutzer auf einer Web-Seite ‚verkräften‘ oder überhaupt wahrnehmen können. Gegenüber der Menge dürfte aber die Qualität von Hilfen zur Orientierung im System die größere Rolle spielen. Das heißt, dass für die Benutzerfreundlichkeit die Strukturlinks, ihre Benennung (s.o.) und Präsentation am wichtigsten sind. Hingegen vermitteln die anderen Linktypen, insofern sie die Vernetzung des Wortschatzes repräsentieren, von der schon Hermann Paul gesprochen hat, dasjenige lexikologische Wissen, das über die wörterbuchtypisch lexemzentrierte Betrachtung hinaus greift und das die Form von wortartikelinternen Angaben sprengt.

Grundsätzlich muss, wie oben erwähnt, die Struktur, in der die Nutzer mithilfe von Strukturlinks navigieren, mit der Mikrostruktur, die – auch – die Navigation auf Autorenseite leitet, nicht identisch sein. Die Sicht der Nutzer und die Sicht der Autoren können aber dann identisch sein, wenn beide Gruppen dergleichen Fachdisziplin angehören, d.h. wenn die Herangehensweise an den Gegenstandsbereich bei den Adressaten gleich oder ähnlich strukturiert ist wie bei den Autoren. Im Projekt LEKSIS stellt die sprachwissenschaftliche Nutzung einen von zwei Benutzungssituationstypen dar; der andere ist der inter- und extra-disziplinäre Benutzungssituationstyp. Für letzteren ist die Herangehensweise an den Gegenstandsbereich Wortschatz, seine Einheiten und ihren Zusammenhang, aber sehr heterogen und wenig begrifflich gefasst.

Wie die Struktur, über der nicht-linguistisch interessierte Nutzer navigieren, beschaffen sein sollte, muss wenigstens ansatzweise empirisch erprobt werden, etwa durch Fragebogen. Die Erforschung echter Benutzungssituationen kann erst nach Fertigstellung dieses oder eines ähnlichen Systems geschehen. Die von uns in Erwägung gezogene Fragebogen-Methode muss kombiniert werden mit den Ergebnissen allgemeiner Rezeptionsforschung zum Hypertext. In dieser unbefriedigenden Situation mag es verlockend erscheinen, die Situation der Hypertextnutzung an den weit besser bekannten Benutzungssituationen der Printwörterbücher auszurichten und die Nutzer von vorneherein in ihren am Papierwörterbuch geschulten Erwartungen zu bestätigen und lediglich die dort bekannten Mängel – zu starke Textverdichtung und typografische Unübersichtlichkeit – zu beheben.

Doch dieser Weg verschenkt die neuen Möglichkeiten und verschläft sozusagen die unaufhaltsam in Gang befindliche ‚Evolution‘ der Lexikografie. Wenn wir, die Linguistinnen und Lexikografen, uns jetzt nicht um angemessene Hypertext-Konzepte lexikologischen Wissens kümmern, werden Fachfremde dies tun und dabei Standards schaffen, die wir später kaum wieder werden außer Kraft setzen können.

## 5 Zwang zu theoretischer Reflexion

Insgesamt gesehen führt die Erarbeitung der Mikrostruktur für LEKSIS wiederholt zur Auseinandersetzung mit einigen sprachtheoretischen Problemen, von denen hier das wichtigste erläutert sei:

Was ist der zentrale Gegenstand lexematischer Information – die gesamte Inhaltsseite des Lemmazeichens oder die jeweilige Einzelbedeutung?

In gedruckten Wörterbüchern stehen alle formbezogenen und von Einzelbedeutungen unabhängigen, d.h. orthografischen und morphosyntaktischen Angaben im oft „Kopf“ genannten, d.h. ausgelagerten Formkommentar der Mikrostruktur; danach erst kommen die mittels Zahlziffern geordneten einzelbedeutungsspezifischen, meist semantisch-pragmatischen Angaben. Dabei kommt es nicht selten zu Problemen mit dem Skopus des Formkommentars. Unmarkierter Weise umfasst der Skopus alle lesartenspezifischen Angaben; im markierten Ausnahmefall enthalten aber auch die semantisch-pragmatischen Angabenklammern ‚unten‘ Informationen zu morphosyntaktischen und anderen Besonderheiten, wodurch der Skopus des Formkommentars aufgehoben oder eingeschränkt wird. Vor allem grammatische Informationen, z.T. aber auch historisch-diachrone und enzyklopädische werden auf diese Weise in zwei oder mehr Elemente aufgespalten. Außerdem ist diese gängige Praxis, die in gedruckten Wörterbüchern aus Platzgründen ja nirgends explizit erläutert werden kann, Quelle von Schwer- und Missverständlichkeit. Es stellt sich also die Frage, ob die konventionelle Auslagerung des Formkommentars im Hypertext-Wörterbuch nicht aufgehoben bzw. durch eine konsequentere Zuordnung von formbezogener und inhaltsbezogener Information ersetzt werden könnte.

Die durch die Zweidimensionalität des Druckraums erzwungene Strukturierung suggeriert, dass das Lemmazeichen der eigentliche, zentrale Gegenstand ist, über den das Wörterbuch etwas aussagt, obwohl das Lemmazeichen aus zeichentheoretischer, textlinguistischer und auch semantisch-kognitiver Perspektive im Normalfall keine Einheit darstellt. Man hat es in der einen (text- und korpuslinguistischen) Perspektive vielmehr mit potenziell unendlich vielen, durch Kontext monosemierten Wortzeichen zu tun, die abstrahierend-klassifizierend zu Lesarten zusammengefasst werden. Der Unterschied zwischen Glossar und Wörterbuch besteht in eben dieser Abstraktion. Die Lesarten/Einzelbedeutungen<sup>6</sup> besitzen darüber hinaus aber auch im allgemeinen Sprecherbewusstsein eine mehr oder weniger deutliche Identität. In kognitionslinguistischer Perspektive spricht man daher von Klassenkonzepten als den Einheiten des mentalen Lexikons.

<sup>6</sup> *Einzelbedeutung* ist die in der Lexikografie gängigste Bezeichnung für das, was in der Semantik(theorie) meist *Lesart* genannt wird. Die beiden Ausdrücke beziehen sich auf das selbe Denotat.

Es kann hier nur angedeutet, nicht alles aufgeführt werden, was dafür spricht, die einzelne Lesart zum zentralen Gegenstand der Mikrostruktur, d.h. zu der auf das Lemmazeichen unmittelbar folgenden Ebene zu machen und jeweils alles, was es an grammatischen, semantisch-pragmatischen, historischen und enzyklopädischen Informationen gibt, lesartenbezogen zuzuordnen. Dafür spricht u.a., dass paradigmatische und syntagmatische Vernetzungen ausschließlich zwischen Lesarten, nicht zwischen Lemmazeichen bestehen. Im Falle eines z.B. 42-fach polysemen Lemmazeichens, bei dem Schreibung, Aussprache und morphosyntaktische Angaben 42mal bzw. 39mal identisch und 3mal spezifisch sind, hieße dies, 42- bzw. 39mal die gleiche Angabe zu speichern oder von 42 bzw. 39 Positionen aus automatisch die gleiche Angabe zu erzeugen.

Auf der anderen Seite hängen die Lesarten einer Wortform meistens in irgendeiner Weise zusammen, so dass von einer Gesamtbedeutung gesprochen werden kann, auch wenn diese gemessen an der Textbedeutung noch stärker konstruiert ist und eine weitergehende Abstraktionsstufe darstellt. Die lesartenzentrierte Mikrostruktur muss demnach besondere Angaben zum Zusammenhang zwischen den Lesarten eines Lemmazeichens enthalten. Dennoch suggeriert diese Struktur eine weitgehende Autonomie jeder Einzelbedeutung. Bei polysemen Wörtern wie *Absatz* scheint dies angemessen, weniger hingegen bei Konzeptfamilien wie *Schule*.

Alternativ könnte man diesen semasiologischen Zusammenhang zum zentralen Gegenstand der Mikrostruktur machen, dessen Einheit durch weitgehend einheitliche Schreibung, Aussprache und Grammatik „erwiesen“ schiene, wenn man nicht gar eine – dann hochgradig konstruierte – „Kern“- oder „Grund“-Bedeutung formulieren wollte. Bei den semantisch-pragmatischen Angaben wie bei den diversen semantischen Vernetzungen müsste dann nach Lesarten differenziert werden – und bei zahlreichen weiteren Ausnahmen. Diese Struktur suggeriert demnach eine quasi natürliche, weil in den seltensten Fällen (nämlich nur bei Konzeptfamilien) begründbare, einheitliche Inhaltsseite der Lemmazeichenform.

Nicht nur aus linguistischer Perspektive dominieren die Argumente für ein lesartenbezogenes Datenmodell; auch datenbanktechnologisch scheint diese Alternative die weniger komplizierte zu sein.

## 6 Literatur

- Fraas, Claudia/Haß-Zumkehr, Ulrike (1999): Vom Wörterbuch zum Informationssystem. Über ein neues Projekt des Instituts für Deutsche Sprache. In: Deutsche Sprache 4/1998 (1999), 289–303.
- Lemberg, Ingrid/Petzold, Sybille/Speer, Heino (1998): Der Weg des Deutschen Rechtswörterbuchs in das Internet. In: Wiegand, Herbert E. (Hg.) (1998): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium, Tübingen, 262–284.
- Reichmann, Oskar (1986): Lexikographische Einleitung, in: Frühneuhochdeutsches Wörterbuch, hg. von Robert K. Anderson, Ulrich Goebel und Oskar Reichmann, Bd. 1, Berlin, 10–164.
- Storrer, Angelika (2000): Was ist „hyper“ am Hypertext? In: Werner Kallmeyer (Hg.) (2000): Sprache und neue Medien. Berlin, New York.
- und Bettina Harriehausen (Hrsg.) (1998): Hypermedia für Lexikon und Grammatik. – Tübingen 1998 (Studien zur deutschen Sprache 12).
- Wiegand, Herbert Ernst (1989): Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. – Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. Ed. Franz Josef Hausmann, Oskar Reichmann, Herbert E. Wiegand, Ladislav

- Zgusta. 1. Teilbd. – Berlin, New York (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1), 462–501.
- (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd. – Berlin, New York.
  - (1998[99]): Neuartige Mogelpackungen: Gute Printwörterbücher und dazu miserable CD-ROM-Versionen. Diskutiert am Beispiel des LEXIKONS DER INFESTIONSKRANKHEITEN DES MENSCHEN. In: Lexicographica 14/1998 (1999), 239–253.

*Ulrike Haß-Zumkehr, Mannheim*

