

# Design eines Internet-Lexikons zwischen Recherche und Rezeption

Sonja MÜLLER-LANDMANN, Mannheim, Germany

## Zusammenfassung

Am Institut für Deutsche Sprache wird mit LEKSIS ein hypermediales Nachschlagewerk zum Wortschatz der deutschen Sprache entwickelt. Die Stichwörter werden im Lauf der Zeit von Projektgruppen mit unterschiedlicher lexikografischer Zielsetzung bearbeitet. Zielgruppe sind sowohl interessierte Leser, die Datenmaterial entlang der Wörterbuchartikel rezipieren, als auch Experten, die Informationseinheiten quer über den gesamten Wortschatz recherchieren. Diesen unterschiedlichen Ansprüchen trägt das Design des Systems sowohl bei der Datenerfassung, i.e. der lexikografischen Arbeit, als auch bei der Darstellung der Inhalte, i.e. der Benutzungssituation, Rechnung durch ein lexikografisches Maximalprogramm: die äußerst umfassende deskriptiv-elementare Mikrostruktur.

## 1 Das LEKSIS-Projekt: Systemarchitektur

Das lexikalisch-lexikologische Informationssystem des Instituts für Deutsche Sprache, LEKSIS, ist konzipiert als multimediales, hypertextuelles, interaktives und kooperatives Nachschlagewerk zum Wortschatz der deutschen Sprache<sup>1</sup>. Zur Zeit noch in der Projektphase der Konzeption wird es im Endausbau 250 000 – 300 000 Stichwörter umfassend dokumentieren, linguistisch-lexikologisch erläutern und auch für sprachinteressierte Laien ohne linguistische Vorbildung verständlich darstellen.

Die drei zentralen Bausteine des Gesamtsystems sind eine XML-Dokumenttypdefinition (DTD), eine Datenbank und eine Sammlung von Textkorpora:

1. Mit der eXtensible Markup Language (XML; vgl. [Bray et al. 1998]) werden Elemente als Daten eines bestimmten Typs, i.e. eines bestimmten Inhalts, markiert. Die DTD liefert somit das Gerüst für die Formulierung von Inhalten.

Die für LEKSIS entwickelte *XML-DTD* modelliert die umfassende, feinkörnige und linguistisch-lexikologisch fundierte Mikrostruktur der Wörterbuchartikel. Sie liefert die Grundlage für das Datenbankschema und bildet die Schnittstelle zwischen Lexikograf und Datenbank sowie zwischen Datenbank und im Browser angezeigtem Hypertext. Die zugrunde liegende Struktur wird in Abschnitt 3 erläutert.

2. Die objektrelationale *Datenbank*, die das "Wissen über Wörter" verwaltet, wird von verschiedenen lexikologisch-lexikografischen Projekten im Laufe der Zeit entlang der XML-Mikrostruktur gefüllt werden. So werden in einer Pilotanwendung die "Neologismen der 90er Jahre" von einer Arbeitsgruppe des IDS bearbeitet, während zeitgleich der Kernwortschatz des Deutschen beschrieben wird. Weitere Teilprojekte können z.B. Synsematika behandeln oder Wortschatzbereiche, deren Vernetzung innerhalb übergeordneter Wissens-Schemata durch korpusbasierte Kollokationsanalysen konstituiert und bestätigt werden. Das Datenbankschema orientiert sich an der Struktur der XML-DTD.

3. Um die Aktualität des Datenmaterials zu gewährleisten, wird der Wortgebrauch auf der Grundlage der *Textkorpora* des IDS von z.Z. ca. 500 Millionen in elektronischer Form erfassten Textwörtern in klassifizierten Texten belletristischer, (populär-)wissenschaftlicher, journalistischer und sonstiger Art des Zeitraums von Ende des siebzehnten Jahrhunderts bis heute kontinuierlich dokumentiert und fortgeschrieben. Dazu stehen den Lexikografen wie auch den Benutzern von LEKSIS diese Korpora und entsprechende Recherche- und Analysewerkzeuge zur Verfügung. Für die (halb-)automatische Erzeugung von lexikalischen Angaben sind zunehmend auch morphosyntaktisch getaggte Textdaten verfügbar. Zum Beispiel kann man diese einsetzen, um einerseits vollautomatisch einzelnen Lemmata die Wortart zuzuweisen oder andererseits Vorschläge zu generieren zu Valenzrahmen für die manuelle Weiterbearbeitung durch den Lexikografen.

Die benutzergerechte Präsentation der Daten als Hypertext im Internet übernimmt das *Konstanzer Hypertextsystem KHS*<sup>2</sup>. Die jeweils angefragte Information wird individuell aus der Datenbank zusammengestellt und als Hypertext organisiert. Die Ausarbeitung der Präsentationsform(en) der Daten ist getrennt von deren Strukturierung als eigene informationstechnologische und lexikografische Handlung geplant.

Das Medium des Hypertexts ermöglicht – neben Schnittstellen zu anderen Projekten und Informationssystemen – die Vernetzung des lexikalischen Materials durch typisierte Links innerhalb der LEKSIS-Mikrostruktur

- zu anderen lexikalischen Einheiten,
- zu einzelnen lexikologischen Angaben,
- zu nicht-sprachlichen multimedialen Erläuterungen (z.B. Audio-, Video- oder Bilddateien),
- zu terminologischer Information (z.B. Erläuterung linguistischer Begriffe im Glossar),
- zu lexikologischen Abhandlungen über ausgewählte Teile des Wortschatzes,
- zu zur Laufzeit mit den aktuell gegebenen Parametern auszuführenden Prozessen,
- zu Informationen zu den beteiligten Projektgruppen und deren jeweiligem lexikografischen Ausgangspunkt,
- zu einem kommentierten Primär- und Sekundärquellenverzeichnis etc.

## 2 Anwendungsprofile: Recherche, Rezeption

Vor allem dient die Hypertextstruktur der flexiblen, individuellen Informationsauswahl für Experten und Sprachinteressierte. Sie erlaubt die Navigation auf unterschiedlichen Ebenen eines mehrdimensionalen Suchraums. Das ist zum einen die Suche nach einzelnen lexikalischen Einheiten wie Einwort-, Mehrwort- und Wortelementlemmata mit den jeweils assoziierten Angaben, wahlweise in kompakter Kurzdarstellung und/oder in wissenschaftlich elaborierter und diskutierender Form. Zum anderen besteht die Möglichkeit, entlang der linguistischen Mikrostruktur, Untermengen des Wortschatzes zusammenzustellen, die sich durch bestimmte Merkmale/Eigenschaften auszeichnen - z.B. für wortschatzbezogene IDS-interne oder -externe Projekte.

Für kontrastive Studien könnte z.B. die Vernetzung eines semasiologischen Feldes interessant werden und für kulturgeschichtliche Untersuchungen die - möglicherweise sogar im Korpus belegten – Bedeutungsverschiebungen, die sich aus den wortgeschichtlichen Angaben zu einem Lemma ergeben. Ein Nicht-Muttersprachler will möglicherweise die Bedeutungsparaphrasen und Kollokationen aller Synonyme im onomasiologischen Feld eines Lemmas sehen, um die beste Wortwahl für seinen speziellen Kontext zu treffen, oder an den Angaben zu den transformationellen Anomalien des Mehrwortlemmas *die Nase voll haben* feststellen, dass er eben nicht *seine* Nase voll haben kann. Für die maschinelle Sprachanalyse sind unter anderem die Angaben zu Valenz, Flexion und Wortbildung, für Textverarbeitungsprogramme Angaben zu Schreibvarianten und Silbentrennung, für die maschinelle Übersetzung Angaben zu den Kollokationspartnern und deren jeweilige semantische Klassifizierung von Belang. Im Text-Retrieval und bei natürlichsprachlichen Abfragesystemen sind die Angaben zu Synonymen, Antonymen, Hypero- und Hyponymen, etc. essentiell. Oder aber ein Benutzer will – einer traditionellen Wörterbuchbenutzungssituation entsprechend – ganz allgemein etwas über ein Wort wissen, ohne diese Information aus einer Menge detaillierter linguistischer Sachverhalte induzieren zu müssen: Er wird wahrscheinlich die Kompaktinformation zur Bedeutung und zur Grammatik wählen.

Damit eröffnet sich für die Konzeption eines lexikographischen Projekts im Medium Internet, hier des LEKSIS-Informationssystems, ein Spannungsfeld zwischen einerseits der *Rezipierbarkeit* durch den interessierten Leser, der die einzelnen Informationseinheiten aus ansprechenden, daher eher diskursiven, Textpassagen intellektuell erschließt, und andererseits der *Recherchierbarkeit* kleinster Informationseinheiten über den gesamten erfassten Wortschatz durch den an möglichst fein strukturierten und vollständigen Angaben interessierten Experten.

Dieses Zusammentreffen unterschiedlichster lexikografischer Zielsetzungen innerhalb eines Systems mit einem – während der Konzeption und Entwicklung -

- nicht festlegbaren Autorenkreis und
- mit einem Adressatenkreis mit stark variierenden Ansprüchen,
- ohne zeitliche Befristung und
- ohne quantitative Einschränkung der lexikologischen Inhalte
- bei gleichzeitiger ständiger Aktualität und dennoch
- frühzeitiger Verfügbarkeit der schon erarbeiteten Daten

erfordert ein lexikografisches Maximalprogramm:

LEKSIS vereint beliebig viele Wörterbuchtypen – je nach dem Fokus der jeweils involvierten Projektgruppen – und muss dabei in konsistenter Weise ständig ergänzt und korrigiert bzw. aktualisiert werden können.

Ein solches System muss die zu beschreibenden Objekte über eine äußerst umfassende und elementare Mikrostruktur erfassen: Nicht nur die Informationsauswahl und -tiefe bei der Suche für die Adressaten von LEKSIS, sondern auch die Informationsauswahl und -tiefe bei der Bearbeitung für die Autoren muss flexibel, individuell *und* kontrollierbar sein.

## 3 Wortschatzwissen

### 3.1 Informationsdimensionen

Die in LEKSIS abzubildende Lemmastrecke ist in Einträge zu Einwort-, Mehrwort- und Wortelementlemmata unterteilt mit einer jeweils partiell eigenständigen Informationsstruktur, die sich von den anderen Lemmatypen in gesondert ausgezeichneten syntaktischen, lexikalisch-semanticen und pragmatischen Angaben und teilweise in der Obligatorik unterscheidet.

Lemmata, die aus mehreren Einzelwörtern bestehen, werden dann aufgenommen, wenn es sich um usuelle Kookkurrenzen – meist syntaktischer, lexikalisch-semanticer oder pragmatischer Art – handelt, die einen gewissen Einheitsstatus besitzen, in diesem Sinne als komplexere Einheiten reproduzierbar und also als sprachliche Erscheinungen lexikalisiert sind. Darunter fallen vor allem Phraseologismen, phrasen- oder satzwertige Formeln, Funktionsverbgefüge und syn-semantiche Fügungen. Gegenüber Einträgen zu Einwortlemmata sind hier vor allem Informationen zur syntaktischen und semanticen Binnenstruktur zusätzlich erforderlich.

Unter dem Sammelbegriff der Wortelementlemmata werden Erläuterungen zu unselbstständigen Ein- oder Mehrwortlemmabestandteilen wie Affixen, Konfixen oder Unikalen gegeben. Dementsprechend sind die morphosyntaktischen Angaben grundsätzlich verschieden von den auf kategorialer Klassifizierung basierenden grammatischen Angaben zu Einzelwort- und Mehrwortlemmata.

Die gewählte Perspektive ist jedoch die der Lesart als – wenn möglich nicht hypothetisches, sondern – korpusanalytisch erschlossenes Konzept. Die Informationsstruktur bezieht sich also nur mittelbar auf das Formativ oder das Wortschatzganze. Vielmehr werden deskriptive Angaben zu denjenigen Einheiten unterhalb der Lemmastrecke gemacht, die sich (nach den Belegen) durch ihre Position in unterschiedlichen Paradigmen bzw. ihre Position in unterschiedlichen Syntagmen/Kollokationsbezügen abgrenzen: Das heißt unter anderem, dass sie einen unterschiedlichen Referenzbereich und/oder unterschiedliche Schreibung bei gleicher Lautung und umgekehrt aufweisen oder - in bestimmten Fällen – unterschiedlichen Wortarten zugehören.

Die lexikalisch-lexikologischen Informationen zu jeder Lesart bündelt LEKSIS in vier strukturierte Dimensionen von Einzelangaben:

- *Bedeutung und Verwendung*
  - Semantik (Bedeutungsparaphrase, semantiche Klasse, u.Ä.)
  - Syntagmatik (Kollokationen und Syntagmen)
  - Paradigmatik (Synonyme, Antonyme, Hyperonyme, Hyponyme, Kohyponyme, Komplenyme und Meronyme)
  - Pragmatik (Belegungssituation, Angaben zur Diatopik, Textspezifik, Fachsprache, Registerspezifk u.Ä.)
- *Grammatik*
  - Flexion
  - Valenz
  - Wortbildung

- syntaktisches Verhalten
- *Geschichtliches und Sachliches*
  - Formgeschichte
  - Bedeutungsgeschichte
  - Enzyklopädisches
- *Kritisches und Normatives*

(kommentierte) sprachkritische Normierungen<sup>3</sup> und Kategorisierungen in den Bereichen

  - Semantik-Pragmatik
  - Orthographie
  - Stil etc.

Das Lemma bzw. Formativ selbst trägt die Informationen der fünften Informationsdimension *Schreibung und Aussprache* mit Angaben zu

- Schreibung und Schreibvarianten
- Silbentrennung
- Aussprache
- Abkürzungsmöglichkeiten bzw. Auflösung der Abkürzung, die das Lemma selbst bildet
- Reimwörtern

sowie die Polysemieangabe, die die Beziehungen der dem Lemma angehörenden Lesarten innerhalb der Konzeptfamilie mit den jeweils vorliegenden Konzeptverschiebungen enthält.

### 3.2 Datenstruktur

Während die eigentliche inhaltliche Ausarbeitung eines Wörterbuchartikels natürlich in der Verantwortung des Lexikografen bleibt, benötigt LEKSIS für die lexikografische Arbeit in dem beschriebenen vielschichtigen Umfeld also einen klar strukturierten, lexikologisch begründeten Arbeitsplatz: Die (partiell) gleiche Lemmastrecke kann überlappend von mehreren Forschergruppen bearbeitet werden. Je nachdem wird z.B. ein Projekt zur “Synonymik“ vertiefende Angaben zu den paradigmatischen Relationen eines Lemmas machen, während eine Gruppe zu “Kollokationen“ – für dasselbe Lemma - insbesondere die syntagmatischen Angaben intensiv bearbeitet. Kritisch sind dabei vor allem diejenigen Angaben, die entweder den Arbeitsbereich mehrerer Gruppen betreffen oder aber die inhaltlich mit anderen Angaben benachbart sind und somit kompatibel bleiben müssen. (Dieses Problem stellt sich übrigens schon bei der Auswahl der Lemmastrecke: Es wäre z.B. nicht wünschenswert, wenn eine Gruppe eine Partizipialform als Adjektiv lexikalisiert, während eine andere Gruppe davon unabhängig – möglicherweise abweichende – Angaben zur adjektivischen Verwendung des Partizips beim entsprechenden Verb macht.)

Daher muss es Aufgabe einer Autorenoberfläche sein, von den Lexikografen aus diversen Projekten konsistente, kompatible und möglichst vollständige Angaben zu fordern und diesen Zustand der Daten zu bewahren. Zu diesem Zweck muss die Möglichkeit einer (Selbst-)Kontrolle

der Lexikografen sowie der Administration und Versionenkontrolle während der Erstellung und Aktualisierung der Inhalte auf weitestgehend automatisierte Weise gegeben sein.

Diesem Ziel kommen wir mit XML als angemessenem Beschreibungsformat nahe: Die Erstellung einer DTD fördert – erzwingt teilweise sogar – logische Konsequenz und damit Entscheidungseindeutigkeit und Redundanzfreiheit. Mit inhaltsorientiertem Markup wird dem Autor eine Oberfläche zur Verfügung gestellt, die je nach dem definierten Typ einer Angabe und (teilweise) projektspezifischer Obligatorik

- textuelle Elemente (z.B. kommentierende Diskussion des linguistisch-lexikologischen Hintergrunds einer konkreten lexikografischen Entscheidung, diskursive Passagen zur Bedeutungsgeschichte etc.),
- atomare Elemente (z.B. Flexionsformen, Valenzangaben) und/oder Attribute (z.B. Genußangaben, Frequenzangaben)

fordert. Das bedeutet, dass jeweils entweder freie Formulierung oder Angabe eines Wertes freien Inhalts, aber in vorgegebener Form, oder Auswahl aus vorgegebenen Werten erlaubt ist. So wird verhindert, dass (verschiedene) Lexikografen Angaben zur einer konkreten lexikalisch-lexikologischen Information an einer oder an mehreren Stellen unterschiedlich formulieren.

Wo immer möglich werden Wertemengen vorgegeben, die – jeweils erweiterbar oder auch nicht – als selbstständige Angaben oder als Attribute zu Fließtextpassagen die notwendige Eindeutigkeit für Recherchezwecke bieten.

So ist die Zeitangabe (*ZeitA*) innerhalb des diskursiven Beispieltexts zur Entlehnungsgeschichte von *alternativ* als “etwa Ende des 17. Jahrhunderts“ formuliert, die Angaben zur Identifizierung der Sprache (*SpIA*) als “aus dem Französischen“ und “lateinische“. So wird der Text auch zu lesen sein – was dem Aspekt der Rezipierbarkeit Rechnung trägt:

*Deutsch alternativ wurde etwa Ende des 17. Jahrhunderts aus dem Französischen (alternativ: Adjektiv, maskuline Form; alternative: Adjektiv, feminine Form) ins Deutsche übernommen.*

*Die französischen Formen gehen auf das lateinische alternare in der Bedeutung 'abwechseln' zurück. Die Verkehrssprache der europäischen Diplomaten jener Zeit war das Französische; dadurch wurde die Entlehnung des französischen Adjektivs ins Deutsche begünstigt.*

Diese angenehm lesbaren, freien Formulierungen jedoch sind in solch hohem Maße variierbar, dass eine Induktion von lexikologischen Zusammenhängen über diese Angaben nicht möglich ist: Wie könnte ein automatisches Verfahren z.B. erkennen/errechnen, dass eine Zeitangabe “um 1690“ den selben Zeitraum umfasst wie “etwa Ende des 17. Jahrhunderts“?

Erst die zusätzlichen (für den Leser unsichtbaren) Subelemente und Attribute `<Zeitraum zeitpunkt1 = '16800000' zeitpunkt2 = '17000000' ca = 'ja'>` oder

<SpIA sprache = 'franzoesisch'> und <SpIA sprache = 'lateinisch'> des XML-modellierten Wörterbucheintrags (siehe den entsprechenden Ausschnitt unten) bieten die Möglichkeit, beispielsweise alle zwischen 1650 und 1750 aus dem Französischen entlehnten Lemmata automatisch zu extrahieren – was die Recherchierbarkeit optimiert.

```

<FormLehnA>Deutsch <TXTobj>alternativ</TXTobj> wurde
<ZeitA><Zeitraum zeitpunkt1='16800000' zeitpunkt2='17000000' ca='ja'>etwa Ende des 17. Jahrhunderts</Zeitraum></ZeitA>
<SpIA sprache='franzoesisch'>aus dem Französischen</SpIA>
(<TXTobj>alternatif</TXTobj>: Adjektiv, maskuline Form; <TXTobj>alternative</TXTobj>: Adjektiv, feminine Form) ins Deutsche übernommen. </FormLehnA>
<FormLehnA>Die französischen Formen gehen auf das <SpIA sprache='lateinisch'>lateinische</SpIA> <TXTobj>alternare</TXTobj> in der Bedeutung 'abwechseln' zurück.</FormLehnA>

```

Solchermaßen erzeugte Wörterbucheinträge sind zumindest formal, wenn auch nicht vollständig inhaltlich, validierbar, d.h. sie bieten einen gewissen (Selbst-) Kontrollmechanismus und bestehen auf der definierten Informationsstruktur. Vor allem für die Übertragung der Daten in ein dem Objektmodell entsprechendes Datenbankschema ist das eine essentielle Voraussetzung: Die Vorgabe der formal stringent konzipierten DTD-Struktur kann direkt in die Datenbankkonzeption eingehen und erlaubt daher den automatisierten Datenaustausch zwischen den XML-Dateien und den entsprechenden objektrelationalen Datenbanktabellen (vgl. [Bourret 1999] und [Wait 1999]). Eine regelmässige formale und inhaltliche Überprüfung der Datenbanktabellen sowie inhaltliche und technische Dokumentation der Mikrostruktur und lexikografische Anleitung dient der Kontrolle und Erhaltung der Konsistenz.

Ein weiteres Desiderat von LEKSIS ist, dass die enthaltenen Informationen zum Wortschatz möglichst therietolerant zugeordnet werden können und damit über die Zeit hinweg gültig bleiben und für Autoren und Benutzer vieler lexikologisch-lexikografischer Strömungen verwendbar sind und bleiben. Das erfordert eine elementare und deskriptive Strukturierung der Inhalte.

Diese Konzeption steht im Gegensatz erstens zu einer Strukturierung nach Kategorien, die einer speziellen linguistischen Theorie folgen, und zweitens auch zu einer – potenziell zeitkritischen – normierenden Kategorisierung. Zum dritten wird so auch eine (fast) nicht strukturierte, hauptsächlich diskursive Beschreibung der Inhalte vermieden, die zwar intellektuell erschließbar, aber nicht automatisch recherchierbar ist.

Für die lexikologischen Angaben, wie sie in der Mikrostruktur von LEKSIS definiert sind, bedeutet das, dass theoretisch-kategorisierende Klassifizierungen so weit wie möglich auf die

zugrunde liegenden minimalen Informationseinheiten und deren Wertebereiche zurückgeführt und im lexikografischen Arbeitsplatz als einzelne Angaben abgefragt werden. Selbst wo solche Klassifizierungen vorgenommen werden (z.B. die Zuordnung einer semantischen Klasse PFLANZE zu *Kartoffel*), werden die diese Klassifizierung motivierenden Informationseinheiten zusätzlich aufgeführt, in diesem Fall die zentralen, peripheren und sonstigen semantischen Merkmale (im Beispiel: OBJEKT, NATÜRLICH, LEBEND, ESSBAR etc.).

Im Ergebnis ist jeder Lesart jeden Lemmas eine scharf strukturierte Menge an möglichst elementar-deskriptiven lexikologischen Angaben zugeordnet (wann immer möglich auch empirisch belegt<sup>4</sup>), die dennoch größtenteils in freier Formulierung dargeboten werden.

## 4 Mehrwert und Ausblick

Bei der Konzipierung des hypermedialen Informationssystems LEKSIS für die Anwendung in vielfältigsten Benutzungssituationen spielt die Repräsentation des Wissens vor dem Hintergrund übergeordneter komplexer Zusammenhänge, die Entschlüsselung der Prinzipien der Klassifikation der lexikalischen Elemente und die Strukturierung und Vernetzung des Informationszusammenhangs durch Markierung einzelner Informationseinheiten eine zentrale inhaltliche Rolle.

Damit erreichen wir zum einen die Explizitheit eben dieser klassenimmanenten Informationen, zum zweiten eine inhaltliche Kompatibilität von Projekten, die sich auf sich überschneidende, verschiedene oder auch gleiche theoretische Kategorisierungen stützen, und zum dritten eine konsistente Zuordnung der lexikologischen Angaben zu den einzelnen Lemmata und -lesarten.

Die lexikologische Datensammlung und/oder maschinelle Sprachverarbeitung kann so auf eine Informationsstruktur zugreifen, die das Wissen vollständig und explizit abfragt und darstellt, also möglichst feinkörnig recherchierbar ist. Dadurch öffnet sich nicht zuletzt auch ein neuer empirischer Zugang für die linguistische Theorie: Bei einer Indizierung aller Stichwörter nach unterschiedlichen Kriterien werden implizite und explizite Vernetzungen innerhalb des Wortschatzes sichtbar: Klassifizierungen ergeben sich, empirisch nachweisbar, aus der individuellen Zusammenstellung bzw. Entdeckung von charakteristischen Mustern an zugeordneten Angaben (Merkmalcluster), die einer bestimmten Untermenge der Lemmastrecke gemeinsam sind. Neben der Induktion von neuen Klassen bietet sich natürlich auch die empirische Fundierung oder auch Evaluierung bestehender lexikologischer Klassifizierungen an. Ergebnisse solcher externer wissenschaftlicher Wortschatzforschung werden - gebündelt und in Beziehung zueinander gesetzt - ein integraler, mit der eigentlichen Lemmastrecke durch Verlinkung verwobener Bestandteil des LEKSIS-Informationssystems sein.

Mit den Anwendern eines solchen Informationssystems und dem potenziell sehr weiten Spektrum ihres Vorgehens bei der Benutzung, ihrer Anforderungen an dieses System, gibt es noch keine Erfahrungen. Das neue Medium fordert so eine ganz eigene Wörterbuchbenutzungsforschung. Gleichzeitig bietet es aber auch die wissenschaftlich hochinteressante Möglichkeit, diese Forschung auf neue Weise optimal zu betreiben durch das Verfolgen, Untersuchen und Klassifizieren des konkreten Navigationsverhaltens *aller* Anwender.

Durch dieses Wechselspiel von Benutzung und Anpassung des Systems kann für den Leser durch die diskursiven, intern mit semantisch getaggtten Angaben versehenen Passagen und



durch bei Bedarf in Lückentexte eingebettete atomare Angaben eine hypermediale, bequem rezipierbare Darstellung generiert werden, die das ausgewählte Datenmaterial konsistent, den jeweiligen Anforderungen angemessen und ansprechend aufbereitet, ihm aber gleichzeitig die Möglichkeit bietet, jeweils zu komplexeren Inhalten und Zusammenhängen vorzustoßen.

## Anmerkungen

<sup>1</sup>Weitere Information zum Projekt ist in [Fraas/Haß-Zumkehr 1998], [Haß-Zumkehr 2000] und unter <http://www.ids-mannheim.de/leksis> zu finden.

<sup>2</sup>Weitere Information dazu in [Hammwöhner 1997] und unter <http://www.inf-wiss.uni-konstanz.de/FG/IV/KHS/index.htm>.

<sup>3</sup>LEKSIS selbst will nicht normieren, aber über existierende Normen informieren bzw. eine Normendiskussion erlauben.

<sup>4</sup>Kompetenzbasierte und korpusbasierte Angaben werden in LEKSIS für Autor und Leser sichtbar unterschieden. So wird z.B. für jedes Belegbeispiel (im Gegensatz zu einem Kompetenzbeispiel) oder jede korpusbasierte Synonymangabe die Quelle vermerkt, so dass bei Bedarf der jeweilige Kontext (v.a. im Korpus) recherchiert werden kann.

## Literatur

- [BOURRET 1999] R. Bourret: „XML and Databases“, Technische Universität Darmstadt, September 1999 <http://www.informatik.tu-darmstadt.de/DVS1/staff/bourret/xml/XMLAndDatabases.htm>
- [BRAY ET AL. 1998] T. Bray, J. Paoli und C-M- Sperberg-McQueen (Hrsg.): „Extensible Markup Language (XML) 1.0“, W3C Recommendation 10. Februar 1998 <http://www.w3.org/TR/REC-xml>.
- [FRAAS/HASS-ZUMKEHR 1998] C. Fraas und U. Haß-Zumkehr: „Vom Wörterbuch zum lexikalischen Informationssystem. LEKSIS – ein neues Projekt des Instituts für deutsche Sprache“, in *Deutsche Sprache* 26, 1998:289-303.
- [HAMMWÖHNER 1997] Hammwöhner, R.: *Offene Hypertextsysteme. Das Konstanzer Hypertextsystem (KHS) im wissenschaftlichen und technischen Kontext*, (Konstanz: Universitätsverlag) 1997.
- [HASS-ZUMKEHR 2000] U. Haß-Zumkehr: „Zur Mikrostruktur im Hypertext-Wörterbuch“, in: I. Lemberg, B. Schröder und A. Storrer (Hg.): *Chancen und Perspektiven computerunterstützter Lexikographie*, im Druck
- [WAIT 1999] B. Wait: „Using XML in Oracle Database Applications“, 1999 [http://technet.oracle.com/tech/xml/info/htdocs/otnwp/about\\_xml.htm](http://technet.oracle.com/tech/xml/info/htdocs/otnwp/about_xml.htm)

