Erschienen in: Haß, Ulrike (Hrsg.): Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz. - Berlin/New York: de Gruyter, 2005. S. 20-54
(Schriften des Instituts für Deutsche Sprache 12)

Carolin Müller-Spitzer

Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess

1. Einleitung

Das Ziel von *elexiko* ist die Neuerstellung eines elektronischen Wörterbuchs bzw. Wortschatzinformationssystems.[1] Dabei ist nicht nur der Aufbau der lexikografischen Inhalte zu klären, d. h. welches Konzept beispielsweise hinter der semantischen oder grammatischen Beschreibung steht. Auch die Aufbereitung dieser lexikografischen Inhalte sind bei einem so langfristig angelegten Projekt mit dem Ziel einer elektronischen Publikation ein wichtiges Thema. Dabei geht es zum einen darum, den lexikografischen Prozess so aufzubauen, dass die erarbeiteten Inhalte möglichst flexibel adaptiv zu Benutzungssituationen präsentiert werden können, und zum anderen darum, diese Inhalte für die Wörterbuchbenutzer in möglichst innovativer Weise recherchierbar zu machen. Daneben ist es wichtig, die Lexikografen in der Erarbeitung der Artikel bestmöglich zu unterstützen, damit die lexikografischen Daten auch später so konsistent vorliegen, dass eine elektronische Verarbeitung möglich ist.

---

[1] Zum Begriff des *Wortschatzinformationssystems* siehe Müller-Spitzer (2004).

Diese texttechnologischen Aspekte des Aufbaus eines lexikografischen Prozesses sind dabei keine technischen Fragen am Rande, sondern gehören bei einem Projekt mit den Zielen von *elexiko* (siehe U. Haß, *elexiko* – Das Projekt, in diesem Band) zum Kern der Projektarbeit. In diesem Artikel soll daher beschrieben werden, wie das Konzept für die Modellierung der lexikografischen Daten in *elexiko* aussieht und wie diese Modellierung im lexikografischen Prozess angewandt wird.

Da wir in *elexiko* weder auf vorhandene Daten zurückgreifen konnten oder mussten, wurde auch die Modellierung völlig neu entwickelt. Dabei war es wichtig, die Anforderungen an die Aufbereitung lexikografischer Daten, so wie sie in der neueren Wörterbuchforschung diskutiert werden, zur Kenntnis zu nehmen und für den konkreten Aufbau der Modellierung möglichst umfassend zu berücksichtigen. Um daher den theoretischen Hintergrund bestimmter Modellierungsprinzipien besser verstehen zu können, sollen diese Anforderungen hier zunächst erläutert werden.

2. Anforderungen an die Modellierung lexikografischer Daten

2.1 Funktionalitäten von Wortschatzinformationssystemen

In der Literatur werden Anforderungen an die Modellierung lexikografischer Daten meist nicht direkt unter diesem Thema diskutiert sondern eher im Zusammenhang mit der erwünschten Funktionalität von elektronischen Wörterbüchern bzw. Wortschatzinformationssystemen. Einige dieser Funktionalitätsanforderungen implizieren jedoch eine bestimmte Form der Datenmodellierung und -aufbereitung. Der wichtigste Punkt für die Datenmodellierung ist dabei der Zugriff auf die lexikografischen Daten. Denn bei der Modellierung der Daten und einer entsprechenden Datenaufbereitung wird der Grundstein dafür gelegt, wie gezielt und genau auf diese Daten später zugegriffen werden kann. „What you mark is what you get" (Gloning/Welter 2001, 128), so formulieren es Gloning/Welter bezogen auf eine XML-basierte Modellierung.

Der Zugriff auf die lexikografischen Daten wird auch immer wieder als einer der wichtigsten Punkte aufgeführt, in denen sich elektronische Wörterbücher von gedruckten unterscheiden (sollten). „From a user's perspective, the most innovative aspect of EDs (electronic dictionaries; Anm. d. Verf.) is probably the retrieval system" (Schryver 2003, 146). Dabei geht es nicht allein um die Formulierung von bestimmten Suchanfragen, sondern auch um die flexible Präsentation des Suchergebnisses. Es gilt dabei als nicht besonders benutzerfreundlich, im elektronischen Wörterbuch zwar flexible äußere Zugriffsstrukturen auf die Wörterbuchartikel zu bieten, aber als Ergebnis nach wie vor den gesamten Wörterbuchartikel zu zeigen. Denn damit reicht die Nachschlagehandlung – wie beim gedruckten Wörterbuch – bis zum Artikelan-

fang. Der Wörterbuchartikel selbst muss gelesen werden; der Lesevorgang wird bestenfalls durch Suchzonen erleichtert.[2] Der Bildschirm ist zum Lesen jedoch schlecht geeignet. Daher wird gefordert, abhängig von Benutzungssituationen aus einem Wörterbuchartikel nur die gesuchten Angaben anzeigen zu können. „It is one thing to be able to store ever more data, but another thing entirely to present just the data users want to response to a particular look-up." (Schryver 2003, 178) Ähnlich auch Klosa (2001, 97):

> „Ein gutes CD-ROM-Wörterbuch lässt die Benutzer(innen) aus einer Fülle von Informationen diejenigen auswählen, die momentan für sie am wichtigsten sind, z. B. einmal die Orthographie eines Wortes, einmal seine Herkunft, seine Aussprache [...]."

„Die Benutzerschnittstelle von digitalen Wörterbüchern" sollte daher „an Typen von Benutzungssituationen adaptierbar sein." (Storrer 2001, 64; ähnlich auch Büchel/Schröder 2001, 8, Thielen et al. 1998, 191 und Steindler 1995, 238)

Die Voraussetzungen für diese flexiblen Zugriffs- und Präsentationsmöglichkeiten müssen bei der Modellierung geschaffen werden.

> „Lexikalische Daten können so modelliert werden, dass in Abhängigkeit von Nutzerinteressen und Nutzungssituationen die jeweils relevanten lexikografischen Angaben und Verweise herausgegriffen und in ästhetisch ansprechender Weise am Bildschirm dargestellt werden." (Storrer 2001, 53 f.)

In ähnlicher Weise auch Gloning/Welter (2001, 118):

> „Elektronische Datenbasen erlauben es nunmehr, auf einen Datenbestand, z. B. eine Wortschatzdokumentation, nach unterschiedlichen Kriterien zuzugreifen und dabei auch ganz unterschiedliche ‚Ansichten' des Datenbestandes je nach Interesse und Fragestellungen von Benutzern hervorzubringen. Voraussetzung ist, dass die entsprechenden Informationen [...] in expliziter Weise im Datenbestand enthalten sind."

Mit einer solchen Modellierung kann damit der

> „Widerspruch zwischen der Multifunktionalität des Produkts und der monofunktionalen Nutzung dadurch aufgelöst werden, dass, entsprechende Strukturierung vorausgesetzt, der Wörterbuchbenutzer eine seiner Verwendung gerechte Sichtweise auf des Wörterbuchmaterial auswählen oder sogar selbst definieren kann." (Feldweg 1997, 112)

Wie man aus diesen Stellungnahmen unabhängig von z. T. unterschiedlichen Terminologien herauslesen kann, wird eine deutliche Trennungslinie zwischen der lexikografischen Datenbasis und der Präsentationsebene eines Wortschatzinformationssystems gezogen. Die Datenbasis enthält Daten, die so strukturiert sein sollen, dass auf der Präsentationsebene eine spezifische Auswahl präsentiert werden soll. Genau dies wird als eine der wesentlichen Vor-

---

2 Zu Suchbereichsstrukturen siehe Bergenholtz et al. (1999).

teile von Wortschatzinformationssystemen gesehen. Dazu z. B. Storrer (2001, 60): Es wurde

> „bereits gezeigt, dass die zunächst nahe liegende Analogie von Wörterbuch und Hypertext einen zentralen Mehrwert des digitalen Mediums außer acht lässt: Die Art und Weise, wie Daten in einem Datenbank- oder Hypertextsystem strukturiert sind, muss nicht der Art und Weise entsprechen, wie diese Daten dem Benutzer am digitalen Lesegerät [...] präsentiert werden. Die Zielsetzung der Informationsmodellierung besteht vielmehr gerade darin, Daten so zu strukturieren, dass aus ein und demselben Datenpool für verschiedene Anwendungszwecke und Nutzungskontexte die jeweils relevanten Informationen herausgegriffen und in geeigneter Weise präsentiert werden können."

Voraussetzung für eine solche Modellierung ist dabei auch, dass die Funktionalitäten des Computers von Anfang an mitgedacht werden. „Electronic Dictionaries would be most effective if they were designed from scratch with computer capabilities and computer search mechanisms in mind." (Nesi 2000a, 140; zitiert nach Schryver 2003, 163) Das heißt in aller Regel auch, dass nicht allein ein bereits gedrucktes Wörterbuch ohne angestrebte Ergänzung oder Überarbeitung Ausgangspunkt der Überlegungen sein sollte. Hierin liegt einer der Gründe dafür, dass eine solche Modellierung und eine darauf aufbauende Entwicklung von Wortschatzinformationssystemen noch wenig in die Praxis umgesetzt ist.[3] Denn gedruckte Wörterbücher sind nach wie vor v. a. für die kommerzielle Verlagslexikografie sehr viel wichtiger als elektronische Wörterbücher. (Klosa 2001, 98). Daher stellt auch Storrer zu der Umsetzung der o.g. Zugriffsmöglichkeiten fest:

> „Dass es sich bislang nur um Prototypen handelt, liegt weniger daran, dass nicht bekannt wäre, für welche usuellen Benutzungssituationen typischerweise welche Klassen von Angaben relevant werden. Die Ursache liegt vielmehr darin, dass eine kontextadaptive Präsentation lexikalischer Informationen eine linguistisch motivierte und feinkörnige Modellierung der lexikografischen Daten voraussetzt. Eine der-artige Modellierung erfordert, wenn sie auf der Grundlage eines gedruckten Wörterbuchs erfolgt, einen relativ hohen Auf- und Nachbearbeitungsaufwand und lässt sich deshalb am schnellsten realisieren, wenn ein digitales Wörterbuch unabhängig von einer vorhandenen Printvorlage konzipiert werden kann." (Storrer 2001, 64)

Aus diesen erwünschten Funktionalitäten von elektronischen Wörterbüchern, die in der Forschung und Praxis diskutiert werden, lassen sich bestimmte Anforderungen an die Modellierung in *elexiko* ableiten. Daneben leiten sich weitere Anforderungen aus der erwünschten Unterstützung im lexikografischen Prozess ab.

---

[3] Eine Ausnahme im Bereich der Datenaufbereitung stellt dabei sicherlich das Projekt der „Duden ontology" dar (vgl. Alexa et al. 2002).

2.2 Unterstützung im lexikografischen Prozess

> „Die Computertechnik verändert auch den lexikografischen Arbeitsprozess. Der Computereinsatz in der Wörterbuchwerkstatt erlaubt es bei entsprechender technischer Infrastruktur, Abläufe effizienter und flexibler zu gestalten und damit gerade umfangreiche Wörterbuchprojekte schneller, qualitätvoller und kostengünstiger abzuschließen." (Lemberg et al. 2001, 2)

Dabei sind auch hier flexible und gezielte Zugriffsmöglichkeiten auf die lexikografischen Daten für die Lexikografen ein wichtiger Punkt. Denn je besser man in dem bereits erarbeiteten Material suchen kann, desto besser lassen sich z. B. inhaltliche Inkonsistenzen vermeiden. Beispielsweise kann es für einen Lexikografen hilfreich sein, einen Überblick darüber zu bekommen, wo er überall das Hyperonym „Gebäude" angesetzt hat. Auch können gezielte Querschnittsabfragen nach z. B. allen Artikeln, in denen zur Lemmazeichengestaltangabe eine orthografische Variante in alter Rechtschreibung angegeben ist, einen besseren Überblick und bessere Korrekturmöglichkeiten verschaffen. Dabei arbeiten die Lexikografen – anders als die Wörterbuchbenutzer – direkt auf der lexikografischen Datenbasis.

Ein weiterer wichtiger Punkt im lexikografischen Prozess, aus dem Anforderungen an die Modellierung abgeleitet werden können, ist der vielfach formulierte Wunsch nach Unterstützung der Lexikografen in der formalen Einhaltung der Artikelstruktur. „Die Verantwortung für die Einhaltung des Artikelformats kann an ein Computerprogramm abgetreten werden. Bei anspruchsvollen Wörterbüchern bedeutet dies eine Entlastung für den Lexikografen." (Wiegand 1998, 232) Traditionell werden Schreibanweisungen für die Artikelarbeit in Instruktionsbüchern oder Manuals festgehalten. (Engelberg/Lemnitzer 2001, 211) Das Problem in der lexikografischen Praxis ist jedoch oft, dass besonders eine komplexe Artikelstruktur ohne umfangreiche technische Unterstützung kaum einzuhalten ist.

> „Konventionen, die die standardisierten Wörterbuchartikel erfüllen müssen, werden normalerweise in Instruktionsbüchern festgehalten. Das Einhalten dieser Konventionen wird manuell und oft nur stichprobenartig überprüft. In Anbetracht der hochkomplexen Textform ‚Wörterbuchartikel' ist eine redaktionelle Betreuung eines Wörterbuchs mit mehr als 60000 Artikeln allerdings sehr schwierig und letztendlich sind inhaltliche und formale Mängel kaum auszuschließen." (Heyn 1992, 187)

Neben der Anforderung, die Lexikografen in der Einhaltung der Artikelstruktur zu unterstützen, ist in Wiegand (1998, 217 f.) ein zweites Desiderat formuliert: die lexikografischen Daten sollen ohne Strukturanzeiger[4]

[4] Mit Strukturanzeigern werden in der Printlexikografie allgemein formuliert sowohl die verschiedenen standardisierten typografischen Auszeichnungen einzelner Angaben in einem Wörterbuchartikel als auch die Interpunktionszeichen, die die einzelnen Angaben voneinander abtrennen, bezeichnet. Vgl. Wiegand (1989a, 428).

eingegeben werden; diese sollen für die Präsentation automatisch generiert werden können. Diese Unterstützungen können eine wesentliche Erleichterung für die lexikografische Arbeit darstellen. In diesem Sinne formuliert Heyn (1992, 192):

> „Ein integrativer Bestandteil zukünftiger Instruktionsbücher muss eine formale Grammatik für Wörterbuchartikel im Sinne einer Dokumenttypdefinition sein, mit der für jeden Typ von Artikel festgehalten wird, wie die erlaubten Bausteine seiner Architektur aussehen können. Das legt dem Lexikografen nicht nur zusätzliche Fesseln an, sondern erleichtert im Gegenteil die Arbeit zur Konsistenzhaltung und erlaubt eine Konzentration auf die eigentliche deskriptive Arbeit."

Diese bis jetzt genannten Punkte sind die wesentlichen Anforderungen, die für die Entwicklung der Modellierung in *elexiko* beachtet wurden. In der Projektpraxis von *elexiko* sind diese theoretisch wünschenswerten Eigenschaften einer Modellierung allerdings manchen Einschränkungen unterlegen, wie später zu zeigen sein wird. Zunächst sollen jedoch die wesentlichen Eckpfeiler des Modellierungskonzeptes von *elexiko* erläutert werden.

3. Das Modellierungskonzept von *elexiko*

a) XML-basierte Modellierung

Die Modellierung erfolgt in *elexiko* XML-basiert. XML (eXtensible Markup Language) ist eine international festgelegte Syntax, eine Metasprache, die die Entwicklung von Auszeichnungssprachen zur Beschreibung von Daten hinsichtlich ihrer hierarchischen Struktur und ihrer inhaltlichen Einheiten ermöglicht. (XML Standard [Third Edition]) Neben grundsätzlichen Vorteilen, die die Anwendung von XML bietet wie Nachhaltigkeit, Systemunabhängigkeit und gesicherte Softwareunterstützung (vgl. Büchel/Schröder 2001), ist im Zusammenhang mit der Modellierung lexikografischer Daten v. a. diese Trennung von Inhalt und Layout der Daten von zentraler Wichtigkeit. XML ist so angelegt, dass die inhaltliche Struktur von Texten separat von der gestalterischen Umsetzung deutlich gemacht werden kann. Mit XML kann man daher – wie mit SGML (Standard Generalized Markup Language)[5] – die eigentlichen Daten sowohl von ihrer Struktur als auch von ihrer Gestalt trennen. Stefan Freisler spricht in diesem Zusammenhang auch von einer „Explizierung" bzw. „Formalisierung der logischen Textstruktur":

> „Durch SGML wird der Prozess der Explizierung der logischen Struktur eines Textes, den ich als ‚Entlinearisierung' bezeichne, noch eine Stufe weiterentwickelt. Mit einem Schlagwort könnte man diesen Schritt als den ‚Übergang von der Explizierung zur Formalisierung der logischen Textstruktur' bezeichnen. Die

[5] SGML ist die (mächtigere) Vorgängersprache von XML. Die prinzipiellen Eigenschaften, wie sie in den folgenden Zitaten erwähnt werden, gelten sowohl für SGML wie für XML.

> klassische Typografie verlässt sich bei der Auszeichnung und Identifikation von Textteilen und deren Beziehungen auf bestimmte Traditionen und gewisse gestaltpsychologische Wahrnehmungsgesetze. Hierbei ist das ‚Was' des Textes eng mit dem ‚Wie' des Textes verknüpft. Diese beiden Ebenen sollen mit SGML streng getrennt werden. Mit den ‚Markups' von SGML lässt sich die logische Struktur eines beliebigen Textes – unabhängig von Soft- und Hardware-Basis – deskriptiv definieren. Das konkrete Aussehen des Textes wird erst in einem weiteren davon unabhängigen Prozess festgelegt." (Freisler 1994, 41).

Gerade Wörterbuchtexte sind in ihrer gedruckten Form durch typografische und nichttypografische Strukturanzeiger in zahlreiche kleine Einheiten aufgeteilt, die die Wörterbuchbenutzer so voneinander differenzieren können. Möchte man diese Einheiten elektronisch gezielt zugreifbar machen, müssen sie für den Computer direkt zu identifizieren sein. Nur so können die in Abschnitt 2 beschriebenen gezielten Zugriffsmöglichkeiten für Lexikografen und Wörterbuchbenutzer in die Praxis umgesetzt werden. In einem Projekt wie *elexiko* ist daher eine solche möglichst granulare Kennzeichnung der einzelnen Angaben – unabhängig von ihrer konkreten Darstellung – von zentraler Wichtigkeit. Ziel ist die Erstellung *einer Datenbasis*, aus der *mehrere Präsentationen* entwickelt werden können. Dies kann mit dem Einsatz XML sehr gut realisiert werden.

Manche der in Abschnitt 2 erwähnten erwünschten Funktionalitäten von Wortschatzinformationssystemen könnten darüber hinaus noch besser oder umfassender durch den Aufbau eines semantischen Netzes realisiert werden. In aller Regel setzt dies jedoch den Einsatz einer weiteren Software voraus, auch wenn die Modellierung selbst zum Beispiel als Topic Map[6] standardbasiert erfolgen kann. So sprachen in *elexiko* u. a. Kostengründe gegen den Aufbau eines semantischen Netzes. Außerdem kann das Potenzial eines semantischen Netz erst dann wirklich genutzt werden, wenn große Datenmengen durch diese Zugriffsstruktur erschlossen werden. Da im Projekt jedoch erst mit dem Aufbau von Artikeln begonnen wird, wäre momentan der Nutzen eines semantischen Netzes begrenzt.

Ein weiterer Vorteil von XML gegenüber anderen Modellierungssprachen, der m. E. im linguistischen Anwendungsbereich zu wenig betont wird, ist, dass XML von Menschen lesbar und gleichzeitig von Maschinen interpretierbar ist. (Vgl. Gennusa 1999, 30) Die Syntax von XML-DTDs ist nach kurzer Einarbeitung leicht zu verstehen, was ein elementarer Vorteil für die Strukturentwicklung ist. Gerade diese leichte Verständlichkeit ermöglicht die gemeinsame Diskussion über die zu entwickelnde Struktur unter allen Projektbeteiligten, auch und gerade mit denen, die für die Inhalte maßgeblich verantwortlich sind,

---

6 Kurze informative Beiträge zum Topic-Map-Standard bieten Pepper (1999) und Rath (1999). Ein umfassenderer Einblick ist in Widhalm/Mück (2002) zu finden. Ein mögliches Anwendungsszenario wird in Schmidt/Müller (2000) beschrieben. Die Spezifikation von XML-Topic-Maps ist im Internet unter <www.topicmaps.org/xtm/1.0/> einzusehen.

für die die technische Umsetzung jedoch von geringem Interesse ist. Diese Kommunikation läuft nicht über Alltags- oder Fachsprache, sondern anhand der Struktur selber. Es ist in Projekten häufig zu beobachten, dass erst dann die Verständigung über die Modellierung detaillierter wird, wenn man gemeinsam den Entwurf einer DTD diskutiert. Erst die formale Syntax lässt es oft augenscheinlich werden, dass hier und da die entworfene Struktur doch den Inhalten nicht angemessen ist, dass es an bestimmter Stelle eine nicht berücksichtigte aber wohl begründetet Ausnahme gibt etc. Dies ist auch der Grund, weshalb in *elexiko* die Modellierung in XML-DTDs und nicht in XML-Schemas festgehalten ist, da DTDs besser zu ‚lesen' sind.

b) Maßgeschneiderte Modellierung

Allein mit dem Einsatz von XML ist das Modellierungskonzept jedoch noch kaum beschrieben, da XML auf viele Weisen eingesetzt werden kann. Modellierungskonzepte unterscheiden sich von ihrer Intention her u. a. darin, ob sie allgemeine Richtlinien für eine Modellierung festlegen, die dann maßgeschneidert für ein Projekt umgesetzt werden kann oder einen konkreten Modellierungsvorschlag entwickeln, der für möglichst viele Projekte anzuwenden ist. Es gilt daher die Vor- und Nachteile einer maßgeschneiderten vs. einer Standard-Modellierung abzuwägen. Der bekannteste Vorschlag für eine Standard-Modellierung ist die Wörterbuchstruktur der TEI (Text Encoding Initiative, vgl. Sperberg-McQueen/Burnard 2002), einen neuen Vorschlag dazu gibt Franziskus Geeb mit leXeML (Geeb 2001). Ein Vorteil dieser Initiativen oder Vorschläge besteht darin, dass keine eigene Modellierung im Projekt entwikkelt werden muss. Nachteilig ist jedoch, dass entweder projektspezifisch starke Anpassungen vorgenommen werden müssen oder die Struktur sehr allgemein und – im Fall der TEI – sehr weich ist (siehe u. a. Schmidt/Müller 2001, 37 ff.) Eine solche Struktur bietet daher meist nicht in dem Maße eine Unterstützung im lexikografischen Prozess, wie das eine maßgeschneiderte Modellierung kann und wie sie in *elexiko* notwendig ist.

c) Konzeptuelle Inhaltsmodellierung

In *elexiko* wurde angestrebt, das Inhaltsstrukturenprogramm so genau wie möglich in der Modellierung abzubilden, um den Lexikografen eine bestmögliche Strukturführung zu bieten. Ein Grund dafür ist, dass voraussichtlich über einen längeren Zeitraum hinweg in verschiedenen personellen Zusammensetzungen an Wörterbuchartikeln gearbeitet wird, sodass eine möglichst umfängliche automatische Konsistenzkontrolle der Daten sehr wichtig ist. Für *elexiko* wurde daher eine maßgeschneiderte Modellierung entwickelt. Erklärtes Ziel dieses Modellierungskonzeptes ist es, den inhaltlichen Gehalt der lexikografischen Daten und damit auch den genuinen Zweck (i. S. v. Wiegand 1989a,

426), weshalb sie von den Lexikografen für die potenziellen Benutzer angesetzt wurden, so genau wie möglich zu kodieren und transparent zu machen. Die Modellierung stellt damit eine konzeptuelle Inhaltsmodellierung dar. Unsere Hypothese ist dabei, dass potenziell jede angesetzte Angabe für einen gezielten Zugriff durch potenzielle Benutzer interessant sein kann. Die Modellierung ist darüber hinaus möglichst streng, d. h. in ihr wird so genau wie möglich festgelegt, welche Angaben in welcher Reihenfolge in den *elexiko*-Artikeln zu erarbeiten sind. Dies wird in folgenden Abschnitten noch näher erläutert. Diese strenge Modellierung bietet die in Abschnitt 2 geforderte Unterstützung im lexikografischen Prozess, da formal geprüft werden kann, ob die Artikelstruktur eingehalten wurde oder nicht.

d) Schnittstelle zu sprachtechnologischen Anwendungen

Darüber hinaus sollte diese granulare Modellierung – wie sie jetzt vorliegt – und entsprechende Aufbereitung der lexikografischen Daten eine gute Schnittstelle zu sprachtechnologischen Anwendungen bieten. Diese Schnittstelle ist allerdings im Moment für die Projektpraxis noch nicht relevant.

4. Skizze der technischen Redaktionsumgebung

Bevor nun die XML-Modellierung für *elexiko* genauer erläutert wird, soll die gesamte technische Umgebung, in der im Projekt Artikel erstellt werden, kurz skizziert werden.[7] Die XML-Modellierung stellt – um in der Terminologie der Printlexikografie zu sprechen – das formal festgehaltene Mikrostrukturenprogramm dar. Die Lexikografen erstellen ihre Artikel daher zunächst in einem XML-Editor[8], durch den sie durch die XML-Struktur geführt werden (siehe Abschnitt 5.3.5). Ist ein Artikel im XML-Editor validiert, wird er im objektrelationalen Datenbank-Managementsystem Oracle 9i gespeichert. Alle weiteren redaktionellen Zugriffe der Lexikografen und Anfragen der Wörterbuchbenutzer werden an die Datenbank gerichtet. Wird ein Artikel erneut von einer Lexikografin überarbeitet, muss er aus der Datenbank ausgecheckt und erneut im XML-Editor bearbeitet werden.

Auch für die Datenbank stellt die XML-Struktur die Modellierung dar, allerdings werden die Daten nicht so granular in einzelne Tabellenspalten abgelegt, wie es von der XML-Modellierung her möglich ist. Aus Performanzgründen werden im Moment nur diejenigen Elemente in einzelne Tabellenspalten abgelegt, die für den direkten Zugriff bei der Recherche durch Wörterbuchbenutzer notwendig sind. Die eigentlichen XML-Inhalte werden mit Hilfe eines

---

7 Die Konzeption und erste Implementierung dieser technischen Umgebung in der jetzigen Form wurde von Roman Schneider vorgenommen.

8 Im Moment setzen wir XMetaL ein; denkbar ist aber auch (fast) jeder andere XML-Editor.

speziellen Datentyps (XMLType) abgespeichert, welche XML-spezifische Operationen – beispielsweise unter Verwendung von XPath – unterstützt. Die Darstellung der Artikel im Internet wird über ein XSL-Stylesheet spezifiziert (siehe Abschnitt 6).

Perspektivisch wäre es wünschenswert, ein umfassenderes Redaktionssystem zur Verfügung zu haben, um den gesamten Erstellungsprozess homogener zu gestalten und v. a. auch besser die Verweisbeziehungen innerhalb der lexikografischen Daten verwalten zu können. Von der technischen Umgebung her ist in elexiko daher noch einiges zu verbessern.

5. Die XML-Struktur von *elexiko*

5.1 Grundsätzliche Aufteilung der DTDs

Das Projekt *elexiko* soll, wie in der Einleitung (siehe U. Haß, *elexiko* – Das Projekt, in diesem Band) beschrieben wurde, auf der einen Seite ein allgemeines, einsprachiges Wörterbuch werden, welches von der *elexiko*-Projektgruppe erarbeitet wird. Auf der anderen Seite soll *elexiko* auch für andere Projekte der Abteilung Lexik des IDS, perspektivisch auch für Projekte außer Haus die Möglichkeit bieten, sich in die Projektarchitektur von *elexiko* integrieren zu können und so ihre Inhalte elektronisch darzustellen und recherchierbar zu machen. Die Modellierung muss daher zweierlei leisten: auf der einen Seite für jedes Modul eine möglichst genaue, maßgeschneiderte Modellierung bieten, die eine gezielte Recherche und flexible Darstellung erlaubt, auf der anderen Seite so ausgelegt sein, dass diese einzelnen Module möglichst gut integrativ in einer Oberfläche behandelt werden können. Demnach muss die Modellierung modular aufgebaut sein.

Eine XML-basierte Modellierung erfolgt in Form einer DTD (Document Type Definition) oder einem XML-Schema. Aus Gründen der ‚Lesbarkeit' haben wir uns – wie oben bereits erwähnt – für die Modellierung in DTDs entschieden. In einer solchen XML-DTD werden die Regeln festgelegt, der alle dazugehörigen Instanzen, d. h. in lexikografischen Projekten meist alle Artikel, zu gehorchen haben. Dies kann über Parsing-Prozeduren in XML-basierter Software geprüft werden. (siehe Abschnitt 5.3.5). Bisher gibt es DTDs für das *elexiko*-Kernmodul, welche für den Demonstrationswortschatz angewandt wurden, und daneben modulspezifisch angepasste DTDs für die Neologismen-Gruppe.[9] Veranschaulicht in einer Grafik kann das wie in Abbildung 1 gezeigt dargestellt werden. Die XML-Strukturen für die einzelnen Module sollten dabei eine möglichst große Schnittmenge bilden, denn je

[9] Die Neologismen der 90er Jahre sind zunächst als gedrucktes Wörterbuch (Herberg et al. 2004) erschienen; sollen aber baldmöglichst auch über *elexiko* elektronisch recherchierbar sein.

größer die Anzahl der gemeinsamen Angaben ist, desto mehr gemeinsame Zugriffsstrukturen (also Recherchemöglichkeiten) gibt es auf die Artikel.

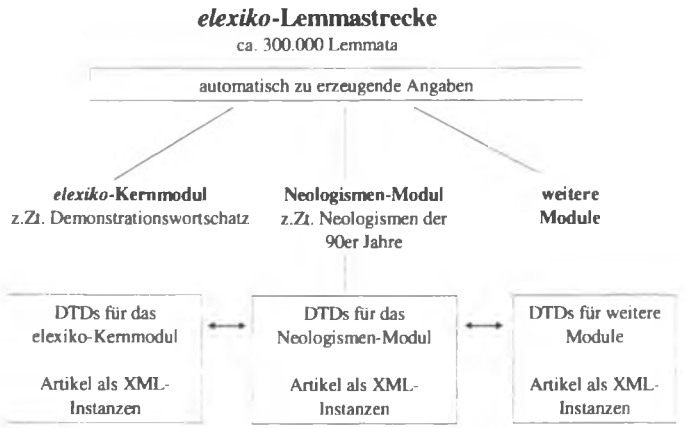

Abb. 1: Veranschaulichung der modularen Aufteilung der DTDs
„ ⟷ " bedeutet soviel wie „sollten eine möglichst große Schnittmenge bilden"

Anfangs wurde im Projekt *elexiko* versucht, die Bedürfnisse aller beteiligten Projekte in einer einzigen XML-Struktur abzubilden. Dies führte jedoch zu einer sehr weichen XML-Modellierung. Diese (neue) Aufgliederung der DTDs ist daher durch folgende Punkte motiviert:

- Alles, was in einzelnen Modulen jetzt oder in Zukunft benötigt wird, kann nicht in einer Struktur abgebildet werden. Eine modulare Aufgliederung der DTDs ist daher sinnvoll.
- Nur durch die modulare Aufgliederung können die DTDs für die jeweiligen Module als Strukturhilfe beim Erarbeiten der Artikel dienen und damit eine Konsistenzsicherung gewährleisten. Gerade diese Strukturhilfe und die konsistente Datenhaltung sind wichtige Vorteile des Einsatzes von XML im lexikografischen Prozess.

5.2 Grundstruktur eines *elexiko*-Artikels

Die DTDs für das *elexiko*-Kernmodul, die hier v. a. beschrieben werden sollen, sind in sich nochmals aufgeteilt in DTDs für Einwortlemmata, Mehrwortlemmata wie „toter Hund" und Wortelementlemmata wie „-lich" oder „-heit". Da der Demonstrationswortschatz nur aus Einwortlemmata besteht, sind diese DTDs bisher am intensivsten getestet und werden daher im Zentrum der Erläuterung stehen.

Die Struktur für Einwortlemmata ist unterteilt in lesartenübergreifende Angaben wie Angaben zum Lemmazeichen, zur Orthografie etc. und lesartenbezogene Angaben zur Semantik, Verwendungsspezifik und Grammatik. Allgemeine Angaben wie Kommentare und Hinweise, die in vielen Zusammenhängen eingesetzt werden, sind in einer DTD für „allgemeine Objekte" zusammengefasst. Diese Aufteilung in mehrere DTDs ermöglicht einen besseren Überblick, da allein die *elexiko*-Kernstruktur für Einwortlemmata aus mehr als 400 Elementen und dazugehörigen Attributen besteht. Die einzelnen Bestandteile werden in einer Kopf-DTD für den Einwortlemma-Artikel zusammengefasst. Demnach besteht ein einzelner Artikel (`ewl-artikel`) zunächst aus lesartenübergreifenden (`ewl-allgemein`) und lesartenbezogenen Angaben (`ewl-lesart`).[10]

Abb. 2: Kontextdiagramm zum Element `ewl-artikel`

Die lesartenübergreifenden Angaben bestehen aus Angaben zum Lemma, zur Frequenz des beschriebenen Wortes, zur Orthografie, zur Morphologie, zum Kookkurrenzverhalten, zur Diachronie, zu möglichen regionalen Markierungen und zu übergreifenden Informationen aller angesetzten Lesarten.[11]

---

[10] Einzelne Element- und Attributnamen der XML-Struktur werden in gesonderter Schriftart dargestellt. Die Teile der XML-Struktur, die hier direkt erläutert werden sollen, werden hier zunächst nicht in XML-Syntax gezeigt, sondern zur leichteren Verständlichkeit anhand der Darstellungskonventionen unserer DTD-Dokumentation (siehe Abschnitt 5.4) veranschaulicht.

[11] Die inhaltlichen Aspekte der einzelnen Angaben werden in den verschiedenen Artikeln dieses Bandes erläutert

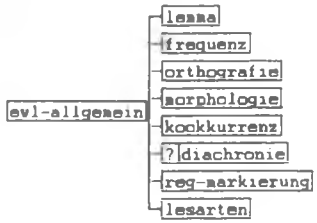

Abb. 3: Kontextdiagramm zum Element `ewl-allgemein`

Die lesartenbezogenen Angaben sind Angaben zur Aussprache, zu möglichen Abkürzungen (wie „Prof" für eine Lesart von „Professor") und Abkürzungsauflösungen (wie „Europäische Union" bei „EU"), wiederum zu möglichen regionalen Markierungen und zu Bedeutung & Verwendung und Grammatik.

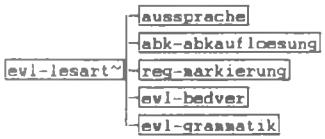

Abb. 4: Kontextdiagramm zum Element `ewl-lesart`

Unter Bedeutung & Verwendung finden sich die semantische Paraphrase, Angaben zur Enzyklopädik, zu den Disambiguierungskriterien, zur Argumentstruktur, zur Paradigmatik, zu den typischen Verwendungsmustern, zur Verwendungsspezifik und – falls vorhanden – zu Lesarten-Spezifizierungen.

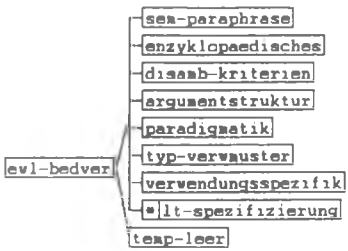

Abb. 5: Kontextdiagramm zum Element `ewl-bedver`

Unter `ewl-grammatik` muss der Lexikograf zunächst die Wortart des zu beschreibenden Stichwortes bestimmen. Unter den einzelnen Wortarten öffnen sich dann die Elemente für die entsprechenden Angabeklassen.[12]

[12] Siehe A. Klosa, Grammatik, in diesem Band

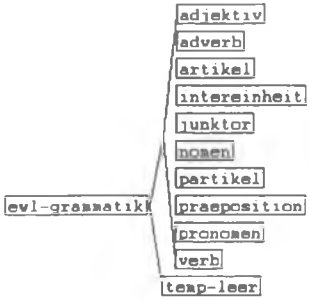

Abb. 6: Kontextdiagramm zum Element `ewl-grammatik`

Unter Bedeutung & Verwendung und Grammatik ist in der XML-Struktur ein Element zu sehen, welches noch nicht erläutert wurde: `temp-leer` für „temporär leer". Dieses Element kann dann von Lexikografen ausgewählt werden, wenn ein Artikel angelegt werden soll, jedoch noch nicht alle Angaben gemacht werden können, d. h. einige Angabegruppen noch zeitlich begrenzt leer sind. Damit der Artikel trotzdem validiert werden kann, muss dieses Ausweichelement an den jeweiligen Stellen auswählbar sein.

Wie bereits oben erwähnt wurde, gibt es neben dieser *elexiko*-Kernstruktur die angepasste Struktur für das Neologismen-Modul. Die Strukturen sind möglichst ähnlich aufgebaut worden, um viele gemeinsame Recherchemöglichkeiten entwickeln zu können. Eine modulspezifische Anforderung ist jedoch beispielsweise, dass unter den lesartenübergreifenden Angaben bei den Neologismen auch eine klassifizierende Angabe zum Neologismus zu machen ist:

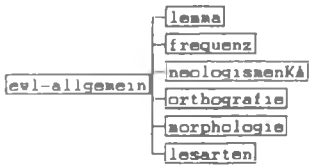

Abb. 7: Kontextdiagramm zum Element `ewl-allgemein` im Neologismenmodul

In dieser Angabe wird über ein Attribut festgelegt, ob es sich beim Stichwort um ein Neulexem, einen Neuphraseologismus, eine Neubedeutung oder eine andere Kategorie handelt. Ein weiterer Unterschied ist z. B. auch, dass im Element `ewl-grammatik` für die Neologismen die Möglichkeit besteht, das Element `keine-wortart` auszuwählen. Dies war notwendig, da zu den Neologismen der 90er Jahre (Herberg et al. 2004) auch das Stichwort „@" gehört. In dieser Weise könnten auch andere Module in das Gesamtsystem der

XML-Strukturen von *elexiko* integriert werden, auch wenn ihre Angaben sich stärker von denen des *elexiko*-Kernmoduls unterscheiden.

Somit ist die Grundstruktur eines *elexiko*-Artikels und seine mögliche modulare Anpassung gezeigt. Da im Folgenden nicht die gesamte Angabestruktur von *elexiko* näher erläutert werden kann, werden nun grundlegende Aspekte der XML-Modellierung anhand ausgewählter Beispiele veranschaulicht.

5.3 Erläuterung grundlegender Aspekte der *elexiko*-DTDs

5.3.1 Granulare Modellierung

Alle Angaben sollen durch die XML-Modellierung direkt zugreifbar sein. Dies setzt eine granulare, maßgeschneiderte Modellierung voraus. Die terminalen Elemente der XML-Struktur werden daher in *elexiko* in drei Typen unterteilt: in Angaben, Kommentare und Hinweise. Diese Differenzierung wirkt sich auf die Art des Inhaltsmodells aus, z. B. sollte innerhalb eines Elementes, welches eine Angabe ist, ausschließlich der Text stehen, auf den der Rechner bei einer Recherche nach dieser Angabe zugreift, d. h. *innerhalb* der Angabe steht kein Kommentar o. Ä. Dieser ist der Angabe als Angabe-Zusatz zugeordnet. Angaben sind damit die granular herauszugreifenden, rein inhaltlich ausgezeichneten Elemente in *elexiko*, auf denen eine gezielte formalisierte Recherche primär aufbaut. Lay-out-orientierte Auszeichnungen sind in Angaben nicht möglich.

Die Kommentare bieten dagegen die Möglichkeit, aufgrund der nicht vorhandenen Platzbeschränkung im elektronischen Medium Angaben ausführlich zu kommentieren. Kommentare beziehen sich damit immer auf eine Angabe.[13] Hinweise können von Kommentaren insofern abgegrenzt werden, dass Lexikografen mit ihnen nicht selber eine Angabe kommentieren, sondern auf andere Texte hinweisen und sie sich deshalb meist auf andere Artikel oder *elexiko*-externe Texte, Werke etc. beziehen. Für Hinweise und Kommentare (und ihre Unterelemente) wird als Elementinhalt eine freiere Textstruktur eingesetzt.

Angaben, Kommentare und Hinweise werden in den *elexiko*-DTDs jeweils mit einem „A", „K" und „H" am Ende des Elementnamens gekennzeichnet.

---

[13] Damit wird hier der Terminus *Kommentar* anders verwendet als bei Wiegand in seiner Theorie lexikografischer Texte, der immer die unmittelbaren Konstituenten des gesamten Wörterbuchartikels, also z. B. den Formkommentar oder den semantischen Kommentar, als Kommentar klassifiziert. (vgl. u. a. Wiegand 1989b) Das, was hier als Kommentar benannt wird, wird bei Wiegand spezifischer als Angabe-Kommentar (AK) bezeichnet. Unsere Verwendung entspricht insgesamt eher der von Reichmann, der Angaben und Kommentare von der Art der Beschreibungsmethode (geschlossene vs. offene Beschreibungsmethode) und dem Grad der Standardisierung unterscheidet (Reichmann 1986, 152 ff).

5.3.2 Arten von Angaben

Die Kennzeichnung eines Elementes als Angabe wird in den DTDs – wie oben schon gesagt – vor allem deshalb gemacht, um eine bestimmte Modellierung des Inhaltsmodells damit zu verbinden. Das Inhaltsmodell von Angaben besteht dabei vom grundsätzlichen Prinzip her immer aus Fließtext oder ist leer, zumindest enthält es keine weiteren Auszeichnungen. Deshalb sind nur die terminalen Elemente der Artikelstruktur, also die Elemente, die inhaltlich als Angabe aufzufassen sind und in denen etwas eingetragen wird, als Angabe aufzufassen. Damit weicht die Klassifizierung eines Elementes als Angabe von der allgemeinsprachlichen Verwendung ab. Z. B. wird man sagen, dass man eine Angabe zum Nominativ Singular macht. In der DTD ist jedoch das Element `nom-sg` noch nicht als Angabe gekennzeichnet, sondern erst die `formA`, die innerhalb des Elementes eingetragen wird. Die Vorteile dieser genauen Modellierung überwiegen unserer Erfahrung nach gegenüber dem Nachteil, dass beim Bearbeiten mehr Elemente ausgewählt werden müssen.

Viele Elemente der Artikelstruktur tragen daher keinerlei Kennzeichnung; nicht als „A" für Angabe, „K" für Kommentar oder „H" für Hinweis. Dies sind in der Regel Klammerelemente, also Teile der hierarchischen Inhaltsstruktur. Ein Beispiel ist die Modellierung der Semantischen Paraphrase (siehe P. Storjohann, Semantische Paraphrasen und Kurzetikettierungen, in diesem Band).

Als Klammerelement wird das Element Semantische Paraphrase definiert:[14]

```
<!ELEMENT sem-paraphrase        (paraphraseA, definitionsbeleg?) >
```

Innerhalb dieses Klammerelements steht die eigentliche Paraphrasen-Angabe, die ausschließlich aus Fließtext besteht:

```
<!ELEMENT paraphraseA           (#PCDATA) >
```

In der Semantischen Paraphrase ist kein Angabe-Zusatz vorgesehen. Jedoch wird auch bei Angaben, denen ein Angabe-Zusatz zugeordnet werden soll, dieses Prinzip verfolgt. Ein Beispiel aus den Angaben zur Paradigmatik (siehe P. Storjohann, Paradigmatische Relationen, in diesem Band).

Für den einzelnen Relationspartner gibt es zunächst ein Klammerelement:

```
<!ELEMENT relpartner            (relpartnerA, angabe-zusatz?) >
```

---

[14] Im Folgenden werden die Beispiele aus den XML-DTDs in XML-Syntax gezeigt, da es sich um nicht so komplexe Inhaltsmodelle handelt.

Die eigentliche Angabe des Partners ist vom dazugehörigen Angabe-Zusatz getrennt. Durch das Klammerelement ist der Angabe-Zusatz jedoch der adressierten Angabe zuzuordnen. Die Angabe selbst enthält wiederum nur Fließtext, d. h. keine weiteren Auszeichnungen:

```
<!ELEMENT relpartnerA        (#PCDATA) >
```

Damit ist der eigentliche Angabetext klar von Zusatzinformationen zu der Angabe getrennt. Die Adressierungs- und Skopusbeziehungen werden daher so gut wie möglich in der Modellierung abgebildet. Außerdem kann so klarer modelliert werden, dass in der Angabe selbst keine rein Lay-out-orientierten Auszeichnungen erwünscht sind. In diesem Modell besteht also immer eine Folgebeziehung zwischen Angabe und Zusätzen zu dieser Angabe. In sehr narrativ aufgebauten Angaben ist das unpassend. Hier gibt es ein gesondertes Modell (siehe unten).

Das Element `angabe-zusatz` gruppiert Kommentare, Hinweise und Belege, die zu Angaben gegeben werden können.

```
<!ELEMENT    angabe-zusatz    (kommentar | hinweis | belege)+ >
```

Eine besondere Form von Angaben sind klassifizierende Angaben, in denen kein Text eingetragen, sondern durch die Auswahl leerer Elemente bestimmte Eigenschaften zugeordnet werden. Diese Angaben sind mit „KA" gekennzeichnet. Eine Sonderform sind diese Angaben deshalb, weil hier nicht die terminalen Elemente als Angabe gekennzeichnet werden, sondern die Eigenschaft, die beschrieben werden soll. Ein Beispiel sind die Angaben zur semantisch-satzfunktionalen Klasse (siehe U. Haß, Das Bedeutungsspektrum, in diesem Band):

```
<!ELEMENT sem-satzfunktklasseKA      (praedikator | quantor |
                                      referenzwort) >
```

Hier wird eine Angabe zur semantisch-satzfunktionalen Klasse gemacht, nicht zum Prädikator. Deshalb ist das Oberelement mit „A" gekennzeichnet.

Falls die eigentliche Angabe im Attribut gemacht wird, wird trotzdem die gesamte Angabe, welcher das Attribut zugeordnet wird, mit „A" gekennzeichnet. Ein Beispiel ist die Angabe zur nationalen Variante innerhalb der regionalen Markierung (siehe P. Storjohann, Das *elexiko*-Korpus, in diesem Band):

```
<!ELEMENT nat-varianteA    (angabe-zusatz) >
<!ATTLIST nat-varianteA     brd-deutsch       (b-ja | b-nein)
                                               #REQUIRED
                            ddr-deutsch       (d-ja | d-nein)
                                               #REQUIRED
                            oesterreichisch   (o-ja | o-nein)
```

```
                                                    #REQUIRED
                          schweizerisch     (s-ja | s-nein)
                                                    #REQUIRED>
```

Hier ist innerhalb der Angabe zwar ein Angabe-Zusatz möglich; da der eigentliche Inhalt der Angabe jedoch über die Attribute gemacht wird, ist die Angabe und der Zusatz hierzu immer noch klar getrennt.

Genauso gibt es Angaben, in denen die eigentliche Angabe schon mit der Auswahl des Elementes gemacht wird. Ein Beispiel dafür sind die Angaben zum Geltungsbereich bei Adjektiven (siehe A. Klosa, Grammatik, in diesem Band):

```
<!-- Adjektiv: Angabe des syntaktischen Geltungsbereiches -->
<!ELEMENT adj-geltbereich     ((adj-attributivA | praedikativA |
                                adverbialA)+, angabe-zusatz?) >
<!-- syntaktischer Geltungsbereich: Attributiv (Adjektiv) -->
<!ELEMENT adj-attributivA     (angabe-zusatz?) >
<!ATTLIST adj-attributivA      stellung    (praenominal |
                                           postnominal | prae-post)
                                                    #REQUIRED >
<!-- syntaktischer Geltungsbereich: Adverbial -->
<!ELEMENT adverbialA          (angabe-zusatz?) >

<!-- syntaktischer Geltungsbereichs:  Praedikativ ====-->
<!ELEMENT praedikativA        (angabe-zusatz?) >
```

Hier wird zwar auch von dem Prinzip abgewichen, dass innerhalb von Angaben keine Angabe-Zusätze gemacht werden sollen, doch auch hier ist die Angabe und der Zusatz zur Angabe klar voneinander zu trennen, da das Inhaltsmodell ausschließlich aus einem Angabe-Zusatz besteht.

An diesem Beispiel ist auch etwas anderes gut zu erkennen: Die Kennzeichnung nur der terminalen Elemente als Angaben hat für die Lexikografen den Vorteil, dass sie sehen, an welchen Stellen der Artikelstruktur wirklich etwas einzutragen ist, d. h. wann sie am Ende des jeweiligen Astes in der Baumstruktur angelangt sind.

5.3.3 Arten von Inhaltsmodellen für Angaben

Angaben haben als Inhaltsmodell in der Regel Fließtext (d. h. #PCDATA) oder sie sind als leer (d. h. EMPTY) definiert. Darüber hinaus gibt es zwei Typen von Inhaltsmodellen für Angaben:

- eines für narrative Angaben
- eines für Elemente, die als Inhalt mehrere Formangaben haben können, d. h. vor allem für schwankende Formen innerhalb der Grammatik.

Das Inhaltsmodell für narrative Angaben ermöglicht zum einen, Absätze auszuzeichnen und zum anderen, einen Angabezusatz (Kommentare, Belege etc.) in den Fließtext einzubauen. Der Angabezusatz muss bei diesen Angaben damit nicht – wie bei ‚normalen' Angaben – von dem eigentlichen Angabetext getrennt werden. Narrative Angaben kommen v. a. innerhalb der Verwendungsspezifik und der Diachronie vor. Mit diesem speziellen Inhaltsmodell soll der freieren Formulierbarkeit und dem möglichen Umfang dieser Angaben Rechnung getragen werden.

Das Inhaltsmodell für v. a. grammatische Angaben ermöglicht die Auszeichnung mehrerer Formangaben innerhalb einer Klammerangabe. Den einzelnen Formangaben kann dabei eine Verwendungshäufigkeitsangabe zugeordnet werden (siehe A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band). Schwankende Formen können insgesamt in der sie umschließenden Klammerangabe kommentiert werden.

Beispiel für ein solches Modell sind die Angaben aus dem Flexionsparadigma, z.B. der Genitiv Singular:

```
<!ELEMENT gen-sg                (form+, angabe-zusatz?) >
```

Innerhalb des Klammerangabe `gen-sg` können mehrere Formen ausgezeichnet werden, denen zusätzlich ein Angabe-Zusatz, d. h. Kommentare, Hinweise etc., zugeordnet werden kann. Im einzelnen Element `form` wird eine Formangabe und optional eine Verwendungshäufigkeitsangabe eingetragen:

```
<!ELEMENT form                  (formA, verw-hkeitA?) >
```

5.3.4 Arten von Kommentaren und Hinweisen

Es gibt in *elexiko* fünf Arten von Hinweisen:

```
<!ELEMENT   hinweis     (verwendungH | sprachreflexionH |
                         grammisH | literaturH | woerterbuchH)+ >
```

Dabei sind diese Hinweise nochmals in zwei Gruppen zu unterteilen: Der Verwendungshinweis und der Sprachreflexionshinweis sind umfangreichere Hinweise, die ihrerseits auch die weiteren verschiedenen Hinweise auf Literatur u. Ä. enthalten und in denen Absätze gemacht werden können. Die anderen drei Hinweise – Grammishinweis, Hinweis auf Sekundärliteratur und auf

Wörterbücher – bestehen aus Fließtext, der Möglichkeit, ein Zitat auszuzeichnen und einem Nachweis.
Kommentare werden momentan in zwei Typen unterteilt:

- den lexikografischen Interpretationskommentar
- den lexikografischen Begründungskommentar

Ein Kommentar ist immer einer dieser beiden Typen zuzuordnen (siehe A. Klosa, Belege in *elexiko*, in diesem Band).

Diese Unterscheidung in einzelne Arten von Hinweisen und Kommentaren fördert unserer Erfahrung nach neben der strikt inhaltsorientierten Modellierung auch die Reflexion beim Schreiben der Artikel seitens der Lexikografen. Denn so muss immer klar Stellung dazu genommen werden, welche Art von Hinweis oder Kommentar hinzugefügt werden soll. In diesem Sinne gehört eine Modellierung in der Art, wie sie für *elexiko* entwickelt wurde, auch zur selbstreflexiven Komponente der Lexikografie als einer eigenständigen, kulturellen Praxis (vgl. Wiegand 1998, 77).

5.3.5 Strenger Aufbau der Modellierung als Unterstützung für die Erarbeitung der Artikel

Anfangs wurde gesagt, dass die Lexikografen bei der Einhaltung der formalen Artikelstruktur unterstützt werden sollen. Dies wird in *elexiko* v. a. durch den strengen Aufbau der XML-Modellierung erreicht. Dies soll kurz demonstriert werden am Beispiel des Artikels „international". Die Abbildung 8 zeigt einen Screenshot des Artikels „international" im XMetaL-Editor.
Auf der linken Seite des Bildschirms sieht man einen Strukturüberblick über die Teile des Artikels. Es handelt sich dabei um die Struktur für ein Einwortlemma. Die Angaben zu einem `ewl-artikel` sind wie oben beschrieben in zwei große Gruppen aufgeteilt: in lesartenübergreifende Angaben, die unter dem Element `ewl-allgemein` gefasst werden, und lesartenbezogene Angaben, die mit dem Element `ewl-lesart` ausgezeichnet werden. In *elexiko* wird auch die Grammatik lesartenbezogen angegeben; deshalb ist das Element `ewl-grammatik` unter den lesartenbezogenen Angaben angeordnet. Innerhalb von `ewl-grammatik` sieht man hier die einzelnen Angabegruppen, die zu Adjektiven gegeben werden: Angaben zur Deklinierbarkeit, zur Steigerung, zur Valenz und zur Syntax. In der Mitte sieht man den Artikel selbst. Die meisten Inhalte sind auf dieser Abbildung ausgeblendet, wie an dem kleinen „+" an den Elementen zu sehen ist. Rechts oben sieht man Attribute, die Elementen zugeordnet werden, und rechts unten nach den Regeln der DTD auszuwählende Elemente.

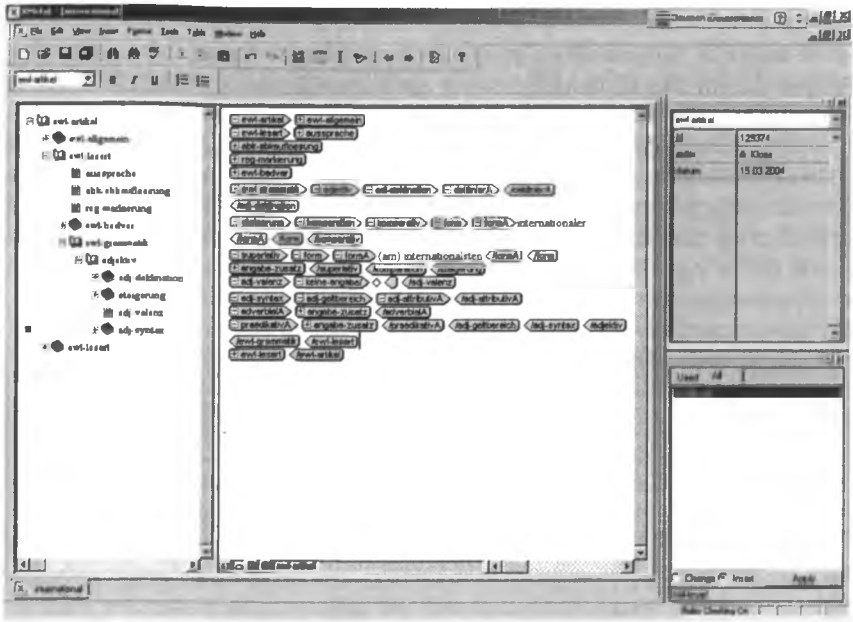

Abb. 8: Artikel „international“ im Bearbeitungsstadium im XML-Editor

Nun wurde schon mehrfach herausgestellt, dass die Lexikografen beim Eingeben der Artikel durch die Modellierung in dem Sinne unterstützt würden, dass die formale Einhaltung der Artikelstruktur geprüft wird. Dies funktioniert in der *elexiko*-Umgebung folgendermaßen: Die Lexikografen bekommen an der jeweiligen Stelle im Artikel, an dem sie gerade arbeiten, nur die Angaben im Editor angeboten, die durch die DTD an der jeweiligen Stelle erlaubt sind. Will ein Lexikograf beispielsweise die Grammatik zu „international“ in der Lesart ‚überstaatlich' im Artikel ausfüllen, muss er zunächst die Wortart bestimmen, hier also `adjektiv`. Abhängig von der Wortart öffnen sich dann jeweils unterschiedliche Angabeklassen. Unter der `deklinierA` muss beispielsweise über ein Attribut ausgewählt werden, ob „international“ in der Lesart ‚überstaatlich' deklinierbar ist oder nicht. Als Demonstration kann man annehmen, dass der bearbeitende Lexikograf vergisst, den passenden Attributwert auszuwählen. Daraufhin muss er etwas zur Steigerung aussagen: entweder er kann `keine-angabe` geben oder er wählt das Element aus, falls das Adjektiv steigerbar ist. Wählt er das Element `komparation` aus, muss er den Komparativ und den Superlativ angeben. Zur Valenz gibt es im Fall von „international“ `keine-angabe`. Unter Syntax werden die von Annette Klosa beschriebenen Angaben zum Geltungsbereich gegeben (siehe A. Klosa, Grammatik, in diesem Band). Man wählt also das Element `adj-`

`geltbereich` aus und sieht rechts unten die Auswahl von attributiv, adverbial und praedikativ. Im Falle von „international" sind alle drei Möglichkeiten des Geltungsbereiches nacheinander auszuwählen. Mit dieser Auswahl sind die Angaben zum Geltungsbereich gemacht.

Ist der Lexikograf mit seinen Angaben zur Grammatik fertig und will den Artikel abschließen, muss er den Artikel zunächst validieren. Das bedeutet, dass der eingegebene Artikel gegenüber der in der DTD festgelegten Struktur geprüft wird. Da die Modellierung in *elexiko* sehr genau und streng ist, kann an dieser Stelle geprüft werden, ob der Lexikograf die festgelegte Artikelstruktur eingehalten hat oder nicht. Führt er diese Validierung im eben erläuterten Artikel durch, bekommt er die in Abbildung 9 gezeigten Fehlermeldungen.

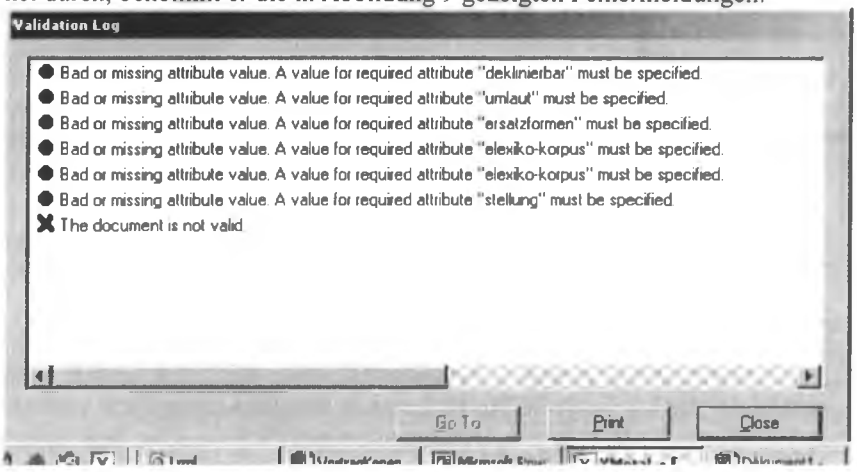

Abb. 9: Fehlermeldungen beim Validieren des Artikels „international"

Zunächst muss er also den Attributwert zur Angabe der Deklinierbarkeit ausfüllen, die er vorhin vergessen hatte. Dann wurde übersehen, dass zum Element `komparation` Attribute zu `umlaut` und `ersatzformen` auszufüllen sind. Diese sind im Falle von „international" beide auf „nein" zu setzen, weil die Steigerungsformen regelmäßig sind, also ohne Umlaut und Ersatzformen gebildet werden. Außerdem wurde bei den Steigerungsformen noch nicht vermerkt, ob sie im *elexiko*-Korpus (vgl. P. Storjohann, Das *elexiko*-Korpus, in diesem Band) belegt sind oder nicht. Dies muss aber in *elexiko* wegen des Prinzips der Korpusbasiertheit (vgl. U. Haß, *elexiko* – Das Projekt, in diesem Band) bei allen Formen gesagt werden. Die Steigerungsformen von „international" sind im *elexiko*-Korpus belegt, also werden die Attributwerte auf „`ja`" gesetzt. Die letzte Fehlermeldung besagt schließlich, dass für den Geltungsbereich „attributiv" etwas zur Stellung ausgesagt werden muss: für „international" ist dabei `praenominal` auszuwählen. Validiert der Lexikograf den

Artikel nun erneut, ist die Prüfung erfolgreich. So werden die Lexikografen unserer Erfahrung nach sehr gut in der formalen Einhaltung der Artikelstruktur unterstützt.

An diesem Beispiel kann man auch erkennen, dass in der Modellierung unterschieden wird zwischen absolut obligatorischen, relativ obligatorischen und fakultativen Angaben. Diese Unterscheidung geht zurück auf Wiegand (1989a). Absolut obligatorische Angaben sind dabei solche Angaben, die immer gegeben werden sollen und auch gegeben werden können, wie z. B. eine Lemmazeichengestaltangabe oder hier bei Adjektiven die Angabe zur Deklinierbarkeit. Relativ obligatorische Angaben sind dagegen Angaben, die nicht zu jedem Lemmazeichen gegeben werden können, wie z. B. Komplemente bei Adjektiven, die aber stets angegeben werden sollen, wenn es möglich ist. In der Modellierung werden diese relativ obligatorischen Angaben von rein fakultativen Angaben folgendermaßen unterschieden werden: Die Angabe selbst wird in der DTD, d. h. im ‚Regelwerk' für die Erstellung der Artikel, – wie die absolut obligatorischen Angaben – als obligatorisch definiert. Innerhalb dieser Angabe oder Angabegruppe gibt es allerdings ein Ausweichelement `keine-angabe`. Damit wird deutlich gemacht, dass diese Angabe oder Angabegruppe immer bearbeitet werden soll und Lexikografen Stellung dazu nehmen müssen; anders als dies bei fakultativen Angaben der Fall ist. Mit dieser Modellierung soll gewährleistet werden, dass diese Angaben nicht versehentlich vergessen werden.

5.4 Die Dokumentation der XML-Struktur

In Abschnitt 3 wurde als ein Vorteil von XML herausgestellt, dass die Syntax dieser Beschreibungssprache von Menschen leicht lesbar ist. Dieser Vorteil kann allerdings nur dann genutzt werden, wenn diejenigen, die die DTD schreiben, sich Mühe geben, diese auch übersichtlich zu gestalten. Dementsprechend schreibt Tommie Usdin (1991, 1):

> „There are several reasons to dress up a DTD; the same reasons good computer programmers have always dressed up programms. The primary reasons are to: Ensure that it is understood, make it easier to maintain, show off your brilliant analysis, and lighten the shroud of mystery that hangs over densely coded material. [...] DTDs are only human-readable if the creator takes little time to make them readable."

Es gibt verschiedene Aspekte in einer DTD, die die Übersichtlichkeit und Verständlichkeit der Struktur verbessern. Dies sind die optische Gestaltung der DTD, die Verteilung von Elementen, Attributen und Entities, die Modularisierung von DTDs, die Kommentierung der DTD, die Dokumentation der Struktur und schließlich die Wohlüberlegtheit in der Auswahl der Sprache, in der die DTD entwickelt wird.

Allen Elementdefinitionen sind in den *elexiko*-DTDs deshalb Kurzkommentare vorangestellt, die die Elementnamen vollständig auflösen, z. B. „Semantische Paraphrase" beim Element `sem-paraphrase`. Wichtiger ist aber noch die DTD-Dokumentation. Diese hat in *elexiko* verschiedene Funktionen: Sie soll

- innerhalb der Projektgruppe als Kurzreferenz zur Erläuterung der Modellierung dienen,
- Vereinbarungen hinsichtlich der Beschreibungssprache analog zur XML-Struktur soweit wie möglich integrieren,
- für diejenigen, die die Modellierung entwickeln, bestimmte Modellierungsentscheidungen auch nach einiger Zeit noch nachvollziehbar machen, und
- für neue Projektgruppenmitglieder den Einstieg in die Modellierung erleichtern.

Zur Erstellung der Dokumentation wurde die Software *DTDhelp* der Ovidius GmbH eingesetzt, die es ohne großen Aufwand ermöglicht, die Dokumentation sowohl als Kurzreferenz zu benutzen, d. h. gezielt nach bestimmten Strukturelementen zu suchen, als auch systematisch die Modellierung zu verfolgen, d. h. die Elemente im hierarchischen Aufbau durchzugehen. Zur Demonstration ist in Abbildung 10 der Eintrag zum Element `wortbildung` gezeigt.

Auf der linken Seite des Bildschirms ist die Suche-Funktion zu sehen, in der in diesem Beispiel das Suchwort „Wortbildung" eingegeben wurde. Als Suchergebnis werden dabei alle Einträge angezeigt, in denen die Zeichenkette „Wortbildung" vorkommt, d. h. das gesuchte Element, das Oberelement, alle Unterelemente und alle zugehörigen Attribute. Auf der rechten Seite befindet sich der Eintrag zum Element `wortbildung`. Unter „Modellierung" finden sich Hinweise sowohl zur Modellierung als auch redaktioneller Art. Im Fall von Wortbildung steht in diesem Abschnitt beispielsweise, dass – falls ein Stichwort gebildet ist – immer nur ein Wortbildungstyp ausgewählt werden kann und soll. Bestehen Zweifel in der Zuordnung, soll ein Wortbildungstyp eingetragen und dieser Zweifel kommentiert werden. Hinweise dieser Art sind hilfreich beim Erarbeiten der Artikel, da die Dokumentation direkt aus dem XML-Editor aufgerufen werden kann und zwar genau an dem Element, an dem der Lexikograf gerade arbeitet. Will daher jemand innerhalb der Wortbildung

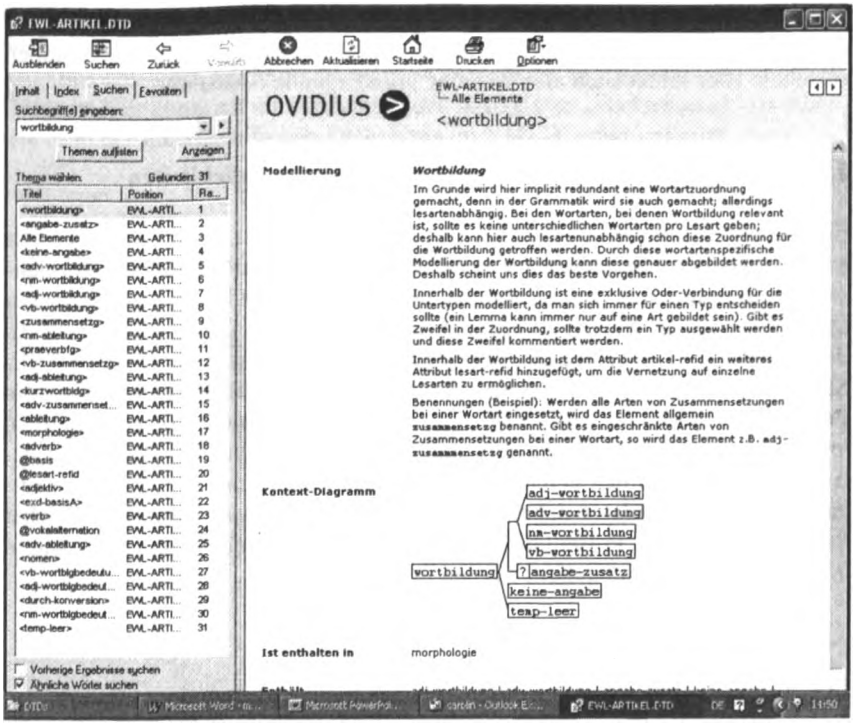

Abb. 10: Eintrag zum Element `wortbildung` in der DTD-Dokumentation

zwei Wortbildungstypen eintragen und kann dies aber nach den Vorgaben der DTD nicht, kann er die Dokumentation aufrufen und die redaktionelle Richtlinie finden, die die Modellierung erklärt. Daher ist es sinnvoll, das Redaktionshandbuch in die DTD-Dokumentation zu integrieren. Diese redaktionellen Hinweise befinden sich in *elexiko* allerdings noch im Aufbau.

Das „Kontextdiagramm" zeigt zum einen das Inhaltsmodell von `wortbildung`, zum anderen kann darüber zu den Unterelementen navigiert werden. Genauso dienen die Felder „Ist enthalten in" und „Enthält" der Navigation durch die hierarchische Struktur. Unter diesen Abschnitten befinden sich Informationen zu Attributen, falls dem Element Attribute zugeordnet sind.[15] Hier werden automatisch der Name des Attributs, der Wertebereich und die Einstellungen der Obligatorik gezeigt. Auch zu den Attributen können darüber hinaus redaktionelle Hinweise oder Erläuterungen zur Modellierung hinzugefügt werden. Eine solche DTD-Dokumentation ist bei der Rolle, die die Modellierung in *elexiko* im lexikografischen Prozess spielt, unerlässlich. Aller-

[15] Dieser Abschnitt ist in der Abbildung nicht zu sehen

dings ist es das Schicksal vieler Dokumentationen, dass sie im Projektalltag stiefmütterlich behandelt werden, da scheinbar immer dringendere Aufgaben anstehen. Hier bildet auch *elexiko* keine grundlegende Ausnahme.

6. Eine Datenbasis – mehrere Präsentationsmöglichkeiten

Ein anfangs formuliertes Ziel bei der Modellierung ist, dass die Daten der lexikografischen Datenbasis auf verschiedene Weise im Wortschatzinformationssystem präsentiert werden können, ohne die Daten selbst zu verändern. Zur Darstellung der in XML vorliegenden lexikografischen Daten wird in *elexiko* die Extensible Style Language, kurz XSL, eingesetzt. Mit ihr lassen sich XSL-Stylesheets erstellen, in denen die Darstellung der Daten festgelegt wird. Im Prinzip kann dabei für jedes XML-Element und -Attribut spezifiziert werden, wie es den potenziellen Benutzern dargestellt werden soll. Der Vorteil ist dabei, dass separat zu den eigentlichen Daten die Darstellung der Daten spezifiziert wird. Je granularer dabei die Datenauszeichnung ist, desto spezifischer kann demgemäß die Präsentation definiert werden. Dieser Prozess kann wie in Abbildung 11 veranschaulicht werden.

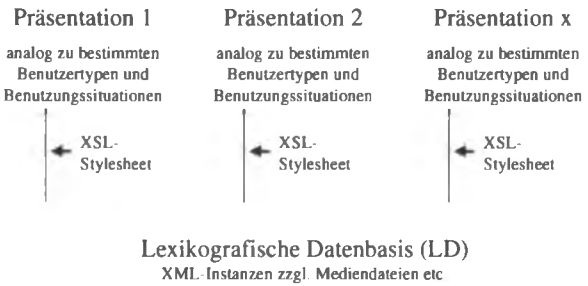

Abb. 11: Veranschaulichung der Herstellung mehrerer Präsentationen aus einer Datenbasis „⟶" bedeutet soviel wie „aus der Datenbasis wird entwickelt" „⟵" bedeutet soviel wie „mit dem Stylesheet wird spezifiziert"

Die Kombination von granular ausgezeichneter Datenbasis und Spezifikation der Darstellung mittels XSL bietet also die Möglichkeit, beispielsweise bestimmte Formen von Kommentaren immer nur in bestimmten Benutzungssituationen anzuzeigen oder für unterschiedliche Benutzergruppen eine unter-

schiedliche Terminologie, unter der die Angaben gezeigt werden, zu entwickeln.

Dies soll kurz verdeutlicht werden anhand der grammatischen Angaben. Dafür ist hier zunächst ein Ausschnitt des Artikels zu „international" zu sehen, wie er in der lexikografischen Datenbasis im XML-Format vorliegt.

```
<adj-syntax>
<adj-geltbereich>

<adj-attributivA stellung="praenominal"></adj-attributivA>

<adverbialA>
<angabe-zusatz><belege><ek-beleg><zeitung-beleg>
<belegtextA>Wie von der Bundesregierung geplant und von Warschau
nicht gewollt, verhandeln Kohl und Mazowiecki morgen parallel über
den Grenzvertrag und die Rechte der Vertriebenen - beides wird
gekoppelt, obwohl Bonn <belegwortA>international</belegwortA>
zugesichert hat, die Grenze ohne Wenn und Aber anzuerken-
nen.</belegtextA>
<zt-belegnachweisA name="taz" datierung="07.11.1990">die tageszei-
tung, 07.11.1990, S. 4, Polen-Vertrag: Bonn setzt sich durch.</zt-
belegnachweisA></zeitung-beleg></ek-beleg>
</belege></angabe-zusatz>
</adverbialA>

<praedikativA>
<angabe-zusatz><belege><ek-beleg><zeitung-beleg>
<belegtextA>Und da im Internet nicht nach Nationalitäten unter-
schieden wird, könnte aus dem Netz-Dauer-TED eine Nationalhymne
entstehen, die wahrhaft <belegwortA>international</belegwortA> ist
        so wie ja einst die Internationale die sowjetische Hymne
gewesen ist, bevor der Zweite Weltkrieg dazwischen
kam.</belegtextA>
<zt-belegnachweisA name="BZ" ressort="Feuilleton" datie-
rung="23.11.2000">Berliner Zeitung, 23.11.2000, Tagebuch,
S. 13.</zt-belegnachweisA></zeitung-beleg></ek-beleg>
</belege></angabe-zusatz>
</praedikativA>

</adj-geltbereich>
</adj-syntax>
```

Abb. 12: Ausschnitt der XML-Instanz des Artikels „international"

Zur kurzen Erläuterung der XML-Struktur: Die Angaben zum Geltungsbereich werden über die Auswahl der Elemente `adj-attributivA`, `adverbialA` und `praedikativA` gemacht, d. h. wenn diese Elemente von den Lexikografen ausgewählt werden, ist eine entsprechende Aussage über den Geltungsbereich gemacht. Zu jeden dieser Element kann ein `angabe-zusatz` gegeben werden, der beispielsweise wie hier Belege enthalten kann. Dies entspricht dem Konzept von *elexiko*, keinen Belegblock anzulegen, sondern mit Belegen

immer einzelne Angaben zu dokumentieren. (siehe A. Klosa, Belege in *elexiko*, in diesem Band)

In der folgenden Abbildung ist der entsprechende Ausschnitt des Artikels „international" in der Darstellung zu sehen, wie sie im jetzt angewandten Stylesheet festgelegt ist.

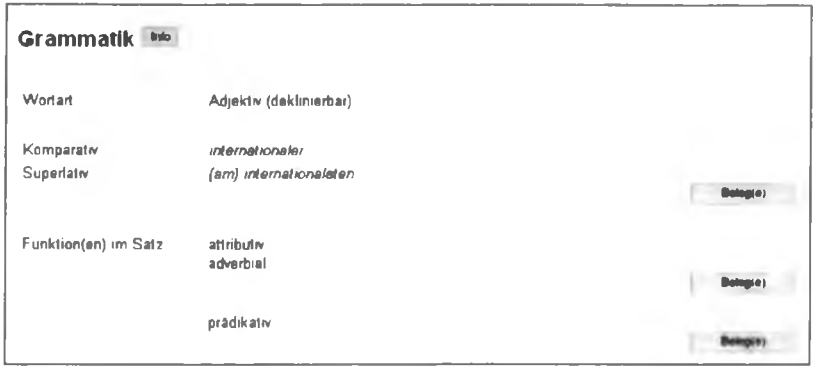

Abb. 13: Ausschnitt der aktuellen Bildschirmansicht des Artikels „international"

Im Stylesheet wird dabei – wie oben erwähnt – für jedes XML-Element und Attribut festgelegt, wie es darzustellen ist. Hier ist beispielsweise zu sehen, dass die Belege nur auf Wunsch der Benutzer ‚aufgeklappt' werden und für den besseren Überblick nicht immer zu sehen sind. Die Darstellung wird also in einer separaten Schicht spezifiziert. Zu Demonstrationszwecken wurde nun das Stylesheet verändert, und zwar im ersten Beispiel so, dass die Darstellung für Deutsch-als-Fremdsprache-Nutzer benutzerfreundlicher sein sollte. Hier sind die Angaben zu Steigerung mehr ausformuliert und die Angaben zum syntaktischen Geltungsbereich unter einer etwas veränderten Terminologie angezeigt. Darüber hinaus wäre denkbar, gerade zu den Angaben zum Geltungsbereich die Belege standardmäßig zu öffnen, da diese für DaF-Nutzer wahrscheinlich besonders wichtig sind.

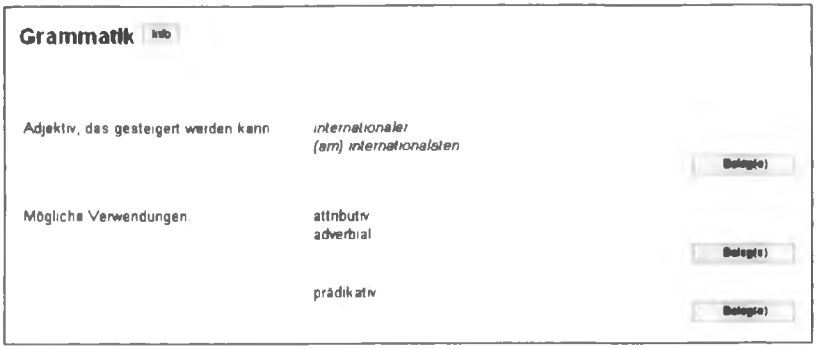

Abb. 14: Bildschirmansicht 2 des Artikels „international"

Der entsprechende Stylesheet-Auszug zeigt, wie ein Teil dieser Darstellung festgelegt wurde:

```
<td class="layout-cell1">Mögliche Verwendungen:</td>
       <td class="ids-color">
       <xsl:for-each select="./adj-syntax/
                             adj-geltbereich/node()">
              <xsl:choose>
              <xsl:when test="name(.) = 'adj-attributivA'">
                             attributiv
              <xsl:apply-templates select="./angabe-zusatz"/>
              </xsl:when>
              <xsl:when test="name(.) = 'praedikativA'">
                             prädikativ
              <xsl:apply-templates select="./angabe-zusatz"/>
              </xsl:when>
              <xsl:when test="name(.) = 'adverbialA'">
                             adverbial
              <xsl:apply-templates select="./angabe-zusatz"/>
```

Abb. 15: Stylesheet-Auszug zur Darstellung der Angaben zum Geltungsbereich in der Bildschirmansicht 2

Im zweiten Beispiel wurde das Stylesheet so verändert, dass die Ansicht mehr auf Experten ausgerichtet ist. Hier ist die Deklinier- und Steigerbarkeit nur anhand der zwei entsprechenden Termini beschrieben; die Steigerungsformen selbst brauchen Experten nicht. Auch in den Angaben zum syntaktischen Geltungsbereich wird eine mehr fachorientierten Terminologie verwendet. Die Belege sind in dieser Ansicht unterdrückt.

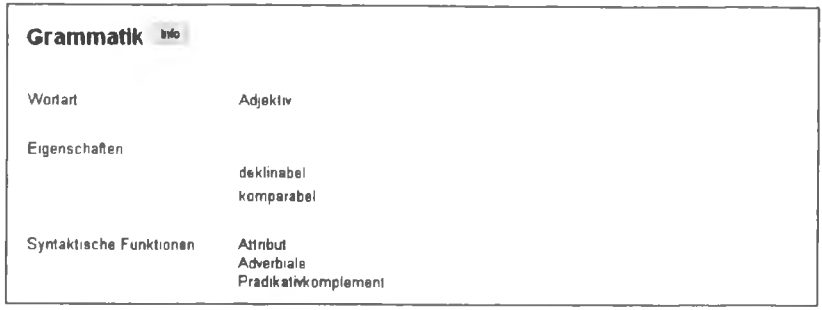

Abb. 16: Bildschirmansicht 3 des Artikels „international"

Auch hier zeigt der Stylesheet-Auszug zu den Angaben des Geltungsbereiches die Veränderung:

```
<td class="layout-cell1">Syntaktische Funktionen:</td>
       <td class="ids-color">
       <xsl:for-each select="./adj-syntax/
                             adj-geltbereich/node()">
              <xsl:choose>
              <xsl:when test="name(.) = 'adj-attributivA'">
                                    Attribut
              <xsl:apply-templates select="./angabe-zusatz"/>
              </xsl:when>
              <xsl:when test="name(.) = 'praedikativA'">
                                    Prädikativkomplement
              <xsl:apply-templates select="./angabe-zusatz"/>
              </xsl:when>
              <xsl:when test="name(.) = 'adverbialA'">
                                    Adverbiale
              <xsl:apply-templates select="./angabe-zusatz"/>
```

Abb. 17: Stylesheet-Auszug zur Darstellung der Angaben zum Geltungsbereich in der Bildschirmansicht 3

Analog dazu ist es beispielsweise auch möglich festzulegen, dass im Falle, dass ein nichtmuttersprachlichen Benutzer einen Text produziert, die typischen Verwendungsmuster und Mitspielerangaben (siehe P. Storjohann, Typische Verwendungen, in diesem Band) prominenter erscheinen, als wenn ein muttersprachlicher Nutzer einen Text rezipiert. Denn für den nichtmuttersprachlichen Benutzer ist es beim Schreiben von Texten wichtig zu wissen, dass es nicht beispielsweise ‚internationale Gruppe' sondern „internationale Gemeinschaft" heißt und dass nicht auf dem ‚internationalen Platz' aber „auf der internationalen Bühne" oder „auf dem internationalen Parkett" gehandelt wird.

Für einen muttersprachlichen Nutzer sind dagegen diese Informationen beim Rezipieren eines Textes überflüssig.

Wichtig ist dabei, dass allen drei gezeigten Ansichten derselbe Wörterbuchartikel auf Ebene der lexikografischen Datenbasis zu Grunde liegt. Die flexible Darstellbarkeit lexikografischer Daten analog zu Benutzungssituationen – so wie sie in Abschnitt 2 gefordert wurde – lässt sich also gut mit dem Modellierungskonzept von *elexiko* und mit den entsprechend strukturierten Daten umsetzen. Allerdings ist die Entwicklung verschiedener Stylesheets für unterschiedliche Benutzungssituationen in *elexiko* noch nicht in die Tat umgesetzt.

7. Perspektiven für die Recherche

Eine weitere zentrale Anforderung an die Modellierung, die in Abschnitt 2 aufgeführt wurde, ist der gezielte Zugriff auf einzelne Angaben in den Artikeln. Über das Prinzip der granularen, inhaltsorientierten Modellierung ist diese Möglichkeit auf Ebene der lexikografischen Datenbasis geschaffen. Prinzipiell kann auf alle XML-Elemente und -Attribute und damit auf die lexikografischen Inhalte direkt bei der Recherche zugegriffen werden. Von der Aufbereitung der Daten wäre es daher jetzt schon möglich, beispielsweise nach allen Nomina zu suchen, zu denen Hyperonyme angegeben sind. Oder nach allen Stichwörtern zu suchen, die Kurzwortbildungen sind und nur in bestimmten Domänen verwendet werden.

Allerdings ist die Realisierung dieser Zugriffsmöglichkeiten ein Programmieraufwand, der in *elexiko* bisher nur teilweise umgesetzt werden konnte. Einige Angaben – beispielsweise zur Orthografie und Paradigmatik – können jetzt schon in Rechercheanfragen eingebunden werden. (siehe A. Klosa, Orthografie und morphologische Varianten, und P. Storjohann, Paradigmatische Relationen, in diesem Band). Allerdings werden diese Funktionalitäten erst schrittweise erweitert werden können. Das, was durch die Modellierung und entsprechende Datenaufbereitung in diesem Zusammenhang geleistet werden kann, ist allerdings in *elexiko* schon realisiert.

8. Schlussbemerkung

Die Entwicklung einer XML-Modellierung in der Weise, wie sie für *elexiko* vorgenommen wurde, ist eine langfristige und dauerhafte Projektaufgabe. Denn die Modellierung muss nicht nur den lexikografischen Inhalten wirklich angemessen sein und deshalb intensiv besprochen und getestet werden, sondern sie muss auch bei der Weiterentwicklung der inhaltlichen Konzeption immer wieder angepasst werden. In diesem Sinne sind die *elexiko*-DTDs nie

fertig. Trotzdem lohnt sich unserer Erfahrung nach dieser intensive Aufwand für die DTD-Enwicklung, da so die Arbeit an den Inhalten und die technische Umsetzung durch die Schnittstelle der XML-Struktur sowohl sachlich wie personell besser verzahnt sind.

9. Literaturverzeichnis

9.1. Wörterbücher

Herberg, Dieter/Kinne, Michael/Steffens, Doris (unter Mitarbeit von Tellenbach, Elke/al-Wadi, Doris) (2004): Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Berlin/New York. (Schriften des Instituts für Deutsche Sprache 11).

9.2 Forschungsliteratur

Alexa, Melina/Kreissig, Bernd/Liepert, Martina/Reichenberger, Klaus/Rautmann, Karin/Rostek, Lothar/Scholze-Stubenrecht, Werner/Stoye, Sabine (2002): The Duden ontology: An Integrated Representation of Lexical and Ontological Information. In: Workshop at IREC2002. Las Palmas, Gran Canaria (27.5.2002); zitiert nach <www.darmstadt.gmd.de/~rostek/alexa-etal-Irec2002.pdf>, 8 Seiten.

Bergenholtz, Henning/Tarp, Sven/Wiegand, Herbert Ernst (1999): Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffman, Lothar/Kalverkämper, Hartwig/Wiegand, Herbert Ernst (Hg.) (1999): Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. 2. Halbbd. Berlin/New York. S. 1762-1832.

Büchel, Gregor/Schröder, Bernhard (2001): Verfahren und Techniken in der computergestützten Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.) (2001): Chancen und Perspektiven computergestützter Lexikographie. Tübingen. S. 7-28. (Lexicographica. Series Maior 107).

Engelberg, Stefan/Lemnitzer, Lothar (2001): Lexikographie und Wörterbuchbenutzung. Tübingen. (Stauffenburg Einführungen; Bd. 14).

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

Freisler, Stefan (1994): Hypertext – Eine Begriffsbestimmung. In: Deutsche Sprache 22, S. 19-50.

Geeb, Franziskus (2001): leXeML – Vorschlag und Diskussion einer (meta-)lexikographischen Auszeichnungssprache. In: Sprache und Datenverarbeitung 2/2001, S. 27-61.

Gennusa, Pamela L. (1999): Evolution and use of generic markup languages. In: Möhr, Wiebke/Schmidt, Ingrid (1999): SGML und XML. Anwendungen und Perspektiven. Berlin. S. 27-50.

Gloning, Thomas/Welter, Rüdiger (2001): Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe Wörterbuch. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (2001): Chancen und Perspektiven computergestützter Lexikographie. Tübingen. S. 117-132. (Lexicographica. Series Maior 107).

Heyn, Matthias (1992): Zur Wiederverwendung maschinenlesbarer Wörterbücher. Eine computergestützte metalexikographische Studie am Beispiel der elektronischen Edition des „Oxford Advanced Learner's Dictionary of Current English" Tübingen. (Lexicographica. Series maior 45).

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

Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (2001): Einführung. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (2001): Chancen und Perspektiven computergestützter Lexikographie. Tübingen. S. 1-4. (Lexicographica. Series Maior 107).

Müller-Spitzer, Carolin (2004): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: Lexicographica 19/2003, S. 140-168.

Pepper, Steve (1999): Euler, Topic Maps and Revolution. In: XML Europe 1999. Conference Proceedings, S. 135-150.

Rath, Holger (1999): Mozart oder Kugel. Mit Topic Maps intelligente Informationsnetze aufbauen. In: iX 12/1999, S. 149-155.

Reichmann, Oskar (1986): Lexikographische Einleitung. In: FWB 1, S. 10-164.

Schmidt, Ingrid/Müller, Carolin (2001): Entwicklung eines lexikographischen Modells: Ein neuer Ansatz. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (2001): Chancen und Perspektiven computergestützter Lexikographie. Tübingen. S. 29-52. (Lexicographica. Series Maior 107).

Schmidt, Ingrid/Müller, Carolin (2000): Zaubernetz. Inhaltsstrukturen und Topic Maps als Potenzial neuer Informationstechnik. In: iX 11/2000, S. 100-107. / In: http://www.heise.de/ix/artikel/2000/11/100/.

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

Sperberg-McQueen, C.M./Burnard, Lou (2002): TEI P 4: Guidelines for Electronic Text and Interchange. Text Encoding Initiative Consortium. XML Version. Oxford u.a. (siehe auch www.tei-c.org/Guidelines2/index.html)

Steindler, Larry (1995): Voraussetzungen und Perspektiven für ein Informationssystem: Österreich – Gegenwart eines Kulturraumes. In: Lexicographica 11/1995, S. 219-241.

Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (2001): Chancen und Perspektiven computergestützter Lexikographie. Tübingen. S. 53-69. (Lexicographica. Series Maior 107).

Thielen, Christine/Breidt, Elisabeth/Feldweg, Helmut (1998): COMPASS: Ein intelligentes Wörterbuchsystem für das Lesen fremdsprachiger Texte. In: Storrer, Angelika/Harriehausen, Bettina (1998): Hypermedia für Lexikon und Grammatik. Tübingen. S. 173-194. (Studien zur deutschen Sprache 12).

Usdin, Tommie (1991): The Well Dressed DTD. In: Tag 14/1990, S. 1-5.

Widhalm, Richard/Mück, Thomas (2002): Topic Maps. Semantische Suche im Internet. Berlin u.a.

Wiegand, Herbert Ernst (1989 a): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilbd. Berlin/New York. S. 409-462. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1).

Wiegand, Herbert Ernst (1989 b): Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilbd. Berlin/New York. S. 462-501. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1).

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

9.3 Internetressourcen

XML Standard [Third Edition]: www.w3.org/TR/REC-xml/. (letzter Zugang Oktober 2004)

XTM: www.topicmaps.org/xtm/1.0/. (letzter Zugang Oktober 2004)