



INSTITUT FÜR
DEUTSCHE SPRACHE

OPAL

Online publizierte Arbeiten zur Linguistik

ISSN 1860-9422

2/2006

Carolin Müller-Spitzer

Das Konzept der Inhaltsstruktur

Ein Ausschnitt aus einer neuen Konzeption für die
Modellierung lexikografischer Daten

OPAL – Online publizierte Arbeiten zur Linguistik
Herausgegeben vom Institut für Deutsche Sprache



Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
opal@ids-mannheim.de

Technische Redaktion: Norbert Volz

© 2006 IDS Mannheim – Alle Rechte vorbehalten

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechts ist ohne Zustimmung der Copyright-Inhaber unzulässig und strafbar. Das zulässige Zitieren kleinerer Teile in einem eigenen selbstständigen Werk (§ 51 UrhG) erfordert stets die Angabe der Quelle (§ 63 UrhG) in einer geeigneten Form (§ 13 UrhG). Eine Verletzung des Urheberrechts kann Rechtsfolgen nach sich ziehen (§ 97 UrhG). Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Die zugänglichen Daten dürfen von den Nutzern also nur zu rein wissenschaftlichen Zwecken genutzt werden. Eine darüber hinausgehende Nutzung, gleich welcher Art, oder die Verarbeitung und Bearbeitung dieser Daten mit dem Zweck, sie anschließend selbst oder durch Dritte kommerziell zu nutzen, bedarf einer besonderen Genehmigung des IDS (Lizenz). Es ist nicht gestattet, Kopien der Textdateien auf externen Webservern zur Verfügung zu stellen oder Dritten auf sonstigem Wege zugänglich zu machen. Bei der Veröffentlichung von Forschungsergebnissen, in denen OPAL-Publikationen zitiert werden, bitten die Autoren und Herausgeber um eine entsprechende kollegiale Information an opal@ids-mannheim.de.

Carolin Müller-Spitzer

Das Konzept der Inhaltsstruktur

Ein Ausschnitt aus einer neuen Konzeption für die Modellierung lexikografischer Daten¹

Alles Gescheite ist schon gedacht
worden, man muss nur versuchen, es
noch einmal zu denken.

Johann Wolfgang von Goethe

1. Einführung

Wenn man heute neue lexikografische Projekte plant, ist es unerlässlich, sowohl den Computer als Arbeitswerkzeug, als auch das elektronische Medium als Publikationsmedium in die Konzeption einzubeziehen. Dabei ist es oft erwünscht, die einmal erarbeiteten lexikografischen Daten mehrfach zu verwerten, sei es als gedruckte und elektronische Version eines Wörterbuchs, sei es einmal als eine lange und kurze Ausführung eines lexikografischen Produkts oder als eine ausschließlich elektronische, aber benutzeradaptiv unterschiedliche Darstellung der gleichen lexikografischen Daten. (Vgl. u.a. Schryver 2003, Storrer 2001, Engelberg/Lemnitzer 2001, S. 220ff.) Solche Wünsche sind mit den heutigen Möglichkeiten des Computereinsatzes zu erfüllen, doch müssen die lexikografischen Prozesse entsprechend geplant werden. Vor allem die Frage der Strukturierung und Aufbereitung der lexikografischen Daten muss sorgfältig überlegt sein, denn hierbei wird der Grundstein dafür gelegt, wie flexibel auf die Daten zugegriffen werden kann und wie unterschiedlich darstellbar sie sind. Für einen inhaltlich wie gestalterisch möglichst flexiblen Umgang mit lexikografischen Daten ist daher eine bestimmte Form der Datenmodellierung und -auszeichnung erforderlich. Das Stichwort, das in diesem Zusammenhang immer wieder fällt, ist das der medienneutralen Datenhaltung. Doch wie müssen lexikografische Daten ausgezeichnet werden, damit diese Kodierung losgelöst von den Eigenschaften einer bestimmten Präsentation ist? Welche Richtlinien können hier projektübergreifend angewandt werden? Die moderne Wörterbuchforschung gibt darauf noch keine hinreichenden Antworten. Vor allem fehlen klare, theoretisch fundierte und intersubjektiv nachvollziehbare Richtlinien für die maßgeschneiderte Modellierung lexikografischer Daten. Denn oft ist es in lexikografischen Projekten nicht möglich, eine Standard-Modellierung wie die TEI² oder leXeML³ anzuwenden, da diese kein ausreichend granulares und maßgeschneidertes Modellierungsinventar bieten, um die lexikografischen Daten später sehr flexibel darstellen zu können und verschiedenartige Zugriffsstrukturen zu bieten. Doch auch bei der Entwicklung einer maßgeschneiderten Modellierung ist es wichtig, einen klaren Leitfaden für die Datenmodellierung zu haben, um nicht bei jedem Phänomen neu entscheiden zu müssen, wie die Modellierung aussehen soll. Ziel dieses Aufsatzes ist es daher, ein Konzept vorzustellen, welche Eigenschaften lexikografischer Daten Gegenstand für eine Modellierung einer lexikografischen Da-

¹ Dieser Artikel wird voraussichtlich in leicht modifizierter Form auch in der Zeitschrift *Lexicographica* erscheinen.

² Die TEI (= Text Encoding Initiative) wurde Ende der 80er-Jahre als Forschungsprojekt mit dem Ziel ins Leben gerufen, Richtlinien für die Auszeichnung verschiedener Texttypen aus dem geisteswissenschaftlichen Bereich zur Verfügung zu stellen. 1994 erschienen die „Guidelines for Electronic Text and Interchange“, bekannt als TEI P 3 (= TEI Proposal number 3) (Sperberg-McQueen/Burnard 1995), worin auch ein Inventar für die Kodierung von Wörterbuchdaten enthalten ist. Mittlerweile liegen die TEI-Richtlinien bereits als Proposal number 5 vor (vgl. www.tei-c.org/P5). Da die TEI-Wörterbuchstruktur allerdings den Anspruch hat, für alle Wörterbücher der westlichen Welt anwendbar zu sein, ist sie für die meisten Projekte, die viele individuelle Zugriffsmöglichkeiten auf lexikografischen Daten entwickeln wollen, zu allgemein gehalten. (Vgl. auch Müller-Spitzer 2005b und Schmidt/Müller 2001.)

³ leXeML ist ein neuerer Vorschlag für eine Standard-Modellierung von Geeb (vgl. Geeb 2001 und www.lexeml.org), die sich von den Ansätzen her deutlich von denen der TEI unterscheidet, und vor allem potenzielle Benutzer in die Modellierung einbezieht. Insofern ist sie aber für eine Modellierung mit dem Anspruch, zunächst möglichst unabhängig von einer potenziellen Benutzungssituation zu sein, nicht passend.

tenbasis sein sollen, wie eine solche Modellierung aussehen kann und schließlich auch, welche Folgen diese Art der Modellierung für die lexikografische Praxis hat. Ein wichtiger Teil dieses Konzepts der Inhaltsstrukturen ist dabei auch die Entwicklung und Ausgestaltung einer eigenen Terminologie; auch in Abgrenzung zu Wiegands Konzept der Mikrostrukturen (vgl. Abschnitt 5).

2. Zu Grunde gelegte Rahmenbedingungen

Das Konzept der Inhaltsstrukturen ist Teil eines Gesamtkonzeptes für die Modellierung einer lexikografischen Datenbasis (vgl. Müller-Spitzer 2005b). Dieses Modellierungskonzept ist primär für die Neuerstellung lexikografischer Daten gedacht, also für zukünftige lexikografische Projekte.⁴ Im Fokus liegen v. a. Sprachwörterbücher. Die Anwendung dieses Konzeptes soll dabei dazu dienen,

- die Lexikografen im lexikografischen Prozess, d.h. bei der Datenerarbeitung und Dateneingabe so zu unterstützen, dass sich eine Instruktionenbuch hinsichtlich der Einhaltung der formalen Artikelstruktur erübrigt (vgl. Wiegand 1998; Engelberg/Lemnitzer 2001, S. 211ff.; Heyn 1992, S. 192);
- die so erarbeiteten Daten in innovativer Weise für potenzielle Benutzer zugreifbar zu machen (dies betrifft vor allem die elektronische Präsentation), und
- den Weg offen zu halten, die erarbeiteten lexikografischen Daten auch in gedruckter Form präsentieren zu können.

Um dieses Ziel erreichen zu können, muss die Ebene der Datenbasis strikt von der der Präsentation(en) getrennt werden. Die Modellierung und die entsprechende Datenstrukturierung erfolgt dabei auf Ebene der lexikografischen Datenbasis.

Das Modellierungskonzept soll allerdings nicht nur (theoretisch) beschreiben, welche Eigenschaft der lexikografischen Daten wie modelliert werden kann, sondern auch einen konkreten Leitfaden für Anwender bereitstellen. Dafür ist es notwendig, konkrete Beispiele für Modellierungen aufzuzeigen, d.h. auch eine bestimmte Modellierungssprache auszuwählen. Hierbei bietet sich die Anwendung von XML an, da XML ein softwareunabhängiger (Industrie-)Standard ist, der eine langlebige Datenhaltung garantiert und mit dessen Anwendung eine strikte Trennung von inhaltlicher Strukturierung und Präsentation erreicht werden kann. „The use of a single, internationally agreed syntax was not the only reason why a generic markup language was the appropriate solution for publishers that wanted to publish to multiple media. A fundamental concept in the use of generic markup languages is a move from formatting markup to content markup.“ (Gennusa 1999, S. 28). Stefan Freisler spricht in diesem Zusammenhang auch von einer „Explizierung“ bzw. „Formalisierung der logischen Textstruktur“ (Freisler 1994, S. 41). Dabei kann XML als Metasprache gelten, mit der eigene Auszeichnungssprachen definiert werden können. „Zu XML gehört ein Grammatikformalismus, mit dem man in Form von so genannten *Document Type Definitions (DTDs)* kontextfreie Grammatiken [...] erstellen kann, um die Struktur einer Klasse von Dokumenten formal zu beschreiben. Eine DTD legt die erlaubten Tags fest und gibt Regeln an, wie sie verschachtelt werden dürfen. Dadurch kann mit einem Parser die *Gültigkeit* von Dokumenten bezüglich einer DTD-Grammatik festgestellt werden, um nicht nur die Wohlgeformtheit zu garantieren, sondern die Zuverlässigkeit zur Verarbeitung durch eine eingegrenzte Klasse von Softwaresystemen.“ (Hedtstück 2003, S. 132). Mit der Bestimmung der Tags wird dabei in der DTD auch das Auszeichnungsvokabular festgelegt. In diesem Sinne ist eine DTD wiederum eine Metasprache zur Beschreibung einer Klasse von Dokumenten. „A document type definition spells out exactly what the allowable language is. A DTD is thus a metalanguage for a certain type document.“ (O'Meara/Reeder 2005). Somit ist es mit XML möglich, eine individuell maßgeschneiderte Modellierung für ein konkretes lexikografisches Projekt zu entwickeln.

Für die Anwendung von XML im Rahmen des Modellierungskonzeptes ist auch das Zusammenspiel von XML und verwandten Standards wie XSL(T) wichtig, da mit diesen flankierenden Standards bei-

⁴ Im Projekt *ellexiko*, einem elektronischen, lexikologisch-lexikografischen, korpusbasierten Informationssystem des Instituts für Deutsche Sprache, wird das Modellierungskonzept bereits angewendet (vgl. Müller-Spitzer 2005a).

spielsweise die Darstellung der XML-Daten auf standardisierte Weise festgelegt werden kann. „The key to understanding the revolutionary potential of XML is that it is just one piece of a larger picture. XML by itself can provide standardized interchange formats for databases and spreadsheets. This is significant. But XML and XSL together can replace existing word processing and desktop publishing formats as well. It can give us, in effect, a single, completely internationalized format of almost unlimited power for both print and online publishing that is fully interoperable across all products and all platforms. The implications of this go far beyond data exchange and far beyond the Web.“ (Bosak 1998, S. 3). Gerade im Gegensatz zu SGML ist diese leichte Anwendbarkeit der flankierenden Standards ein zentraler Vorteil von XML.

Das Modellierungskonzept ist vor allem auf die Anwendung in der wissenschaftlichen Lexikografie ausgelegt. Dabei ist es von großem Vorteil, dass für den Einsatz von XML zunächst einmal nicht unbedingt aufwändige Software-Anschaffungen notwendig sind, anders als beispielsweise beim Einsatz semantischer Netze. „The point of XML – and the reason that it's becoming so popular – is that all the various special-purpose languages that can be defined using it can be parsed by a single standardized processor small enough to be built into every Web browser.“ (Bosak 1998, S. 2). Die Einbindung einer Datenbank in die Redaktionsumgebung ist in den meisten lexikografischen Projekten jedoch unerlässlich.

Mit der Anwendung von XML ist allerdings nicht automatisch eine Trennung von Inhalt und Layout in der Strukturierung erreicht. Vor allem führt eine XML-basierte Kodierung von lexikografischen Daten nicht automatisch dazu, dass die Daten so aufbereitet sind, dass sie flexibel zugreifbar und darstellbar werden. Vielmehr sind hierfür gewissen Richtlinien erforderlich. Für die Art der Modellierung, wie sie hier vorgeschlagen wird, wurde deshalb der Begriff der *konzeptuellen Inhaltsmodellierung* entwickelt (vgl. Müller-Spitzer 2005b). Eine solche Modellierung setzt eine inhaltsorientierte Datenauszeichnung von menschlicher Hand, d.h. intellektuell reflektiert, voraus. Außerdem sollte die Modellierung selbst ein Konzept aufweisen, d.h. bestimmten Prinzipien folgen, bestimmte Regeln des Aufbaus haben etc. Welche Prinzipien und Richtlinien dabei für lexikografische Daten gelten können, wird im Folgenden ausgeführt.

3. Das Konzept der Inhaltsstruktur

3.1 Inhaltliche Konzeption

Für die Modellierung der Inhaltsstruktur müssen die einzelnen inhaltlichen Einheiten in der lexikografischen Datenbasis konstituiert und klassifiziert werden. Wichtig ist dabei, dass alle möglichen Eigenschaften der Daten, die ihre Präsentation betreffen, dabei ausgegrenzt werden. Im Zusammenhang mit der Modellierung einer lexikografischen Datenbasis soll daher von einer *funktionalen Aufgliederung* der lexikografischen Daten gesprochen werden. Da keine Darstellungsaspekte Gegenstand der Modellierung sein können, darf auch die mögliche Reihenfolge der Daten für eine bestimmte Präsentation nicht schon in der Modellierung festgehalten werden. Allerdings wird im gedruckten Wörterbuch über die Präzedenzrelation, d.h. über die Vorgänger-Nachfolger-Relation, nicht nur die Reihenfolge der Angaben abgebildet, sondern auch inhaltliche Zusammenhänge verdeutlicht wie Skopus- oder Adressierungsbeziehungen. Diese inhaltlichen Zusammenhänge, die durch den Textverdichtungsprozess in standardisierten gedruckten Wörterbüchern nicht mehr explizit sind, sind für potenzielle Benutzer jedoch zum Verständnis der Artikeltexte unabdingbar. „Bei dem Textverdichtungsprozess geht die Kohäsion (i.S.v. Wiegand 1988, S. 79f.) weitestgehend verloren [...]. Für den durch die Textverdichtung entstehenden Verlust der kohäsiven Voraussetzungen, um die Konstituenten aufeinander zu beziehen, muss es einen lexikografischen Ersatz geben. Dieses liefern die artikelinternen Adressierungsbeziehungen. Ihre Kenntnis erlaubt, dass die Angaben richtig aufeinander bezogen werden; dies ist eine der notwendigen Voraussetzungen für die Gewinnung von lexikografischen Informationen anhand von Artikeltexten.“⁵ (Wiegand 2002b, S. 129). Diese inhaltlichen Zusammenhänge sind auch für

⁵ Anmerkung der Autorin: Sämtliche Zitate wurden an die neue Rechtschreibung angepasst. Hervorhebungen im Original.

die Modellierung der Inhaltsstrukturen entscheidend, denn diese sind inhaltlich begründet und unabhängig von einer einzelnen Präsentation. Umso wichtiger ist es, bei der funktionalen Aufgliederung der lexikografischen Daten die Teil-Ganzes-Beziehungen so vollständig wie möglich inhaltlich abzubilden, damit die inhaltlichen Zugehörigkeiten explizit gemacht sind und die Modellierung insgesamt so hierarchisch aufzubauen, dass diese inhaltlichen Zusammenhänge rekonstruierbar sind. Dieser Anspruch hat zur Folge, dass die lexikografischen Daten auf Ebene der Datenbasis teilweise anders geordnet werden müssen, als man das von gedruckten Wörterbüchern gewohnt ist. Zum Beispiel kann es sinnvoll sein, grammatische Angaben einzelbedeutungsbezogen anzulegen, damit ihr Geltungsbereich immer klar ist. Die hierarchische Inhaltsstruktur soll also so genau wie möglich schon durch ihren hierarchischen Aufbau die Skopusbeziehungen deutlich machen. Dies ist auch deshalb möglich, weil die Form und Reihenfolge, in der die lexikografischen Daten in der lexikografischen Datenbasis abgelegt werden, nicht der Form und Reihenfolge ihrer Präsentation entsprechen muss.

Die *hierarchische Inhaltsstruktur* kann in eine *konkrete hierarchische Inhaltsstruktur* und in eine *abstrakte hierarchische Inhaltsstruktur* unterschieden werden. Die Trägermenge einer konkreten hierarchischen Inhaltsstruktur sind *Vor-Angaben*, die inhaltlich einem konkreten Lemmazeichen zuzuordnen sind. Von *Vor-Angaben* wird deshalb gesprochen, da Angaben in der Theorie lexikografischer Texte nach Wiegand wohlbestimmt sind, und zwar als funktionale Textsegmente, die aus einer Angabeform und wenigstens einen genuinen Zweck bestehen (Wiegand 1989, S. 412). Die Einheiten in der lexikografischen Datenbasis bestehen jedoch nur aus einem genuinen Zweck und natürlich einer schriftlichen Realisierung, die jedoch nicht mit der Angabeform auf Präsentationsebene zu verwechseln ist. Den Terminus *Vor-Angabe* verwendet auch Wiegand in seinem Szenario eines „Mikrostrukturengenerators“ (Wiegand 1998, S. 215ff.), in dessen Artikelformulare die Lexikografen „nur *Vor-Angaben*“ eingeben; „das sind Angaben mit einer vorläufigen Angabeform“ (Wiegand 1998, S. 217). Auf Ebene der Datenbasis wird entsprechend der Ausdruck *Lemmazeichen* statt *Lemmazeichengestaltung* verwendet, da das Lemmazeichen nur als Ordnungselement dient und andere Formeigenschaften wie die korrekte orthografische Form gemäß den Grundsätzen konzeptueller Inhaltsmodellierung gesondert kodiert werden.⁶ *Vor-Angaben* werden im Folgenden – im Sinne einer kürzeren Ausdrucksweise – mit *Angabe_V* bezeichnet. Trägermenge der abstrakten hierarchischen Inhaltsstruktur sind entsprechend *Klassen von Angaben_V*, die inhaltlich einem Lemmazeichen zuzuordnen sind, und zwar Klassen von *Angaben_V* gleichen genuinen Zwecks. Konkrete und abstrakte hierarchische Inhaltsstrukturen lassen sich als Baumstruktur darstellen und auch in einem Strukturgraphen zusammenführen (vgl. auch Abschnitt 5).

Die Modellierung soll so aufgliedert werden, dass die *terminalen Elemente*, d.h. die Blätter der erzeugten Bäume, nicht weiter aufgliederbare Klassen von funktionalen Textsegmenten mit mindestens einem genuinen Zweck sind. Wie bei den Mikrostrukturen nach Wiegand soll auch bei den Inhaltsstrukturen der jeweilige genuine Zweck des funktionalen Textsegments – der darin besteht, dass es anhand bestimmter Eigenschaften dem potenziellen Benutzer dazu dient, diejenigen Benutzungsziele erreichen zu können, um deren Erreichung willen der Lexikograf das funktionale Textsegment formuliert hat – die Benennung für die jeweilige Klasse liefern. Dies entspricht dem Prinzip für die Datenauszeichnung nach einer konzeptuellen Inhaltsmodellierung (vgl. Müller-Spitzer 2005b). Die Benennung ist damit gleich motiviert wie im Rahmen der funktional-positionalen Segmentation von Wiegand, trotzdem werden sie Benennungen anders anders realisiert (vgl. Abschnitt 5). Bei Wiegand sind die einzelnen Klassen von Angaben in Kürzeln benannt wie beispielsweise „BPA“ für „Bedeutungsparaphrasenangabe“. Dies ist im Rahmen der Theorie auch günstig, da hier die Benennungen entsprechend dem wissenschaftlichen Ansatz möglichst kurz, genau und in Schaubildern darstellbar sein sollen. Außerdem wird mit dieser Art der Benennung das Ziel verfolgt, wörterbuchspezifische Termini, die sich in einzelnen Projekten etabliert haben, in eine wörterbuchübergreifende, einheitliche Terminologie zu überführen (vgl. auch Storrer 1996, S. 249). In der Modellierung ist mit der Benen-

⁶ Zur Motivation des Terminus *Lemmazeichengestaltung* vgl. u.a. Wiegand (2002b, S. 134f.).

nung der Angaben_v jedoch ein anderer Handlungszweck verbunden: Die Modellierung wird im lexikografischen Prozess eingesetzt, d.h., die Lexikografen arbeiten direkt mit diesen Benennungen. Insofern sollten diese möglichst nicht stark verkürzt und so selbsterklärend wie möglich sein. Es kann beispielsweise im Rahmen der Modellierung auch hilfreich sein, den Lexikografen durch die Benennung in der Modellierung zu signalisieren, dass ein Element ein terminales Element in der Baumstruktur ist. Außerdem ist es im Rahmen dieses Modellierungskonzeptes gerade sinnvoll, wenn diese Benennungen wie die Modellierung insgesamt maßgeschneidert sind. Sie müssen allein unter den Projektbeteiligten eindeutig sein. Allerdings sollte die grundsätzliche Benennungsmotivation projektübergreifend dem hier ausgeführten Ansatz entsprechen.

In Abb. 1 ist zur Veranschaulichung ein fiktiver Ausschnitt aus einer abstrakten hierarchischen Inhaltsstruktur als Baumgraph dargestellt. Dieser Ausschnitt soll vor allem der Veranschaulichung der möglichen Benennung der Trägerelemente einer abstrakten hierarchischen Inhaltsstruktur dienen. Hier ist zu sehen, dass der genuine Zweck (i.S.v. Wiegand 1989, S. 426) die Benennung für die jeweilige Angabe_v-Klasse liefert. Die Benennungen sind dabei ausgeschrieben, sodass sie unmittelbar zu verstehen sind. Die terminalen Elemente sind jeweils entweder mit einem „A“ für Angabe oder „K“ für Kommentar gekennzeichnet. Diese Kennzeichnung signalisiert den Lexikografen beim Erarbeiten der Artikel im XML-Editor, dass in diese Felder etwas einzutragen ist. Das Wurzelement ist `artikel` benannt. Wichtig ist dabei zu beachten, dass damit ein Artikel auf Ebene der lexikografischen Datenbasis bezeichnet ist, und kein Wörterbuchartikel in einem gedruckten Wörterbuch oder Wortschatzinformationssystem (vgl. Müller-Spitzer 2003). Denn wie oben ausgeführt wurde, können aus einem Artikel aus der lexikografischen Datenbasis, d.h. allen Angaben, die inhaltlich einem Lemmazeichen zuzuordnen sind, mehrere Wörterbuchartikel auf Ebene der Präsentation werden.

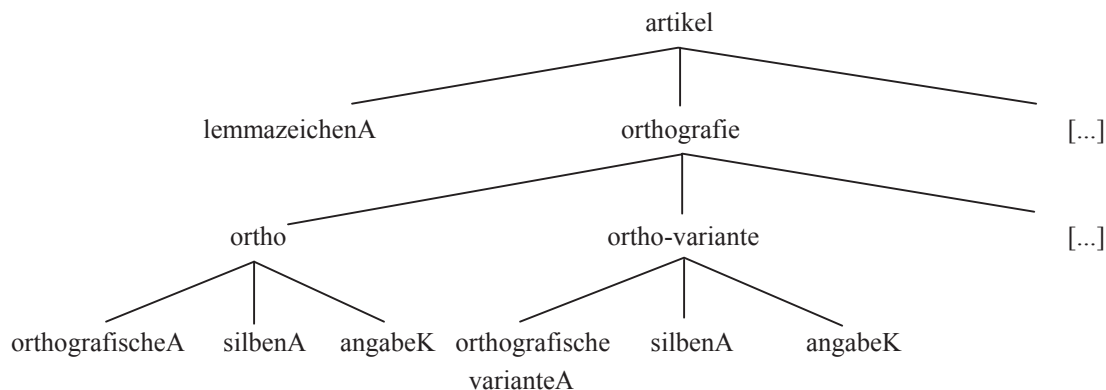


Abb. 1: Veranschaulichung zur Benennung einzelner Trägerelemente einer abstrakten hierarchischen Inhaltsstruktur

In diesem Beispiel ist auch zu erkennen, dass in einer konzeptuellen Inhaltsmodellierung verschiedene Angabe_v-Klassen, die in gedruckten Wörterbüchern in einer Form zusammengeführt werden, in jeweils einzelne Elemente einzuordnen sind. Beispielsweise ist es in allgemeinen, einsprachigen Wörterbüchern üblich, die Lemmazeichengestaltangabe, die korrekte Schreibung und die Silbenangabe in einer Form verdichtet darzustellen. Der Vorteil bei einer granularen Modellierung und dementsprechenden Datenauszeichnung ist jedoch, dass u.a. für die Ansetzung des Lemmazeichens eigene Richtlinien angewendet werden können. Es ist beispielsweise denkbar, dass in einem strikt korpusbasierten Wortschatzinformationssystem immer die Schreibung, die in den Korpora am häufigsten nachzuweisen ist, als Lemmazeichen angesetzt wird. Als orthografisch korrekte Form wird jedoch die nach der neuen Rechtschreibung gültige Form angegeben. Ein weiterer Vorteil bei einer solchen granularen Inhaltsmodellierung ist, dass Angabe-Kommentare von ihrer Adressierung genau zuzuordnen sind. Beispielsweise ist im oben gezeigten Ausschnitt in der abstrakten hierarchischen Inhaltsstruktur vorgehen, dass Angabe-Kommentare zu den einzelnen orthografischen Formen gegeben werden können.

Genauso wie die Unterscheidung von abstrakten und konkreten hierarchischen Inhaltsstrukturen aus dem Konzept der Mikrostrukturen übernommen wurde, kann auch der Terminus des Mikrostrukturen-

programms fruchtbar angewendet werden. „Die linguistisch bestimmten Eigenschaften legen fest, welche Angabeklassen (mit Angaben, die sich auf den Wörterbuchgegenstand beziehen) zu berücksichtigen sind; zu jedem Lemmzeichentyp gehört daher eine Menge von Angabeklassen, auf der bestimmte Ordnungsrelationen definiert sind, so dass zu jedem Lemmzeichentyp eine abstrakte *hierarchische Mikrostruktur* gehört [...]. Die Menge aller abstrakten Mikrostrukturen bildet das *Mikrostrukturenprogramm* des in P₂ [ein in Gang befindlicher lexikografischer Prozess, Anm. d. Verf.] zu erarbeitenden Wörterbuchs; dieses ist im Instruktionbuch festgelegt.“ (Wiegand 1998, S. 215). Genauso bildet die Menge aller abstrakten hierarchischen Inhaltsstrukturen das *Inhaltsstrukturenprogramm* einer lexikografischen Datenbasis. Mit dem Inhaltsstrukturenprogramm wird damit die Erzeugung aller zulässigen abstrakten hierarchischen Inhaltsstrukturen lizenziert. Daher ist genau dieses Inhaltsstrukturenprogramm Gegenstand der Modellierung. Man kann somit auch von der Erstellung einer *Inhaltsstrukturengrammatik* für eine lexikografische Datenbasis sprechen, d.h. der Erstellung einer kontextfreien Grammatik, die festlegt, welche abstrakten hierarchischen Inhaltsstrukturen erzeugt werden können.⁷

Die Stufen der funktionalen Aufgliederung der lexikografischen Daten können in einer Baumstruktur dargestellt werden. Dabei werden in dieser Baumstruktur nicht nur relativ zu Lemmzeichentypen die zugehörigen abstrakten hierarchischen Inhaltsstrukturen festgelegt, sondern an jeder beliebigen Stufe der Hierarchie können Alternativen relativ zu bestimmten zu beschreibenden Eigenschaften des Lemmzeichens abgebildet werden. Ein Beispiel: In einem lexikografischen Projekt wird festgelegt, dass die Grammatik immer einzelbedeutungsbezogen angegeben werden soll. Die Angaben in der Grammatik unterscheiden sich natürlich abhängig davon, welche Wortart dem Lemmzeichen zugeordnet wird. Dementsprechend sind im Inhaltsstrukturenprogramm hier unterschiedliche Äste im Baum angelegt.

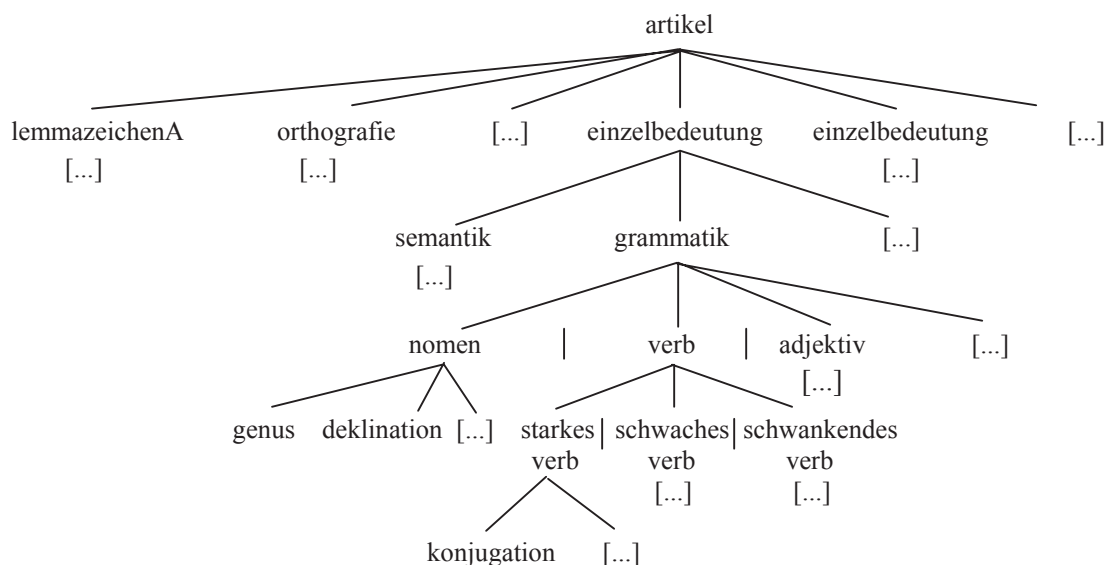


Abb. 2: Veranschaulichung zu einem beispielhaften Ausschnitt aus einem Inhaltsstrukturenprogramm.
 „x | y“ bedeutet soviel wie: „x steht in einer exklusiven Oder-Verbindung zu y“

Genauso können die Verästelungen aufgrund anderer Kriterien innerhalb der Semantik etc. differenziert werden. Diese Differenzierung an den jeweils inhaltlich relevanten Stellen ist sinnvoll, da es viele verschiedene Eigenschaften eines Lemmzeichens gibt, die eine spezielle hierarchische Inhaltsstruktur verlangen. Daher können nicht nur auf der Ebene des Lemmzeichens Kriterien zur Differenzierung

⁷ Vgl. Storrer (1996, S. 247). Sie sieht einen weiteren Abstraktionsschritt von abstrakten Mikrostrukturen zur „*Mikrostrukturengrammatik*, die idealiter alle wohlgeformten abstrakten hierarchischen Mikrostrukturen [aller] Wörterbuchartikel erzeugt, bzw. lizenziert“.

angewandt werden. Die Lexikografen arbeiten sich dabei beim Erarbeiten der lexikografischen Daten durch dieses Inhaltsstrukturenprogramm und die damit abgebildete Hierarchie hindurch.

Damit ist die Modellierung des Inhaltsstrukturenprogramms zusammenfassend folgendermaßen konzipiert: Die Modellierung des *Inhaltsstrukturenprogramms* erfolgt in Form einer XML-DTD (bzw. XML-Schemas), in der alle zulässigen *abstrakten* hierarchischen Inhaltsstrukturen festgelegt sind. Die lexikografischen Daten zu einem Lemmazeichen in einer zu dieser DTD validen XML-Instanz bilden eine *konkrete hierarchische Inhaltsstruktur*. Für die Modellierung findet eine *funktionale Aufgliederung* der zu erstellenden lexikografischen Daten in *Klassen von Angaben_v* mit mindestens einem genuinen Zweck statt. Dieser genuine Zweck bietet auch die Bezeichnung für die jeweilige Klasse. Die terminalen Elemente in der Modellierung, d.h. die Vaterknoten der Blätter des Baumes, sollten *nicht weiter zerlegbare* Klassen von Angaben_v sein. So wird auch sichergestellt, dass alle einzelnen inhaltlichen Einheiten so granular modelliert sind, dass sie direkt maschinell zugreifbar sind. Dabei sollen die *Stufen dieser funktionalen Aufgliederung*, und damit die inhaltlichen Teil-Ganzes-Beziehungen, so *vollständig wie möglich abgebildet* werden.

Da die Modellierung in XML erfolgt, können die o.g. Grundsätze folgendermaßen präzisiert werden: Durch die XML-DTD bzw. durch das DTD-Modulsystem werden alle zulässigen abstrakten hierarchischen Inhaltsstrukturen lizenziert. Diese Menge aller zulässigen abstrakten hierarchischen Inhaltsstrukturen soll *Inhaltsstrukturenprogramm* heißen. Die Erarbeitung dieses Inhaltsstrukturenprogramms muss damit von den konzeptionellen Grundzügen her der DTD-Entwicklung vorgelagert sein. Das Inhaltsstrukturenprogramm wird dann wechselseitig mit der DTD-Entwicklung verändert und verfeinert. Die XML-DTD(s) werden anschließend für die Erarbeitung der lexikografischen Daten eingesetzt, i.d.R. werden die Daten in einem XML-Editor geschrieben. Damit entstehen konkrete hierarchische Inhaltsstrukturen, zu deren Trägermenge Angaben_v und die zugehörigen XML-Auszeichnungen gehören.

Die Modellierungsrichtlinien sind in Form von Leitsätzen formuliert. Alle Leitsätze hier aufzuführen, wäre in diesem Kontext zu umfangreich. Damit trotzdem ein Eindruck vermittelt werden kann, welche Form diese Leitsätze haben, sollen im Folgenden die Richtlinien zur Modellierung des hierarchischen Aufbaus des Inhaltsstrukturenprogramms beispielhaft ausgeführt werden. Entsprechend dazu gibt es Richtlinien zur Modellierung der terminalen Elemente (vgl. Müller-Spitzer 2005b).

3.2 Eine Auswahl von Richtlinien für die Modellierung des Inhaltsstrukturenprogramms: die Modellierung des hierarchischen Aufbaus

Beim grundsätzlichen Aufbau der Baumstruktur eines Inhaltsstrukturenprogramms kommt es zunächst vor allem darauf an, die inhaltlichen Teil-Ganzes-Beziehungen so vollständig wie möglich abzubilden. Denn je genauer diese abgebildet sind, desto besser lassen sich für die Präsentation benutzerfreundliche Angabestrukturen entwickeln, d.h. (abstrakte und isomorphe konkrete) hierarchische Mikrostrukturen mit Adressierung (vgl. Wiegand 2002b, S. 150). Die Strukturgraphen nach Wiegand legen dabei den gedruckten Wörterbuchartikel zu Grunde. Die Modellierung der Teil-Ganzes-Beziehungen auf Ebene der lexikografischen Datenbasis bezieht dagegen allein auf die inhaltlichen Zusammenhänge und noch nicht auf eine mögliche Präsentation der Daten. Deshalb ist es sinnvoll, im Inhaltsstrukturenprogramm allein das in Angabegruppen zusammenzufassen, was inhaltlich zusammengehört.

Für die Daten von allgemeinen, einsprachigen Wörterbüchern bedeutet dies vor allem eine Trennung in zwei Angabegruppen: in einzelbedeutungsübergreifende Angaben und einzelbedeutungsbezogene Angaben.⁸ Neben diesen Angabegruppen gibt es als dritten grundsätzlichen Bestandteil die Angabe des Lemmazeichens, die vor allem Adressierungsfunktion hat. Die Angabe der korrekten Schreibung

⁸ In manchen Projekten wird der Terminus *Lesart* statt *Einzelbedeutung* verwendet.

wird in einer konzeptuellen Inhaltsmodellierung – anders als bei gedruckten Wörterbüchern üblich – nicht gleichzeitig mit der Angabe des Lemmzeichens gemacht.

Das Wurzelement eines Inhaltsstrukturenprogramms ist das Element `artikel`, wobei auch hier wiederum der Artikel auf Ebene der lexikografischen Datenbasis gemeint ist, der für die Präsentation in mehrere Wörterbuchartikel aufgeteilt sein kann. Ein Artikel enthält dabei alle Angaben, die inhaltlich dem Lemmzeichen zuzuordnen sind. Damit ergibt sich als Kopfstruktur eines Artikels folgende Modellierung (die Benennungen der Elemente sind eine Möglichkeit):

```
<!ELEMENT artikel (lemmzeichenA, allgemein, einzelbedeutung+) >
```

Sind nicht die Daten für ein allgemeines, einsprachiges Wörterbuch die Grundlage der Modellierung, sondern andere Daten, muss diese Kopfstruktur angepasst werden; allerdings bleibt das grundsätzliche Prinzip bestehen. Wichtig ist, dass die Modellierung vom Wurzelement an so aufgegliedert wird, dass die Adressierungsbeziehungen so allgemein wie möglich formuliert werden können. In der obigen Modellierung kann beispielsweise vereinheitlicht werden, dass alle Angaben, die unter dem Element `allgemein` zusammengefasst sind, an das Lemmzeichen adressiert sind. Alle Angaben, die unter eine `einzelbedeutung` eingeordnet sind, sind zunächst an die Einzelbedeutung und erst im zweiten Schritt an das Lemmzeichen adressiert. Daher sollten beispielsweise Formangaben, die für die einzelnen Bedeutungen eines Lemmas verschieden sind, auch unter diesen Einzelbedeutungen eingeordnet werden. Dies bedeutet für ein Lemma wie „Moos“, dass die Angaben zur Pluralbildung grundsätzlich einzelbedeutungsbezogen angegeben werden müssen, da es zu „Moos“ im Sinne von ‘Geld’ keinen Plural gibt. Der Skopus ist also entscheidend: Unter allgemeinen Angaben sollen *nur* Angaben gruppiert werden, die sich auf *alle* Einzelbedeutungen beziehen, unter einzelbedeutungsbezogenen Angaben jeweils alle Angaben zu dieser Einzelbedeutung. Zusätze zu Angaben, die an die Angabe selbst adressiert sind, werden gesondert gekennzeichnet.

In gedruckten Wörterbüchern werden die Formangaben allerdings meist ausgelagert, d.h. neben die semantischen Angaben zu den einzelnen Bedeutungen gestellt. Dies ist eine Verdichtung der Artikeltexte zum Einsparen von Druckraum. Bei einer Modellierung, wie sie hier vorgeschlagen wird, muss daher für eine gedruckte Präsentation geprüft werden, welche Verdichtungsoperationen für die Darstellung im gedruckten Wörterbuch vorgenommen werden können. Dabei wird das Ergebnis nach den hier empfohlenen Richtlinien sicherlich nicht vollständig dem entsprechen, wie gedruckte Wörterbücher heute in der Regel verdichtet sind. Der Mehrwert einer solchen Modellierung für die Möglichkeiten einer elektronischen Publikation sind jedoch so ausschlaggebend, dass diese Herangehensweise als sinnvoll erachtet werden kann. Verdichtungsoperationen können dabei über die gleiche Technologie gemacht werden, mit der alle Präsentationen der Daten entwickelt werden: über XSL-Stylesheets (vgl. Abschnitt 4).

Die Benennung der einzelnen Knoten im Baum sollte so genau wie möglich sein. Sind zum Beispiel zu Verben Angaben_v zur Konjugation angegeben, sollten diese nicht nur unter dem Oberlement `grammatik`, sondern mit einem Element `konjugation` ausgezeichnet werden. Außerdem sollen die Benennungen möglichst sprechend sein, da die Lexikografen beim Erarbeiten der Artikel direkt mit diesen Benennungen im XML-Editor arbeiten. Die Benennungen der nicht-terminalen Knoten im Inhaltsstrukturenprogramm sollen daher so gewählt werden, dass gilt:

Leitsatz zur Benennung der nicht-terminalen Knoten des Inhaltsstrukturenprogramms

Das Element E enthält weitere Angaben_v zum Themengebiet x [aufgelöster Name des Elements].
x ist dabei so sprechend und genau wie möglich zu wählen.

Diese Richtlinie kann zu einer relativ umfangreichen Schachtelung von Elementen führen, bis als terminales Element die eigentliche Angabe_v folgt. Als ein mögliches Beispiel ist hier die Modellierung

von orthografischen Angaben_V zu sehen.⁹ Im Klammerelement `orthografie` wird dabei zunächst unterschieden in Angaben_V zur orthografisch korrekten Form (mit dem Klammerelement `ortho` auszuzeichnen) und Angaben_V zu möglichen Schreibvarianten (`ortho-variante`). Außerdem sind mögliche Zusätze zu dieser Angabe_V-Gruppe wie Belege, Hinweise oder Kommentare vorgesehen. Diese sind in der Entity `zusatz` zusammengefasst.

```
<!-- ===== Orthografie ===== -->
<!ELEMENT orthografie      (ortho, ortho-variante+, (%zusatz;)* ) >
```

Zu den Angaben_V der orthografisch korrekten Form gehört die orthografische Angabe_V, die zugehörige Silbenangabe_V und wiederum mögliche Zusätze.

```
<!-- = orthografisch korrekte Form = -->
<!ELEMENT ortho            (orthoA, silbenA, (%zusatz;)* ) >
```

Die Modellierung der Angaben_V zu Schreibvarianten sieht vergleichbar aus.

```
<!-- = orthografische Variante = -->
<!ELEMENT ortho-variante  ((ortho-varianteA, silbenA, (%zusatz;)* ) |
                           keine-angabe) >
```

Die orthografisch korrekte Form, mögliche orthografische Varianten und die zugehörigen Silbenangaben_V werden dabei in die mit „A“ gekennzeichneten Elemente eingetragen. Diese Schachtelung, in der sehr genau die einzelnen Angaben_V in Gruppen zusammengefasst werden, hat den Nachteil, dass sich die Lexikografen bei der Artikelbearbeitung im XML-Editor durch die Schachtelung ‘durchklicken’ müssen. Der Vorteil ist jedoch, dass die inhaltlichen Zusammenhänge sehr genau abgebildet werden. Dieser Vorteil überwiegt – auch nach der praktischen Erfahrung der Lexikografen in *ellexiko* – den Nachteil der umfangreichen Schachtelung.

Im oben gezeigten Beispiel ist ein weiterer Punkt zu sehen, der für die Modellierung des hierarchischen Aufbaus des Inhaltsstrukturenprogramms wichtig ist: In der Modellierung soll zwischen den Angaben_V unterschieden werden, die zu jedem Lemmazeichen angegeben werden können, und denen, bei denen dies nicht der Fall ist. Diese Unterscheidung findet sich auch bei Wiegand: „Es ist zu beachten, dass man bei den obligatorisch zu bearbeitenden Angabeklassen unterscheiden muss zwischen denen, deren Bearbeitung stets zu einer formulierten Angabe im Artikel führt, und denen, bei denen dies nicht der Fall ist. Die erstgenannten Klassen sowie ihre Elemente heißen *absolut obligatorisch*, die letztgenannten *relativ obligatorisch*.“ (Wiegand 1989, S. 455). Die Angabe_V der orthografisch korrekten Form ist somit eine absolut obligatorische Angabe_V. Relativ obligatorische Angaben_V sind dagegen solche Angaben_V, die nicht zu jedem Lemmazeichen bzw. jeder Einzelbedeutung gegeben werden können, wie z.B. orthografische Varianten oder Synonyme, die aber stets angegeben werden sollen, wenn es möglich ist. In der Modellierung sollen diese relativ obligatorischen Angaben_V von rein fakultativen Angaben_V folgendermaßen unterschieden werden: Die Angabe_V-Gruppe selbst wird in der DTD – wie die absolut obligatorischen Angaben_V – als obligatorisch definiert (vgl. die Modellierung des Elementes `orthografie` oben). Innerhalb dieser Angabe_V oder Angabe_V-Gruppe gibt es allerdings ein Ausweichelement, was beispielsweise `keine-angabe` benannt werden kann (vgl. die Modellierung des Elementes `ortho-variante` oben). Damit wird deutlich gemacht, dass diese Angabe_V oder Angabe_V-Gruppe immer bearbeitet werden soll und Lexikografen Stellung dazu nehmen müssen; anders als dies bei fakultativen Angaben_V der Fall ist. Diese Modellierungsrichtlinie gewährleistet eine inhaltlich angemessene strenge Modellierung.

Die Abbildung der Obligatorik kann in folgendem Leitsatz zusammengefasst werden:

⁹ Diese Modellierung ist angelehnt an die Modellierung orthografischer Angaben in *ellexiko* (siehe www.ellexiko.de und Haß (Hg.) 2005), ist jedoch auch vorstellbar für andere Projekte.

Leitsatz zur Abbildung der Obligatorik im Inhaltsstrukturenprogramm

Bei jedem Knoten des Inhaltsstrukturenprogramms muss entschieden werden, ob es sich bei der Angabe_v oder Angabe_v-Gruppe um absolut obligatorische, relativ obligatorische oder fakultative Angaben_v handelt. Absolut obligatorische Angaben_v und relativ obligatorische Angaben_v sind als obligatorisch zu modellieren (d.h. ohne Operator oder mit „+“-Operator). In relativ obligatorischen Angaben_v muss ein Ausweichelement vorgesehen werden, sodass die Angabe_v ausgewählt wird, aber ohne Inhalt bleiben kann. Fakultative Angaben_v sind als fakultativ zu modellieren (d.h. mit „*“- oder „?“-Operator).

Diese Grundsätze zur Modellierung des hierarchischen Aufbaus eines Inhaltsstrukturenprogramms können einen Eindruck davon vermitteln, wie die Modellierungsrichtlinien angelegt sind.

4. Folgen für die lexikografische Praxis – ein Beispiel

Eine Anwendung dieses Modellierungskonzeptes hat Auswirkungen in ganz unterschiedlichen Bereichen eines lexikografischen Prozesses: zum einen betrifft es die Erarbeitung der Daten, insbesondere die Unterstützung der Lexikografen in der Einhaltung der formalen Artikelstruktur, es hat Folgen für die redaktionelle Bearbeitung der Daten (beispielsweise können Abfragen nach verschiedensten Eigenschaften der lexikografischen Daten programmiert werden), die lexikografischen Daten sind sehr flexibel darstellbar, ohne dass sie auf Ebene der Datenbasis angetastet werden müssen und es können neue Formen des Zugriffs auf Ebene der Präsentation für potenzielle Benutzer geschaffen werden. Wie diese Auswirkungen im Einzelnen aussehen können, kann hier nicht umfassend ausgeführt werden. Deshalb soll beispielhaft ein Aspekt herausgegriffen werden: die Flexibilität in der Darstellung der entsprechend modellierten Daten.

Das standardbasierte Verfahren für die Darstellung von XML-Daten ist die „Extensible Stylesheet Language Family“.¹⁰ Zu dieser Sprachfamilie gehören die *Extensible Stylesheet Language* (vgl. XSL 2001), die *XSL Transformations* (vgl. XSLT 1999), die zu Transformationen von XML-Daten dienen und die *XML Path Language* (vgl. XPath 1999), eine Adressierungssprache, die es ermöglicht, Teile von XML-Dokumenten zu adressieren. In der Einleitung zur XSL-Spezifikation heißt es: „XSL is a language for expressing stylesheets. Given a class of arbitrarily structured XML [...] documents or data files, designers use an XSL stylesheet to express their intentions about how that structured content should be presented; that is, how the source content should be styled, laid out, and paginated onto some presentation medium, such as a window in a Web browser or a hand-held device, or a set of physical pages in a catalog, report, pamphlet, or book.“ (XSL). Angewendet auf lexikografische Prozesse und die Rahmenbedingungen des hier vorgestellten Modellierungskonzeptes, kann die Herstellung mehrerer Präsentationen aus einer lexikografischen Datenbasis wie in Abb. 3 gezeigt veranschaulicht werden.

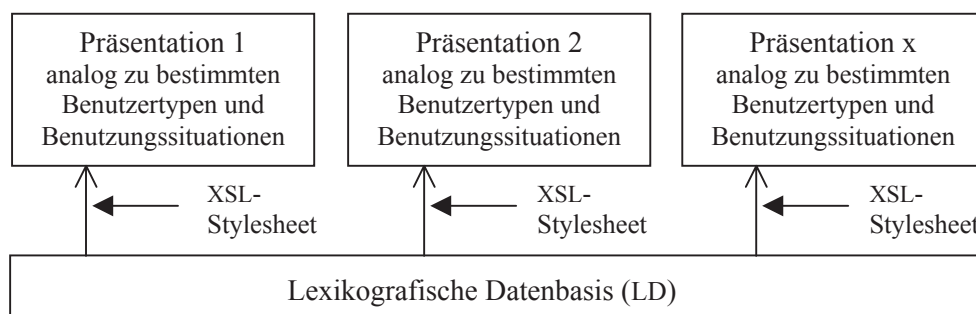


Abb. 3: Veranschaulichung der Herstellung mehrerer Präsentationen aus einer Datenbasis
 „→“ bedeutet soviel wie „aus der Datenbasis wird entwickelt“
 „←“ bedeutet soviel wie „mit dem Stylesheet wird spezifiziert“

¹⁰ Siehe www.w3c.org/Style/XSL/.

Dies soll beispielhaft demonstriert werden anhand eines (fiktiven) Ausschnitts aus einem Artikel zum Lemmazeichen „knickrig“, dessen einzelbedeutungsübergreifende Angaben_v analog zum Modellierungskonzept ausgezeichnet sind.

Zu einer kurzen Beschreibung des Inhaltsstrukturenprogramms: Nach der DTD ist vorgesehen, dass zusätzlich zur Lemmazeichen-Angabe_v separat eine orthografische Angabe_v mit Silbenangabe_v gemacht wird (als absolut obligatorische Angaben_v). Als relativ obligatorische Angaben_v sind Angaben_v zu orthografischen und/oder morphologischen Varianten vorgesehen. Diesen Angaben_v wird jeweils eine eigene Silbenangabe_v zugeordnet. Alle genannten Angaben_v können fakultativ durch Belege illustriert oder um Hinweise und Kommentare ergänzt werden. Zum Lemmazeichen „knickrig“ gibt es in der XML-Instanz keine orthografische, aber eine morphologische Variante, nämlich „knickerig“. Sowohl die orthografische Angabe_v, als auch die morphologische Variante sind hier durch einen Beleg illustriert.

Die Erstellung verschiedener Ansichten dieser XML-Instanz werden hier daher nur für die einzelbedeutungsübergreifenden Angaben_v gezeigt. Bei diesen verschiedenen Präsentationsmöglichkeiten geht es nicht um Details, sondern darum, ein Prinzip zu verdeutlichen. Es handelt sich dabei erstens um eine Ansicht, wie sie für Lexikografen am lexikografischen Arbeitsplatz zur Verfügung stehen soll, zweitens um eine Ansicht für eine Präsentation der Daten in einem – beispielsweise im Internet zugänglichen – Wortschatzinformationssystem und drittens um eine mögliche Darstellung im gedruckten Wörterbuch.

Für die Lexikografen am lexikografischen Arbeitsplatz ist es wichtig, die von ihnen erstellten Artikel ohne die XML-Auszeichnungen sehen zu können. Eine solche Ansicht bietet einen besseren Überblick über die Inhalte. Damit aber nach wie vor ersichtlich ist, mit welchen Elementnamen die Inhalte ausgezeichnet sind, stehen diese in aufgelöster Form vor den Inhalten.



Abb. 4: Ansicht der einzelbedeutungsübergreifenden Angaben aus „knickrig.xml“ am lexikografischen Arbeitsplatz

Im ersten Stylesheet ist festgelegt, dass alle Angaben in Arial 10 pt. dargestellt werden sollen, dass die aufgelösten Elementnamen, die für jedes Element einzeln im Stylesheet spezifiziert sind, immer in schwarz gezeigt werden sollen, wobei die Element- und Attributinhalt in blau dargestellt werden. Für die Lexikografen bietet diese Ansicht den Vorteil, im Prinzip alle Auszeichnungen, auch die Attribut-

werte, mit Inhalt zu sehen, und trotzdem einen besseren Überblick zu erhalten als wenn die Daten ausschließlich in der XML-Instanz anzuschauen sind. Dieses Stylesheet ist dabei natürlich auf alle Instanzen anzuwenden, die nach der Beispiel-DTD ausgezeichnet sind und gilt nicht nur für den Artikel „knickrig“. Die Darstellung wird durch eine Umwandlung der Daten nach HTML erreicht und ermöglicht damit das Ansehen der XML-Daten in jedem beliebigen Browser. Auch die folgenden Beispiele sind auf diese Art erstellt.

Für eine elektronische Präsentation, zum Beispiel für ein online zugängliches Wortschatzinformationssystem, könnte die Präsentation der gleichen XML-Instanz ganz anders aussehen. Für die folgende Ansicht ist dabei im Stylesheet festgelegt, dass als Überschrift die Lemmazeichen-Angabe in Arial 16 pt. fett dargestellt wird, darunter die orthografischen Angaben unter der Überschrift „Orthografie“ und die morphologischen Angaben entsprechend unter der Überschrift „Morphologie“. Die Inhalte sind hier für die potenziellen Benutzer teilweise anders benannt als in der Modellierung, z.B. wird die `orthoA` (orthografische Angabe) als „normgerechte Schreibung“ benannt. Die Belege sind über einen entsprechend beschrifteten Button durch einen Mausklick zu erreichen. Dies bewirkt eine übersichtliche Gestaltung des Artikels und ermöglicht es potenziellen Benutzern, nur bei Bedarf die Belege anzusehen. Klickt ein potenzieller Benutzer auf den Beleg-Button, öffnet sich der jeweilige Beleg.

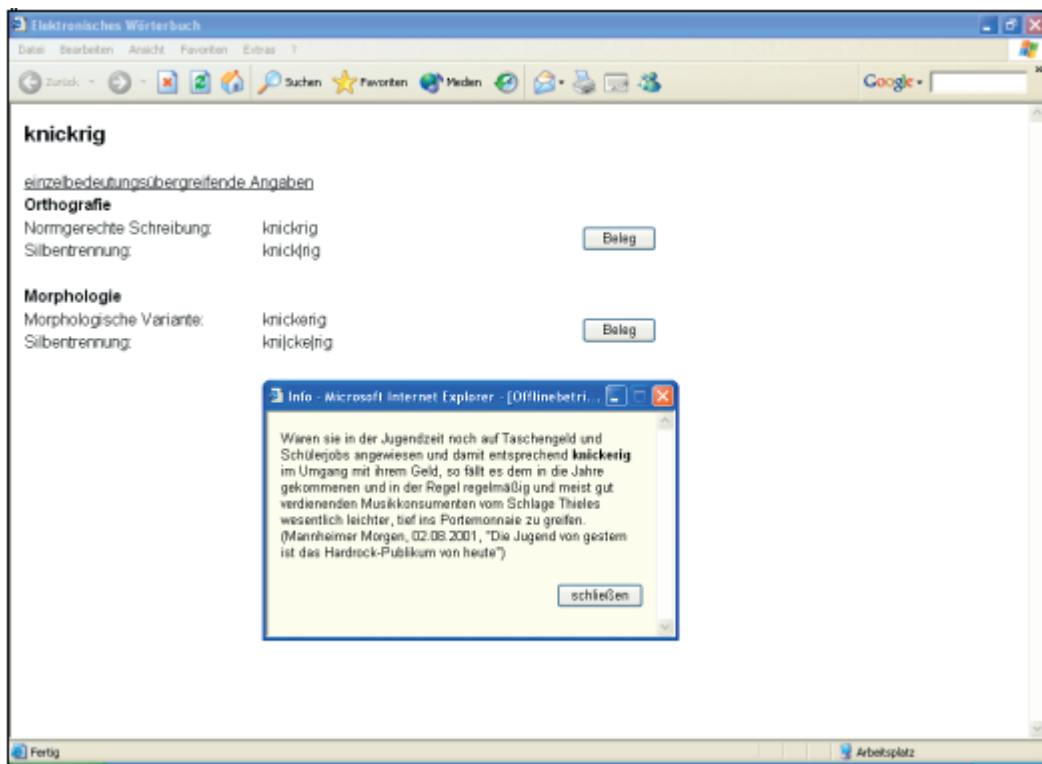


Abb. 5: Ansicht der einzelbedeutungsübergreifenden Angaben aus „knickrig.xml“ im Wortschatzinformationssystem (mit geöffnetem Belegfenster)

Selbstverständlich kann den Lexikografen auch diese Ansicht für ein Wortschatzinformationssystem am lexikografischen Arbeitsplatz zur Verfügung stehen, so wie auch die folgende Darstellung für ein gedrucktes Wörterbuch. Dies kann wichtig sein, wenn schon bei der Artikelerstellung die Präsentationsansicht mit einbezogen werden soll. So fordert beispielsweise Wiegand in seinem Szenario des „Mikrostrukturengenerators“, dass eine „druckgetreue Anzeige“ auf „speziellen Befehl hin möglich sein“ sollte (Wiegand 1998, S. 173).

Die dritte Ansicht zeigt nun eine mögliche Präsentation der Daten für ein gedrucktes Wörterbuch. Hier werden nur die Silbenangabe zur orthografischen Angabe und die Silbenangabe zur morphologischen

Variante, letztere mit einem vorangestellten „auch:“ dargestellt, alle anderen Angaben werden unterdrückt. Dies ist damit eine Form der verdichteten Darstellung.

knick|rig, auch: kni|cke|rig

Abb. 6: Ansicht der einzelbedeutungsübergreifenden Angaben aus „knicrig.xml“ im gedruckten Wörterbuch

Diese Darstellung wird hier zur Demonstration auch durch eine Umwandlung nach HTML erreicht. Für die professionelle Herstellung gedruckter Wörterbücher ist jedoch vermutlich die Anwendung von XSL-FO empfehlenswert, da daraus wiederum andere Formate wie PostScript oder PDF erzeugt werden können.

Wie anhand der Beispiele zu erkennen ist, ist die granulare, inhaltliche Auszeichnung der lexikografischen Daten analog zum Modellierungskonzept die Voraussetzung dafür, dass mit Hilfe von XSLT-Stylesheets so verschiedene Ansichten einer Instanz erstellt werden können. Denn im Stylesheet wird für jedes XML-Element eine Darstellungsregel festgelegt; stehen also mehrere Angaben_v in einem Element, sind diese Angaben_v nicht einzeln zugreifbar und damit auch nicht in der Weise unterschiedlich darzustellen. Die Auszeichnung lexikografischer Inhalte nach dem Modellierungskonzept zusammen mit der XSL-Technologie bietet damit eine sehr gute, standardbasierte Möglichkeit, lexikografische Inhalte flexibel zugreifbar und darstellbar zu machen. Denn alle gezeigten Ansichten greifen auf dieselbe XML-Instanz zu, ohne dass die Auszeichnungen und Inhalte angetastet werden. Damit ist eine zentrale Forderung der neueren Wörterbuchforschung in Bezug auf lexikografische Prozesse mit dem Aufbau des Modellierungskonzeptes erfüllt.

5. Abgrenzung zum Konzept der Mikrostrukturen

Auch im Bereich der printlexikografischen Forschungen wurden Methoden entwickelt, die für die Modellierungsrichtlinien als Basis dienen können. In diesem Zusammenhang wurde bereits mehrfach die Teiltheorie der Mikrostrukturen aus der Theorie lexikografischer Texte nach Wiegand (vgl. u.a. Wiegand 1989, Wiegand 1998) genannt, da hier eine formale Analyseverfahren für die Segmentation lexikografischer Daten entwickelt wurde (vgl. u.a. Wiegand 1989). Um deutlich zu machen, welche Teile der Analyseverfahren für die Modellierungsrichtlinien angewandt werden konnten und welche nicht, sollen zunächst die Grundzüge dieses Teilbereiches der Theorie lexikografischer Texte vorgestellt werden.

Für die Modellierungsrichtlinien ist die Methode der *funktional-positionalen Segmentation* von besonderem Interesse. Die korrekte Anwendung dieser Methode „führt stets zu funktionalen Textsegmenten sowie zu einer expliziten Kennzeichnung ihrer Positionen innerhalb der linearen Reihenfolge aller Textsegmente eines Wörterbuchartikels“ (Wiegand 1989, S. 438). Wichtig ist dabei, dass die Segmentation stufenweise erfolgt und auch so abgebildet wird, sodass „das, was relativ zu einem möglichst einheitlichen sprachtheoretischen Hintergrund zusammengehört, auch als zusammengehörig erfasst wird. Eine Strukturdarstellung, die alle unterschiedlichen Textelemente nur als unmittelbare Textkonstituenten des ganzen Artikels darstellt, wäre ganz unbefriedigend, weil sie zu wenig Einsichten in die Struktur vermittelt.“ (ebd., S. 437). Die *funktionalen Textsegmente* eines Wörterbuchartikels sind dabei folgendermaßen definiert: „Ein funktionales lexikografisches Textsegment ist ein Teil eines Wörterbuchartikels, bestehend aus einer Form und wenigstens einem genuinen Zweck, höchstens aber endlich vielen genuinen Zwecken, der (bzw. die) der Form in ihrer Ganzheit vom Lexikografen zugeordnet wurde(n).“ (ebd., S. 425). Unter der Form eines funktionalen Textsegmentes ist dabei seine spezifische grafische Gestalt zu verstehen. Der *genuine Zweck* „eines funktionalen Textsegmentes [...] besteht darin, dass es anhand bestimmter Eigenschaften dem potenziellen Benutzer dazu dient, diejenigen Benutzungsziele erreichen zu können, um deren Erreichung willen der Lexikograf das funktio-

nale Textsegment formuliert hat.“ (ebd., S. 426). Die wichtigsten Klassen von funktionalen Textsegmenten sind Angaben und Strukturanzeiger, wobei die Strukturanzeiger wiederum in typografische und nichttypografische zu unterteilen sind. Typografische Strukturanzeiger sind beispielsweise Kursivierungen oder Fettungen, nichttypografische Strukturanzeiger sind z.B. Kommata zwischen zwei Angaben. Der Begriff der Angabe ist dabei im Rahmen der Theorie sehr weit gefasst, sodass sein Gebrauch als Terminus nicht mehr unbedingt mit dem pragmatisch eingespielten innerhalb der lexikografischen Werkstattsprache übereinstimmt; z.B. sind auch lexikografische Hinweise oder Verweise als Angaben zu verstehen, die zu zwei besonderen Angabeklassen zusammengefasst werden können.

Die *Mikrostruktur* eines Wörterbuchartikels besteht ausschließlich aus Angaben im Gegensatz zur Artikelstruktur, zu deren Trägermenge jede Klasse von funktionalen Textsegmenten gehören darf, d.h. auch Strukturanzeiger. Im Rahmen der Mikrostruktur wird dabei der Wörterbuchartikel formal als größte Angabe aufgefasst, ohne die Unterschiede hinsichtlich der Textualität zu beachten. „Die Mikrostruktur eines standardisierten Wörterbuchartikels [...] ist daher diejenige Ordnungsstruktur, die festlegt, welche Angaben aus welchen Angabeklassen in welcher Reihenfolge der Artikel enthält. Die Mikrostruktur kann daher als besonders ausgezeichnete Teilstruktur der vollständigen Artikelstruktur [...] gelten, da sie – neben der Makrostruktur – die ‘Informationsverteilung’, d.h. die Verteilung der Angaben über den Wörterbuchgegenstandsbereich, regelt.“ (Wiegand 1989, S. 443). Die Mikrostruktur ist damit die zentrale Struktur für den inhaltlichen Aufbau eines Wörterbuchartikels, wogegen beispielsweise die Makrostruktur allgemein gesagt den Weg bis zu einem Wörterbuchartikel beschreibt. Das Ergebnis einer mikrostrukturellen Analyse eines einzelnen Wörterbuchartikels nach der Methode der funktional-positionalen Segmentation ist die *konkrete hierarchische Mikrostruktur*. Wird von den konkreten Angaben zu Klassen von Angaben abstrahiert, wird von der *abstrakten hierarchischen Mikrostruktur* gesprochen. „Der Unterschied zwischen konkreten und abstrakten Mikrostrukturen besteht darin, dass bei ersteren die Elemente der Trägermenge Angaben eines ganz bestimmten Artikels sind, aus dessen Analyse seine konkrete Mikrostruktur resultiert [...]. Die Elemente der Trägermenge von abstrakten Mikrostrukturen dagegen sind Mengen von Angaben (und zwar Klassen von Angaben gleichen Zwecks).“ (ebd., S. 440). Diese Differenzierung wurde für das Konzept der Inhaltsstrukturen übernommen. Die Benennung von Angaben und auch ihre Zuordnung zu Angabeklassen werden – im Konzept der Mikrostrukturen wie in dem der Inhaltsstrukturen – aus ihrem genuinen Zweck abgeleitet, so wie er oben bestimmt wurde. „Sucht man nach einer Benennung für diese Klasse, also nach einem Klassennamen (einem einstelligen Prädikat), dann ist es günstig, das Klassifikationskriterium, also hier den Zweck Z, zu berücksichtigen.“ (ebd., S. 415).

Zur Darstellung des Ergebnisses der stufenweisen funktional-positionalen Segmentation eines Wörterbuchartikels bei gleichzeitiger Strukturierung und Klassifizierung eignen sich nach Wiegand die Mittel der Graphentheorie, genauer die Darstellung als Baumgraphen. Dabei sind die Baumgraphen so nach Konventionen erstellt, dass sich aus ihnen „drei Arten von Informationen über die Struktur eines Wörterbuchartikels erschließen“ lassen: „(i) Informationen über die Zugehörigkeit einer Angabe zu einer Klasse von Angaben von gleichem genuinem Zweck, (ii) Informationen über die Teil-Ganzes-Beziehungen, in denen die Angaben zueinander und zum Wörterbuchartikel stehen und (iii) Informationen über die lineare links-rechts-Ordnung der Angaben.“ (Wiegand 1989, S. 439f.). Dabei können abstrakte und konkrete Strukturen durch ihr Verhältnis der Isomorphie in einem Strukturgraphen dargestellt werden. „Der Isomorphismus – und auf Ebene der Strukturdarstellung die formalen Eigenschaften von Baumgraphen – sind die Voraussetzung dafür, dass eine abstrakte hierarchische zusammen mit der isomorphen konkreten hierarchischen Textkonstituentenstruktur eines Artikels in *einem* Strukturgraphen dargestellt werden können, in dem zwei Baumgraphen kombiniert sind. [...] Bei ihrer Präsentation werden von der konkreten Textkonstituentenstruktur lediglich die terminalen Elemente explizit berücksichtigt, sodass nur die präzedentive Textkonstituentenstruktur direkt erkennbar ist; alle nichtterminalen konkreten Textkonstituenten lassen sich aufgrund des Isomorphismus erschließen.“ (ebd., S. 121).

Ergebnis dieser Methode zur mikrostrukturellen Analyse ist also eine vollständige Segmentierung der Daten in einem Wörterbuchartikel in Angaben mit mindestens einem genuinen Zweck. Gleichzeitig wird die lineare Anordnung der Angaben im gedruckten Wörterbuchartikel abgebildet (vgl. Abb. 7).

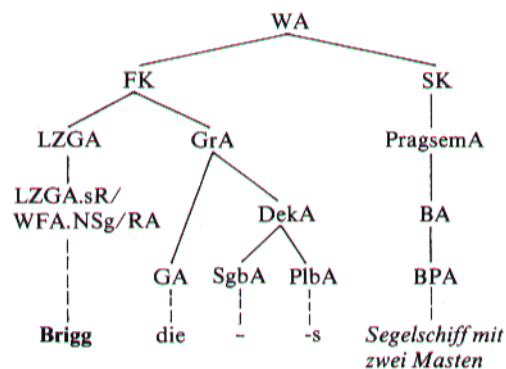


Abb. 38 a.21: Darstellung des Ergebnisses einer funktional-positionalen Segmentierung: für Angaben mit Klassifizierung und Strukturierung am Beispiel von wa_{19} mittels eines Baumgraphen; $x \text{---} y = x \in y$; (hierarchische Mikrostruktur von wa_{19} , vgl. 4.2.2.).

Abb. 7: Beispiel für die Darstellung des Ergebnisses einer funktional-positionalen Segmentierung nach Wiegand (Wiegand 1989, S. 439)

Die Analysemethode der funktional-positionalen Segmentierung wurde anhand von Printwörterbüchern entwickelt und legt die Artikel im Druckraum zu Grunde. Deshalb galt es zu prüfen, ob sie für die Festlegung von Richtlinien für die Modellierung lexikografischer Daten auf Ebene der Datenbasis nutzbar gemacht werden kann, die eine strikte Trennung von Datenbasis und Präsentation voraussetzt. Außerdem ist der „Zweck einer Methode zur Segmentierung von Wörterbuchartikeln [...] u.a. darin zu sehen, intuitives Wissen über Teile von Artikeln und ihre Zusammenhänge in ein möglichst explizites und schriftlich dargebotenes (also exteriorisiertes) Wissen zu überführen.“ (Wiegand 1989, S. 437). Das Ziel ist dabei eine „präzise und nachvollziehbare Segmentierung von Wörterbuchartikeln“ zu leisten, die eine der „elementaren Voraussetzungen für eine metalexikografische Textdeskription“ ist (Wiegand 1989, S. 438). Mit dem Modellierungskonzept wird dagegen das Ziel verfolgt, theoretisch fundierte Richtlinien für die Modellierung einer lexikografischen Datenbasis festzulegen, die praktisch im lexikografischen Prozess anzuwenden sind. Der formalisierte Ansatz der mikrostrukturellen Analyse ist jedoch viel versprechend für ein fundiertes Konzept zur Modellierung der hierarchischen Inhaltsstruktur. Denn durch formale Methoden werden Regeln für eine nachvollziehbare Modellierung festgelegt. Dies führt wiederum zu einer konsistenten Auszeichnung der lexikografischen Inhalte, die ihrerseits eine wichtige Voraussetzung für deren automatisierte Handhabung ist.

Daher wurden bestimmte Aspekte der Methode, wie ein Teil der funktionalen Segmentierung lexikografischer Daten, für das Konzept der Inhaltsstrukturen übernommen. Auch eingeführte Termini wie die des genuinen Zweckes, eines funktionalen Textsegments und weitere Aspekte konnten fruchtbar angewendet werden. Eine direkte Übertragung der Methode der funktional-positionalen Segmentierung auf die Modellierung lexikografischer Daten ist jedoch nicht möglich. Denn dann könnte weder der Unterschiedlichkeit der Gegenstandsbereiche und der Handlungsziele Rechnung getragen werden und noch wäre eine strikte Trennung von Datenbasis und Präsentation möglich, da das Konzept der Mikrostrukturen eindeutig der Ebene der Präsentation zuzurechnen ist.¹¹

¹¹ Deshalb kann auch nicht – zumindest nicht im Rahmen dieses Modellierungskonzeptes – die Methode der funktional-positionalen Segmentierung direkt auf die Modellierung übertragen werden. Diesen Weg schlägt Kammerer jedoch in seinem Aufsatz „XML-getaggte Wörterbuchartikel“ (Kammerer 2001) vor. „Liegen die im Wörterbuch vorgesehenen abstrakten hierarchischen Mikrostrukturen erst einmal fest, kann von diesen jeweils systematisch zu einer DTD über-

Die unterschiedliche Ausgangsbasis – bei der Modellierung die lexikografische Datenbasis, bei der Theorie lexikografischer Texte einzelne Wörterbücher – führt darüber hinaus zu einer verschiedenen Ordnung in Teilbereiche. Dies soll an einem Beispiel verdeutlicht werden: Im Inhaltsstrukturenprogramm für ein bestimmtes lexikografisches Projekt ist vorgesehen, dass – falls es sich beim Lemmazeichen um ein starkes Verb handelt – innerhalb der Grammatik die Stammformen in der 3. Person Singular angegeben werden sollen. Wird dann beispielsweise der Artikel zu „gehen“ verfasst, wird die Form „ging“ ein Element der konkreten hierarchischen Inhaltsstruktur des Lemmazeichens „gehen“ sein. Wird für die Präsentation im gedruckten Wörterbuch entschieden, dass alle Stammformen starker Verben als Verweisartikel angelegt werden sollen, dann kann aus dieser Modellierung ein Verweisartikel z. B. in der Form „ging: 3.P.Sg.Imp. → gehen“ automatisch generiert werden. Auf Ebene des gedruckten Wörterbuchs handelt es sich dann um zwei Artikel mit jeweils einer konkreten hierarchischen Mikrostruktur, auf Ebene der lexikografischen Datenbasis ist es jedoch nur ein Lemmazeichen mit einer zugehörigen konkreten hierarchischen Inhaltsstruktur.

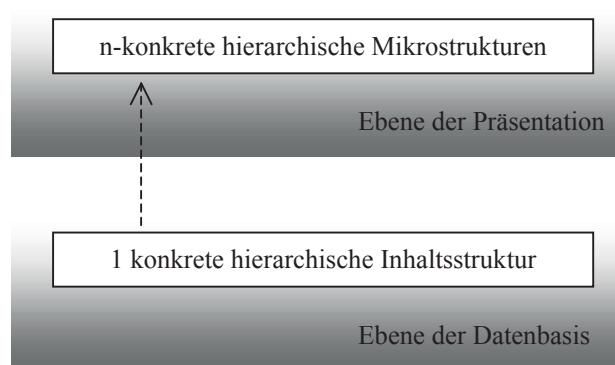


Abb. 8: Veranschaulichung des Verhältnisses von hierarchischen Inhaltsstrukturen (auf Ebene der Datenbasis) und hierarchischen Mikrostrukturen (auf Ebene der Präsentation) ($n \geq 1$)

„ ----> “ bedeutet soviel wie „aus der Datenbasis wird entwickelt“

Umso deutlicher dürfte damit sein, dass genau zu prüfen ist, welche Teile aus dem Mikrostrukturenkonzept direkt zu übernehmen sind. Wie in vielen anderen Bereichen geht es daher auch bei einer relativ neuen Aufgabe wie der Modellierung einer lexikografischen Datenbasis darum, bereits Gedachtes zur Kenntnis zu nehmen, wenn möglich aufzunehmen und davon ausgehend Neues zu entwickeln. Denn vieles „Gescheite“ ist – wie schon Goethe wusste – bereits gedacht worden.

gegangen werden.“ (Kammerer 2001, S. 250). Damit nicht für jede abstrakte hierarchische Mikrostruktur eine separate DTD zu definieren sei, schlägt er als Grundlage die „Maximaldatenstruktur“ vor. „Die Maximaldatenstruktur kann – grob gesehen – als die Überführung von mehreren abstrakten hierarchischen Mikrostrukturen in eine einzige Struktur angesehen werden. Diese Struktur kann weitere Knoten besitzen, die nicht Elemente der Trägermenge der zusammengeführten abstrakten hierarchischen Mikrostrukturen sind. Diese Knoten können beispielsweise dazu genutzt werden, um technische oder administrative Daten zu hinterlegen.“ (ebd., S. 257). Sinn und Zweck dieser Maximaldatenstruktur ist es, „die Anzahl der in einem Wörterbuch potenziell möglichen abstrakten hierarchischen Mikrostrukturen auf ein handhabbares Maß [...] zu reduzieren“ (ebd.). Abgesehen davon, dass viele der in diesem Beitrag vorgeschlagenen Grundsätze nicht den Anforderungen des hier in Teilen beschriebenen Modellierungskonzeptes genügen und dass kein wechselseitiger Prozess zwischen Modellierung und Festlegung der potenziellen Angaben im Artikel angedacht ist, müssen die Ansätze grundsätzlich gegeneinander abgegrenzt werden. Bei Kammerer werden die abstrakten hierarchischen Mikrostrukturen zur Grundlage der Modellierung gemacht. Eine strikte Trennung von Datenbasis und Präsentation ist damit aufgehoben, da die Mikrostruktur gerade auf der Ebene der Präsentation anzusiedeln ist. Ein solcher direkter Weg vom Mikrostrukturenprogramm zur Modellierung kommt hier daher nicht in Frage.

6. Schlussbemerkung

Dieser Aufsatz zeigt einen Ausschnitt aus einem Konzept für die maßgeschneiderte Modellierung lexikografischer Daten, das projektübergreifend anzuwenden ist. Das Konzept der Inhaltsstrukturen wird im Gesamtkonzept flankiert von einem Konzept zur Modellierung der Vernetzung von lexikografischen Daten (vgl. Müller-Spitzer 2005b). Eine Modellierung, wie sie hier vorgeschlagen wird, benötigt Zeit und muss daher in einem lexikografischen Prozess als ein wichtiger Teil mit eingeplant werden. Sicherlich ist dies eine Änderung gegenüber lexikografischen Prozessen, die allein auf ein gedrucktes Wörterbuch angelegt waren. Hier konnten allein die lexikografischen Inhalte im Zentrum stehen, da sich der Prozess der Buchherstellung über Jahrzehnte bzw. Jahrhunderte kulturell herausgebildet hat und daher (meist) in routinierter Weise abläuft. Heute besteht dagegen in der Regel die Notwendigkeit, sich über die Ausgestaltung der Modellierung lexikografischer Daten sorgfältige Gedanken zu machen. Denn für elektronische Wörterbücher ist die Art der Modellierung absolut entscheidend für die Funktionalität, die man potenziellen Benutzern später anbieten kann. Insofern muss sorgfältig abgewägt werden, ob es sinnvoll ist, eine einfachere Modellierung zwar schnell zu entwickeln, aber dafür langfristig sehr viel weniger Flexibilität hinsichtlich der Darstellung und der Zugriffsmöglichkeiten zu haben. Auch in medienneutral konzipierten lexikografischen Prozessen oder in computerlexikografischen Prozessen (vgl. Müller-Spitzer 2003) stehen daher die lexikografischen Inhalte im Zentrum – es ist nur elementar wichtig, wie sie ‘verpackt’ und aufbereitet werden, damit man die Vorteile der neuen Medien auch wirklich ausnutzen kann.

7. Literatur

- Bosak, Jon (1998): Media-Independent Publishing: Four Myths about XML. Internet: www.ibiblio.org/pub/sun-info/standards/xml/why/4myths.htm (Stand: April 2006; 4 Seiten).
- Engelberg, Stefan/Lemnitzer, Lothar (2001): Lexikographie und Wörterbuchbenutzung. (= Stauffenburg Einführungen 14). Tübingen.
- Freisler, Stefan (1994): Hypertext – Eine Begriffsbestimmung. In: Deutsche Sprache 1, S. 19-50.
- Geeb, Franziskus (2001): leXeML – Vorschlag und Diskussion einer (meta-)lexikographischen Auszeichnungssprache. In: Sprache und Datenverarbeitung 2, S. 27-61.
- Gennusa, Pamela L. (1999): Evolution and use of generic markup languages. In: Möhr, Wiebke/Schmidt, Ingrid (Hg.): SGML und XML. Anwendungen und Perspektiven. Berlin/Heidelberg/New York. S. 27-50
- Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York.
- Hedtstück, Ulrich (2003): Einführung in die theoretische Informatik. Formale Sprachen und Automatentheorie. 2. Aufl. München/Wien.
- 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“. (= Lexicographica. Series maior 45). Tübingen.
- Kammerer, Matthias (2001): XML-getaggte Wörterbuchartikel. Ein Bericht aus der Praxis des Wörterbuchs zur Lexikographie und Wörterbuchforschung. In: Lexicographica 17, S. 249-301.
- Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika Storrer (Hg.) (2001): Chancen und Perspektiven computergestützter Lexikographie. (= Lexicographica. Series Maior 107). Tübingen.
- Müller-Spitzer, Carolin (2003): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: Lexicographica 19, S. 140-168.
- Müller-Spitzer, Carolin (2005a): Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess. In: Ulrike Haß (Hg.): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz, Berlin/New York. (= Schriften des Instituts für Deutsche Sprache 12). S. 21-54.
- Müller-Spitzer, Carolin (2005b): Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis. Unveröff. Diss. Heidelberg (ersch.).

- O'Meara, Dick/Reeder, Beverly (2005): meta. In: *whatis.com*. Internet: http://searchsqlserver.techtarget.com/sDefinition/0,,sid87_gci212555,00.html (Stand: April 2006).
- Schmidt, Ingrid/Müller, Carolin (2001): Entwicklung eines lexikographischen Modells: Ein neuer Ansatz. In: Lemberg/Schröder/Storrer (Hg.), S. 29-52.
- Schryver, Gilles-Maurice de (2003): Lexicographer's Dreams in the Electronic-Dictionary Age. In: *International Journal of Lexicography* 16/2, S. 143-199.
- Sperberg-McQueen, Carl Michael/Burnard, Lou (1995): The Design of the TEI Encoding Scheme. In: Ide, Nancy/Véronis, Jean (Hg.): *Text Encoding Initiative. Background and Context*. Dordrecht. S. 17-39.
- Storrer, Angelika (1996): Metalexikographische Methoden in der Computerlexikographie. In: Wiegand, Herbert Ernst (Hg.): *Wörterbücher in der Diskussion II. (= Lexicographica. Series Maior 70)*. Tübingen. S. 239-255.
- Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg/Schröder/Storrer (Hg.), S. 53-69.
- Wiegand, Herbert Ernst (1988): Wörterbuchartikel als Text. In: Harras, Gisela (Hg.): *Das Wörterbuch. Artikel und Verweisstrukturen. (= Jahrbuch des Instituts für Deutsche Sprache 1987)*. Düsseldorf. S. 30-120.
- Wiegand, Herbert Ernst (1989): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilbd. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1). Berlin/New York. S. 409-462.
- Wiegand, Herbert Ernst (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd. Berlin/New York.
- Wiegand, Herbert Ernst (2002a): Altes und Neues zur Mediostruktur in Printwörterbüchern. In: *Lexicographica* 18, S. 168-252.
- Wiegand, Herbert Ernst (2002b): Adressierung in zweisprachigen Printwörterbüchern. In: Wiegand, Herbert Ernst (Hg.): *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*. Hildesheim/Zürich/New York. S. 111-175.
- XPath (1999) = Clark, James/deRose, Steve (Hg.) (1999): *XML Path Language (XPath) Version 1.0*. World Wide Web Consortium, W3C Recommendation 16. November 1999. Internet: www.w3.org/TR/xpath. Deutsche, kommentierte Übersetzung: www.edition-w3c.de/TR/1999/REC-xpath-19991116 (Stand: April 2006).
- XSL (2001) = Adler, Sharon/Berglund, Anders/Caruso, Jeff et al. (Hg.) (2001): *Extensible Stylesheet Language Version 1.0*. World Wide Web Consortium, W3C Recommendation 15. October 2001. Internet: www.w3.org/TR/xsl/ (Stand: April 2006).
- XSLT (1999) = Clark, James (Hg.) (1999): *XSL Transformations (XSLT) Version 1.0*. W3C Recommendation 16. November 1999. Internet: www.w3.org/TR/xslt. Deutsche Übersetzung: www.edition-w3c.de/TR/1999/REC-xslt-19991116 (Stand: April 2006).