

*Carolin Müller-Spitzer**

Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung

Abstract

Lexicographic data are normally linked with each other in a complex manner. Especially, within the electronic lexicographic context, the following issues are addressed: How to encode these cross-reference structures so that both the lexicographers' editorial work with the linking-up is easy to handle and the options of the presentation are adequately flexible. The objective of this paper is to elucidate the presentation of an XML-modelling of cross-reference structures as part of a complete modelling concept. Thereby, the modelling potential of the XML-connected standard XLink and a new lexicographic concept will be brought together with cross-project guidelines for the modelling of link-structures.

1. Aufgabe

Lexikografische Produkte haben in der Regel eine ausgeprägte Verweisstruktur. Dabei sind es bei gedruckten Wörterbüchern vor allem zwei Umstände, die zu dieser Verweisstruktur führen: Zum einen orientiert sich die Anordnung der Elemente (der äußeren Zugriffsstruktur) auch bei semasiologischen Wörterbüchern an ihrer grafischen Form, d. h. die inhaltlichen Beziehungen über einzelne Artikel hinweg lassen sich nur über Verweise herstellen, und zweitens führen die engen räumlichen Begrenzungen bei Printwörterbüchern dazu, dass Informationen immer nur an einer Stelle genannt sind, auch wenn sie an mehreren relevant wären. (Vgl. Engelberg/Lemnitzer 2001, 154) Dabei gestaltet sich die manuelle Konsistenzkontrolle bei Verweisen immer schwieriger, je präziser das Verweisziel angegeben wird. „Präzision der Verweisziele und Konsistenz des Verweissystems sind also Anforderungen an ein Wörterbuch, denen bei manueller Bearbeitung nur sehr schwer

* *Carolin Müller-Spitzer*
Institut für Deutsche Sprache
Postfach 10 16 21
D-68016 Mannheim
mueller-spitzer@gmx.de

zugleich entsprochen werden kann.“ (Blumenthal et al. 1988, 365) Die Computerunterstützung im lexikografischen Prozess kann da eine wesentliche Verbesserung bewirken. Das Ziel sollte dabei sein, auf Ebene der lexikografischen Datenbasis die Voraussetzungen dafür zu schaffen, dass zum einen die Konsistenz der Daten gewährleistet wird und dass zum anderen die Verweise möglichst präzise den Verweisursprung und das Verweisziel verknüpfen. Wie diese Voraussetzungen im Rahmen einer XML-basierten Modellierung im lexikografischen Prozess geschaffen werden können, soll im Folgenden erläutert werden.

2. Rahmenbedingungen

Es wird hier von einer XML-basierten Modellierung ausgegangen. Das heißt nicht, dass die grundsätzlichen Überlegungen nicht auch auf andere Formate zu übertragen sind, doch wenn das dargestellte Konzept nicht zu allgemein bleiben sollen, muss eine bestimmte Modellierungssprache ausgewählt werden. Das ist hier allgemein XML und speziell für die Modellierung der Vernetzungen der XML-verwandte (Industrie-)Standard XLink. (Vgl. Behme/Mintert 2000, XLink; vgl. auch Abschnitt 4) Dieses Konzept zur Modellierung von Vernetzungsstrukturen ist Teil einer neu entwickelten, allgemeinen Konzeption zur Modellierung lexikografischer Daten. (Vgl. Müller-Spitzer 2006a) Deshalb sind vorweg einige Rahmenbedingungen und vorgenommene Grundlegungen zu erläutern.

▪ Ebene der Datenbasis vs. Ebene der Präsentation

In allen lexikografischen Prozessen, in denen lexikografische Daten nicht allein zielgerichtet auf ein Produkt erarbeitet werden, sollte immer zwischen der Ebene der Datenbasis und der der Präsentation(en) unterschieden werden. (Vgl. auch Müller-Spitzer 2003, 164ff.) Die hier dargestellte Modellierung der Vernetzungsstrukturen bezieht sich dabei auf die Ebene der Datenbasis. Hier befinden sich alle lexikografischen Daten inkl. aller Daten für mögliche Umtexte, Benutzerhinweise, interne Kommentare usw. Für eine Präsentation werden von dieser Datenmenge eine Untermenge definiert und – falls die Daten auf Ebene der Datenbasis ohne Gestaltungsaspekte kodiert sind – bestimmte typografische und weitere gestalterische Eigenschaften festgelegt. Falls die Da-

ten im XML-Format in der Datenbasis abgelegt sind, werden die Präsentationen in der Regel über XSLT-Stylesheets¹ spezifiziert.

▪ **Inhaltsorientierte Datenauszeichnung**

Bevor auf die Vernetzung der lexikografischen Daten eingegangen werden kann, muss kurz ausgeführt werden, wie die lexikografischen Inhalte kodiert werden sollen. Allgemein wird zwischen *inhaltsorientierter* und *layoutorientierter Datenauszeichnung* unterschieden. Hier wird davon ausgegangen, dass die lexikografischen Daten inhaltsorientiert ausgezeichnet werden. Dies soll in diesem Zusammenhang bedeuten, dass ein Mensch seine Kenntnisse über den Inhalt der Daten so maschinenlesbar kodiert, dass der Computer diese Kodierung in datenverarbeitenden Prozessen ausnutzen kann. Der Rechner kann dabei nur rein zeichenorientiert arbeiten; die Zeichenketten der Tags sind aber so gewählt, dass sie Aufschluss über den Inhalt der umschlossenen Zeichenketten geben. Mit Weber kann diese Art der Datenauszeichnung auch als ein ‚Explizit-Machen‘ des inhaltlichen Gehalts der Daten bezeichnet werden: „‚Explizit machen‘ ist zu verstehen als: einem menschlichen Interpretanden IM oder einem künstlichen Simulationssystem IS semantische Interpretationen ermöglichen. Eine gelungene Interpretation kann mit Ausdrücken wie ‚(Wieder)erkennen‘, ‚Realisieren‘, ‚Verstehen‘ bei Menschen und ‚kommunikativ adäquate Verwendung‘ bei künstlichen Systemen belegt werden.“ (Weber 1999, 258)

▪ **Konzeptuelle Inhaltsmodellierung**

Es können zwei Wege zu einer inhaltsorientierten Datenauszeichnung unterschieden werden: die *layoutbasierte Inhaltsmodellierung* und die *konzeptuelle Inhaltsmodellierung*. Die erst genannte Art der Modellierung kann z. B. bei der Retrodigitalisierung gedruckter Wörterbücher gegeben sein, bei der beispielsweise die Satzbänder geparkt werden. (Vgl. u. a. entsprechende Beiträge aus Lemberg et al. 2001 oder Raymond/Tompa 1988). Der Weg geht dabei über zunächst layoutorientiert ausgezeichnete Daten hin zu einer inhaltsorientierten Auszeichnung.

Ein anderer Weg ist es, die Daten von Anfang an von menschlicher Hand, d. h. auch intellektuell reflektiert, inhaltsorientiert auszuzeichnen. Eine Strukturentwicklung kann dann auf der Konzeption der Wörterbuchartikel aufbauen. Die Modellierung sollte darüber hinaus selbst

¹ S. <www.w3.org/Style/XSL/>.

ein Konzept aufweisen, d. h. bestimmten Prinzipien folgen, genaue Regeln des Aufbaus haben etc. Eine solche Modellierung soll *konzeptuelle Inhaltsmodellierung*² heißen. (S. auch Müller-Spitzer 2006a und 2006b) Im Rahmen des hier vorgeschlagenen Modellierungskonzeptes wird eine solche konzeptuelle Inhaltsmodellierung vorausgesetzt.

▪ Inhaltsstrukturen

Auch wenn eine konzeptuelle Inhaltsmodellierung schon bestimmte Grundsätze für eine XML-DTD- oder XML-Schema-Entwicklung beinhaltet, muss trotzdem geklärt werden, welche Eigenschaft der lexikografischen Daten genau Gegenstand der Modellierung sein soll. Dabei sind im Rahmen dieses Modellierungskonzeptes zwei Aspekte entscheidend: Auf Ebene der Datenbasis soll die Voraussetzung dafür geschaffen werden, dass die Daten erstens flexibel darstellbar – z.B. adaptiv zu Benutzertypen und Benutzungssituationen – und zweitens gezielt zugreifbar sind. Dies verlangt eine sehr granulare Modellierung, die gestalterische Eigenschaften der Daten außen vor lässt. Für die Entwicklung von Richtlinien für eine solche Modellierung lexikografischer Daten wurde das Konzept der *Inhaltsstruktur* entwickelt. (Vgl. Müller-Spitzer 2006a und 2006b) Die Inhaltsstruktur eines Artikels (in der Datenbasis) beinhaltet als Trägermenge alle Vor-Angaben, die inhaltlich einem Lemmazeichen zugeordnet sind. Sie ist damit allgemein gesagt von Ebene der Datenbasis aus gesehen das Pendant zur Mikrostruktur auf Ebene der Präsentation. Von Vor-Angaben – oder kurz Angaben_v – wird deshalb gesprochen, weil Angaben auf Ebene der Datenbasis noch nicht über eine bestimmte (Präsentations-)Gestalt verfügen.³ Die granulare inhaltsorientierte Auszeichnung der lexikografischen Daten ist da-

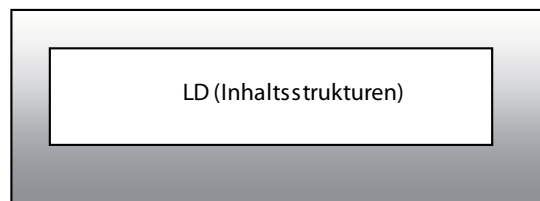
² Hier ist die Rede von *konzeptueller* statt von *konzeptioneller* Inhaltsmodellierung (was analog zur üblichen Bezeichnung konzeptionelle Datenmodellierung wäre), da konzeptionell laut DFWB bedeutet: „die Konzeption betreffend“, konzeptuell dagegen: „ein Konzept aufweisend“. (DFWB 1994) Da eine Inhaltsmodellierung nach dem hier zu Grunde gelegten Modellierungskonzept selbst ein Konzept aufweisen sollte, ist daher die Bezeichnung konzeptuelle Inhaltsmodellierung treffender.

³ Der Terminus *Angabe* ist in der Theorie lexikografischer Texte nach Wiegand wohlbestimmt, und zwar als funktionales Textsegment, das aus einer Angabeform und wenigstens einem genuinen Zweck besteht. (Wiegand 1989, 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.

bei eine zentrale Voraussetzung für die Modellierung der Vernetzungsstrukturen, wie sie hier vorgeschlagen wird.

▪ Lexikografische Datenbasis – Lexikografische Datenbasis (Inhaltsstrukturen)

Gegenstand der hierarchischen Inhaltsstruktur sind allgemein gesagt die lexikografischen Daten in der Datenbasis, die in einem einzelnen Wörterbuch oder Wortschatzinformationssystem in Wörterbuchartikeln präsentiert werden, d. h. eine Teilmenge der gesamten lexikografischen Datenbasis. Diese Teilmenge soll als *Lexikografische Datenbasis (Inhaltsstrukturen)*, abgekürzt *LD (IS)*, benannt werden. Trägermenge dieser Teilmenge der lexikografischen Datenbasis sind die lexikografischen Daten, die zur Trägermenge aller konkreten hierarchischen Inhaltsstrukturen der Datenbasis gehören. Weitere Daten, die zwar in der lexikografischen Datenbasis sind, aber nicht zur hierarchisch strukturierten LD (IS) gehören, können z.B. Abbildungen sein.



Lexikografische Datenbasis (LD)

Abb. 1: Veranschaulichung zum Verhältnis der Lexikografischen Datenbasis und der Lexikografischen Datenbasis (Inhaltsstrukturen)

Zum Abschluss dieser Erläuterung von Rahmenbedingungen soll kurz auf die Diskussion in der Wörterbuchforschung um Wörterbuchfunktionen eingegangen werden. Im Moment herrscht ein Forschungsstreit zwischen Wiegand und seiner „allgemeinen Theorie der Lexikografie“ (vgl. u.a. Wiegand 1998 und 2001) und der modernen lexikografischen Funktionslehre, die v.a. von Bergenholtz und Tarp entwickelt wurde. (Vgl. u.a. Bergenholtz/Tarp 2002 und 2003) Ohne dass zu diesem Forschungsstreit Position gezogen werden soll, ist in Bezug auf die Modellierung von lexikografischen Daten festzuhalten, dass sie meiner Meinung nach unabhängig von diesen unterschiedlichen Positionen vor-

zunehmen ist. Zwar ist es nach der modernen Funktionslehre so, dass „[...] functions are the basic elements of lexicographic theory and practice and constitute the leading principle of all dictionaries. Everything in a dictionary, absolutely everything, is to greater or smaller extent influenced by its respective functions. Neither the content nor the form of a dictionary can be conceived without taking the functions into account.“ (Bergenholtz/Tarp 2003, 177, s. auch Bergenholtz/Tarp 2002) In einem Szenario von lexikografischem Prozess, wie es hier vorgesehen ist, ist es jedoch denkbar, dass entweder nach Wiegands Paradigma oder dem der modernen lexikografischen Funktionslehre die Daten ausgesucht werden, diese dann strikt nach inhaltlichen Gesichtspunkten – d.h. auch unabhängig von anvisierten Funktionen – modelliert werden, um dann beispielsweise abhängig von anvisierten Wörterbuchfunktionen, Benutzern- und Benutzungssituationen präsentiert zu werden. Dabei werden die inhaltlich klassifizierten Daten den jeweils relevanten Benutzungssituationen zugeordnet. Insofern ist das Modellierungskonzept sowohl für die Inhalts- wie für die Vernetzungsstrukturen unabhängig von diesem Theorienstreit. Allerdings bietet sich zu einer Fundierung der Richtlinien für eine strikt inhaltsorientierte Auszeichnung der lexikografischen Daten die Theorie Wiegands an, da diese sich in detaillierter Weise mit den formalen Eigenschaften dieser Daten auseinandersetzt. Dies hat jedoch keine einschränkenden Auswirkungen auf den weiteren Umgang mit den Daten.

3. Inhaltliches Konzept

Zunächst einmal gilt es zu klären, was im Rahmen des Modellierungskonzeptes zum Gegenstandsbereich der Vernetzungsstrukturen (im Gegensatz zu dem der Inhaltsstrukturen) gehört. Dafür soll geprüft werden, ob Arbeiten aus der Wörterbuchforschung – auch rein printlexikografisch orientierte – hierfür fruchtbar gemacht werden kann. Denn auch wenn im Rahmen der computerlexikografischen Praxis manche Ergebnisse von Wörterbuchforschungen, die sich allein auf Printwörterbücher beziehen, nicht übernommen werden können, so sind doch auch immer wieder Aspekte auszumachen, deren Prüfung sinnvoll und nutzbringend sein kann. Für die Frage der Modellierung lexikografischer Daten bietet sich u. a. die Theorie lexikografischer Texte nach Wiegand deshalb an, da sie lexikografische Daten in formalisierter Wei-

se analysiert. Für die Modellierung von Vernetzungsstrukturen ist dabei der Teilbereich der Mediostrukturen von Interesse. (Vgl. Wiegand 1996, Wiegand 2002)

Entsprechend der im ersten Abschnitt genannten Eigenschaften von Wörterbüchern, die zu ihrer ausgeprägten Verweisstruktur führen, lassen sich mit Wiegand drei Arten von Verweisvoraussetzungen differenzieren: „Es lässt sich daher feststellen: Für jede lexikografische Verweishandlung, die sich auf Elemente des Wörterbuchgegenstandes bezieht (und nicht z. B. auf Wörterbuchquellen, andere Wörterbücher oder wissenschaftliche Literatur [...]) müssen grundsätzlich drei Verweisvoraussetzungen berücksichtigt werden: wörterbuchgegenstands-, wörterbuchform- und funktionsbedingte Verweisvoraussetzungen.“ (Wiegand 2002b, 173) Wörterbuchgegenstandsbedingte Verweisvoraussetzungen bilden „Beziehungen [...], die im Wörterbuchgegenstandsbereich vorliegen und zum Wörterbuchgegenstand gehören und für die weiterhin gilt, dass sie mit theoriebestimmten terminologischen Mitteln der Linguistik als ganz bestimmte sprachliche Beziehungen [...] eingeordnet werden“. (Wiegand 2002b, 171) Die wörterbuchformbedingten Verweisvoraussetzungen sind abhängig vom Datendistributionsprogramm, denn sie sind nur dann gegeben, wenn die Daten an „zugriffsverschiedenen textuellen Orten“ (Wiegand 2002b, 172) präsentiert sind. Funktionsbedingte Verweisvoraussetzungen sind abhängig von den anvisierten Wörterbuchfunktionen. (Vgl. auch Tarp 1999) Wörterbuchgegenstands- und wörterbuchfunktionsbedingte Verweisvoraussetzungen spielen auch auf der Ebene der lexikografischen Datenbasis eine Rolle. Allerdings können sich in Wörterbüchern hinter einer Art der Verweisangabe oft verschiedene Beziehungen auf der Ebene des Objektbereichs verbergen. Auf „Ebene der konzeptionellen Datenmodellierung“ sollte man sich daher „darum bemühen, Beziehungen so anzusetzen, dass sie für die Zwecke der Wörterbucherstellung hinreichend atomar sind.“ (Blumenthal et al. 357) Polyseme Verweise sind deshalb zu disambiguieren; auf der Ebene des Wörterbuchs können sie dann nach Bedarf wieder zusammensetzt werden. Darüber hinaus wird diskutiert (u. a. von Kammerer/Lehr 1996), ob es in elektronischen Wörterbüchern überhaupt noch Verweise geben soll. M. E. führen Engelberg/Lemnitzer dazu allerdings richtig aus: „Kammerer & Lehr (1996) gehen sogar so weit, generell Verweise im elektronischen Wörterbuch zu Anachronismen zu erklären, die durch andere Formen von Datenverknüpfungen abgelöst

werden. Wir teilen diese Auffassung nicht. Zumindest die wörterbuchgegenstandsbedingten Verweise werden auch in den Wörterbüchern der Zukunft ihre Berechtigung haben.“ (Engelberg/Lemnitzer 2001, 160) Allerdings ist es für die Zukunft nicht ausgeschlossen, dass sich andere Ordnungssysteme und Datenverknüpfungen bzw. -einbettungen bei Wortschatzinformationssystemen entwickeln. In jedem Fall spielen wörterbuchformbedingte Verweisvoraussetzungen auf Ebene der lexikografischen Datenbasis keine Rolle, da diese eindeutig der Ebene der Präsentation zuzurechnen sind.

Neben dieser Differenzierung der verschiedenen Verweisvoraussetzungen, die für die Abgrenzung zwischen den Ebenen der Datenbasis und der Präsentation wichtig sind, besteht ein großer Bereich der Teiltheorie der Mediostrukturen in der Analyse von Verweisphänomenen im Printwörterbuch, d. h. darin zu untersuchen, welche Arten von Verweise es gibt, aus welchen Teilen sie bestehen, und darin, genaue Benennungen für diese Phänomene zu entwickeln etc. Dabei ist der Analysegegenstand das gedruckte Wörterbuch und damit die Ebene der Präsentation. Im Rahmen des Modellierungskonzeptes ist allerdings die Ebene der lexikografischen Datenbasis Gegenstand der Betrachtung, was ein ganz anderer Phänomenbereich ist. Sehr anschaulich haben Blumenthal et al. diese unterschiedlichen Ebenen mit Hilfe eines Vergleichs von Straßenschildern und Landkarten illustriert: „Zur Verdeutlichung sei dieses spezielle Wechselverhältnis zwischen der Ebene des Wörterbuchs und der Ebene der konzeptionellen Datenmodellierung mit dem Verhältnis zwischen Straßenschildern und Landkarten verglichen: Zunächst zur Ebene der Straßenschilder, die in unserem Vergleich mit der Ebene des Wörterbuchs parallelisiert werden soll: Dem Verweissymbol auf der Ebene des Wörterbuchs entspricht in diesem Vergleich ein Straßenschild, das in eine bestimmte Richtung weist, entkleidet von seiner Beschriftung, die den Namen des Ortes angibt, in dessen Richtung das Schild zeigt. Dem Verweiszielrepräsentanten entspricht der Schriftzug auf dem Straßenschild, der den Namen des Ortes nennt, der auf dem angegebenen Wege zu erreichen ist. Der Verweisangabe entspricht das ganze Schild einschließlich Beschriftung. Dem Verweisursprung schließlich entspricht der Ort, von dem aus der auf dem Schild angegebene andere Ort in der bezeichneten Richtung zu erreichen ist. Und genauso, wie auf der Ebene des Wörterbuchs der Verweisursprung nicht symbolisch repräsentiert ist, ist nun auch auf der

Ebene der Straßenschilder der Ort, von dem aus der andere Ort zu erreichen ist, nicht symbolisch repräsentiert, sondern qua Lokalität des Verkehrszeichens faktisch gegeben. Nun zur Ebene der Landkarte, die in unserem Vergleich der Ebene der konzeptionellen Datenmodellierung entsprechen soll: Genauso, wie auf der Ebene der konzeptionellen Datenmodellierung Verweisursprung, Verweisziel und Verweisrelation symbolisch repräsentiert sind, so sind auch auf der Ebene der Landkarte sowohl die beiden Orte als auch die zwischen diesen beiden Orten bestehende Beziehung der Verbundenheit durch eine Straße symbolisch repräsentiert.“ (Blumenthal et al. 1988, 356f.) Und wie es vermutlich eine wesentlich größere Vielzahl von Straßenschildern gibt – unterschieden nach Farbe, Schrift, Aufstellungsort und genauer Platzierung – so ist auch die Untersuchung des Phänomenbereichs der Verweise auf Ebene der Präsentation (d. h. eines Wörterbuchs oder Wortschatzinformationssystem) vielfältiger als das, was auf Ebene der Datenbasis modelliert wird bzw. was grundlegend zur Modellierung beachtet werden muss. Die Mediostruktur auf Ebene des gedruckten Wörterbuchs bzw. Wortschatzinformationssystem und die Vernetzungsstruktur auf Ebene der lexikografischen Datenbasis sind damit zwei sehr unterschiedliche Phänomenbereiche.

Dies kann leicht an einem Beispiel veranschaulicht werden: Im Inhaltsstrukturenprogramm für ein allgemeines, einsprachiges Wörterbuch sei vorgesehen, dass die Stammformen von starken Verben anzugeben sind. Somit ist z. B. die Angabe_v „ging“ Teil des Artikels „gehen“ auf Ebene der lexikografischen Datenbasis. Für die elektronische Präsentation wird daraufhin festgelegt, dass potenzielle Benutzer, die die Form „ging“ in das Suchfeld eingeben, automatisch auf den Artikel „gehen“ geleitet werden und dem Artikel eine automatisch generierte Zeichenfolge wie „ging: 1./3. Person Singular Imperfekt“ vorangestellt wird. Voraussetzung dafür ist natürlich, dass die Form „ging“ analog zu den anfangs kurz skizzierten Richtlinien des zu Grunde liegenden Modellierungskonzeptes inhaltlich granular ausgezeichnet ist. Für die gedruckte Präsentation wird dagegen festgelegt, dass ein Verweisartikel beispielsweise der Form „**ging** ↑gehen“ automatisch generiert wird. Auf Ebene der lexikografischen Datenbasis ist die Angabe_v „ging“ also Teil des Artikels „gehen“, d. h. in der gleichen XML-Instanz abgespeichert, ohne dass Vernetzungsstrukturen eine Rolle spielen. Im Wortschatzinformationssystem gehört die Angabe „ging“ – wenn man die Termino-

logie der Printlexikografie überträgt – zur äußeren Zugriffsstruktur, im gedruckten Wörterbuch entstehen Angaben, die Gegenstand der Mediosstruktur sind. Genauso könnte festgelegt werden, dass Literaturhinweise, beispielsweise mit dem Element `literaturH` ausgezeichnet, immer mit vorangestelltem Pfeil dargestellt werden sollen, egal ob eine elektronische Verknüpfung zum Ziel vorliegt oder nicht. Damit wären diese Angaben (oder Teile davon) auf Präsentationsebene Teil der Trägermenge der Mediosstrukturen, auf Ebene der lexikografischen Datenbasis sind es jedoch normale Angaben.^v Trotz dieses unterschiedlichen Phänomenbereichs ist die Beschäftigung mit der Teiltheorie der Mediosstrukturen sinnvoll und Einzelheiten können – wie im weiteren Verlauf zu sehen sein wird – angewandt und übertragen werden.

Laut Wiegand ist ein Verweis eine lexikografische Information, d. h. eine kognitive Entität und aus Angaben zu erschließen. (Wiegand 2002, 180) Diese Abgrenzung soll übernommen werden; deshalb wird hier von *Vernetzungsstruktur* und nicht von Verweisstruktur gesprochen. Außerdem trifft die Bezeichnung Vernetzung m. E. eher die Verbindung zweier Punkte auf der ‚Landkarte‘, denn mit dieser Verbindung allein ist noch nicht entschieden, ob diese Vernetzung in einen Verweis umgesetzt werden soll. Für eine Definition von Vernetzung ist die Auseinandersetzung Kammerers mit der Definition von Hyperlinks nach Kuhlen (Kuhlen 1991) und seine eigenen alternativen Vorschläge hierzu hilfreich. (Vgl. Kammerer 1998) Dabei will Kammerer „Hyperlinks“ verstanden wissen „als ein Tupel, bestehend aus einem Verknüpfungsanzeiger und einer Adressenangabe“ (Kammerer 1998, 151). Die verschiedenen Eigenschaften von Hyperlinks, die Kammerer dann nennt, zeigen allerdings, dass sein Hyperlink-Konzept nicht vollständig auf die Vernetzungsstruktur übertragen werden kann, da es auf Ebene der Präsentation angelegt ist. Beispielsweise spielt die Position des Verknüpfungsanzeigers oder die Anzeigart der adressierten informationellen Einheit eine Rolle. (Vgl. Kammerer 1998, 152)

Trotzdem kann der Ansatz der Definition übertragen werden: Demnach sollen Vernetzungen nach dem Modellierungskonzept zunächst verstanden werden als ein Tupel, bestehend aus einer elektronisch adressierten Quellressource und einer elektronisch adressierten Zielressource, also:

Erste Definition von Vernetzung

| | |
|----------------|---|
| Vernetzungen = | {(quelle, ziel) ∈ Q x Z ziel ist mit quelle vernetzt} |
| | Q = Menge aller möglichen elektronisch adressierten Quellressourcen |
| | Z = Menge aller möglichen elektronisch adressierten Zielressourcen |
| | Dabei gilt: |
| | Q ∈ LD (IS) [= Lexikografische Datenbasis (Inhaltsstrukturen)]; vgl. Abb. 1 |

Diese Definition lehnt sich auch an eine Definition von Wiegand an, der zu Verweisbeziehungen Folgendes bemerkt: „Das zweistellige Prädikat ‚x verweist auf y‘ (mit ‚x‘ und ‚y‘ als Variablen für die behandelten Angaben) trifft dann auf Paare von Angaben zu, und alle diese Paare sind Elemente einer zweistelligen Verweisrelation, welche transitiv und asymmetrisch ist.“ (Wiegand 1996, 1192) Präzisierend kann man hier für Q auch sagen, dass Q ein XML-Element ist, welches Teil einer konkreten hierarchischen Inhaltsstruktur ist. Mit dieser Definition ist klar, dass nur solche Beziehungen zur Vernetzungsstruktur gehören, die auch explizit als solche modelliert sind. Beispielsweise will ein Lexikograf mit einem Literatur-Hinweis den Benutzer auf eine weiterführende Literatur hinweisen, er nimmt daher eine Verweisungs*handlung* vor. Besteht dieser Literaturhinweis jedoch nur aus einer bibliografischen Angabe_v, gibt es keine elektronisch adressierte Zielressource. Daher gehört eine solche Angabe_v nicht zur Trägermenge der Vernetzungsstruktur.

Bei Vernetzungsstrukturen sind hinsichtlich der *Lokalität der Zielressource* drei Typen zu unterscheiden:

a) *Inhaltsstrukturen-interne Vernetzungen*

Ein Knoten einer konkreten hierarchischen Inhaltsstruktur wird mit einem Knoten einer anderen hierarchischen Inhaltsstruktur vernetzt, d. h. die Vernetzung findet innerhalb von Artikeln aus der lexikografischen Datenbasis (Inhaltsstrukturen) statt;

d. h. $Q \in \text{LD (IS)}$ und $Z \in \text{LD (IS)}$.

Paradigmatische Relationen sind ein Beispiel von Vernetzung, die in der Regel diesem Typ zuzuordnen sind.

b) *Datenbasis-interne Vernetzungen*

Ein Knoten einer konkreten hierarchischen Inhaltsstruktur wird mit einer elektronischen Ressource außerhalb der LD (IS), aber innerhalb der lexikografischen Datenbasis verbunden;

d. h. $Q \in \text{LD (IS)}$ und $Z \in \text{LD}$.

Unter diesen Typ können beispielsweise Abbildungen fallen, die aus technischen Gründen nicht Teil der XML-Instanz eines Artikel sind oder Verweise auf Texte, die als Umtexte im Wortschatzinformationssystem eingebunden werden (z. B. eine Wörterbuchgrammatik).

c) *Datenbasis-externe Vernetzungen*

Ein Knoten einer konkreten hierarchischen Inhaltsstruktur wird mit einer außerhalb der lexikografischen Datenbasis liegenden elektronischen Ressource verbunden;

d. h. $Q \in \text{LD (IS)}$ und $Z \notin \text{LD}$.

Beispiele können Hinweise auf Internetadressen oder Vernetzungen zu Korpusbelegen sein.

Mit diesen Beispielen wird auch deutlich, dass die Abgrenzung von Inhaltsstrukturen und Vernetzungsstrukturen im Grunde rein formal getroffen wird: Alles, was von der Modellierung her aus der Baumstruktur der Inhaltsstrukturen explizit hinausweist, gehört zur Vernetzungsstruktur. Dies ist die einzig sinnvolle Unterscheidung für ein Modellierungskonzept, da es hier um Richtlinien für die Modellierung einer lexikografischen Datenbasis geht und nicht um Richtlinien für die Formulierung lexikografischer Texte. Und die Modellierung muss nur dann anders erfolgen, wenn die Angaben_v auf andere Weise miteinander verbunden werden müssen als in einer hierarchischen Baumstruktur. Für Vernetzungen sollte daher auch nicht das gelten, was für gedruckte Wörterbücher gilt, nämlich dass man einem Verweisziel nicht ansieht, dass es ein solches ist. Edmund Weiner, einer der Mitarbeiter der Computerisierung des Oxford English Dictionary, hat diesen Sachverhalt für Printwörterbücher sehr treffend folgendermaßen ausgedrückt: „[...] crossreferences [...] are like arrows that have not yet left the bow [...].“ (Weiner 1985, 70; zitiert nach Blumenthal et al. 1988, 365) Bei Vernetzungen soll dagegen der Pfeil den Bogen sehr wohl verlassen, sodass Anfangs- und Endpunkt explizit miteinander verknüpft sind.

Neben der Lokalität der Zielressource gilt es für die Modellierung zwei Typen von Vernetzungen dahingehend zu unterscheiden, ob sie in eine oder beide Richtungen bestehen. Diese zwei Typen sind:

- *unidirektionale Vernetzungen*, z. B. Vernetzungen zu Abbildungen, Korpusbelegen, und
- *bidirektionale Vernetzungen*, z. B. Vernetzungen zwischen sinnverwandten Wörtern, Kookkurrenzen. (Vgl. auch Abbildung 2)

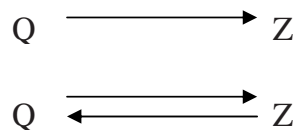


Abb. 2: Unidirektionale vs. bidirektionale Vernetzungen

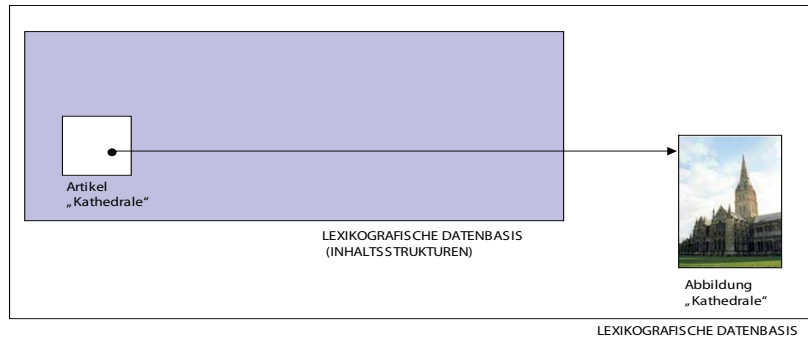
„ → “ bedeutet soviel wie „ist in diese Richtung vernetzt mit“

Unidirektionale Vernetzungen bestehen in eine Richtung, d. h. von der Zielressource wird nicht auf die Quellressource zurückverwiesen. Der Terminus bidirektionale Vernetzungen steht in Kurzform für zwei Vernetzungen, in denen die gleichen Ressourcen abwechselnd die Rolle als Quell- bzw. Zielressource einnehmen. D. h. eine bidirektionale Vernetzung ist nicht eine Vernetzung in beide Richtungen – denn eine Vernetzung ist eine zweistellige, gerichtete Relation – sondern steht für zwei Vernetzungen.

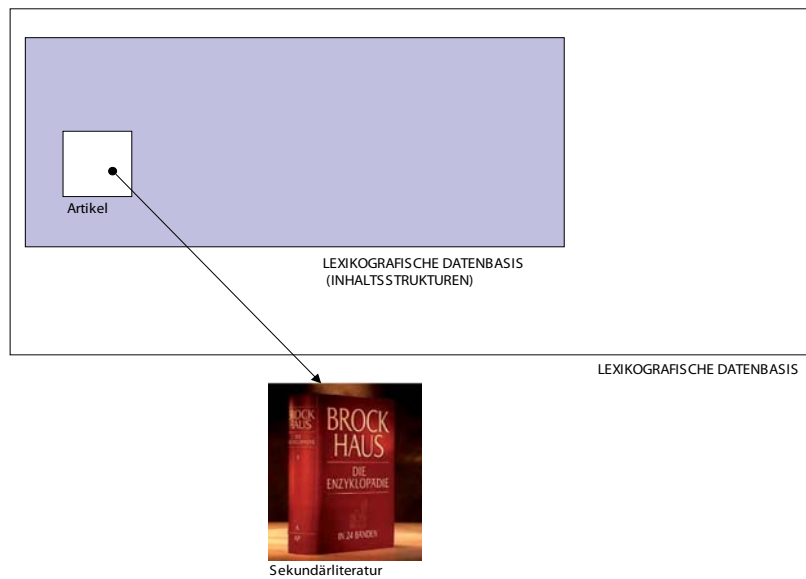
Bei den unidirektionalen Vernetzungen handelt es sich bei Daten für allgemeine, einsprachige Wortschatzinformationssysteme in der Regel um Vernetzungen aus Artikeln heraus auf Sekundärliteratur, Wörterbuchquellen, Korpusbelege oder Abbildungen. Meist ist die Quellressource also Teil der lexikografischen Datenbasis (Inhaltsstrukturen) und die Zielressource ist entweder auch Teil der lexikografischen Datenbasis (beispielsweise bei Abbildungen) oder liegt außerhalb dieser (beispielsweise bei Wörterbuchquellen), d. h. es gilt in der Regel:

$$Q \in \text{LD (IS)} \wedge Z \in \text{LD} \text{ oder } Q \in \text{LD (IS)} \wedge Z \notin \text{LD. (Vgl. Abbildung 3 und 4)}$$

Unidirektionale Vernetzungen sind also meist Datenbasis-interne oder Datenbasis-externe Vernetzungen.



**Abb. 3: Veranschaulichung zu unidirektionalen Vernetzungen vom Typ: $Q \in LD$
 $(IS) \wedge Z \in LD$**
 „ \longrightarrow “ bedeutet soviel wie „ist in diese Richtung vernetzt mit“



**Abb. 4: Veranschaulichung zu unidirektionalen Vernetzungen vom Typ: $Q \in LD$
 $(IS) \wedge Z \notin LD$**
 „ \longrightarrow “ bedeutet soviel wie „ist in diese Richtung vernetzt mit“

Unidirektional sind Vernetzungen deshalb, weil sie nur aus den Artikeln heraus stattfinden, da nicht von Korpusbelegen oder von Sekundärliteratur auf Wörterbuchartikel, in denen diese Literatur erwähnt wird, zurückverwiesen wird. Sind Abbildungen nicht unidirektional in einem Artikel eingebunden, sondern bieten beispielsweise einen onomasiologischen Zugriff auf lexikografische Inhalte – wie in der elektronischen Version des *Oxford Advanced Learner's Dictionary* (OALD) – dann würden diese Vernetzungen unter die bidirektionalen Vernetzungen fallen. Denn dann sollte sowohl von der Abbildung auf den Wörterbuchartikel als auch aus dem Artikel auf die Abbildung verwiesen werden. An diesem Beispiel ist zu sehen, dass man Vernetzungen nicht auf Grund der Art der Zielressource von vorneherein einem Typ zuordnen kann, sondern dass allein die Verankerung der Quell- und Zielressource und ihre Vernetzungsrichtung(en) geprüft werden müssen.

Bidirektionale Vernetzungen können nur im Verantwortungsraum der Lexikografen angelegt werden, da nur dann zwei Ressourcen einmal die Rolle als Quellressource und einmal die Rolle als Zielressource einnehmen können. Für bidirektionale Vernetzungen gilt daher in der Regel, dass beide Ressourcen innerhalb der lexikografischen Datenbasis liegen, d. h.:

$$Q \in LD (IS) \wedge Z \in LD (IS) \text{ oder } Q \in LD (IS) \wedge Z \in LD.$$

Bei den bidirektionalen Vernetzungen handelt es sich daher meist um Inhaltsstrukturen-interne oder um Datenbasis-interne Vernetzungen. Wenn Q und Z Teile der lexikografischen Datenbasis (Inhaltsstrukturen) sind, handelt es sich um Vernetzungen zwischen Angaben_v von Artikeln. Ist die Zielressource nicht Teil der LD (IS), kann es sich um Vernetzungen zu einer Wörterbuchgrammatik oder anderen Texten handeln, die als Umtexte präsentiert werden sollen. Denn oft wird nicht nur aus den Artikeln in die Umtexte verwiesen, sondern auch von den Umtexten in die Artikel, sodass diese Vernetzungen unter die bidirektionalen Verknüpfungen fallen. Daher kann umgekehrt auch gelten, dass $Q \in LD \wedge Z \in LD (IS)$ ist.

Die Gesamtheit der Vernetzungen sind im *Vernetzungsstrukturenprogramm* festgehalten. Dieses Vernetzungsstrukturenprogramm lizenziert die möglichen (abstrakten) Vernetzungsstrukturen. Das Vernetzungsstrukturenprogramm ist damit Gegenstand der Modellierung. Das Mediostrukturenprogramm (oder eine entsprechende Bezeichnung für

elektronische Wörterbücher) legt dagegen fest, wie diese Vernetzungen präsentiert werden. In diesem Sinne sind die Vernetzungen auf Ebene der lexikografischen Datenbasis auch als Vor-Vernetzungen zu verstehen; adäquat zu den Angaben_v bei den Inhaltsstrukturen. Beispielsweise kann im Vernetzungsstrukturenprogramm festgelegt werden, dass paradigmatische Partner immer als Vernetzungen miteinander zu verbinden sind. Im Mediostrukturenprogramm wird jedoch eingeschränkt, dass diese Vernetzungen bei bestimmten Benutzungssituationen im elektronischen Wörterbuch nicht als Links aktiviert werden. Auf der anderen Seite gibt es auch Vernetzungen, die rein aus technischen Gründen Teil des Vernetzungsstrukturenprogramms sind, wie z. B. die Verknüpfungen zu Bildern. Diese können nicht Teil einer XML-Instanz sein, sondern müssen von dort aus referenziert werden. Für den Benutzer sollte dies nicht sichtbar sein, da die Bilder bei der Präsentation in die Textinhalte eingebettet werden können. Genauso gibt es aber auch auf Ebene eines Wörterbuchs oder Wortschatzinformationssystems mediostrukturelle Adressierungsbeziehungen, die nicht auf eine Vernetzung in der lexikografischen Datenbasis zurückgreifen. Z. B. könnte für die Darstellung eines elektronischen Wörterbuchs programmiert werden, dass automatisch jedes Wort einer Paraphrase anklickbar ist, welches selbst lemmatisiert ist (im günstigsten Fall also jedes).⁴ Dies ist jedoch nicht Teil des Vernetzungsstrukturenprogramms auf Ebene der lexikografischen Datenbasis, sondern allein des Mediostrukturenprogramms auf Präsentationsebene. In diesem Sinne sprechen Blumenthal et al. davon, dass nicht jedem Verweis (in einem Wörterbuch) eine eigene Beziehung (auf Ebene der Datenmodellierung) entspricht und nicht jedem Verweisursprung und -ziel eine eigene Objektidentität. „Vielmehr wird man sich bemühen, Verweise zu dekomponieren, d. h. aus (in der Regel linguistisch motivierten) Objektbereichsbeziehungen zusammenset-

⁴ Es ist allerdings die Frage, wie sinnvoll eine solche Praxis ist. Vgl. dazu Wiegand 2004, 353.: „Es ist überflüssig und nützt niemandem, die sozusagen permanente Befolgung dieses Prinzips durch Tausende von überflüssigen Verweispfeilen, die das Lesen behindern, zu dokumentieren. Diese fachlexikografische Praxis [...] wird auch dadurch nicht besser, dass sie in verschiedenen Praxisvarianten in sehr vielen Fachwörterbüchern zu finden ist. Es scheint sich um eine ansteckende fachlexikografische Textkrankheit zu handeln, die ausgelöst wird durch ein meistens pfeilförmiges Virus aus der Familie der mediostrukturellen Printviren (die mit gewissen elektronischen Linkviren verwandt sind), ein Virus, das sich an alle Angaben, die Fachtermini genannt werden, immer dann andockt, wenn diese Fachtermini auch mit einem Lemma genannt werden.“

zen. Wörterbuchform- und erst recht wörterbuchdesignabhängige Entitäten und Beziehungen sollten also auf der Ebene der konzeptionellen Datenmodellierung nicht sichtbar sein.“ (Blumenthal et al. 1988, 356)

Ziel der weiteren Ausführungen dieses Kapitels soll es also sein, Richtlinien für die XML-basierte Modellierung des Vernetzungsstrukturenprogramms zu entwerfen. Es geht damit nicht um die ausführliche, erschöpfende Darstellung möglicher Phänomene in diesem Bereich und wie mit ihnen umgegangen wird, sondern um einen Leitfaden zur grundsätzlichen Behandlung von Vernetzungen zunächst von Daten für allgemeine, einsprachige Wortschatzinformationssysteme (vgl. Müller-Spitzer 2003) bzw. Wörterbücher. Diese Richtlinien sind allerdings so allgemein gehalten, dass sie auch auf die Entwicklung einer maßgeschneiderten Modellierung für andere lexikografische Projekte übertragen werden können. Dabei wird vorausgesetzt, dass die lexikografischen Inhalte analog zu den Richtlinien des anfangs in Ansätzen beschriebenen Modellierungskonzeptes ausgezeichnet sind. (Vgl. Abschnitt 2)

4. Standardbasierte Ausdrucksmöglichkeit: XLink

„Durch die XML-Grundauffassung einer Welt vernetzter unabhängiger Dokumente ergibt sich die natürliche Notwendigkeit zur Schaffung eines Verweismechanismus, der für beliebige XML-Sprachen gleichermaßen universell eingesetzt werden kann. Dieser Anforderung soll die XML Linking Language (XLink) genügen.“ (Jeckle 2004, 83) XLink ist als XML-verwandter Standard seit Juni 2001 als Empfehlung (Recommendation) des W3C veranschiedet.⁵ XLink bietet eine Untermenge der Möglichkeiten von HyTime (Hypermedia/Time Based Structuring Language), einem SGML-basierten Standard zur Modellierung von Vernetzungen, und erweitert wesentlich die Möglichkeiten von HTML in Bezug auf Verlinkungen. In der XLink-Spezifikation (XLink) wird die grundsätzliche Aufgabe des Standards folgendermaßen beschrieben: „XLink bietet einen Rahmen für die Erzeugung sowohl grundlegender unidirektionaler Links als auch komplexerer Link-Strukturen. Die Sprache erlaubt XML-Dokumenten,

⁵ S. <www.w3.org/XML/Linking>.

- Link-Beziehungen zwischen mehr als zwei Ressourcen herzustellen
- Metadaten mit einem Link zu assoziieren
- Links auszudrücken, die sich an einem Ort außerhalb der verlinkten Ressourcen befinden.“ (XLink, 4)

Für die Modellierung des Vernetzungsstrukturenprogramms scheint XLink also der geeignete XML-basierte Standard. (Unabhängig davon, in wie vielen Tools XLink bereits implementiert ist, empfiehlt es sich meiner Meinung nach, solche Standards aufzugreifen, da sich in ihnen meist eine ganze Reihe Überlegungen zum Thema – wie hier zur Verlinkung von Daten – spiegeln, die nicht immer wieder neu angestellt werden müssen. Falls dann ein Werkzeug eingesetzt wird, das XLink nicht direkt verarbeiten kann, kann dies durch daran angelehntes maßgeschneidertes XML-Vokabular ersetzt werden.) Mehr in die Richtung semantischer Netze geht der Topic-Map-Standard.⁶ Eine Anwendung von Topic Maps oder von semantischen Netzen bietet sich allerdings eher dann an, wenn für eine große, oft auch heterogene Datenbasis gemeinsame Zugriffsmöglichkeiten geschaffen werden sollen. So sind beispielsweise im Projekt der „Duden ontology“ (Alexa et al. 2002) semantische Netze von zentraler Bedeutung, um die verschiedenen lexikografischen Produkte gemeinsam verwalten und pflegen zu können. Für die Modellierung von Vernetzungsstrukturen im Rahmen dieses Modellierungskonzeptes ist jedoch XLink das Mittel der Wahl. Falls das Modell erweitert werden soll oder eine lexikografische Datenbasis mit anderen Datenbasen verbunden werden soll, können darauf aufbauend mit einem semantischen Netz gemeinsame Zugriffsstrukturen und weitergehende Vernetzungen entwickelt werden. Dabei ist eine granulare Inhaltsmodellierung bei den lexikografischen Inhalten und ihren Vernetzungen, so wie sie nach dem Modellierungskonzept vorgesehen sind, von entscheidendem Vorteil. Die Modellierung des Vernetzungsstrukturenprogramms, wie sie hier entwickelt ist, ist also bei Bedarf weiter ausbaubar.

⁶ S. <www.topicmaps.org/xtm/1.0/> sowie die XML-Topic-Maps-Seiten der *Organization for the Advancement of Structured Information Standards (OASIS)* unter <<http://xml.coverpages.org/topicMaps.html>>. Weiterführende Informationen und Anwendungsprojekte zeigen (in Auswahl) Widhalm/Mück 2002, Pepper 1999 und Rath 1999.

Zur weitergehenden Charakterisierung des Standards heißt es in der XLink-Spezifikation: „Neben der Bereitstellung von Link-Datenstrukturen bietet XLink ein minimales Link-Verhaltensmodell; Anwendungen höherer Ebene, die auf XLink aufsetzen, werden oft alternative oder ausgefeiltere Darstellungs- und Verarbeitungsverfahren spezifizieren.“ (XLink, 4) Wie XML ein grundsätzliches Mittel für die Modellierung und Auszeichnung von hierarchisch aufgebauten Inhalten ist, so bietet auch XLink eine standardbasierte syntaktische Ausdrucksmöglichkeit für deren Vernetzungen untereinander. Dabei haben weitergehende softwaretechnische Funktionalitäten, die projektgebunden zugeschnitten werden, nichts mit dem Standard zu tun.

In diesem Zusammenhang soll darauf verzichtet werden, die Möglichkeiten von XLink gesondert vorzustellen. In den Richtlinien zur Modellierung der Vernetzungsstrukturen werden stattdessen entsprechende Verweise auf die Abschnitte im XLink-Standard oder dazugehörige Literatur gegeben. (S. dazu auch Müller-Spitzer 2006a)

5. Richtlinien für die XML-basierte Modellierung des Vernetzungsstrukturenprogramms

Die vorgestellte inhaltliche Konzeption für die Modellierung des Vernetzungsstrukturenprogramms soll nun zusammen mit den standardbasierten Ausdrucksmöglichkeiten von XLink in Richtlinien zur Modellierung des Vernetzungsstrukturenprogramms zusammengeführt werden.

Die wichtigste grundsätzliche Modellierungsrichtlinie für das Vernetzungsstrukturenprogramm ist, dass die

- *unidirektionalen Vernetzungen mit einfachen Links* und die
- *bidirektionalen Vernetzungen mit erweiterten Links* modelliert werden.⁷

Die einfachen Links für unidirektionale Vernetzungen werden an der gewünschten Stelle *in der XML-DTD* bzw. *im XML-Schema*, die die Modellierung des Inhaltsstrukturenprogramms jeweils beinhalten, *eingebunden*. Damit lautet der

⁷ Zu einfachen vs. erweiterten Links in XLink s. XLink, 12ff.

Leitsatz zur Modellierung von unidirektionalen Vernetzungen

Unidirektionale Vernetzungen werden als einfache Links modelliert und in die DTDs bzw. Schemata, die die Modellierung des Inhaltsstrukturenprogramms beinhalten, eingebunden.

Eine Modellierung für einen Hinweis auf elektronische Wörterbücher, d. h. für eine Datenbasis-externe Vernetzung, könnte daher folgendermaßen aussehen:

```
<!-- ===== Hinweis auf elektronische Woerterbuecher ===== -->
<!ELEMENT el-woerterbuchH (%hinweistext;) >
<!ATTLIST el-woerterbuchH
  xmlns:xlink          CDATA          #FIXED
                        "http://www.w3c.org/1999/xlink"
  xlink:type           (simple)        #FIXED "simple"
  xlink:href           CDATA          #REQUIRED
  xlink:title          CDATA          #REQUIRED
  xlink:show           (new | replace | embed | other |
                        none)         #REQUIRED
  xlink:actuate        (onLoad | onRequest)
                        #FIXED
                        "onRequest" >
```

Das erste Attribut verweist auf den XLink-Namensraum.⁸ Sind mehrere XLink-spezifische Elemente in einer DTD vorhanden, bietet es sich an, dieses Attribut einmalig auf höherer hierarchischer Ebene anzugeben. Als Linktyp ist `simple` angegeben, d. h. es ist ein einfacher Link. Das Attribut `href` ist obligatorisch gesetzt, da immer eine Zielressource angegeben werden soll. Auch ein für Menschen lesbarer Titel soll immer angegeben werden; daher ist auch dieses Attribut auf `REQUIRED` gesetzt. Die Vergabe des Titels soll deshalb vorgeschrieben werden, da dieser Titel die Möglichkeit eröffnet, den potenziellen Benutzern des Wortschatzinformationssystems schon am Verweisursprung deutlich zu machen, was sie am Verweisziel erwartet. Diese Funktionalität ist eine oft formulierte Forderung in der Wörterbuchforschung. Beispielsweise schlagen Engelberg/Lemnitzer vor, dass die vielfältigeren Möglichkeiten der Darstellung im elektronischen Wörterbuch dazu genutzt

⁸ Um die XLink-eigenen Elemente und Attribute eindeutig als solche auszuweisen, wird – wie bei anderen XML-verwandten Standards auch – der sogenannte Namensraummechanismus angewendet. „XML-Namensräume bieten eine einfache Möglichkeit, um Element- und Attributnamen, die in ‚Extensible-Markup-Language‘-Dokumenten verwendet werden können, eindeutig zu benennen. Die Element- und Attributnamen werden mit Namensräumen verknüpft, die durch URI-Verweise identifiziert werden.“ (XML-Namensräume, 1) Dabei bezeichnen diese URIs nicht tatsächliche Internetquellen, sondern dienen allein der eindeutigen Bezeichnung. (Vgl. Jeckle 2004, 29)

werden sollen, „den Benutzern bereits am Verweisursprung zu signalisieren, was sie am Verweisziel erwartet: eine Grafik, eine Übersichtstafel, eine Regel, ein Videoclip, ein kompletter Wörterbuchartikel etc.“ (Engelberg/Lemnitzer 2001, 159) Die Anzeige über das Attribut `show` muss in diesem Beispiel auch immer spezifiziert werden. Der Auslösevorgang über das Attribut `actuate` ist schon in der DTD festgesetzt auf `onRequest`, da hier von vorneherein festgelegt wurde, dass ein Wörterbuchhinweis immer nur bei Mausklick des potenziellen Benutzers aktiviert werden soll. Genauso ist auch vorstellbar, dass der Wert für `show` schon in der DTD auf „new“ festgesetzt wird, d. h. dass immer ein neues Fenster geöffnet werden soll.⁹

Der zugehörige Ausschnitt aus einer XML-Instanz könnte dann folgende Form haben:

```
<el-woerterbuchH
xmlns:xlink="http://www.w3c.org/1999/xlink"
xlink:type= "simple"
xlink:href= "http://germa63.uni-trier.de:8080/Projects/WBB/woerterbuecher/
          dwb/wbgui?lemid=GA05417"
xlink:title='Artikel "Arbeit" im Deutschen Wörterbuch (Grimm)'
xlink:show= "new">
...
</el-woerterbuchH>
```

Möchte man in seiner Modellierung keine Elemente und Attribute aus dem XLink-Namensraum verwenden, könnte eine eigens entwickelte adäquate Modellierung folgendermaßen aussehen:

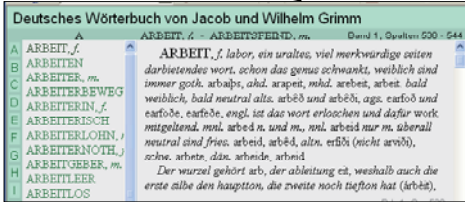
```
<!-- ===== Hinweis auf elektronische Woerterbuecher ===== -->
<!ELEMENT el-woerterbuchH (%hinweistext;) >
<!ATTLIST el-woerterbuchH
  typ                (einfacherLink)          #FIXED
                    "einfacherLink"
  href               CDATA                    #REQUIRED
  titel              CDATA                    #REQUIRED
  anzeige            (neu | ersetzend | eingebettet |
                    andere | keine)          #REQUIRED
  ausloesen          (aufAnforderung)        #FIXED
                    "aufAnforderung">
```

Diese unidirektionale Vernetzung kann wie in Abbildung 5 gezeigt veranschaulicht werden:

⁹ Zu den Verhaltens- und Anzeigeattributen in XLink s. XLink, 33ff.

XML-INSTANZ

```
<?xml version="1.0"?>
<!DOCTYPE artikel SYSTEM "C:\...">
<artikel>
<allgemein>
<lemmazeichenA>beispiel</lemmazeichenA>
...
<el-woerterbuchH
  xmlns:xlink="http://www.w3c.org/1999/xlink"
  xlink:type="simple"
  xlink:href="http://germa63.uni-trier.de:8080/Projects/WBB/woerterbuecher/dwb/wbgui?lemid=GA05417"
  xlink:title="Artikel "Arbeit" im Deutschen Wörterbuch (Grimm)"
  xlink:show="new">
</el-woerterbuchH>
...</allgemein>...
</artikel>
```



ELEKTRONISCHES WÖRTERBUCH

Abb. 5: Veranschaulichung zu unidirektionalen Vernetzungen
 „ \longrightarrow “ bedeutet soviel wie „ist in diese Richtung vernetzt mit“

Ein anderes Beispiel von unidirektionalen Vernetzungen sind Verknüpfungen zu Abbildungen, die aus rein technischen Gründen nicht Teil der XML-Instanz für einen Artikel sein können, also Datenbasis-interne Vernetzungen. Eine Modellierung für eine solche Vernetzung könnte folgendermaßen aussehen:

```
<!-- ===== Abbildung ===== -->
<!ELEMENT abbildung EMPTY >
<!ATTLIST abbildung
  xmlns:xlink          CDATA          #FIXED
                    "http://www.w3c.org/1999/xlink"
  xlink:type           (simple)        #FIXED "simple"
  xlink:href           CDATA          #REQUIRED
  xlink:show           (new | replace | embed | other |
                        none)         #FIXED "embed"
  xlink:actuate        (onLoad | onRequest | other | none)
                    #FIXED "onLoad" >
```

Es handelt sich auch hier um einen einfachen Link, für den schon in der DTD festgelegt wird, dass er automatisch beim Laden die Zielressource einbetten soll, denn dies ist für Abbildungen, die Teil eines Artikels

im elektronischen Wörterbuch sind, in der Regel die sinnvollste Festlegung. Auch hier könnte eine eigene Modellierung, die auf XLink-spezifisches Vokabular verzichtet, als Ersatz entwickelt werden. Wichtig ist, dass bei unidirektionalen Vernetzungen die Vernetzung aus der XML-Instanz des Artikels heraus erfolgt, d. h. dass die Quellressource der Vernetzung Teil des Artikels ist.

Die wichtigste Richtlinie für die Modellierung bidirektionaler Vernetzungen ist, dass diese als *erweiterte Links* in Form einer *Linkbank* modelliert werden sollen.¹⁰ Dabei werden beide Vernetzungen (aus denen die bidirektionale Verknüpfung besteht) in einem Vernetzungsobjekt, d. h. einem erweiterten Link, modelliert. Dies ist möglich über die Modellierung mehrerer an dem Link teilnehmender entfernter Ressourcen und der Spezifizierung des Traversierungsverhaltens zwischen ihnen.¹¹ Der Vorteil einer solchen Modellierung ist, dass das Vernetzungsstrukturenprogramm frei von Redundanzen ist, sodass ein effektives Vernetzungsmanagement erheblich erleichtert wird. Zum Beispiel wird eine Vernetzung durch Synonyme in einem Objekt festgehalten, sodass einfacher geprüft werden kann, dass diese Vernetzung wirklich in beide Richtungen angelegt ist. Die Form einer Speicherung in einer Linkbank hat zudem den Vorteil, dass unabhängig von der Positionierung der einzelnen Angaben_v in den Artikeln, von denen die Vernetzung zunächst ausgeht, die Vernetzungen an die inhaltlich relevanten Teile der Artikel adressiert werden kann. Der erste Leitsatz zur Modellierung bidirektionaler Vernetzungen lautet demnach:

¹⁰ Ein erweiterter Link ist in der Spezifikation folgendermaßen definiert: „Ein *erweiterter Link* ist ein Link, der eine beliebige Anzahl von Ressourcen assoziiert.“ (XLink, 13) „Üblicherweise werden erweiterte Links getrennt von den Ressourcen gespeichert, die sie assoziieren (zum Beispiel in einem ganz anderen Dokument).“ (XLink, 13) Ist dies der Fall, spricht man im Kontext von XLink von einer *Link-Bank*, in der die Links gespeichert sind. (XLink, 7) Der Vorteil beim Aufbau solcher Link-Banken ist, dass die Links verändert werden können, ohne dass die Dokumente selbst angetastet werden. Außerdem erleichtert eine Link-Bank das Vernetzungsmanagement, was für lexikografische Anwendungen von großem Vorteil ist.

¹¹ *Traversierung* bezeichnet dabei „das Benutzen oder Verfolgen eines Links zu welchem Zweck auch immer“. (XLink, 6) „Informationen darüber, wie ein Paar von Ressourcen zu traversieren ist, inklusive der Traversierungsrichtung und möglichen Anwendungsverhaltens, nennen wir eine *Kante*.“ (XLink, 6)

Erster Leitsatz zur Modellierung von bidirektionalen Vernetzungen

Bidirektionale Vernetzungen werden in einem erweiterten Link modelliert. Die beteiligten Quell- und Zielressourcen werden als entfernte Ressourcen modelliert und die sie verbindenden Kanten als Third-Party-Kanten. Die bidirektionalen Vernetzungen sollen in Form einer Linkbank gespeichert werden.

Diese Modellierung von bidirektionalen Vernetzungen soll an den folgenden Beispielen veranschaulicht werden.

Angst wgerm. (nicht engl.) von dem in *†eng* erhaltenen Adj. abgel. (mit *-st-* wie *Dienst, Ernst* usw.); ahd. *angust, angest* F., mhd. fnhd. oft M.; anders als die meisten Sinnverwandten (s.u.) seit alters auch im Pl. (evtl. von dem auch etym. verw. lat. Pl. *angustiae* ›Enge, Beklemmung‹ beeinflusst); *Wenn mein Herz in engsten ist* (Lu. Ps. 77,4); rdtal. *in tausend Angsten* schweben (Ad. 1774). I.1 Im Mhd. und noch im 17./18.Jh. häufig konkreter als heute ›Bedrängnis, Not, Leid‹ (vgl. Mhd.Wb), bes. in Verbindungen wie *A. und Not, A. und Pein*; I.2 dann auch durch die Affektenlehre stärker psychologisiert: *A. Der höchste Grad der Furcht, und also eine sehr wichtige Leidenschaft .. nicht so plötzlich und so vorübergehend .. wie der Schrecken .. Von allen Künstlern kann der tragische Dichter den besten Gebrauch davon machen* (Salzer, Allg. Theorie d. Schönen Künste T.1,1771,144); ein Zentralbegriff moderner Psychologie: *Zumeist versteht man unter A. den subjektiven Zustand, in den man durch die Wahrnehmung der »Angstentwicklung« gerät, und heißt diesen einen Affekt* (Freud I,382); das Subjektive der *A.* betonend: *A. vor Atomnüll, Überbevölkerung, Hungerkatastrophen* usw.? *Nein. Es gibt keine reale A. vor einem kollektiven Schicksal. Das ist Sorge, politisches Gewissen, allenfalls Verzweiflung* (BStrauß, Paare 164); I.3 abgeschwächt ›Besorgnis, Befürchtung, dann häufig mit abhängigem Satz: *ich hatte schon A., ich würde ..* (GoeWb). – In der Gebrauchshäufigkeit hat *A.* erst in den letzten 100 Jahren *Furcht* überflügelt (Bergenholtz, Das Wortfeld »A.«, 1980,247). Als Grundwort z.B. in den verstärkenden Bildungen *Höllten-* (vgl. Lu. Ps. 116,3 *a. der Hellen*), *Todesa.* (beide GoeWb), jünger *Heidol.*; neuere spezielle Bezz. *Platzsa.* (Duden ¹¹1934; vgl. Sprachdienst 13,136f), *Schwelena.* ›Scheu‹ einen ersten Schritt zu tun‹ (vgl. nl. *drempelvoet*; Sprachdienst 14,23f), *Prüfungsa.* usw. Sinnverwandte Substantive: *Bange* (†*bang[e]*), *Enissetzen* (†*entsetzen*), *Feigheit* (†*feige*), †*Furcht*, *Grauen* (†*grauen*), *Grausen* (†*grausen*), *Grusel* (†*gruseln*), †*Horror*, *Lampenfieber* (†*Lampe*), *Panik* (†*panisch*), *Schauer* (†*schauern*), *Scheu* (†*scheuen*), †*Schock*, *Schüchternheit* (†*schüchtern*), *Schrecken* (†*schrecken*), †*Terror*, *Zaghaftigkeit* (†*zagen*); ferner ugs. †*Bammel*, †*Manschette(n)*, *Muffensausen* (†*Muff*), †*Schiff.* *angst* prädikativ in festen Wendungen, öf-

gruseln mhd. *griuelsen*, Iterativum zu mhd. *griuosen* (†*grausen*); ›vor Angst erschauern, sich fürchten‹, zumeist *inppers.* mit Akk. (urspr. auch mit Dat.); vom 16. bis 18.Jh. kaum bzw. nicht gebräuchlich, seit 1819 wiederbelebt: *Mir gruselt's nicht; das wird wohl eine Kunst sein, von der ich nichts verstehe* (Grimm, Märchen von einem, der auszog), auch subst.: *Das G., das sollst du schon lernen* (ebd.); die Rückbildung *Grusel* M. ›(angenehmer) Angstschauder‹ (†*Angst*) fnhd., dann erst wieder nach der Wiederbelebung des Verbs. Zuss. *Gruselgeschichte, -film, -kabinett*; analog zu *Musical: Grusical* (Duden ¹⁶1967); alte Abl. *grus(e)lig* (15.Jh.).

Abb. 6: Artikel „Angst“ und Artikel „gruseln“ aus dem Deutschen Wörterbuch (DW)

„—————▶“bedeutet soviel wie „ist mediostrukturell adressiert an“
 „-----▶“bedeutet soviel wie „ist artikelintern zugreifbar“

Die Abbildung 6 zeigt die Artikel „Angst“ und „gruseln“ aus der Prin-
 tausgabe des Deutschen Wörterbuchs von Hermann Paul (DW 1992).
 Aus dem Artikel „Angst“ wird auf den Artikel „gruseln“ als Sinnver-

wandte verwiesen, aus dem Artikel „gruseln“ wird innerhalb des Subartikels „Grusel“ auf „Angst“ verwiesen. Nun soll angenommen werden, dass der Datenbestand für dieses Wörterbuch in einer lexikografischen Datenbasis gespeichert ist und analog zu den hier entworfenen Modellierungsrichtlinien aufbereitet ist.

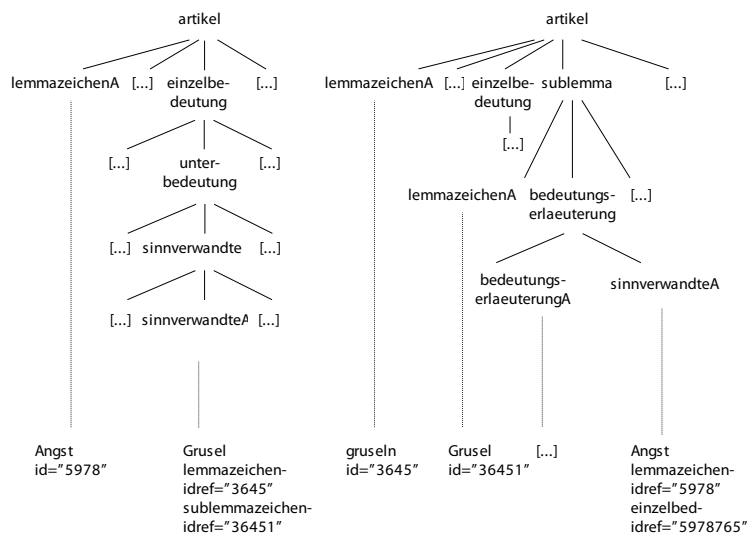


Abb. 7: Mögliche (rudimentäre) konkrete und abstrakte Inhaltsstrukturen der Artikel „Angst“ und „gruseln“ aus dem Deutschen Wörterbuch (DW 1992)

Die Abbildung 7 zeigt in Grundzügen, wie man sich auszugsweise die zugehörigen abstrakten und konkreten Inhaltsstrukturen der Artikel vorstellen kann. Dabei geht es hier nicht um Details dieser möglichen Modellierung oder um die darin verwendete Terminologie, sondern darum, ein anschauliches Beispiel für die Modellierung des Vernetzungsstrukturenprogramms aufzuzeigen.

Innerhalb des Artikels „Angst“ ist die Angabe_v „Grusel“ als *sinnverwandteA* ausgezeichnet. Die zugeordneten *IDREFs*, d. h. Verweise auf *IDs*, verweisen auf den Artikel „gruseln“ (*lemmzeichen-idref*) und auf den Subartikel „Grusel“ (*sublemmzeichen-idref*). Diese Modellierung und zugehörige Auszeichnung würden ermöglichen, dass für die gedruckte Ausgabe die Angabe „Grusel (↑gruseln)“ generiert

wird, auch wenn nur „Grusel“ im Elementinhalt steht. (Im Mediostrukturenprogramm für das DW ist festgelegt, dass mit dem Verweispfeil nur „auf Hauptlemmata [...] verwiesen“ (DW 1992, XIV) wird. Daher ist als Verweisziel hier „gruseln“ angegeben, auch wenn „Grusel“ das sinnverwandte Wort zu Angst ist.) Im Artikel „gruseln“ wird innerhalb des Subartikels „Grusel“ entsprechend auch über eine `sinnverwandteA` und einer `lemmazeichen-idref` auf „Angst“ verwiesen.

Die hier beschriebenen Angaben_v stellen also inhaltlich eine Vernetzung zwischen dem gesamten Artikel „Angst“ und dem Subartikel „Grusel“ her, denn zwischen diesen Objekten besteht die Sinnverwandtschaft. So wie die vernetzungsrelevanten Angaben_v jedoch im Artikel sind, besteht die Vernetzung zwischen der Sinnverwandten-Angabe_v „Grusel“ (innerhalb des Artikels „Angst“) und dem Subartikel „Grusel“; und in der anderen Richtung von der Sinnverwandten-Angabe_v „Angst“ (innerhalb des Subartikels „Grusel“) und dem Artikel „Angst“. Die Quellressource für die Vernetzung ist daher in den konkreten hierarchischen Inhaltsstrukturen inhaltlich nicht genau zugreifbar. (Zugreifbar ist nur die Angabe_v, aber nicht der gesamte Artikel bzw. Subartikel als Quellressource.) Ziel bei der Modellierung der Vernetzungsstruktur sollte daher sein, die Quellressource und die Zielressource in genau richtigem inhaltlichen Umfang kenntlich zu machen. Diese genaue Kodierung von Quell- und Zielressourcen – gesammelt in einer Linkbank – ermöglicht gerade für eine elektronische Präsentation vielfältige Möglichkeiten: Beispielsweise können für bestimmte Benutzungssituationen zu einem gesuchten Artikel gleich am Anfang die bedeutungsverwandten Artikel aufgeführt werden. Oder bedeutungsverwandte Artikel können in eine Suche integriert werden. Diese Auswertung der Vernetzungen für die Programmierung von Zugriffsmöglichkeiten sind nicht in dem Maße gegeben, wenn Quell- und Zielressourcen nicht in ihrem vollen inhaltlichen Umfang adressierbar sind. Bei einem neu zu erstellenden Wortschatzinformationssystem ist es daher empfehlenswert, im Vernetzungsstrukturenprogramm von vorneherein eine möglichst genaue Vernetzung von Quell- und Zielressource vorzusehen. So sind zum Beispiel in *ellexiko*, einem korpusbasierten lexikografischen Informationssystem des Instituts für Deutsche Sprache

(IDS)¹², die paradigmatischen Beziehung auf Ebene der Einzelbedeutungen verankert.

Bei der Sinnverwandtschaft handelt es sich um eine bidirektionale Vernetzung, d. h. es bestehen zwei Vernetzungen, bei denen jeweils die gleichen Ressourcen einmal die Rolle als Quellressource und einmal die Rolle als Zielressource einnehmen. Die Modellierung erfolgt also in Form einer Linkbank, in der die Sinnverwandten in einem erweiterten Link modelliert sind. Dabei werden die beteiligten Artikel bzw. Subartikel als entfernte Ressourcen modelliert, die am Link teilnehmen. Sie werden adressiert und identifiziert durch die XLink-spezifischen Attribute `href` und `label`. Die Kanten sind die Elemente `sinnverwandt_zu1-2` und `sinnverwandt_zu2-1`. Sie sind durch den Wert `"arc"` des `type`-Attributs als Kanten gekennzeichnet. Es handelt sich dabei in der XLink-Terminologie um „Third-Party-Kanten“, da sie ausschließlich entfernte Ressourcen zueinander in Beziehung setzen.¹³ Durch die `from`- und `to`-Attribute werden die Quell- und Zielressourcen identifiziert. Außerdem ist nach dieser Modellierung vorgesehen, dass ein Titel für die Kante vergeben werden soll; Werte zum Anzeige- und Auslöseverhalten sind bereits in der Linkbank festgelegt.¹⁴

Die Instanz zu dieser Linkbank soll automatisch gefüllt werden. Lexikografen müssen somit – außer zum gezielten Vernetzungsmanagement – nicht direkt in der Linkbank arbeiten. Die automatische Füllung funktioniert über die granulare inhaltliche Auszeichnung der vernetzungsrelevanten Angaben_v und über das Ausfüllen vernetzungsrele-

¹² S. www.elexiko.de und zu weiterführenden Informationen Haß 2005 und Müller-Spitzer 2005.

¹³ „Eine Kante, die eine lokale Start- und eine entfernte End-Ressource hat, heißt *ausgehend*, d. h. sie geht fort vom Link-Element. [...] Wenn die End-Ressource einer Kante lokal ist, aber die Start-Ressource entfernt, dann heißt die Kante *eingehend*. [...] Wenn weder die Start- noch die End-Ressource lokal sind, heißt die Kante eine *Third-Party-Kante*.“ (XLink, 7) In HTML gibt es beispielsweise nur ausgehende Kanten, da das `href`-Element immer in der Quellressource enthalten ist.

¹⁴ Auch bei den bidirektionalen Vernetzungen können statt XLink-spezifischem Vokabular eigens definierte XML-Modellierungen eingesetzt werden. Es empfiehlt sich dabei im Sinne der Zukunftsorientiertheit, diese Modellierung an den Möglichkeiten von XLink zu orientieren, soweit das inhaltlich angemessen ist. Den Einsatz einer solchen individuell modellierten Link-Bank in einem laufenden Projekt zeigt beispielsweise Schneider 2004, 280ff.

vanter Attribute. Für einen Ausschnitt aus den XML-Instanzen „Angst“ und „gruseln“ kann das folgendermaßen aussehen:

- Ausschnitt aus der XML-Instanz Angst.xml:

```
<artikel id="5978"><lemmazeichenA>Angst</lemmazeichenA>
...
<sinnverwandteA artikel-idref="3645" subartikel-idref="36451">
Grusel
</sinnverwandteA>
</artikel>
```

- Ausschnitt aus der XML-Instanz (gruseln.xml):

```
<artikel id="3654"><lemmazeichenA>gruseln</lemmazeichenA>
<subartikel id="36541">
<lemmazeichenA>Grusel</lemmazeichenA>
...
<sinnverwandteA artikel-idref="5978">Angst</sinnverwandteA>
...
</subartikel></artikel>
```

Aus diesen Angaben_v kann die zugehörige Linkbank automatisch gefüllt werden. Die Angaben_v in den Artikeln, die zur Füllung der Linkbank dienen, sollen dabei *vernetzungsrelevante Angaben_v* heißen. Demnach lautet der zweite Leitsatz zur Modellierung bidirektionaler Vernetzungen:

Zweiter Leitsatz zur Modellierung von bidirektionalen Vernetzungen

Die Linkbank, in der die bidirektionale Vernetzungen gespeichert werden, soll automatisch gefüllt werden und allein dem Vernetzungsmanagement dienen. Diese automatische Füllung soll über vernetzungsrelevante Angaben_v, die Teil der Artikel sind, erfolgen.

Zur Identifizierung der Ressourcen – genauer des Artikels und des Subartikels – werden die IDs in das Attribut `label` eingetragen. Diese Werte werden dann wiederum in die `from`- und `to`-Attribute der beiden Kanten gefüllt. Die Quell- bzw. Zielressourcen sind damit in diesem Fall der Artikel „Angst“ und der Subartikel „Grusel“. Auch hier wird mit dem Titel der Kante – „Hier gelangen Sie zum Artikel eines sinnverwandten Wortes“ – der Forderung der Wörterbuchforschung entsprochen, die potenziellen Benutzer schon am Verweisursprung über die Art des

Ziele in Kenntnis setzen zu können. Der Zusammenhang von XML-Instanzen und einer Linkbank wird in Abbildung 8 veranschaulicht.

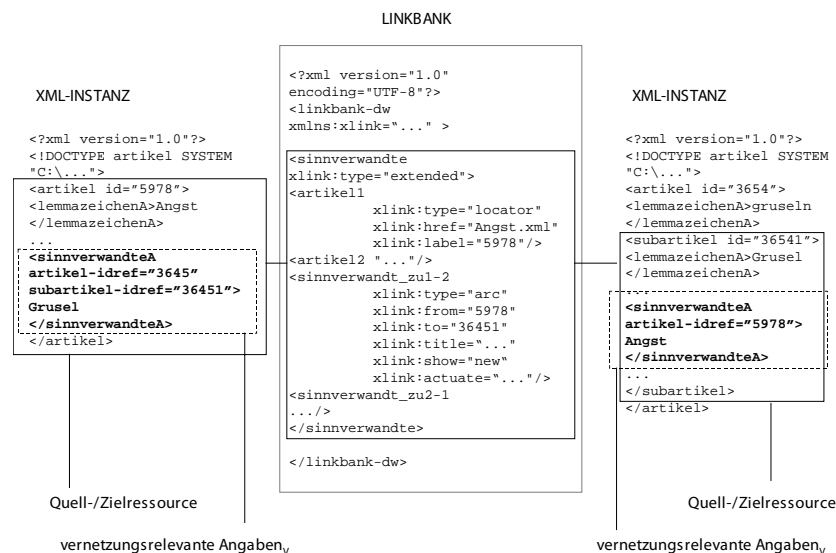


Abb. 8: Veranschaulichung zu bidirektionalen Vernetzungen

Diese oben gezeigte Vernetzung kann – wenn auch in anderer Form – in Grundzügen auch allein über die vernetzungsrelevanten Angaben_v in den Artikeln kodiert werden. Daher stellt sich die Frage, warum der Aufbau einer Linkbank notwendig ist. Für den Aufbau einer Linkbank sprechen mehrere Gründe:

- Wie oben ausgeführt, ist es wichtig, die Teile der Artikel als Quell- und Zielressourcen zu adressieren, zwischen denen inhaltlich die Vernetzung besteht. Dies sind in diesem Fall ein gesamter Artikel und ein Subartikel und nicht einzelne Angaben_v aus den Artikeln.
- Diese genaue Kodierung von Quell- und Zielressourcen – gesammelt in einer Linkbank – eröffnet gerade für eine elektronische Präsentation weitreichendere Möglichkeiten der Einbeziehung der Vernetzungen in das Angebot von Zugriffsmöglichkeiten.
- Die Linkbank ermöglicht eine bessere Konsistenzkontrolle, was bei Vernetzungen elementar wichtig ist. Bidirektionale Vernetzungen sind in einem Objekt gespeichert und können so besser kontrolliert

werden. Außerdem können alle bidirektionalen Vernetzungen, die aus einem Artikel herausgehen oder auf einen Artikel weisen, in ihrer Gesamtheit abgerufen werden.

- Beim Aufbau einer lexikografischen Datenbasis ist der Einsatz einer Linkbank sinnvoll, da beim Neuverfassen eines Artikels zunächst in der Linkbank geprüft werden kann, welche Vernetzungen den neuen Artikel als Zielressource angeben. Diese Angaben_v können beim Schreiben des Artikels berücksichtigt werden. (Dies setzt voraus, dass die Lemmakandidatenliste vorher feststeht und allen Lemmazeichen eine ID zugeordnet ist, damit diese zur Adressierung dienen kann. Ist der Artikel dann verfasst, kann die Vernetzung ggf. auf die Ebene von Einzelbedeutungen o. Ä. verlagert werden.)

Dabei ist es wichtig herauszustellen, dass der Einsatz einer so modellierten Linkbank eine relativ ‚schlanke‘ Lösung für die Kodierung von Vernetzungen ist, die trotzdem ein relativ hohes Potenzial für das Vernetzungsmanagement bereitstellt. Denn der Einsatz einer Linkbank erfordert nicht eine so umfangreiche Softwareunterstützung wie beispielsweise der Einsatz eines semantischen Netzes.¹⁵

Die Richtlinien zur Modellierung der Vernetzungsstrukturen sind damit vorgestellt. Allerdings muss nun die erste Definition der Vernetzungsstrukturen nochmals aufgegriffen werden. Bei der Modellierung der bidirektionalen Vernetzungen in Form der Linkbank war zu sehen, dass die Quell- und Zielressourcen nicht unbedingt Auskunft über den Typ der Vernetzung geben. Dieser Typ ist meist nur über die Kante deutlich gemacht. Anders ist es bei den unidirektionalen Vernetzungen. Da davon ausgegangen wird, dass diese in die XML-Instanz eines Artikels eingebunden sind und von einem Knoten der hierarchischen Inhaltsstruktur ausgehen, der nach den Richtlinien des Modellierungskonzeptes ausgezeichnet ist, können diese über die Benennung der Quellressource einem Typ zugeordnet werden. Im oben gezeigten Beispiel handelt es sich beispielsweise um einen „Hinweis auf ein elektronisches Wörterbuch“, was durch das Ausgangslement `el-worterbuchH` deutlich wird. Unidirektionale Vernetzungen sollen daher weiterhin als eine zweistellige, gerichtete Relation aufgefasst werden:

¹⁵ Dabei kann es in Projekten auch angebracht sein, die Linkbank um die Kodierungen anderer Vernetzungen zu erweitern; vgl. Schneider 2004, 281ff.

Definition unidirektionale Vernetzung (zweistellig)

Unidirektionale Vernetzungen = $\{(quelle, ziel) \in Q \times Z \mid \text{ziel ist mit quelle vernetzt}\}$

Q = Menge aller möglichen elektronisch adressierten Quellressourcen
 Z = Menge aller möglichen elektronisch adressierten Zielressourcen

Dabei gilt:
 Q \in LD (IS) und
 Es gibt kein Paar von Vernetzungen, in denen die gleichen Ressourcen einmal die Rolle als Quell- und einmal die Rolle als Zielressource einnehmen

Die bidirektionalen Vernetzungen sollen dagegen als *dreistellige*, gerichtete Relation aufgefasst werden:

Definition bidirektionale Vernetzung (dreistellig)

Bidirektionale Vernetzungen = $\{(quelle, ziel, typ) \in Q \times Z \times T \mid \text{ziel ist über Kante vom Typ typ mit quelle verbunden}\}$

Q = Menge aller möglichen elektronisch adressierten Quellressourcen
 Z = Menge aller möglichen elektronisch adressierten Zielressourcen
 T = Menge aller möglichen Typen

Dabei gilt:
 Q \in LD (IS)
 Es gibt ein Paar von Vernetzungen, in denen die gleichen Ressourcen einmal die Rolle als Quell- und einmal die Rolle als Zielressource einnehmen

Im obigen Beispiel würde es also heißen: „Artikel ‚Angst‘ ist über Kante vom Typ ‚sinnverwandtschaftlich‘ mit Subartikel ‚Grusel‘ verbunden.“ Ohne den Typ als drittes Argument könnte man diese Vernetzung nicht hinreichend genau ausdrücken.

6. Zusammenfassung

Kurz zusammengefasst lauten damit die wesentlichen Richtlinien zur Modellierung der Vernetzungsstrukturen wie folgt: Für die Modellierung der Vernetzungsstrukturen wird die XML Linking Language bzw. eigenes XML-Vokabular, welches sich an XLink orientieren sollte, eingesetzt. Dabei werden unidirektionale Vernetzungen über einfache Links modelliert, die in die DTDs bzw. XML-Instanzen eingebunden sind. Bidirektionale Vernetzungen werden über erweiterte Links modelliert, die in einer Linkbank gesammelt werden. Der Aufbau einer solchen Linkbank ermöglicht eine präzise Adressierung der Quell- und Zielressourcen der Vernetzung unabhängig von der Verortung der vernetzungsrelevanten Angaben_v in der hierarchischen Inhaltsstruktur und daneben ein effektives Vernetzungsmanagement. Die Linkbank wird über die vernetzungsrelevanten Angaben_v automatisch gefüllt. Durch diese Modellierung wird insgesamt gewährleistet, dass sowohl die Datenkonsistenz hinsichtlich der Vernetzungen sichergestellt ist, als auch die Voraussetzungen dafür geschaffen werden, dass die Vernetzungen auf Ebene der Präsentation vielfältig für Zugriffs- und Navigationsmöglichkeiten ausgewertet werden können. Gleichzeitig ist diese Form des Vernetzungsmanagements eine ‚schlanke‘ Lösung auch hinsichtlich der softwaretechnischen Realisierung – anders als beispielsweise der Einsatz eines semantischen Netzes – und daher praktikabel für viele (wissenschaftliche) lexikografische Projekte. Und um abschließend den Vergleich zwischen der Landkarte und den Straßenschildern nochmals aufzugreifen: So wie man in eine Landkarte ‚reinzoomen‘ kann, um beispielsweise aus einem Stadtplan eine Wanderkarte zu erstellen, so kann man auch die hier vorgestellten Modellierungsvorschläge für Vernetzungen als Grundlage verstehen, die bei Bedarf weiter verfeinert oder ergänzt werden können.

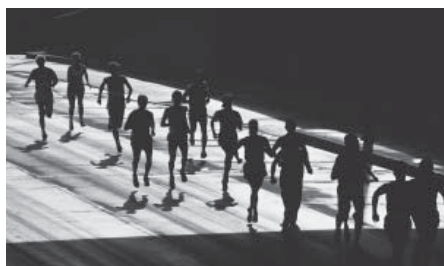
7. Literatur

- Alexa, Melina/Kreissig, Bernd/Liepert, Martina/Reichenberger, Klaus/Rautmann, Karin/Scholze-Stubenrecht, Werner/Stoye, Sabine 2002: *The Duden ontology: An Integrated Representation of Lexical and Ontological Information*. In Workshop at IREC 2002. Las Palmas, Gran Canaria (27.5.2002).
- Behme, Henning/Mintert, Stefan 2000: *XML in der Praxis. Professionelles Web-Publishing mit der Extensible Markup Language*. München: Addison-Wesley.
- Bergenholtz, Henning/Tarp, Sven 2003: Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. In *Hermes* 31 (2003), 171-196.
- Bergenholtz, Henning/Tarp, Sven 2002: Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. In *Lexicographica* 22 (2002), 145-155.
- Blumenthal, Andreas/Lemnitzer, Lothar/Storrer, Angelika 1988: Was ist eigentlich ein Verweis? Konzeptionelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung. In Harras, Gisela (ed.), *Das Wörterbuch. Artikel und Verweisstrukturen* (= Jahrbuch 1987 des Instituts für deutsche Sprache). Düsseldorf/Bielefeld: Pädagogischer Verlag Schwann-Bagel/Cornelsen-Velhagen u. Klasing, 351-373.
- DFWB 1994: *Duden – Das Große Fremdwörterbuch*, hrsg. und bearb. v. Wissenschaftlichen Rat der Dudenredaktion u. Leitung v. Günther Drosdowski, Mannheim et al.: Dudenverlag.
- DW 1992: Hermann Paul, *Deutsches Wörterbuch*, 9., vollständig neu bearbeitete Auflage von Helmut Henne und Georg Objartel unter Mitarbeit von Heidrun Kämper-Jensen. Tübingen: Max Niemeyer.
- Engelberg, Stefan/Lemnitzer, Lothar 2001: *Lexikographie und Wörterbuchbenutzung* (= Stauffenburg Einführungen, Band 14). Tübingen: Stauffenburg.
- Haß, Ulrike (ed.) 2005: *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz* (= Schriften des Instituts für Deutsche Sprache, Bd. 12). Berlin/New York: de Gruyter.
- Jeckle, Mario 2004: Scriptum zur Vorlesung XML. www.jeckle.de/vorlesung/xml/script.html (199 Seiten) (21.7.2006).
- Kammerer, Matthias 1998: Hypertextualisierung gedruckter Wörterbuchtexte. Verweisstrukturen und Hyperlinks. Eine Analyse anhand des „Frühneuhochdeutschen Wörterbuchs“. In Storrer, Angelika/Harriehausen, Bettina (eds.): *Hypermedia für Lexikon und Grammatik* (= Studien zur deutschen Sprache 12). Tübingen: Narr, 145-171.
- Kammerer, Matthias/Lehr, Andrea 1996: Potentielle Verweise und die Wahrscheinlichkeit ihrer Konstituierung. In Wiegand, Herbert Ernst (ed.): *Wörterbücher in der Diskussion*. Tübingen: Niemeyer, 311-354.
- Kuhlen, Rainer 1991: *Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissenschaft*. Berlin: Springer.

- Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (eds.) 2001: *Chancen und Perspektiven computergestützter Lexikographie* (Lexicographica. Series Maior 107). Tübingen: Niemeyer.
- Müller-Spitzer, Carolin 2006a: *Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis*. Unveröffentl. Diss. Heidelberg 2006 (erscheint in: Studien zur deutschen Sprache Tübingen: Narr).
- Müller-Spitzer, Carolin 2006b: Das Konzept der Inhaltsstruktur. Ein Ausschnitt aus einer neuen Konzeption für die Modellierung lexikografischer Daten. In *OPAL – Online publizierte Arbeiten zur Linguistik 2* (2006), 1-18.
- Müller-Spitzer, Carolin 2005: Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess. In: Haß, Ulrike (ed.): *Grundfragen der elektronischen Lexikographie*. elexiko – das Online-Informationssystem zum deutschen Wortschatz (= Schriften des Instituts für Deutsche Sprache, Bd. 12), Berlin/New York: de Gruyter, 21-54.
- Müller-Spitzer, Carolin 2003: Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In *Lexicographica* 19 (2003), 140-168.
- OALD 2002: *Oxford Advanced Learner's Dictionary*. New Edition 2000. Oxford University Press: 2002 (CD-ROM).
- Pepper, Steve 1999: Euler, Topic Maps and Revolution. In *XML Europe 1999. Conference Proceedings*, 135-150.
- Rath, Holger 1999: Mozart oder Kugel. Mit Topic Maps intelligente Informationsnetze aufbauen. In *iX* 12 (1999), 149-155.
- Raymond, Darrell R./Tomba, Frank Wm. 1988: Hypertext and the Oxford English Dictionary. In *Communications of the ACM* 31 (7/1988), 871-879.
- Schneider, Roman 2004: *Benutzeradaptive Systeme im Internet. Informieren und Lernen mit GRAMMIS und ProGr@mm* (= amades. Arbeitspapiere und Materialien zur deutschen Sprache 4/04), Mannheim: Institut für Deutsche Sprache.
- Tarp, Sven 1999: Theoretical foundations of the so-called cross-reference structures. In *Lexicographica* 15 (1999), 114-137.
- Weber, Nico 1999: *Die Semantik von Bedeutungsexplikationen*. Frankfurt a.M.: Lang.
- Widhalm, Richard/Mück, Thomas 2002: *Topic Maps. Semantische Suche im Internet*. Berlin et al.: Springer.
- Wiegand, Herbert Ernst 2004: Überlegungen zur Mediostruktur in Fachwörterbüchern. Auch am Beispiel des „Wörterbuchs zur Lexikographie und Wörterbuchforschung“. In Brdar-Szabó, Rita/Knif-Komlósi, Elisabeth (eds.): *Lexikalische Semantik, Phraseologie und Lexikographie. Abgründe und Brücken. Festgabe für Regina Hessky* (=Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 57). Frankfurt u.a.: Lang, 339-365.
- Wiegand, Herbert Ernst 2002: Altes und Neues zur Mediostruktur in Printwörterbüchern. In *Lexicographica* 18 (2002), 168-252.

- Wiegand, Herbert Ernst 2001: Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neusten Wörterbuchforschung. In *Lexicographica* 17 (2001), 217-248.
- Wiegand, Herbert Ernst 2000: Wissen, Wissensrepräsentation und Printwörterbücher. In *Proceedings of the Ninth Euralex International Congress, Euralex 2000*, Vol. I., Stuttgart, 15-38.
- Wiegand, Herbert Ernst 1998a: *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*, 1. Teilband, Berlin/New York: de Gruyter.
- Wiegand, Herbert Ernst 1998b: Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption. In Kammerer, Matthias/Wolski, Werner (eds.): Herbert Ernst Wiegand, *Kleine Schriften. Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden*, Band 2: 1988-1999 (Erstveröffentlichung des Aufsatzes 1998). Berlin/New York: de Gruyter, 1454-1489.
- Wiegand, Herbert Ernst 1997: Über die gesellschaftliche Verantwortung der wissenschaftlichen Lexikographie. In *Hermes* 18 (1997), 177-202.
- Wiegand, Herbert Ernst (2000): Über die Mediosstrukturen bei gedruckten Wörterbüchern. In Kammerer, Matthias/Wolski, Werner (eds.): Herbert Ernst Wiegand, *Kleine Schriften. Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden*, Band 2: 1988-1999 (Erstveröffentlichung des Aufsatzes 1996). Berlin/New York: de Gruyter, 1163-1192.
- Wiegand, Herbert Ernst 1989: Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1), Berlin/New York: de Gruyter, 409-462.
- XML Linking Language (XLink): Version 1.0, DeRose, Steve/Maler, Eve/Orchard, David/Trafford, Ben (eds.). *World Wide Web Consortium, W3C Recommendation 27. Juni 2001*. Deutsche Übersetzung www.edition-w3c.de/TR/2001/REC-xlink-20010627 (21.7.2006).
- XML-Namensräume (Namespaces in XML): Bray, Tim/Hollander, Dave/Layman, Andrew (eds.). *World Wide Web Consortium, W3C-Recommendation 19. Januar 1999*. Deutsche Übersetzung (18. Juni 2001) www.schumacher-netz.de/TR/1999/REC-xml-names-19990114-de.html (21.7.2006).

**Add a
dimension to
your
sociology
research...**



sociological abstracts

*Comprehensive, cost-effective, timely coverage of current ideas
in sociological research*

Abstracts of articles, books, and
conference papers from nearly 2,000
journals published in 35 countries;
citations of relevant dissertations as
well as books and other media.

Now featuring:

- **Cited references**
- **Additional abstracts
covering 1963-1972**

Available in print or electronically through CSA Illumina
(www.csa.com).

*Contact sales@csa.com for trial Internet access or a sample
issue.*



ILLUMINA

www.csa.com
