



INSTITUT FÜR
DEUTSCHE SPRACHE

OPAL

Online publizierte Arbeiten zur Linguistik

ISSN 1860-9422

Sonderheft

1/2008

Carolin Müller-Spitzer

Der texttechnologische Aufbau von OWID

aus: Klosa (Hg.): Lexikografische Portale im Internet.

(= OPAL-Sonderheft 1/2008), S. 45-55.

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

© 2008 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

Der texttechnologische Aufbau von OWID

Abstract

Die in OWID enthaltenen Wörterbücher sollen nicht unverbunden im Portal nebeneinander stehen, sondern inhaltlich möglichst eng vernetzt werden. Auch werden für jedes einzelne Wörterbuch spezifische Recherchemöglichkeiten angeboten, die einen gezielten Zugriff auf die lexikografischen Inhalte ermöglichen, um so die Möglichkeiten des elektronischen Medium zu nutzen. Die Voraussetzung für die Realisierung dieser Anforderungen werden im Bereich der texttechnologischen Strukturierung der Daten gelegt. Welcher Weg der Modellierung dabei aus welchen Gründen für OWID gewählt wurde, wie er in die Realität umgesetzt wurde und was an Perspektiven daraus abzuleiten ist – dies ist das Thema dieses Beitrags.

The dictionaries being compiled in OWID will not present a random collection of unrelated lexicographic resources, but will form a network of actually related products. At the same time, specific search features are offered to allow flexible access to each individual lexicographic resource. Thus the possibilities of the electronic medium are used extensively. A prerequisite for the realization of these requirements is the texttechnological structure of the lexicographic data. This paper explains which way of modelling was chosen in OWID and which perspectives can be derived from that.

Inhalt:

1. Leitlinien der Modellierung
2. Die Modellierung in Stichworten
3. Der Aufbau des DTD-Modulsystems
 - 3.1 Grundlegende Organisation
 - 3.2 Veranschaulichung anhand von Beispielen
4. Perspektiven
5. Literatur

1. Leitlinien der Modellierung

Mit OWID soll kein unverbundener lexikografischer Gemischtwarenladen präsentiert, sondern tatsächlich ein Verbund von lexikografischen Produkten aufgebaut werden (vgl. auch den Beitrag von Klosa/Lemnitzer/Neumann in dieser Publikation). Damit dieses Ziel erreicht werden kann, muss in den tieferen Schichten unter der Präsentation einiges an Arbeit geleistet werden. Welcher Weg der Modellierung dabei aus welchen Gründen gewählt wurde, wie er in die Realität umgesetzt wurde und was an Perspektiven daraus abzuleiten ist – dies ist Thema dieses Beitrags.

Die einzelnen beteiligten Projekte bzw. die von ihnen erarbeiteten Wörterbücher im Portal sind inhaltlich unabhängig voneinander. Allerdings war von Anfang an klar, dass der Wert von OWID umso höher wird, je mehr gemeinsame Zugriffsstrukturen auf die unterschiedlichen Inhalte entwickelt werden können und je besser die lexikografischen Daten vernetzt sind. Daneben sollten die beteiligten Wörterbücher als einzelne Produkte möglichst vielfältige Zugriffsstrukturen haben, um die Möglichkeiten des elektroni-

schen Mediums angemessen zu nutzen; dies alles natürlich nur, sofern es den Inhalten angemessen ist. Dabei war es eine große Chance, dass alle am Portal beteiligten Projekte im IDS erstellt werden und so der kollegiale Austausch an sich einfach ist. Der Portal-Gedanke war und ist daher in OWID auch wichtig für die einzelnen lexikografischen Prozesse, d.h. für die Erarbeitung der Wörterbücher. Um die Basis für gemeinsame Zugriffsstrukturen auf die Inhalte zu legen, wurden daher einheitliche Prinzipien bei der Modellierung und Strukturierung der Inhalte für alle eingebundenen Produkte angewandt. Unbenommen davon kann das Portal perspektivisch auch um extern erarbeitete Produkte bereichert werden.

Daneben sollte insbesondere das *ellexiko*-Wörterbuch, das als rein elektronisches Wörterbuch geplant war, viele Anforderungen bzw. Wünsche der modernen Wörterbuchforschung vor allem in Bezug auf die Modellierung aufgreifen (vgl. Haß 2005, darin insbesondere Müller-Spitzer 2005). Dazu gehört, die Basis für flexible Zugriffs- und Darstellungsmöglichkeiten in der Modellierung zu legen und die Einhaltung der formalen Artikelstruktur durch eine strenge Modellierung zu unterstützen.

Der Gedanke war also, dass die unter dem Dach von OWID vereinten verschiedenen lexikografischen Produkte nicht nur auf der Ebene der Makrostruktur, also vereinfacht gesagt im Bereich der Stichwörter, sondern auch auf feinerer Ebene zugreifbar und daher vernetzbar sein sollen. Daher wurde versucht, die Art der Modellierung auf der Ebene der Inhaltsstrukturen, d.h. auf der Ebene der einzelnen lexikografischen Angaben, aufeinander abzustimmen und nicht unabhängig voneinander die einzelnen lexikografischen Prozesse unterschiedlich zu organisieren. Somit stellt OWID langfristig eine andere Art von Portal dar als zum Beispiel das von der Berlin-Brandenburgischen Akademie der Wissenschaften und von der Heidelberger Akademie der Wissenschaften betreute [Wörterbuch-Portal](#), das vorwiegend eine Einstiegsseite für verschiedene Wörterbücher darstellt. Auch eine gemeinsame Einstiegsseite auf sehr unterschiedliche lexikografische Projekte ist sinnvoll, doch da in OWID im Moment nur Projekte des IDS vertreten sind, bestand hier die Chance, die Inhalte von Anfang an tiefgehend zu vernetzen und damit die Basis für flexiblere Zugriffsmöglichkeiten zu legen. Denn sehr heterogene Datenbasen können zwar über Metastrukturen wie semantische Netze erschlossen werden, ohne die Datenbasis zu verändern, aber je nachdem, wie die einzelnen lexikografischen Inhalte strukturiert und ausgezeichnet sind, sind die dabei entstehenden Schwierigkeiten nicht zu unterschätzen. Schon verschiedene Stichwortlisten unterschiedlicher Wörterbücher zusammenzubringen, kann – insbesondere bei historischen Wörterbüchern – eine nur mit viel Aufwand zu bewältigende Herausforderung sein, da die Prinzipien des Lemmaansatzes von Wörterbuch zu Wörterbuch differieren. Daneben soll die gleichartige Organisation der einzelnen lexikografischen Prozesse in OWID auch Synergieeffekte hinsichtlich des geforderten Know-how, der eingesetzten Software etc. erzielen. Insgesamt galt es also, bei der Modellierung zum einen die Eigenständigkeit der Produkte zu wahren und zum anderen gleichzeitig die Integrierbarkeit der Daten zu gewährleisten.

2. Die Modellierung in Stichworten

Die Modellierung erfolgt für alle Projekte im Prinzip gleich: Aus Gründen der Langlebigkeit der Daten und aus weiteren Gründen, die hier nicht erläutert werden sollen, wurde eine XML-basierte Modellierung gewählt. Genauer: Es wurde eine maßgeschnei-

derte Modellierung für jedes beteiligte Wörterbuch entwickelt. Dies vor allem deshalb, weil die Standardmodellierung beispielsweise der TEI nicht so granular und passend aufgebaut ist, dass bestmögliche Flexibilität für die Darstellung und Recherche gegeben ist und vor allem die Standardmodellierungen kaum Unterstützung im lexikografischen Prozess bieten (vgl. Schmidt/Müller 2001). Es war aber eine ganz wichtige Anforderung, dass ein Instruktionbuch sich so weit wie möglich erübrigt (vgl. Wiegand 1998, S. 210ff.). D.h., die Lexikografen müssen sich nicht selbst merken, wann wo welche Angabe kommt – dies wird beim Arbeiten im XML-Editor vorgegeben. Unberührt davon sind natürlich Festsetzungen hinsichtlich der Beschreibungssprache bestimmter Angaben, die nicht maschinell kontrollierbar sind.

Es wurden also XML-DTDs bzw. Schemata für jedes einzelne Produkt entwickelt; allerdings wurde im Sinne der Vernetztheit der Projekte ein Modulsystem entwickelt, in dem gleiche Phänomene identisch und auch nur einmal modelliert sind. Die Wörterbuchartikel werden im XML-Editor verfasst und in einem Oracle-Datenbanksystem gespeichert. Um eine Online-Browseransicht zu ermöglichen, werden die XML-Daten mit einem XSLT-Stylesheet transformiert (vgl. den Beitrag von Wolfgang Bock in dieser Publikation).

3. Der Aufbau des DTD-Modulsystems

3.1 Grundlegende Organisation

DTD-Bibliothek für OWID				
Bausteine für das gesamte Portal		allg-entities.dtd	allg-elemente.dtd	
Bausteine für modulübergreifende Objektgruppen	ewl-objekte.dtd	mwl-objekte.dtd	ewl_mwl-objekte.dtd	ewl-grammatik.dtd
Bausteine einzelner Wörterbücher		elexikoBA-allgobj.dtd	neo-allgobj.dtd	
Kopf-DTDs für einzelne Wörterbücher	elexiko elexikoAA-ewl.dtd elexikoBA-ewl.dtd	Neologismen neo-ewl.dtd neo-mwl.dtd	Wortverbindungen mwl.dtd wv.dtd	Schulddiskurs zeitreflexion 1945-55.dtd

Abbildung 1: Tabellarische Übersicht zum DTD-Modulsystem von OWID

Die Organisation der XML-Modellierung kann folgendermaßen skizziert werden: Um vom Inhalt her gleiche lexikografische Angaben aus den verschiedenen Wörterbüchern möglichst auch gleich zu strukturieren, wurde eine DTD-Bibliothek für OWID erstellt, in der in einzelnen DTDs die Entitys, Elemente oder Attribute zusammengefasst sind, die

in verschiedenen Artikelstrukturen gleich sind. Beispielsweise gibt es zwei DTDs für allgemein verwendete Entitys bzw. Elemente und Attribute, die in allen in OWID eingebundenen Projekten verwendet werden (vgl. Abbildung 1). Dabei handelt es sich um so allgemeine Objekte wie IDs zur Identifikation von Elementen oder die Modellierung von Listen, aber auch um inhaltlich wichtige Angaben, die in allen Strukturen in gleicher Weise eingesetzt sind, wie bestimmte Arten von Kommentaren, verschiedene Arten von Hinweisen oder Belege. Unterhalb dieser ganz allgemeinen Ebene, die in allen (Teil-)Produkten von OWID verwendet wird, gibt es in der DTD-Bibliothek Objektgruppen, die beispielsweise im *ellexiko*- und im Neologismenwörterbuch gleich sind. Auch diese sind wieder in einzelnen DTDs zusammengefasst, wie zum Beispiel in der *ewl-objekte.dtd* alle Objekte für Einwortlemmata, die sowohl im *ellexiko*- wie im Neologismenwörterbuch identisch sind, oder in der *ewl-grammatik.dtd* alle Objekte zur Einwortlemma-Grammatik. Dort finden sich beispielsweise alle Angaben aus dem Bereich der Deklination oder Valenz, die im *ellexiko*- und im Neologismenwörterbuch gleich strukturiert sind. Durch diese Aufteilung ist schon anhand der Modellierung zu sehen, dass diese Angaben damit auch gemeinsam über die verschiedenen Wörterbücher hinweg zugreifbar sind.

3.2 Veranschaulichung anhand von Beispielen

3.2.1 Angaben zur Wortbildung (*ellexiko* und Neologismenwörterbuch)

Allerdings ist es anhand dieser Übersicht schwer vorstellbar, was dieses Modulsystem für die Präsentation der einzelnen Wörterbücher bedeutet. Deshalb sollen die Grundprinzipien anhand eines kleinen Beispiels veranschaulicht werden, und zwar an einem Ausschnitt der Angaben zur Wortbildung. Diese Angaben sind Teil der *ewl-objekte.dtd*, d.h. für das *ellexiko*- und das Neologismen-Wörterbuch gleich.

Bevor die Beispielinstanzen und Artikel gezeigt werden, ist in Abbildung 2 ein Ausschnitt aus der DTD-Dokumentation zu sehen, der einen Einblick in die Modellierung der Wortbildungsangaben geben soll.

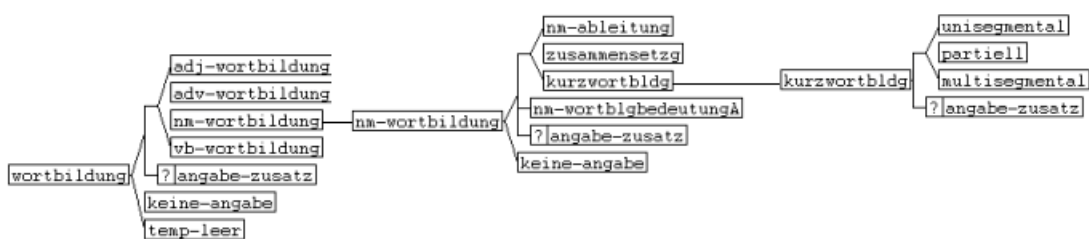


Abbildung 2: Ausschnitt aus der DTD-Dokumentation (Angaben zur Wortbildung – *ellexiko*- und Neologismen-Wörterbuch)

Die Modellierung der Wortbildungsangaben richtet sich – wie in der Grafik zu erkennen ist – zunächst nach der Wortart des Stichworts, denn Nomina können beispielsweise anders gebildet sein als Verben. Deshalb gliedert sich das Element *wortbildung* zunächst in die einzelnen Wortarten. Unter Wortbildung für Nomen kann man dann die einzelnen Wortbildungstypen sehen, die infrage kommen:

- Ableitung (nm-ableitung)
- Zusammensetzung (zusammensetzg)
- Kurzwortbildung (kurzwortbldg)

Da gleich Beispiele für Kurzwortbildungen folgen, ist rechts in Abbildung 2 die Modellierung dieses Elements noch mit aufgezeigt. Man sieht dabei, dass die Kurzwortbildungen nochmals in unisegmentale, partielle und multisegmentale unterschieden werden können.

Zwei Instanzen sollen nun zur Veranschaulichung gezeigt werden: [Video](#) aus dem *lexiko*-Wörterbuch und [Fon](#) aus dem Neologismen-Wörterbuch. (Die folgenden Ausschnitte zeigen Teile der XML-Instanzen mit den Angaben zur Wortbildung.)

* Ausschnitt aus video.xml

```
<wortbildung>

<nm-wortbildung><kurzwortbldg>
<unisegmental typ="kopfwort">
<kurzwortA basistyp="nomen" artikel-refid="108176"
lesart-refid="0" ltspez-refid="0">Videofilm</kurzwortA>

<angabe-zusatz>
<kommentar>
<lex-interpretationK><k-absatz> <obj-text>Video</obj-text> ist
gleichzeitig Kurzwort zu <obj-text>Videoclip</obj-text>,
<obj-text>Videoband</obj-text>, <obj-text>Videokamera
</obj-text>, <obj-text>Videokassette</obj-text>,
<obj-text>Videorekorder</obj-text> und
<obj-text>Videotechnik</obj-text>. Auch in diesen Fällen handelt
es sich um ein unisegmentales Kopfwort (vgl. die folgenden Be-
lege).
<belege>[...]</belege>
</k-absatz></lex-interpretationK>
</kommentar>
</angabe-zusatz>

</unisegmental>
</kurzwortbldg>
<nm-wortbldgbedeutungA bezeichnung="keine"/>
</nm-wortbildung>
</wortbildung>
```

Für [Video](#) ist hier zu sehen, dass der Lexikograf angegeben hat, dass dieses Stichwort eine Kurzwortbildung aus *Videofilm* ist. Die Attribute *artikel-refid* usw. dienen dabei der Vernetzung der Stichworte untereinander. Unter Auswertung dieser Angaben könnten später beispielsweise interessante Wortbildungsnetze erstellt werden. Genauso dient die Angabe des Basistyps für spätere Recherchemöglichkeiten: So könnten z.B. alle Bildungen aus Nomen oder Verben in einer Expertensuche abgefragt werden.

Als Pendant dazu sieht man hier den Artikelausschnitt aus [Fon](#). *Fon* ist eine unisegmentale Kurzwortbildung vom Typ „Endwort“, gebildet aus *Telefon*. Auch hier dienen die Attribute der Vernetzung.

* Ausschnitt aus Fon.xml

```
<wortbildung><nm-wortbildung>
<kurzwortbldg>
<unisegmental typ="endwort">
<kurzwortA
  artikel-refid="105886" lesart-refid="0"
  ltspez-refid="0"basistyp="nomen">
  Telefon
</kurzwortA>
</unisegmental>
<angabe-zusatz><kommentar><lex-interpretationK>
<k-absatz>analog zu <obj-text>Fax</obj-text> gebildet</k-absatz>
</lex-interpretationK></kommentar></angabe-zusatz>
</kurzwortbldg>
</wortbildung>
```

Außerdem sieht man in diesem Ausschnitt weiter unten noch einen Kommentar, nämlich dass diese Bildung analog zu der von *Fax* ist. Solche Kommentare werden im *lexiko*- und im Neologismen-Wörterbuch oft genutzt, da so den Wörterbuchbenutzern interessante zusätzliche Informationen angeboten werden können, die im gedruckten Wörterbuch wegen der Platzbeschränkungen meist wegfallen müssen.¹

Wie sind diese Daten nun von der Benutzeroberfläche her zugreifbar? Es werden im Moment (Ende 2007) erst grundlegende Ansätze dessen gezeigt, was von der Modellierung und Aufbereitung der Daten möglich wäre. Aber es gibt eine Suche nach Eigenschaften, wobei die Suche nach den einzelnen Wortbildungstypen eine davon ist. Man kann also in der [Erweiterten Suche](#) über die Auswahlmöglichkeiten unter „Wortbildung“ nach allen Kurzwörtern in *lexiko* suchen (nur in den bearbeiteten Artikeln). In der Artikelansicht sehen die Angaben zur Wortbildung wie in Abbildung 3 gezeigt aus.

Video

Lesartenübergreifende Angaben

Orthografie Info	
Normgerechte Schreibung:	Video
Worttrennung:	Vi deo
Wortbildung Info	
Wortbildungsart/-typ:	unisegmentales Kurzwort Typ: Kopfwort
Gekürztes Wort:	<u>Videofilm</u> Wortart: Nomen

Abbildung 3: Ausschnitt aus dem Artikel *Video* aus *lexiko*

¹ Solche Analogiebildungen für die Neologismen sollen zukünftig explizit in einem XML-Attribut angegeben werden, sodass sie direkt recherchierbar sind.

Die Angaben zur Wortbildung sind Teil der lesartenübergreifenden Angaben, d.h. Angaben, die nicht zu einer Einzelbedeutung gemacht werden, sondern zum ganzen Stichwort. In der Abbildung ist zu sehen, wie die in XML gemachten Angaben durch Stylesheets umgesetzt werden: Als Wortbildungstyp ist „unisegmentales Kurzwort“ angegeben; darunter das gekürzte Wort, d.h. die Basis, außerdem noch die Wortart, die in der XML-Instanz im Attribut `basistyp` vermerkt ist.

Auch im Neologismenwörterbuch können die Kurzwortbildungen über die [Erweiterte Suche](#) abgefragt werden. Das [Suchergebnis](#) ist hier um einiges länger als in *ellexiko*, da das Phänomen der Kurzwortbildung bei Neologismen häufiger ist, als innerhalb des im *ellexiko*-Wörterbuch schon beschriebenen allgemeinen Wortschatzes. In Abbildung 4 ist dazu der Ausschnitt aus dem Artikel zu sehen, wie er online gezeigt wird.

Fon		
Lesartenübergreifende Angaben		
Neologismtyp:	Neulexem	
Aussprache [fo:n]		
Wortbildung		
Wortbildungsart/-typ:	Kurzwort	(unisegmental; Endwort)
Basis:	Telefon	(Nomen)

Abbildung 4: Ausschnitt aus dem Artikel *Fon* aus dem Neologismenwörterbuch

An diesem Beispiel ist auch zu erkennen, wie die in XML gleich modellierten Angaben auf der Benutzeroberfläche verschieden dargestellt werden:

- In *ellexiko* heißt es „unisegmentales Kurzwort“, im Neologismenwörterbuch „Kurzwort (unisegmental; Endwort)“.
- In *ellexiko* steht „Gekürztes Wort“ als Überschrift, bei den Neologismen „Basis“. Die Wortart der Basis ist im Neologismenwörterbuch in Klammern gesetzt, in *ellexiko* hinter die Überschrift „Wortart“.

Von außen betrachtet ist es vermutlich fraglich, ob eine solche unterschiedliche Terminologie für potenzielle Benutzer sinnvoll ist; begründet ist sie in der inhaltlichen Eigenständigkeit der Projekte. Jedes beteiligte Wörterbuch in OWID hatte daher auch die freie Wahl, wie die XML-Inhalte per Stylesheets umgesetzt werden sollen. Dieses kurze Beispiel sollte aber deutlich machen: Die granulare Modellierung der Daten in OWID ermöglicht eine sehr flexible Darstellbarkeit per Stylesheets, sodass die Daten – ohne sie zu verändern – beispielsweise für Experten und Laien sehr unterschiedlich dargestellt werden können. Dazu am Ende dieses Beitrags noch ein kurzer Ausblick.

3.2.2 Das Beispiel *Schulddiskurs*

Doch zunächst noch ein kurzer Blick auf ein anderes Wörterbuch aus OWID, welches die Verschiedenartigkeit der Produkte im Portal verdeutlichen kann: das Wörterbuch zum „Schulddiskurs im ersten Nachkriegsjahrzehnt“ (vgl. den Beitrag von Kämper/Rothenhöfer in dieser Publikation).

Die Artikel aus diesem Produkt sind sehr viel textlastiger als die Artikel des *ellexiko*- und des Neologismenwörterbuchs; sie sind eher mit kurzen Erörterungen vergleichbar. Daher war es hier unangemessen, eine kleinteilige Modellierung wie die eben dargestellte zu entwickeln. Die XML-Struktur für dieses Wörterbuch ist demnach auch sehr viel flacher, richtet sich aber nichtsdestotrotz nach dem gleichen Modellierungskonzept (vgl. Müller-Spitzer 2007b). Allerdings war die Vorgehensweise hier etwas anders: Die Artikel wurden erst in einem gedruckten Wörterbuch publiziert (Kämper 2006) und nachträglich in die für OWID entwickelte maßgeschneiderte XML-Struktur gebracht. Diese Überführung wurde weitgehend manuell vorgenommen, da dies erstens relativ schnell ging und zweitens nicht automatisch zu vollziehende inhaltliche Einordnungen gleich im ersten Schritt mit vorgenommen werden konnten. Doch auch mit dieser flachen Strukturierung konnte ein echter Mehrwert gegenüber der Buchversion geschaffen werden.

Beispielsweise kann man die Belege – wie z.B. im Artikel [frei](#) – schon im Text in einem Pop-up-Fenster lesen, man kann in den Belegblock und wieder zurück springen und die gesamte Artikelmenge nach bestimmten Eigenschaften durchsuchen. Die Vernetzung dieser Artikel mit den anderen im Portal enthaltenen Wörterbüchern ist natürlich aufgrund der sehr speziellen Inhalte begrenzt. Aber um auch hier mehr als eine Zusammenstellung von Wörterbüchern zu erreichen, wurden beispielsweise in *ellexiko* bereits vorrangig die in diesem Modulprodukt enthaltenen Stichwörter auch in ihrer allgemeinen Bedeutung beschrieben, sofern sie im „[Lexikon zum öffentlichen Sprachgebrauch](#)“ enthalten sind, also zu dem Teil der Stichwortliste gehören, der bereits bearbeitet wurde bzw. in den nächsten Jahren bearbeitet wird.

3.2.3 Kurzer Blick auf *Wortverbindungen online*

Um abschließend kurz den Bogen zum vierten im Portal enthaltenen Produkt – den [Wortverbindungen online](#) – zu schlagen, soll an einem Beispiel die Suche über ganz OWID gezeigt werden. Sucht man über das Portal-Suchfenster² beispielsweise nach „frei“, ist am Suchergebnis (vgl. Abbildung 5) zu sehen, dass hier auch eine Wortverbindung – [frank und frei](#) – mitgefunden wird. Basis dafür ist keine Volltextsuche auf der Lemmaposition oder etwas Ähnliches, sondern die XML-basierte Auszeichnung der Inhaltsstrukturen, genauer die Auszeichnung der Basiselemente der Wortverbindung (vgl. den Beitrag von Steyer/Brunner in dieser Publikation).

² Über das Suchfenster auf der [Startseite](#) wird über das gesamte Portal gesucht, ebenso auf der Suchfläche links oben am Bildschirm, wenn eines der Wörterbücher aufgerufen ist.

Suchergebnis in OWID

Ihre Suchanfrage ergab 4 Treffer. Neue Suche

- **frei**
- **frei**
- frei *Basiselement* zu **frank und frei**
- frei *Sublemma* zu **Freiheit**

schwarz (bzw. grau) sind die Stichwörter aus „*lexiko*“ (dunkelgrau sind in Bearbeitung befindliche Stichwörter, die redaktionell ausgearbeiteten Stichwörter sind fett hervorgehoben)
blau sind die Stichwörter aus dem „Neologismenwörterbuch“
rot sind die Stichwörter von „Feste Wortverbindungen“
grün sind die Stichwörter aus dem „Diskurswörterbuch 1945-55“

Abbildung 5: Suchergebnis zu *frei*

Was ist nun also resümierend der aktuelle Stand? Zum einen ist wichtig, festzuhalten: Die Möglichkeiten, die durch die Modellierung gelegt sind, sind sehr vielfältig. Genau dies war der Grund, so viel Gedanken und Zeit in die Modellierung zu investieren. Im Internetauftritt des Portals sind erste Ansätze dieses Potenzials schon erkennbar. Allerdings können die Benutzer von OWID noch längst nicht in vollem Umfang erkennen, wie feinkörnig die Daten modelliert sind. Auch ist es für uns Lexikografen noch nicht befriedigend, wie wir selbst in den Daten recherchieren können. Wie sind darum die Perspektiven?

4. Perspektiven

Ein Schritt für die nähere Zukunft sollte sein, gleiche Suchanfragen – wie die vorhin gezeigte – gemeinsam über unterschiedliche Wörterbücher starten zu können. Dies ist im Moment noch nicht möglich. So wäre es beispielsweise sinnvoll, nach allen Kurzwörtern im *lexiko*- und im Neologismenwörterbuch gemeinsam suchen zu können.

In der elektronischen Lexikografie ist es außerdem ein sehr wichtiges Thema, die Benutzerschnittstelle von elektronischen Wörterbüchern an die Typen von Benutzungssituationen adaptierbar zu gestalten. Denn Laien suchen beispielsweise ganz andere Informationen als Experten; oder in der Benutzungssituation der Textproduktion sind andere Informationen gefragt als bei der Textrezeption (vgl. u.a. Storrer 2001 und Schryver 2003).

Daher wurde eine rein auf die Inhalte bezogene Modellierung entworfen – frei von Gesichtspunkten der Präsentation, aber auch frei von konkreten Benutzerbezügen. Wie ausgeführt, wurde dies deshalb so gemacht, um ohne Veränderung der Datenbasis die gleichen Daten für unterschiedliche Benutzertypen und Benutzungssituationen verschieden darstellen zu können. Dabei ist bisher für die elektronische Lexikografie kaum systematisch untersucht worden, welche Funktionalitäten für genau welche Benutzergruppen und Benutzungssituationen sinnvoll einzusetzen sind. Vor dem Hintergrund unserer granularen, auf die Inhalte bezogenen Modellierung wäre es daher sehr spannend zu erforschen, wo die Grenze einer solchen Vorgehensweise liegt, d.h., wo zum Beispiel Angaben doppelt erstellt werden müssten, wenn sehr unterschiedliche Benut-

zergruppen angesprochen werden sollen. Die Bedeutungsparaphrase könnte ein Beispiel für eine solche Angabe sein. Denn bei der Entwicklung von Richtlinien für die Paraphrasenformulierung ist unweigerlich ein bestimmter Benutzerbezug gegeben. Trotzdem spricht meiner Meinung nach vieles dafür, die Modellierung grundsätzlich nicht benutzerbezogen vorzunehmen.³

Auch wenn in diesem Bereich also noch viel Forschungsarbeit geleistet werden muss, könnten auch hier in OWID, besonders im *lexiko*-Wörterbuch, erste Schritte gemacht werden. Gerade die granulare Modellierung – verbunden mit der Darstellung über Stylesheets – ermöglicht es, technisch gesehen relativ schnell unterschiedliche Sichten auf den Datenbestand zu definieren.

So ist in der folgenden Abbildung zur Veranschaulichung eine Kurzansicht für die Artikel aus dem *lexiko*-Wörterbuch gezeigt, die nur einen Bruchteil der gesamten Angaben aus einem Artikel über das Stylesheet zur Darstellung bringt.⁴ Dies könnte beispielsweise für Benutzungssituationen, in denen sich die Benutzer nur einen kurzen Überblick über die Schreibung und das Bedeutungsspektrum eines Stichworts verschaffen wollen, hilfreich sein.

The screenshot shows a dictionary entry for the word 'Video'. It is organized into several sections, each with an 'Info' button:

- Video** (Section Header)
- Allgemeine Angaben** (Section Header)
 - Rechtschreibung** (Section Header)
 - Korrekte Schreibung: Video
 - Worttrennung: Vi|deo
 - Wortbildung** (Section Header)
 - Wortbildungsart/-typ: unisegmentales Kurzwort Typ: Kopfwort
 - Gekürztes Wort: Videofilm Wortart: Nomen
 - Nationale Verteilung** (Section Header)
 - Dieses Stichwort ist in Texten des folgenden Sprachraumes belegt: BRD, Österreich, Schweiz
- Einzelbedeutungen** (Section Header)
 - Das Stichwort verfügt über 2 Lesart(en).
 - Lesart 'Film' verfügt über 1 Spezifizierung(en).
 - Lesart 'Film'**
 - Mit Video bezeichnet man einen Film, der auf einer Kassette (oder auch auf einer CD) aufgezeichnet ist und über ein Abspielgerät (Videorekorder oder auch DVD-Spieler) gesehen werden kann.
 - Spezifizierung 'Musikclip'**
 - Mit Video bezeichnet man einen kurzen Musikclip, der zu einem Musiktitel gefilmt wird.
 - Hinweis:**
 - Klicken Sie auf eine Lesart bzw. Spezifizierung, um zu weiteren Angaben zu kommen.

Abbildung 6: Ausschnitt aus einer möglichen Kurzansicht eines Artikels in *lexiko* (Demonstration)

³ Vgl. auch die Bewertungen der Modellierung von OWID bzw. *lexiko* in Kunze/Lemmitzer (2007) und Schlaps (2007).

⁴ Ich danke Wolfgang Schlaupitz für die Erstellung des zugrunde liegenden Stylesheets.

Ein weiteres wichtiges Anliegen für die Zukunft ist der Ausbau von Recherchemöglichkeiten, sodass die Vielfältigkeit der Modellierung gerade für Experten auch auswertbar wird. Außerdem wurde bereits ein texttechnologisches Konzept für den Ausbau der Vernetzungsstrukturen erarbeitet, was vor allem im Bereich der Wortbildung und der sinn- und sachverwandten Wörter innovativere Formen der Darstellung erlauben wird (vgl. Müller-Spitzer 2007a und den Beitrag von Storjohann in dieser Publikation).

Allerdings ist nicht zu vergessen: Im Mittelpunkt der täglichen Projektarbeit steht der Ausbau der vorhandenen Wörterbücher. Denn auch die interessantesten Recherche- und Darstellungsmöglichkeiten bringen wenig, wenn nicht genug Daten zu durchsuchen sind. Deshalb muss für die skizzierten Pläne Geduld aufgebracht werden; sie sollen aber nichtsdestotrotz beharrlich verfolgt werden.

5. Literatur

- Büchel, Gregor/Schröder, Bernhard (2001): Verfahren und Techniken in der computergestützten Lexikographie. In: Lemberg/Schröder/Storrer (Hg.), S. 7-28.
- 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.
- Kämper, Heidrun (2006): Opfer – Täter – Nichttäter. Ein Wörterbuch zum Schulddiskurs 1945-55. Berlin/New York.
- Kunze, Claudia/Lemnitzer, Lothar (2007): Computerlexikographie. Eine Einführung. Tübingen.
- Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.) (2001): Chancen und Perspektiven computergestützter Lexikographie. (= *Lexicographica*, Series Maior 107). Tübingen.
- Müller-Spitzer, Carolin (2005): Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess. In: Haß (Hg.), S. 20-54.
- Müller-Spitzer, Carolin (2006): Das Konzept der Inhaltsstruktur. Ein Ausschnitt aus einer neuen Konzeption für die Modellierung lexikografischer Daten. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2006). Mannheim. Internet: www.ids-mannheim.de/opal.
- Müller-Spitzer, Carolin (2007a): Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung. In: *Hermes* 38, S. 137–171.
- Müller-Spitzer, Carolin (2007b): Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis. (= *Studien zur deutschen Sprache* 42). Tübingen.
- Schmidt, Ingrid/Müller, Carolin (2001): Entwicklung eines lexikographischen Modells: Ein neuer Ansatz. In: Lemberg/Schröder/Storrer (Hg.), S. 29-52.
- Schlaps, Christiane (2007): Grundfragen der elektronischen Lexikographie. *Elexiko* - das Online-Informationssystem zum deutschen Wortschatz. Hg. v. Ulrike Hass. Berlin, New York: de Gruyter 2005. Short review. In: *Lexicographica* 22 (2006), S. 311-314.
- Schryver, Gilles-Maurice de (2003): Lexicographer's dreams in the electronic-dictionary age, in: *International Journal of Lexicography* 16 (2/2003), S. 143-199.
- 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 (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie, 1. Teilbd. Berlin/New York.