



INSTITUT FÜR
DEUTSCHE SPRACHE

OPAL

Online publizierte Arbeiten zur Linguistik

ISSN 1860-9422

Sonderheft

1/2008

Annette Klosa (Hg.)

Lexikografische Portale im Internet

(= OPAL-Sonderheft 1/2008)

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.

Inhalt

Vorwort.....	3
<i>Annette Klosa / Lothar Lemnitzer / Gerald Neumann</i> Wörterbuchportale – Fragen der Benutzerführung.....	5
<i>Wolfgang Bock</i> Technische Aspekte des OWID-Portals	37
<i>Carolin Müller-Spitzer</i> Der texttechnologische Aufbau von OWID	45
<i>Marion Hahn / Annette Klosa / Carolin Müller-Spitzer / Ulrich Schnörch / Petra Storjohann</i> <i>ellexiko</i> – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche.....	57
<i>Heidrun Kämper / Andreas Rothenhöfer</i> Das Online-Wörterbuch zum Schulddiskurs 1945-1955 in OWID.....	87
<i>Doris Steffens</i> Zu den Verlinkungen in den Wörterbüchern des Internetportals OWID – ein Erfahrungsbericht aus Sicht des Neologismenwörterbuches (NWB)	97
<i>Kathrin Steyer / Meike Lauer / Annelen Brunner</i> <i>Usuelle Wortverbindungen in OWID</i> . Lexikografische Aspekte unter besonderer Berücksichtigung der Pragmatik.....	107
<i>Werner Scholze-Stubenrecht</i> Zu den neuen lexikografischen Inhalten in OWID.....	123
<i>Thomas Burch</i> Fragen der Vernetzung in OWID und im <i>Wörterbuchnetz</i>	131
<i>Fons Moerdijk</i> Das „Algemeen Nederlandse Woordenboek“ (ANW) und <i>ellexiko</i> – ein Vergleich.....	143
<i>Jörg Asmussen</i> <i>ordnet.dk</i> und <i>ellexiko</i> – ein Vergleich	153
<i>Andrea Abel</i> ELDIT (Elektronisches Lernerwörterbuch Deutsch – Italienisch) und <i>ellexiko</i> : ein Vergleich	175

Vorwort

Nachschlagewerke, egal ob Enzyklopädien oder Wörterbücher, stehen seit einiger Zeit nicht mehr nur im Bücherregal, sondern sie werden auch in elektronischer Form auf CD-ROM und insbesondere im Internet angeboten. Das Angebot an Online-Wörterbüchern zu sehr vielen Sprachen dieser Welt wächst ständig, sodass es inzwischen zahlreiche Linklisten gibt, in denen Online-Wörterbücher verzeichnet und gegebenenfalls nach Sprachen oder Typen geordnet sind (z.B. [Labyrinth der Links – Dictionaries, Encyclopedia, Glossaries, Online-Wörterbücher](#)).

Etwas umfangreicher sind Seiten, die Links zu verschiedenen Internetwörterbüchern anbieten und daneben unter anderem den Zugriff auf sprachtechnologische Werkzeuge (z.B. Übersetzungstools), sprachbezogene Spiele (z.B. Kreuzworträtsel) und enzyklopädische Nachschlagewerke ermöglichen (z.B. [Dictionary Link](#), [Dictionary.com](#)). Zum Teil kann man auf solchen Seiten auch in ausgewählten Wörterbüchern direkt nachschlagen.

Schließlich werden mehrere Online-Wörterbücher auch über so genannte Portale zugänglich gemacht. Das Besondere hierbei ist, dass gleichzeitig in mehreren Online-Wörterbüchern nachgeschlagen werden kann. Dass es auch hierfür verschiedene Lösungen gibt, zeigt der Vergleich zwischen zwei deutschsprachigen Wörterbuchportalen, nämlich [OWID](#), dem Online-Wortschatzinformationssystem des Instituts für Deutsche Sprache, Mannheim, und dem [Wörterbuch-Portal](#) der Berlin-Brandenburgischen Akademie der Wissenschaften und der Heidelberger Akademie der Wissenschaften.

Der Präsentation eines dieser Projekte, nämlich des neu aufgebauten und gestalteten Portals OWID, war ein Kolloquium vom 10. und 11. Mai 2007 am Institut für Deutsche Sprache in Mannheim gewidmet. Das Portal wurde damals noch als *ellexiko*-Portal präsentiert, seit 2008 trägt das Portal nun den Namen OWID, der Name *ellexiko* beschränkt sich auf eines der innerhalb von OWID publizierten Wörterbücher. Das Kolloquium hatte das Ziel, nicht nur das Portal in seiner neuen Gestalt und mit den verschiedenen lexikografischen Inhalten zu präsentieren, sondern vor allem kritisch zu diskutieren. Dabei standen vor allem Fragen der Benutzerführung wie der Verlinkung im Portal im Vordergrund, aber auch der Vergleich mit anderen Online-Wörterbüchern sollte Rückmeldung dazu geben, ob die für OWID und die darin enthaltenen Wörterbücher gefundenen Lösungen der Online-Präsentation, der Recherchemöglichkeiten usw. gelungen oder verbesserungsfähig sind.

Den Anfang macht ein vergleichender Beitrag zwischen OWID und dem „Wörterbuch-Portal“ (**Annette Klosa / Lothar Lemnitzer / Gerald Neumann**: *Wörterbuchportale – Fragen der Benutzerführung*), wobei der Schwerpunkt auf der Darstellung der Konzeption zur Benutzerführung liegt, die jeweils kritisch hinterfragt wird. Ausführlich wird der technische bzw. texttechnologische Aufbau von OWID in den Beiträgen von **Wolfgang Bock**: *Technische Aspekte des OWID-Portals* und **Carolin Müller-Spitzer**: *Der texttechnologische Aufbau von OWID* dargestellt. Insbesondere in der einheitlichen texttechnologischen Modellierung der einzelnen Wörterbücher des Portals besteht der Hauptunterschied zu anderen lexikografischen Portalen wie z.B. dem „Wörterbuch-Portal“.

Die einzelnen Wörterbücher, die derzeit in OWID online zur Verfügung stehen, werden in den folgenden Aufsätzen präsentiert. **Marion Hahn / Annette Klosa / Carolin Müller-Spitzer / Ulrich Schnörch / Petra Storjohann** stellen in *lexiko – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche* dar, in welchen Bereichen *lexiko* inhaltlich neu konzipiert und ergänzt wurde, woran sich einige grundsätzliche Überlegungen zur Arbeit an einem im Aufbau befindlichen Online-Wörterbuch anschließen. **Heidrun Kämper / Andreas Rothenhöfer** stellen erstmals *Das Online-Wörterbuch zum Schulddiskurs 1945-1955* vor und weisen insbesondere auf die unterschiedlichen Zugriffsstrukturen hin, die natürlich deutlich über das hinausgehen, was das gedruckte Wörterbuch ermöglicht. **Doris Steffens** konzentriert sich in ihrem Beitrag *Zu den Verlinkungen in den Wörterbüchern des Internetportals OWID – ein Erfahrungsbericht aus Sicht des Neologismenwörterbuches (NWB)* auf eine der wichtigsten Fragen zu einem lexikografischen Portal, nämlich der Verlinkung der einzelnen Wörterbücher untereinander. Schließlich präsentieren **Kathrin Steyer / Meike Lauer / Annelen Brunner** in ihrem Aufsatz *Usuelle Wortverbindungen in OWID. Lexikografische Aspekte unter besonderer Berücksichtigung der Pragmatik* die Konzeption für die korpusbasierte lexikografische Beschreibung von Wortverbindungen in OWID.

Werner Scholze-Stubenrecht zeigt in seinem Beitrag *Zu den neuen lexikografischen Inhalten in OWID* an einigen Beispielen, wie sich die vielfältigen Informationen aus den einzelnen Wörterbüchern im OWID-Portal ergänzen. **Thomas Burch** kontrastiert in *Fragen der Vernetzung in OWID und im Wörterbuchnetz* die jeweils gefundenen Lösungen für die Verlinkung der Inhalte untereinander.

Den vorliegenden Band schließen drei Beiträge ab, die das Wörterbuch *lexiko* mit anderen Online-Wörterbüchern vergleichen. Den Anfang macht *Das „Algemeen Nederlandse Woordenboek (ANW)“*, wobei **Fons Moerdijk** insbesondere auf die unterschiedlichen Recherchemöglichkeiten in *lexiko* und dem ANW eingeht. **Jörg Asmussen** vergleicht *ordnet.dk* und *lexiko* und arbeitet dabei heraus, welche Schwerpunkte beide Projekte bei der Konzeption ihres Online-Wörterbuches gesetzt haben. Schließlich zeigt der Vergleich zwischen *ELDIT (Elektronisches Lernerwörterbuch Deutsch – Italienisch)* und *lexiko*, den **Andrea Abel** zieht, deutlich die Unterschiede zwischen einem zweisprachigen und einem einsprachigen Online-Wörterbuch.

Die Beiträge des Kolloquiums sind hier in der dem Medium Internet angemessenen Form der Online-Publikation zusammengefasst, die es vor allem ermöglicht, die vielen Wortartikelbeispiele aus den Wörterbüchern online direkt nachzuschlagen. Die elektronische Publikation soll es uns zukünftig auch ermöglichen, die Links gegebenenfalls zu aktualisieren. Aktualisiert wurden einzelne Beiträge inzwischen schon dahingehend, dass (wo möglich) die aktuellen Seitenansichten in OWID präsentiert werden. Wir zeigen damit auch, dass wir die Kritik, die beim Kolloquium vor einem Jahr geäußert wurde, möglichst umgesetzt haben. Wir hoffen, dass die Beiträge rund um OWID und *lexiko* die Diskussion zu lexikografischen Portalen und Online-Wörterbüchern weiter anregen.

Annette Klosa

Annette Klosa / Lothar Lemnitzer / Gerald Neumann

Wörterbuchportale – Fragen der Benutzerführung

Abstract

In diesem Beitrag geht es um Fragen der Benutzerführung in lexikografisch-lexikologischen Portalen, und zwar insbesondere um die Portale OWID (Mannheim) und „Wörterbuch-Portal“ (Berlin). Diese werden mit ihrer jeweiligen Konzeption sowie ihrem technischen Aufbau vorgestellt und dann aus Benutzersicht bewertet. An Vorschläge für die Weiterentwicklung dieser Angebote schließen sich einige grundsätzliche Überlegungen zur Zukunft lexikografischer Portale an.

This contribution discusses questions of user interaction in dictionary portals. The dictionary portals OWID (Institut für Deutsche Sprache, Mannheim) and „Wörterbuch-Portal“ (Berlin-Brandenburgische Akademie der Wissenschaften) are presented as showcases. Their conceptual as well as their technical layer are discussed in detail. Both portals are evaluated with regard to their interaction with their prospective users. Suggestions for the advancement of these portals as well as some considerations about the future of internet-based dictionary portals conclude this paper.

Inhalt:

1. Was ist ein Portal bzw. was ist ein Wörterbuchportal?
 - 1.1 Definition
 - 1.2 OWID
 - 1.3 Wörterbuch-Portal
2. Struktur und Inhalt von OWID und Wörterbuch-Portal
 - 2.1 OWID
 - 2.2 Wörterbuch-Portal
3. OWID und Wörterbuch-Portal aus der Sicht der Benutzer
 - 3.1 Vorbemerkungen
 - 3.2 OWID
 - 3.3 Wörterbuch-Portal
 - 3.4 Einbeziehung des Benutzers
4. Empfehlungen für die Weiterentwicklung der Portale
 - 4.1 OWID
 - 4.2 Wörterbuch-Portal
5. Wo stehen und was leisten Wörterbuchportale?
6. Literaturverzeichnis
7. Verzeichnis der Nachschlagewerke

In diesem Beitrag werden, ausgehend von den Portalen OWID (www.owid.de; Institut für Deutsche Sprache, Mannheim) und dem Wörterbuch-Portal (www.woerterbuch-portal.de; Berlin-Brandenburgische Akademie der Wissenschaften und Heidelberger Akademie der Wissenschaften; im Folgenden WBP), Fragen der Benutzerführung in lexikografisch-lexikologischen Portalen diskutiert.¹ Dabei wird zunächst der Begriff des Portals definiert, bevor eine Analyse der Struktur von OWID und WBP und der darin angebotenen Inhalte gegeben werden kann. Ein Ausblick auf andere lexikografische Portale zur deutschen Sprache rundet diesen Teil ab. Einen weiteren Hauptteil dieses Beitrags bildet die Untersuchung von OWID und WBP aus der Sicht des Benutzers: Nach einer genauen Bestimmung der Zielgruppen sollen einige Benutzungssituationen für beide Portale durchgespielt werden, woraus sich Empfehlungen für die Weiterentwicklung der Angebote ergeben. Schließlich soll untersucht werden, was Wörterbuchportale im Vergleich zu anderen Anwendungen wie etwa Google leisten.

1. Was ist ein Portal bzw. was ist ein Wörterbuchportal?

1.1 Definition (Lothar Lemnitzer)

Der Terminus *Portal* im engeren Sinn, wie er vor allem in der Informatik verwendet wird, bezeichnet ein Anwendungssystem, das durch folgende Eigenschaften gekennzeichnet ist: a) die Integration von Anwendungen, Prozessen oder Diensten, b) Personalisierung des Zugriffs und c) die Präsentation ausgewählter oder ad hoc zusammengestellter Informationen. In einem weiteren Sinn werden mit diesem Terminus Anwendungen bezeichnet, die thematisch zusammengehörige Informationen bündeln und integrieren, sie dem Benutzer unter einer einheitlichen Oberfläche präsentieren und die Navigation in diesen Informationen erleichtern. Im Folgenden werden wir uns auf Webportale, also Webanwendungen, beschränken, werden jedoch gelegentlich zum Vergleich Anwendungen hinzuziehen, die Charakteristika von Portalen aufweisen, jedoch nicht im WWW verfügbar sind.

Ein notwendiges Merkmal eines Wörterbuchportals ist es natürlich, dass lexikalische Informationen erschlossen und präsentiert werden. Ein weiterer wichtiger Aspekt ist die Integration verschiedenartiger Informationen bzw. lexikalischer Ressourcen. Es ist letztlich diese Integration, die ein Wörterbuchportal von den vielen zum Teil endlosen Auflistungen lexikalischer Ressourcen unterscheidet. Letztere bestehen meist nur aus einem Namen und einem Link zur jeweiligen Ressource. Ein weiterer wichtiger Aspekt ist die Verlässlichkeit der ausgewählten und präsentierten Informationen. Ein in diesem Zusammenhang weniger wichtiges, für den potenziellen Benutzer jedoch bedeutsames Merkmal ist, ob die Benutzung der Informationen kostenfrei oder kostenpflichtig ist. Christa Knapp stellt in ihrer ausführlichen Untersuchung von Übersetzungswörterbüchern im WWW fest, dass ein klarer Zusammenhang zwischen den Nutzungskosten und der Qualität der dargebotenen Informationen besteht (vgl. Knapp 2006, S. 70).

Im Mittelpunkt dieses Artikels steht die **Benutzerführung** in zwei ausgewählten Wörterbuchportalen. Die Verlässlichkeit der dargebotenen Informationen ist deshalb hier

¹ In den Abschnitten 1.2 und 2.1 wird der aktuelle Stand von OWID gezeigt; wegen der in Abschnitt 3.2 und 4.2 beschriebenen Kritikpunkte wurde hier der Stand von Mai 2007 beibehalten. Dies ermöglicht auch, die Dynamik eines Online-Wörterbuchportals nachzuvollziehen.

nur von sekundärem Interesse. Die Benutzung der beiden beschriebenen Wörterbuchportale ist im Übrigen kostenfrei.

1.2 OWID (*Annette Klosa*)

Das Portal OWID (Online-Wortschatzinformationssystem Deutsch) des Instituts für Deutsche Sprache, Mannheim, geht zurück auf eine schon vor einigen Jahren (vgl. Haß 2005, S. 13) getroffene Grundsatzentscheidung, am Institut ein Online-Informationssystem zum deutschen Wortschatz der Gegenwart entstehen zu lassen:

Wissenschaftliches Ziel des Projekts ist, auf dem aktuellen Stand der lexikografischen, lexikologischen und korpuslinguistischen Forschung und Methodik ein langfristig angelegtes Nachschlagewerk zu Fragen des deutschen Gegenwartswortschatzes zu entwickeln, es auszubauen und immer wieder zu aktualisieren. (Haß 2005, S. 1)

Das hiermit beauftragte Projekt hieß zunächst *LEKSIS*, dann *WiW – Wissen über Wörter* und wurde schließlich auf der Suche nach einer geeigneten Internetadresse und, um dem eigentlichen Produkt einen passenden Namen zu geben, in *elexiko* umbenannt. Mit *elexiko* war dabei zunächst sowohl das Portal wie das gegenwartssprachliche Wörterbuch gemeint. Seit 2008 trägt das Portal den Namen OWID, der Name *elexiko* beschränkt sich auf eines der innerhalb von OWID publizierten Wörterbücher.

Von Anfang an war neben der Konzeption eines neuen gegenwartssprachlichen, korpusbasierten Online-Wörterbuches (*elexiko*) mitgedacht worden, im Informationssystem auch Artikelmenge aus anderen Projekten des Instituts zu publizieren; zunächst war dabei von so genannten Modulen die Rede:

elexiko wird auch in Zukunft modular weiterentwickelt und ausgebaut, d. h. nicht entlang dem Alphabet, sondern vertikal nach Teilwortschatzen bzw. horizontal nach bestimmten Angabearbeiten. (Haß 2005, S. 14)

Solche Module sollten einerseits innerhalb des Projektes *elexiko* erarbeitet werden (z.B. der so genannte [Demonstrationswortschatz](#)), andererseits aber auch in anderen Institutsprojekten (z.B. dem [Neologismen-Projekt](#)). Diese Grundidee des modularen Auf- und Ausbaus des Online-Informationssystems ist in erweiterter Form schließlich in das hier vorgestellte Gesamtkonzept für ein lexikografisches Portal des Instituts für Deutsche Sprache mit den nötigen technischen Umstrukturierungen und Programmierungen eingeflossen.

Den Begriff *Portal* gilt es hier für OWID zu bestimmen. Eine mögliche Definition ist die folgende; sie stammt aus dem Neologismenwörterbuch, das Bestandteil des Portals OWID ist:

Webseite, die als Einstieg in einen bestimmten Bereich, ein bestimmtes Themengebiet des Internets dient und den Zugang zu weiteren Webseiten ermöglicht. ([Portal](#); zuletzt eingesehen am 26.10.2007)

„Wikipedia – die freie Enzyklopädie“ ergänzt zum informatischen Begriff des Portals, dass damit ein „Anwendungssystem“ gemeint sei, in dem „Anwendungen, Prozesse und Dienste“ integriert seien und das u.a. die „Suche und Präsentation von Informationen“ ermögliche (Wikipedia 2007).

Für ein Wörterbuchportal wie OWID soll die Definition jedoch noch weiter gefasst werden. Unter „Portal“ wird hier die Webseite www.owid.de verstanden, die nicht ein Tor zu anderen Webseiten ist, sondern ein Zugang zu verschiedenen, eigenständigen lexikografischen Produkten von Projekten des Instituts für Deutsche Sprache innerhalb einer Webadresse. Diese Wörterbücher sind in verschiedener Hinsicht (und insbesondere) lexikografisch miteinander verbunden und sie können über das Portal mithilfe gemeinsamer Zugriffsmöglichkeiten durchsucht werden. Dabei spielt die homogene Benutzeroberfläche eine wichtige Rolle; sie ermöglicht den zentralen Zugriff auf die Daten im Portal (vgl. zum Portal-Begriff genauer auch Müller-Spitzer 2007).

Eine gemeinsame Präsentation und gemeinsame Zugriffsmöglichkeiten sowie eine homogene Benutzeroberfläche sind Charakteristika von OWID, die von den Nutzern an der Oberfläche des Portals wahrgenommen werden können. Ihnen liegen aber auch wichtige texttechnologische Grundüberlegungen zugrunde (vgl. hierzu auch den Beitrag von Carolin Müller-Spitzer in dieser Publikation):

Allerdings spielt der Gedanke des Portals im Fall von *ellexiko* auch eine wichtige Rolle im lexikografischen Prozess, denn die Datenbasen der einzelnen Projekte sollen nach einheitlichen Prinzipien strukturiert werden, damit das Portal mehr als ein Nebeneinander einzelner elektronischer Wörterbücher sein kann. (Müller-Spitzer 2007, S. 2)

Das Portal OWID soll damit nicht nur als Einstieg in lexikografische Produkte des Instituts dienen, sondern es soll den Nachschlagenden ermöglichen, das Angebot insgesamt als Online-Informationssystem zum deutschen Wortschatz zu nutzen. Zusammenfassend können die folgenden Charakteristika von OWID als Wörterbuchportal hervorgehoben werden:

1. In OWID werden Wörterbücher integriert, die am Institut für Deutsche Sprache erarbeitet wurden, wenn auch in verschiedenen Projekten. Die Wortartikel sind deshalb auf makrostruktureller Ebene integriert, d.h., es gibt eine Gesamt-Stichwortliste, die dem Benutzer zugleich als eine einheitliche Zugriffsstruktur bzw. Navigationshilfe zur Verfügung steht.
2. Zugleich bleiben die einzelnen Wörterbücher in OWID identifizierbar und können für eine Suche ausgewählt oder bei einer Suche ausgeschlossen werden.
3. Die homogene Herkunft der Daten garantiert einerseits einen hohen Grad an Verlässlichkeit und Qualität der Informationen und andererseits einen hohen Grad an Standardisierung der Artikelstrukturen.

1.3 Wörterbuch-Portal (*Gerald Neumann*)

Das Wörterbuch-Portal (im Folgenden WBP; <http://www.woerterbuch-portal.de>) geht auf eine Initiative der Heidelberger und der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) aus dem Jahre 2002 zurück. Auslöser dieser Initiative war die Tagung „Internationale intradisziplinäre Tagung zur Lexikographie, 13.-16. Oktober 2002“ (<http://www.bbaw.de/termine/2002/10/16.html> [16.11.2007]) an der BBAW. Ursprünglich gedacht als bloße Informationssammlung zu den am Workshop beteiligten Projekten, entstand schnell der Gedanke, über dieses Portal nicht nur Informationen über lexikografische Projekte zu sammeln, sondern auch die lexikografischen Inhalte

der Projekte auf eine neue Art zu präsentieren. Beide Informationsarten sind nach wie vor im WBP vertreten.

Die zur Verfügung stehenden finanziellen Mittel, von denen die Umsetzung der Portalidee, vor allem aber die mittelfristige Pflege des Portals abhingen, waren denkbar gering. Dies führte dazu, dass schon in der ursprünglichen Konzeption des WBP dieser Voraussetzung Rechnung getragen werden musste. Eine Anschubfinanzierung der Akademien ermöglichte die Bereitstellung der notwendigen Hardware und die Erstellung des WBP als einer Webanwendung. Diese unterteilt sich technologisch, aber auch konzeptionell in zwei Teile, entsprechend den beiden, oben genannten Informationsarten. Der Teil, der Informationen über die Projekte enthält, ist über ein Content-Management-System (CMS) realisiert, das es den einzelnen Projekten erlaubt, ihre eigenen Inhalte selbst zu bearbeiten und zu pflegen. Dieser Ansatz sollte die Arbeit der Aktualisierung verteilen. Der zweite Teil, der einen gemeinsamen Zugriff auf lexikografische Inhalte schaffen soll, ist nicht Bestandteil dieses CMS, sondern wird von den Betreibern des WBP verwaltet.

Im hier verstandenen Sinne eines Portals ist lediglich der zweite Teil derjenige, der relevant ist und beschrieben und untersucht werden soll. Der Begriff des Portals wird also nur für den Teil des WBP benutzt, der lexikografische Inhalte aus verschiedenen Wörterbuchprojekten zusammenführt und gemeinsam recherchierbar macht.

Unter der beschriebenen Einschränkung versteht sich das WBP als eine Webanwendung, die die lexikografischen Ressourcen verschiedener, vorzugsweise akademischer Wörterbuchprojekte über einen zentralen Punkt zugänglich und durchsuchbar macht. Dabei wird ausdrücklich auf das Angebot eigener Ressourcen verzichtet. Selbst in den Fällen, in denen eigene Wörterbuchinhalte zur Verfügung stehen, wie z.B. die digitale Variante des „Wörterbuchs der deutschen Gegenwartssprache“ oder das speziell zu Demonstrationszwecken digitalisierte „Verzeichnis der Idiotismen plattdeutscher Mundart“, werden die Ressourcen wie ‘fremde’ Ressourcen behandelt und auch ebenso in das Portal eingebunden.

2. Struktur und Inhalt von OWID und Wörterbuch-Portal

2.1 OWID (*Annette Klosa*)

Im Portal OWID sind folgende Wörterbücher, die innerhalb von Projekten des Instituts für Deutsche Sprache erarbeitet wurden bzw. werden, veröffentlicht: *ellexiko* (vgl. hierzu den Beitrag von Marion Hahn, Annette Klosa, Carolin Müller-Spitzer, Ulrich Schnörch und Petra Storjohann in dieser Publikation), Neologismenwörterbuch (vgl. hierzu den Beitrag von Doris Steffens in dieser Publikation), Wortverbindungen online (vgl. hierzu den Beitrag von Annelen Brunner, Meike Lauer und Kathrin Steyer in dieser Publikation) und Schulddiskurs im ersten Nachkriegsjahrzehnt (vgl. hierzu den Beitrag von Heidrun Kämper und Andreas Rothenhöfer in dieser Publikation). Ein Ausbau mit weiteren Wörterbüchern, z.B. dem „Handbuch deutscher Kommunikationsverben“² ist geplant. Hierzu können perspektivisch auch externe Wörterbücher treten. Das Portal ist im Internet kostenfrei und ohne Registrierung unter der URL www.owid.de zugänglich.

² Vgl. hierzu <http://www.ids-mannheim.de/ll/eska/> (zuletzt eingesehen am 26.10.2007).

Die Benutzeroberfläche dieses Portals (vgl. hierzu den Beitrag von Wolfgang Bock in dieser Publikation) soll nun näher erläutert werden, wobei folgende Definition des Terminus „Benutzeroberfläche“³ die Darstellung leitet:

Elektronische Wörterbücher sind in eine Benutzeroberfläche eingebunden [...]. Diese besteht aus den Bildschirmfenstern und verschiedenen funktionalen Elementen wie Drop-down-Menüs, Laufleisten, Schaltflächen und Suchmasken. (Engelberg/Lemnitzer 2001, S. 121)

Die Benutzeroberfläche von OWID vereint alle oben genannten funktionalen Elemente. Lexikografische Aufgabe war es, die medienspezifischen Besonderheiten, die bei der Benutzung elektronischer Wörterbücher und eines Portals wie OWID gegeben sind, bei der Neukonzeption zu berücksichtigen. Dabei wurde versucht, möglichst viele verschiedene Nachschlagehandlungen vorauszusehen; diese sollten möglichst alle bei der Konzeption der Führung des Benutzers durch die Oberfläche berücksichtigt werden. Unter Benutzerführung wird hier also die Planung und Umsetzung möglicher Benutzungswegen für verschiedene Nachschlagehandlungen in verschiedenen Wörterbuchbenutzungssituationen verstanden.

2.1.1 Benutzungswege in OWID

Für OWID werden grundsätzlich unterschiedliche Nutzerinteressen angenommen, nämlich zunächst das Interesse daran, direkt nach Aufruf der Portal-Startseite ein gesuchtes Wort eingeben zu können. Andere Nutzer aber wollen ihre Suche unter Umständen auf ein innerhalb des Portals publiziertes Wörterbuch beschränken. Schließlich gibt es Nutzer, die keine Nachschlagehandlung vollziehen möchten, sondern die sich über das Portal oder einzelne darin veröffentlichte Wörterbücher informieren wollen.

Es war also bei der Entwicklung des Portals wichtig, eine Präsentation zu finden, die all dies ermöglicht und dem Nutzer jederzeit zeigen sollte, wo er sich auf seinem (Nachschlage-)Weg durchs Portal gerade befindet. Zugleich sollte der Nutzer aus jeder Position seines Weges, aus einzelnen Wortartikeln oder projektbeschreibenden Seiten, wieder in das Portal zurückkehren können.

2.1.1.1 Recherchen in OWID

Wichtigstes Element der OWID-Startseite (www.owid.de) ist das Eingabefeld für Suchanfragen. Der Zugang zu einzelnen Wörterbüchern wird mithilfe von Schaltflächen, die darunter angeordnet sind, ermöglicht. Informationen zum Portal können über den Menüpunkt [Projekt OWID](#) aufgerufen werden. Im Unterschied zu den Seiten der einzelnen Wörterbücher in OWID (vgl. hierzu die einzelnen Beiträge in dieser Publikation) weisen die Portalseiten selbst einige Besonderheiten auf: Die Navigationsstruktur im Portal ist sehr flach, indem neben der Startseite selbst nur ein weiteres Menü mit Informationen angeboten wird. Die hier enthaltenen Informationen sind wiederum keine Benutzungshinweise oder Projektinformationen, wie sie sich zu den einzelnen Wörterbüchern finden, sondern sollen in Summe in etwa nur Vorwort- bzw. Einleitungsfunktion zum Portal übernehmen (vgl. Abbildung 1).

³ Statt „Benutzeroberfläche“ wird z.T. auch von „Benutzerschnittstelle“ gesprochen (vgl. Engelberg/Lemnitzer 2001, S. 222).

The screenshot shows the OWID homepage. At the top, there is a navigation bar with 'Startseite OWID' and 'Projekt OWID'. The main header area contains the OWID logo and the IDS logo (Institut für Deutsche Sprache). Below this, there is a search section on the left with a search input field and a 'suche' button. A sidebar on the left contains search options: 'Suche in ...' with checkboxes for 'lexiko', 'Neologismen', 'Feste Wortverbindungen', and 'Diskurswörterbuch 1945-55'; and 'Optionen' with radio buttons for 'in Gesamtstichwortliste' (selected) and 'in allen lexikografisch ausgearbeiteten Artikeln'. The main content area features a large heading 'Willkommen in OWID, dem Online-Wortschatz-Informationssystem Deutsch des Instituts für Deutsche Sprache, Mannheim'. Below this, there is a paragraph explaining the system and a list of colored buttons for 'lexiko', 'Neologismenwörterbuch', 'Feste Wortverbindungen', and 'Diskurswörterbuch 1945-55'. The footer contains the copyright information for the Institut für Deutsche Sprache and the website address.

Abbildung 1: Die Startseite von OWID

Die Suche auf der OWID-Startseite erfolgt in der Gesamtstichwortliste, bei der man im metalexikografischen Sinn eigentlich nicht von einer Stichwortliste sprechen kann. Denn diese Liste ist nicht nach homogenen lexikografischen Kriterien zusammengestellt worden,⁴ sondern sie ist ausschließlich eine Zusammenführung der verschiedenen Stichwortlisten der einzelnen Wörterbücher im Portal. Sie dient als gemeinsame Zugriffsstruktur, wenn man innerhalb des Portals sucht.

Wird ein Stichwort über die Suche in OWID aufgerufen, erfolgt die Artikelanzeige innerhalb des Portals und unter Anzeige des jeweiligen Ausschnitts aus der Gesamtstichwortliste. Hierbei liegt die Annahme zugrunde, dass ein Nutzer, der über die Startseite von OWID sucht, sich nicht für ein spezielles Wörterbuch interessiert. Die Nutzer können dabei anhand farbiger Markierungen der Stichwörter in der Stichwortliste erkennen, aus welchem Wörterbuch der Wortartikel stammt, denn jedes Wörterbuch in OWID besitzt, um den Nutzern eine bessere Orientierung zu ermöglichen, eine eigene Farbe. Außerdem erscheint der Wortartikel vor dem entsprechenden farbigen Hintergrund. Der Bildschirmbereich um die Artikelanzeige herum bleibt in der Farbgebung des OWID-Portals erhalten (wie z.B. bei den Wortartikeln *Glück* [*lexiko*] und *Inliner* [Neologismenwörterbuch]). Innerhalb der Suchergebnislisten im Portal wird die Zugehörigkeit

⁴ Für *lexiko* etwa definiert Schnörch (2005, S. 83) folgendermaßen: „*lexiko*-Stichwortliste: Liste aller in *lexiko* gemäß den Lemmatisierungsprinzipien angesetzten Stichwörter“. Die Stichwortlisten der anderen Wörterbücher in OWID wurden nach jeweils eigenen Kriterien zusammengestellt.

eines Stichwortes zu einem Wörterbuch des Portals ebenfalls durch unterschiedliche Farbgebung signalisiert (wie z.B. bei der Suchanfrage nach dem Stichwort *Antifaschist*).

Über die Stichwortliste, die jeweils im linken Bildschirmbereich neben der Wortartikelanzeige erscheint, kann der Nutzer weitere Wortartikel aufrufen. Die Navigation innerhalb der Gesamtstichwortliste ist alternativ auch mithilfe der Alphabetleiste möglich. Der Nutzer kann außerdem das Eingabefeld für eine Suchanfrage im oberen linken grauen Balkenbereich benutzen oder auf die Portal-Startseite zurückkehren, um eine neue Suchanfrage zu starten (vgl. Abbildung 2).

Insgesamt wurde versucht, sowohl mit der übersichtlichen Einstiegsseite wie mit den verschiedenen positionierten und gestalteten Suchmodi als auch mit den knappen Texten die Portalseiten von OWID leicht erfassbar und benutzbar zu gestalten. Dabei sollte natürlich nicht vergessen werden, dass der Aufbau eines Portals wie OWID eine neue und zugleich experimentelle Herausforderung ist, an der auch weiterhin gearbeitet werden muss.

The screenshot shows the OWID lexiko interface. At the top, there is a search bar with the text 'elexiko' and the OWID logo. Below the search bar are navigation links: 'Startseite OWID', 'Projekt OWID', 'Startseite elexiko', 'Wortartikel', 'Stichwortliste', 'Projekt', 'Benutzungshinweise', and 'Erweiterte Suche'. A navigation bar contains letters A through Z, with 'Auswahl:' and a dropdown menu. The main content area is titled 'Monat' and is divided into sections: 'Lesartenübergreifende Angaben' (Orthografie, Normgerechte Schreibung, Worttrennung), 'Herkunft und Wandel', and 'Lesartenbezogene Angaben'. A sidebar on the left lists various related terms, with 'Monat' highlighted. A text box on the right provides additional semantic and grammatical information.

Abbildung 2: Ansicht eines Wortartikels aus *elexiko* in OWID

2.1.1.2 Suchmodi

Für das Portal wurden daneben möglichst verschiedene Suchmodi angedacht, und zwar eingabebasierte Suche (also über Eintippen des gesuchten Wortes in ein Suchfeld), die Möglichkeit, mit Platzhaltersymbolen zu suchen, und die Möglichkeit, auch flektierte

Formen suchen zu können. Andere ebenfalls wünschenswerte Abfragemöglichkeiten (z.B. lautformbasierte Suche) sind in OWID zurzeit (noch) nicht realisiert, die Implementierung einer schreibungstoleranten Suche steht in naher Zukunft bevor.

Auf Portalebene, also in OWID, werden bestimmte Suchmodi, die z.B. in *elexiko* und im Neologismenwörterbuch schon realisiert sind, noch nicht angeboten, und zwar vor allem die indexbasierte Suche (also Suche nicht über ein Eingabefeld, sondern durch Auswahl eines Stichwortes in einer oder mehreren Stichwortlisten) und die filterbasierte Suche (also Suche von Wortartikeln mit bestimmten Eigenschaften). Gerade dies soll aber zukünftig auch ermöglicht werden. Zurzeit (Ende 2007) besteht für die Nutzer schon die Möglichkeit, sich bei der Suche auf alle lexikografisch ausgearbeiteten Artikel zu beschränken, oder in der Gesamtstichwortliste zu suchen, die allerdings auch noch nicht ausgearbeitete Wortartikel enthält.

2.1.1.3 Wechsel in ein lexikografisches Produkt

Über die Portal-Startseite besteht, wie oben erwähnt, auch die Möglichkeit, eines der Wörterbücher direkt aufzurufen. Dies kann für Nutzer von Interesse sein, die sich bei ihrer Suche auf eine bestimmte Stichwortmenge beschränken wollen, die die spezielleren Suchfunktionen in der Artikelmenge eines Wörterbuches nutzen möchten oder die sich über ein bestimmtes Wörterbuch informieren möchten.

The screenshot shows the OWID website interface. At the top left is the 'OWID' logo. In the center, there is a search bar containing the text 'elexiko'. To the right of the search bar is the logo for the 'INSTITUT FÜR DEUTSCHE SPRACHE'. Below the search bar, there are navigation links: 'Startseite OWID', 'Projekt OWID', 'Startseite elexiko', 'Wortartikel', 'Stichwortliste', 'Projekt', 'Benutzungshinweise', and 'Erweiterte Suche'. The main content area has a yellow background and contains the following text:

Willkommen in *elexiko*,
einem Online-Wörterbuch zur deutschen Gegenwartssprache

Der Schwerpunkt des Online-Wörterbuchs *elexiko* liegt auf der Beschreibung von Bedeutung und Verwendung der Stichwörter, daneben gibt es auch Angaben zur Orthografie, zur Worttrennung sowie grammatische Informationen. Sie können in *elexiko* nachschlagen, wenn Sie z. B. beim Schreiben eines Textes nach einem sinnverwandten Ausdruck suchen. Beim Lesen eines Textes helfen Ihnen die Informationen in *elexiko* z. B. dabei, ein Wort richtig zu verstehen.

Sie können Wörter über Eingabe im Suchfeld oben links suchen oder die Suchfunktionen im Menüpunkt "Erweiterte Suche" nutzen.

Da sich *elexiko* noch im Aufbau befindet, können Sie [hier alle bearbeiteten Stichwörter](#) aufrufen. Zu den (überwiegend) noch nicht vollständig bearbeiteten Stichwörtern finden Sie orthografische Angaben (neue und alte Rechtschreibung), Angaben zu morphologischen Varianten, Flexionsinformationen (von canoo.net) und (in Kürze) automatisch ausgewählte Textbelege aus dem *elexiko*-Korpus.

Wenn Sie mit der Stichwortliste, die ca. 300.000 Einträge umfasst und normalalphabetisch oder rückläufig sortiert angeboten wird, arbeiten möchten, öffnen Sie diese bitte über den Menüpunkt "Stichwortliste". Bitte nutzen Sie die Menüpunkte "Benutzungshinweise" und "Informationen", um weitere (umfangreiche) Informationen zum Wörterbuch und zum Projekt *elexiko* aufzurufen.

Wenn Sie Fragen oder Anregungen haben, wenden Sie sich bitte an: elexiko@ids-mannheim.de

At the bottom of the page, there is a copyright notice: © Institut für Deutsche Sprache (http://www.ids-mannheim.de). Below this, there are links to 'Startseite OWID' and 'Projekt OWID'. On the right side of the footer, there are icons for a printer, a document, and a copyright symbol.

Abbildung 3: Ansicht der Startseite von *elexiko*

Nach Anklicken einer Wörterbuch-Schaltfläche öffnet sich die Startseite des jeweiligen Wörterbuches. Außerdem erscheint unterhalb der Portal-Menüleiste eine zweite Wörterbuch-Menüleiste, deren Menüpunkte von Wörterbuch zu Wörterbuch leicht variieren. Für *ellexiko* werden beispielsweise die Menüpunkte „Startseite *ellexiko*“, „Wortartikel“, „Stichwortliste“, „Projekt“, „Benutzungshinweise“ und „Erweiterte Suche“ angeboten. Zugleich ist jederzeit die Suche im gesamten Portal über Eingabe in das Suchfeld im oberen grauen Balkenbereich möglich (vgl. Abbildung 3). Hiermit wird versucht, einerseits speziellere Nutzerinteressen zu befriedigen wie eine Suche auf Portalebene zu ermöglichen.

Anhand der Menüpunkte im *ellexiko*-Wörterbuch zeigt sich auch, dass die Navigationsstruktur innerhalb der einzelnen Wörterbücher tiefer ist als die Navigationsstruktur auf der darüberliegenden Ebene des Portals, wobei jedes Projekt selbst entscheidet, welche Navigationsstruktur es für sein Wörterbuch anbieten möchte. Am Beispiel von *ellexiko*⁵ soll dies nun im Detail gezeigt werden: Zunächst erscheint eine relativ schlanke Eingangssseite zum Wörterbuch, auf der die Nutzer unter dem Menüpunkt [Startseite *ellexiko*](#) die wichtigsten Informationen zu diesem Wörterbuch kurz zusammengefasst erhalten.

Der Menüpunkt [Stichwortliste](#) bedient das Interesse mancher Benutzer, sich ausschließlich in der Stichwortliste bewegen zu können. Diese wird sowohl in normalalphabetischer wie in rückläufiger Sortierung angeboten. Dabei wurde die sehr umfangreiche *ellexiko*-Stichwortliste in Gruppen mit jeweils bis zu 300 Einträgen portioniert; durch Anklicken einer Stichwortgruppe kann man den jeweiligen Ausschnitt der Stichwortliste aufrufen. Von den Einträgen in der Stichwortliste aus lassen sich die entsprechenden Wortartikel öffnen. Ein Export von Stichwortgruppen ist ebenfalls möglich.

Bei den Menüpunkten „Benutzungshinweise“ und „Informationen“ wurde versucht, Vorschläge zur Funktion und Gestaltung von Umtexten in Printwörterbüchern⁶ für ein Online-Wörterbuch umzusetzen. Die [Benutzungshinweise](#) bieten eher an den Laiennutzer gerichtete Erklärungen zu den Nachschlagemöglichkeiten im *ellexiko*-Wörterbuch, erläutern die Angaben im Wörterbuch und die Wortartikelstruktur und ermöglichen das Nachschlagen von Fachbegriffen in einem umfangreichen [Glossar](#). Dieses enthält zum einen grammatische Termini, entspricht daher einer Wörterbuchgrammatik, und erklärt außerdem alle Fachbegriffe, die in den Online-Artikelansichten vorkommen. Durch zahlreiche, mit den jeweiligen Wortartikeln verlinkte Beispiele sind die Benutzungshinweise mit dem Wörterbuch vernetzt (z.B. Glossareintrag [Abbildung](#) mit Link zum Stichwort [Kathedrale](#)). Der Menüpunkt [Projekt](#) enthält eher an ein Fachpublikum gerichtete Texte, die Methoden und Arbeitsweisen in *ellexiko* erklären und Einblicke in Bearbeitungsteilmengen wie in den Projektverlauf vermitteln und die Projektgruppe vorstellen.

Der Menüpunkt [Erweiterte Suche](#) ermöglicht verschiedene Suchmodi ausschließlich im Stichwortbestand des *ellexiko*-Wörterbuchs. Neben der einfachen Stichwortsuche oder der Platzhaltersuche sind besonders filterbasierte Suchen im Stichwortbestand von Interesse, mithilfe deren man eine Menge von Wortartikeln, die dem gleichen lexikalisch-lexikologischen Kriterium genügen, finden kann. So kann der Bestand an ausgearbeiteten Wortartikeln in *ellexiko* etwa nach allen Adjektiven durchsucht werden, die explizite

⁵ Zu den anderen Wörterbüchern im Portal vgl. die Beiträge von Annalen Brunner, Meike Lauer und Kathrin Steyer, von Heidrun Kämper und Andreas Rothenhöfer sowie von Doris Steffens in dieser Publikation.

⁶ Diese Vorschläge legte Herberg (1985) vor.

Derivate sind oder nach allen Verben, die Handlungsprädikatoren sind. Der weitere Ausbau dieser Recherchemöglichkeiten ist geplant.

Im Unterschied zu übergreifenden Recherchen in OWID wird ein Stichwort, das innerhalb eines Wörterbuches aufgerufen wird, nur im Umfeld der Stichwortliste des jeweiligen Wörterbuches angezeigt. Außerdem erscheint der Wortartikel dann in der Farbgebung dieses Wörterbuches und ohne den oben erwähnten Portal-Hintergrund (wie z.B. bei den Wortartikeln [Glück](#) [*lexiko*] und [Inliner](#) [Neologismenwörterbuch]). Dadurch soll dem Nutzer deutlich werden, in welchem Wörterbuch er sich jeweils befindet. Der Wechsel in ein anderes Wörterbuch ist jederzeit über die Auswahl des Menüpunktes „Startseite OWID“ möglich.

2.1.1.4 Verbindungen zwischen den lexikografischen Produkten

Verbindungen zwischen den Wörterbüchern des Portals können auf verschiedene Weise hergestellt werden. In *lexiko* soll beispielsweise zukünftig in der Stichwortliste markiert werden, ob ein Stichwort auch noch in einem anderen Wörterbuch in OWID nachgeschlagen werden kann. Außerdem sind in verschiedenen Wortartikeln der einzelnen Wörterbücher Links angelegt, z.B. im *lexiko*-Artikel [blind](#) auf die Wortverbindungsartikel mit *blind* als Basiselement. Umgekehrt verlinkt der Artikel [blinder Alarm](#) aus dem Wörterbuch „Wortverbindungen online“ auf den *lexiko*-Eintrag zu *blind*. Aus Wortartikeln des Neologismenwörterbuches, die Neubedeutungen beschreiben, erfolgt zukünftig ein Link zum entsprechenden Wortartikel in *lexiko*, in dem das vollständige Bedeutungsspektrum beschrieben wird (zu den Verlinkungen zwischen den Wörterbüchern des Portals vgl. besonders den Beitrag von Doris Steffens in dieser Publikation). Umgekehrt wird vom *lexiko*-Wortartikel aus auf die Neubedeutung im Neologismenwörterbuch verwiesen werden. Solche Vernetzungen sollen künftig noch stärker ausgebaut werden, damit die Nutzer leichter die vielfältigen Informationsmöglichkeiten im Portal nutzen können.

2.1.2 Perspektiven für OWID

OWID soll durch die Aufnahme weiterer (gegebenenfalls auch externer) Wörterbücher bzw. durch die Erweiterung vorhandener Wörterbücher kontinuierlich ausgebaut werden (zum modularen Ausbau von OWID vgl. oben unter 1.1). Wünschenswert bei solchen Integrationen ist, dass diese nicht einen lexikografischen Gemischtwarenladen entstehen lassen, sondern dass tatsächlich ein Verbund von Wörterbüchern innerhalb des Portals entsteht. Es ist deshalb wichtig und zukünftige Aufgabe, die entsprechende Texttechnologie (vgl. hierzu den Beitrag von Carolin Müller-Spitzer in dieser Publikation) bereitzustellen und weiter auszubauen. So können beispielsweise zukünftig auch filterbasierte Recherchen über verschiedene Wörterbücher hinweg angeboten werden.

Eine stärkere Vernetzung der Wortartikelmengen (und natürlich der Artikelmengen eines Wörterbuches) untereinander ist daher ein weiteres wichtiges Ziel. Für *lexiko* liegt bereits ein Konzept der Vernetzungsstrukturen vor, dessen technische Realisierung zeitlich noch nicht genau festgelegt werden kann. Die Ausdehnung dieses Konzeptes auf OWID sowie die technische Realisierung sind aber unabdingbar, wenn OWID seine Nutzer zufriedenstellen soll.

Der Portalgedanke kann zukünftig auch durch die einheitlichere grafische Ausgestaltung der gemeinsamen Benutzeroberfläche von OWID gestärkt werden. Zwar muss jedes Wörterbuchprojekt entscheiden, in welcher Form Wortartikel, Stichwortlisten usw. präsentiert werden sollen, doch wäre ein Ausbau der jetzt erst im Ansatz vorhandenen grafischen Gestaltungsvorgaben nötig, um insgesamt ein einheitlicheres Erscheinungsbild herzustellen. Dieses würde den Benutzern die Rezeption des Portals erleichtern. Zu diesem Zweck wäre außerdem eine Art „Geführte Tour“ durch das Portal hilfreich, die es für andere Online-Wörterbücher (z.B. für das [Oxford English Dictionary](#)) schon gibt.

Zuvor sollten aber durch geeignete Befragung und Untersuchung die Nutzer und ihre Wünsche und Vorschläge selbst zu Wort kommen. Für die wenigsten Online-Wörterbücher ist Benutzerforschung betrieben worden,⁷ für ein Wörterbuchportal wohl noch gar nicht. Gerade hierzu sollte sich zukünftig etwas ändern. Offen ist für OWID auch noch, in welcher Form die Benutzer stärker eingebunden werden können. Denkbar wäre etwa, Vorschläge zu Wörtern, die als Nächstes bearbeitet werden sollen, einzuholen, die Nutzer zur Diskussion fertiger Wortartikel in einem Forum einzuladen oder sie mit wechselnden, aktuellen Angeboten („Wort des Tages“ o.Ä.) stärker an das Portal zu binden. Was sich hiervon zukünftig realisieren lassen wird, ist derzeit (Ende 2007) noch nicht abzusehen.

2.2 Wörterbuch-Portal (*Gerald Neumann*)

2.2.1 Voraussetzungen

Damit das Wörterbuch-Portal (im Folgenden WBP) seine Portalfunktion wahrnehmen kann, müssen einige Voraussetzungen erfüllt sein. Da, wie schon gesagt, das WBP über keine eigenen lexikografischen Ressourcen verfügt, ist es als Erstes notwendig, diejenigen Informationen aus der Artikelstruktur der eingebundenen Wörterbücher zu extrahieren, die durchsucht und präsentiert werden sollen. Es handelt sich dabei bisher überwiegend um die Lemmalisten der Wörterbücher und, im Falle des DWDS-Korpus, um die Ergebnisse der Korpusanfragen.

Die Lemmalisten der eingebundenen Wörterbücher bilden die Grundlage für die derzeit vom WBP bereitgestellten Recherchefunktionalitäten. Prinzipiell ist eine Erweiterung des Zugriffs auch auf andere Teile der Artikelstruktur möglich und gewünscht, allerdings gestaltet sich ein solcher Zugriff, ohne dass es dazu ein standardisiertes Verfahren gäbe, als in der Realisierung zu aufwendig bzw. gänzlich unmöglich.

Neben der Information, welche Lemmata in den externen Ressourcen verfügbar sind, benötigt das WBP für jedes Lemma die Information, unter welcher URL das Lemma verlinkt werden kann. Daneben gibt es noch eine Reihe von Zusatzinformationen, die z.B. Auskunft darüber geben können, ob ein Stichwort zwar im (Print-)Wörterbuch vorhanden ist, aber online noch nicht existiert und damit auch nicht verlinkt werden kann. Genauso ist für Teile des Belegarchivs des Deutschen Wörterbuchs die Information verfügbar, dass es zwar kein Stichwort im Wörterbuch gibt, sehr wohl aber Belege im Belegarchiv.

⁷ In Lemnitzer (2001) z.B. wird über Protokolle der Benutzung von zweisprachigen Internet-Wörterbüchern berichtet.

2.2.2 Ressourcen

Zum jetzigen Zeitpunkt sind acht lexikografische Ressourcen in das WBP eingebunden:

1. Das [Deutsche Rechtswörterbuch](#) (DRW) ist über eine Lemmaliste eingebunden, die die derzeit online verfügbaren Stichwörter enthält.
2. Das [Wörterbuch der deutschen Gegenwartssprache](#) (WDG) ist ebenfalls über eine Lemmaliste eingebunden, die zurzeit alle Hauptstichwörter enthält.
3. Das [Deutsche Wörterbuch](#) von Jacob und Wilhelm Grimm (DWB) ist ebenfalls über eine Lemmaliste eingebunden, die alle Stichwörter der Online-Ausgabe enthält.
4. Die [Neubearbeitung des Deutschen Wörterbuchs](#) (²DWB) ist zum einen mit der Lemmaliste der 23. Lieferung eingebunden und zum anderen mit der Stichwortliste aus dem Belegarchiv für den Bereich *auffassen – Ausbildung*.
5. Das [Goethe-Wörterbuch](#) ist mit einer Lemmaliste eingebunden, die die 1. Lieferung des 5. Bandes umfasst und von *Inhalt* bis *Jammerlitanei* geht.
6. [OWID](#) (vormals *ellexiko*) ist über eine Lemmaliste eingebunden, die alle Stichwörter des *ellexiko*-Wörterbuchs und des Neologismenwörterbuchs enthält.
7. Das [Verzeichnis der Idiotismen plattdeutscher Mundart](#) ist über eine Lemmaliste eingebunden, die alle Stichwörter des Verzeichnisses enthält. Zusätzlich besteht die Möglichkeit, auf sämtliche Faksimiles des Verzeichnisses zuzugreifen.
8. Das [DWDS-Korpus](#) ist über eine Online-Abfrage in das WBP eingebunden, die die Anzahl der Treffer im Korpus zurückliefert.

2.2.3 Suche und Navigation

Das WBP ist so aufgebaut, dass das Sucheingabefeld für die eingebundenen lexikografischen Ressourcen jederzeit sichtbar und benutzbar ist (vgl. Abbildung 4). Diese Ressourcen lassen sich einzeln und in beliebigen Kombinationen auswählen. Beim Aufruf der Seite erscheint eine Standardeinstellung, die das DRW und das WDG als Suchquelle auswählt.

Die einzelnen Ressourcen sind ebenfalls immer auf der linken Seite des Bildschirms sichtbar (vgl. Abbildung 5). Sie sind durch eine Abkürzung gekennzeichnet. Da diese dem normalen Nutzer nicht vertraut sein dürfte, handelt es sich bei diesen Abkürzungen um Links, die entweder zu einer weiterführenden Informationsseite im beschreibenden Teil des WBP führen oder auf die externe Seite des verantwortlichen Projektes. Neben den Abkürzungen der Ressourcen befindet sich in Klammern die Anzahl der Einträge in den verfügbaren Lemmalisten. Diese Angabe soll eine schnelle Einschätzung über den Umfang und das zu erwartende Ergebnis ermöglichen. Wiederum rechts daneben befindet sich ein Fragezeichen, das ebenfalls ein Link ist und im Regelfall zu den bibliografischen Angaben der Ressource führt. Am Ende der Auswahlliste findet sich noch die Information über die Gesamtzahl der Einträge aus allen Lemmalisten.

The screenshot shows the homepage of the WÖRTERBUCH-PORTAL.de. At the top, there is a navigation bar with links for 'HOME', 'Suche', 'Linkliste', 'Übersicht', and 'Impressum'. Below this is a search bar with the text 'Ihr Wort?:' and buttons for 'Suche' and 'Hilfe'. To the right of the search bar is a dropdown menu for 'Akadem. Wb' and a 'Go' button. The main content area is divided into several sections: 'Wörterbuchauswahl' with a list of dictionaries (WVG, DRW, DWB, 2DWB, GWB, ELEXIKO, VIPM) and their respective word counts; 'Wörterbuch-Portal' with news about updates to the lexicon; and a sidebar titled 'Wörterbücher' listing various specialized dictionaries.

Abbildung 4: Startseite des WBP

This screenshot shows a detailed view of the 'Wörterbuchauswahl' section. It features a list of dictionaries with checkboxes and word counts: WVG (89627), DRW (110321), DWB (296612), 2DWB (6904), GWB (926), ELEXIKO (277962), and VIPM (1447). There are also options for 'DWDS-Corpus' and other resources.

Abbildung 5:
Auswahlliste der Res-
ourcen im WBP

Die Suche selbst wird über ein einfaches Suchfeld realisiert. Es handelt sich dabei immer um eine rechtstrunkierte Suche, d.h. um eine Suche, die im Bedarfsfall eine rechtsseitige Erweiterung des Suchbegriffes vornimmt. Platzhalterzeichen wie * oder ? sind nicht notwendig und auch nicht möglich. Im Gegensatz zu herkömmlichen Suchmaschinen, wie z.B. Google, erzeugt die Suche keine Trefferliste. Stattdessen wird im mittleren Hauptteil des Bildschirms eine Übersicht generiert, die das gefundene und dunkel hinterlegte Suchwort mit den 10 vorangehenden und den 10 nachfolgenden Stichwörtern enthält. Ist das Suchwort in keiner der durchsuchten Ressourcen vorhanden, wird das nächste über die Rechtstrunkierung gefundene Stichwort ins Zentrum gerückt, allerdings nicht dunkel hinterlegt (vgl. Abbildung 6).

Rechts neben der Liste der angezeigten Stichwörter findet sich eine Tabelle, die das Vorkommen jedes einzelnen Stichworts in den ausgewählten Ressourcen anzeigt und gegebenenfalls charakterisiert. Im besten Fall ist das Stichwort in der Lemmaliste der Ressource enthalten und auch online verfügbar. Dann wird es mit einem x markiert, wobei dieses x ein direkter Link auf das Stichwort in der Online-Präsentation des Wörterbuchs ist. Neben dieser Markierung gibt es noch eine ganze Reihe anderer möglicher Markierungen, die in einer Legende zusammengefasst und immer unter der generierten Tabelle im Hauptteil der Seite sichtbar sind (vgl. Abbildung 7).

Die Markierungen führen, soweit möglich, in die Online-Präsentationen der Wörterbücher und der Belegarchive sowie auf die Korpusseiten des DWDS. Die Stichwörter selbst sind

Lemma	WDG	DRW	DWB	² DWB	GWB	ELEXIKO	VIPM
↑ Zurückblättern ↑							
Hauptziel	X	0	-	-	-	-	0
Hauptzollpatent	0	X	-	-	-	-	0
Hauptzollstadt	0	X	-	-	-	-	0
Hauptzollstatt	0	X	-	-	-	-	0
Hauptzug	X	0	-	-	-	-	0
Hauptzunft	0	X	-	-	-	-	0
Hauptzunftlade	0	X	-	-	-	-	0
Hauptzusage	0	X	-	-	-	-	0
Hauptzweck	X	0	-	-	-	-	0
Hauptzweig	X	0	-	-	-	-	0
Haurecht	0	X	-	-	-	-	0
Haus	X	0	-	-	-	-	0
Hausachtung	0	X	-	-	-	-	0
Hausacker	0	X	-	-	-	-	0
Hausaltar	X	0	-	-	-	-	0
Hausamt	0	X	-	-	-	-	0
Hausandacht	X	0	-	-	-	-	0
Hausangehörigkeit	0	X	-	-	-	-	0
Hausangesessene	0	X	-	-	-	-	0
Hausangestellte	X	0	-	-	-	-	0
Hausanschlag	0	X	-	-	-	-	0
Hausanschluß	X	0	-	-	-	-	0
↓ Vorblättern ↓							

Abbildung 6: Übersicht zum Suchwort im WBP

Der Hilfe-Button unter dem Suchfeld ist auf eine Seite verlinkt, die einige grundsätzliche Erläuterungen zur Suche und zum erzeugten Ergebnis macht. Wir gehen allerdings davon aus, dass die Benutzeroberfläche des WBP durch die Reduzierung auf wenige Steuer- und Navigationselemente und die sehr flache Hierarchie der Seite ohne Erklärungen verstanden wird und intuitiv bedienbar ist.

Das Ergebnis jeder möglichen Suchanfrage ist durch den Aufbau der dabei entstehenden URL eindeutig und kann über die Benutzung dieser aktuellen URL jederzeit wiederholt und damit auch von beliebiger Stelle direkt durch einen Link erreicht werden. Vergleiche dazu folgende Suchanfrage für den Suchbegriff *Portal*: [Verlinkte Suchanfrage „Portal“](#).

AUFFÄSELN	0	0	X	-	-	
AUFFASERN	0	0	X	-	-	
auffassen / AUFFASSEN	X	X	X	X	X	
auffasslich	0	0	-	A	-	
Auffassung / AUFFASSUNG	X	0	X	X	X	
Auffassungsakt	0	0	-	A	-	
Auffassungsart	0	0	-	A	-	
Auffassungserschwerung	0	0	-	A	-	
Auffassungsfähigkeit	0	0	-	A	-	
Auffassungsform	0	0	-	A	-	
Auffassungsfreiheit	0	0	-	A	-	
AUFFASSUNGSGABE	X	0	X	X	X	
Auffassungssache	0	0	-	-	X	

↓ Vorblättern ↓

Legende:
X - Stichwort existiert im Wörterbuch, **X** ist ein Link zum Eintrag
X - Stichwort existiert im Wörterbuch, ist jedoch nicht online verfügbar
B - Kein Eintrag im Wörterbuch, aber im Belegarchiv vorhanden, **B** ist ein Link zum Beleg
B - Kein Eintrag im Wörterbuch, aber im Belegarchiv vorhanden, ist jedoch nicht online verfügbar
A - im Archiv vorhanden, **A** ist ein Link zu den Archivdaten
0 - Stichwort existiert nicht im Wörterbuch
- - zur Zeit keine Informationen verfügbar
[0-9]+ - Vorkommen des Stichwortes im DWDS-Corpus, Link zur Login-Seite des Corpus

Abbildung 7: Markierung der Treffer in der Übersicht der Stichwörter

2.2.4 Konzept und Struktur

Das hinter den oben beschriebenen Recherche- und Navigationsfunktionen stehende Konzept musste verschiedene Faktoren berücksichtigen.

Die zur Verfügung stehenden Mittel waren und sind, wie eingangs beschrieben, sehr beschränkt. Dies hatte zur Folge, dass programmiertechnische Leistungen zwar für die Startphase zur Verfügung standen, das WBP aber im späteren Verlauf nicht mehr auf solche angewiesen sein durfte.

Da das WBP von einer unbekanntem Anzahl lexikografischer Ressourcen ausging, die in das Portal integriert werden sollten, musste eine Form der Erweiterbarkeit implementiert werden, die ohne großen Aufwand und sogar von externer Stelle aus möglich sein sollte. Dabei durften die schon integrierten Ressourcen natürlich nicht in Mitleidenenschaft gezogen werden.

Im schnelllebigen Medium Internet besteht darüber hinaus auch immer die Gefahr, dass bestehende Angebote ihren Ort und damit auch ihre Adresse ändern. Die notwendigen

Reaktionen auf solche Änderungen durften keine größeren Anpassungsprobleme auf Seiten des WBP aufwerfen.

Das Konzept, das diese Faktoren berücksichtigt und die genannten Funktionen zur Verfügung stellt, sieht folgendermaßen aus:

Jede Ressource, mit Ausnahme des DWDS-Korpus, ist über eine eigene Lemmaliste eingebunden. Diese Lemmaliste beinhaltet die im WBP angezeigten Stichwörter in ihrer Oberflächenform und die notwendigen Informationen zur eindeutigen Identifizierung des Stichworts in seiner Online-Präsentation. Es handelt sich dabei nicht um eine komplette URL, sondern im Regelfall um eine ID-Nummer oder das normalisierte Stichwort selbst. Die URL wird für jede Ressource an zentraler Stelle und erst zum Zeitpunkt der Abfrage zusammgebaut. Damit lassen sich Veränderungen in der Adresse durch eine Korrektur an einer Stelle realisieren. Die Lemmalisten sind in einer relationalen Datenbank abgelegt, in der jede Ressource eine eigene Tabelle besitzt. Die Anzahl der möglichen Tabellen ist nicht beschränkt und zur Erweiterung um neue Ressourcen bedarf es nur der Einrichtung einer neuen Tabelle mit gleicher Struktur.

Wird eine Suchanfrage gestellt, so werden dem System der Suchbegriff und die Auswahl der Ressourcen übergeben. In der Lemmaliste jeder ausgewählten Ressource wird nun nach dem Suchbegriff recherchiert. Das gefundene Stichwort, entweder identisch mit dem Suchbegriff oder nach Rechtstrunkierung des nächsten Stichwortes erhalten, wird mit seinen 10 Vorgängern und 10 Nachfolgern aus der Datenbank geholt. Im zweiten Schritt werden alle jeweils 21 Stichwörter der einzelnen Ressourcen in eine temporäre, kumulative Lemmaliste geschrieben und alphabetisch sortiert. Auf dieser temporären Lemmaliste findet eine erneute Suche mit dem Suchbegriff statt und wiederum werden die 10 Vorgänger und die 10 Nachfolger aus der kumulativen Lemmaliste extrahiert. Im letzten Schritt werden die übrig gebliebenen Stichwörter ausgegeben und die Links zu den externen Wörterbüchern auf der Grundlage der ID-Nummern und der zentralen Adressverwaltung generiert. Auf diese Weise wird es möglich, die Ressourcen einzeln und in beliebiger Kombination zu durchsuchen, ohne die Anfragen entsprechend der Ressourcenauswahl adaptieren zu müssen. Gleichzeitig kann jede Ressource einzeln eingebunden, gepflegt und weiterentwickelt werden. Es besteht die Möglichkeit, die Bearbeitung und Aktualisierung der Lemmalisten in der Datenbank an die inhaltlich Verantwortlichen abzugeben, auch wenn diese Möglichkeit zurzeit nicht genutzt wird.

2.2.5 Ausblicke

Der Nutzen des WBP ergibt sich in erster Linie aus den zur Verfügung stehenden lexikografischen Ressourcen. Daher sind die wichtigsten Ziele für den Ausbau die Hinzunahme weiterer, digital verfügbarer Wörterbücher und Korpora sowie der Ausbau und die Aktualisierung der bereits eingebundenen Ressourcen. Hierfür stehen mehrere Projekte in Aussicht bzw. es existieren die technischen Voraussetzungen für eine Einbindung:

1. Die ersten drei Bände des [Goethe-Wörterbuchs](#) sind vom „Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier“ digitalisiert worden und seit einiger Zeit auch online verfügbar. Dieses Wörterbuch, das im Trierer Wörterbuch-Netz zur Verfügung gestellt wird, kann auf die gleiche Art und Weise in das WBP aufgenommen werden

wie das Deutsche Wörterbuch. Voraussetzung dafür ist allerdings der Zugriff auf eine Lemmaliste, die die Lemmata mit den für den Zugriff notwendigen ID-Nummern enthält.

2. Die Arbeiten zur weiteren Auszeichnung des WDG haben durch die Erfassung der im Wörterbuch angegebenen Beispielkomposita zu einer Erweiterung der Lemmaliste um ca. 40.000 Einträge geführt. Diese können ebenfalls in das WBP aufgenommen werden.
3. Das an der BBAW digitalisierte „Etymologische Wörterbuch des Deutschen“ von W. Pfeiffer durchläuft derzeit die letzten Korrekturgänge. Unter der Voraussetzung, dass es keine rechtlichen Einwände gibt, wird es das erste, frei zugängliche, etymologische Wörterbuch der deutschen Sprache im Internet sein. Die Lemmaliste umfasst ca. 21.000 Haupt- und Nebestichwörter.
4. Ebenfalls ausgebaut wird die elektronische Aufarbeitung des ²DWB-Belegarchivs. In Abhängigkeit von den zur Verfügung stehenden Mitteln werden die Bestände weiterer Lieferungen online verfügbar gemacht.

Ein bislang wenig beachtetes und wenig genutztes Feature des WBP ist die Mailingliste, die, etwas versteckt unter dem Menüpunkt „Übersicht“, den Austausch mit anderen Wörterbuchprojekten organisieren kann. Es handelt sich dabei um eine geschlossene Liste, auf der nur die im WBP eingetragenen Projekte schreibberechtigt sind und bei der auch nur diese die Empfänger der Liste darstellen.

Neben der Ressourcenerweiterung gibt es auch Pläne für eine Funktionserweiterung des WBP. Die oben beschriebene Möglichkeit zur Aktualisierung der Lemmalisten ist bisher nur durch einen direkten SQL-Zugriff auf die Daten der eigenen Lemmaliste möglich. Hierfür existiert bereits ein Konzept, das für diesen Zugriff eine Online-Bearbeitungsoberfläche vorsieht, über die man die eigene Lemmaliste komfortabel bearbeiten, korrigieren und erweitern kann.

Grundsätzlich und abschließend lässt sich sagen, dass sich die Möglichkeiten der gemeinsamen Recherche über mehrere lexikografische Ressourcen in den letzten vier Jahren nicht wesentlich verbessert haben. Es lässt sich zwar eine deutliche Zunahme an digital verfügbaren Wörterbüchern feststellen und ebenso haben große Anstrengungen stattgefunden, die verfügbaren Ressourcen zu zentralisieren, allerdings hat es so gut wie keine Umsetzungen gegeben, die einem Projekt wie dem WBP gestatten, standardisiert, ohne die Programmierung von ‘Einzelfalllösungen’, auf externe Ressourcen zugreifen zu können. Der gemeinsame, erweiterte Zugriff auf mehrere Wörterbücher bleibt nach wie vor den ‘Besitzern’ der elektronischen Wörterbücher vorbehalten.

Wünschenswert wäre eine standardisierte Wörterbuch-Schnittstelle, die wenigstens über die Basisfunktionalitäten eines Zugriffs auf elektronische Wörterbücher verfügt. Dazu gehört u.a. die automatisierte Abfrage der aktuellen Lemmaliste, mit der Information über die elektronische ‘Adresse’ des einzelnen Lemmas, dazu gehören auch ein standardisierter XML-Output für eine einfache Lemmaanfrage und nicht zuletzt standardisierte Abfragemöglichkeiten, die sich aus der Mikrostruktur des Wörterbuchartikels ergeben und eben mehr zulassen, als die Frage nach dem Lemma.

3. OWID und Wörterbuch-Portal aus der Sicht der Benutzer (Lothar Lemnitzer)

3.1 Vorbemerkungen

Die Wörterbuchbenutzungsforschung hat in den letzten zwanzig Jahren als angewandte, empirische Forschung solides methodisches Rüstzeug bereitgestellt und zahlreiche interessante Erkenntnisse zutage gefördert. Als eher methodisch orientierte Arbeit sei hier Wiegand (1998) erwähnt, als Beispiele für fundierte empirische Arbeiten Ripfel (1990) und Battenburg (1991). Engelberg/Lemnitzer (2001, S. 70-71) fassen die wichtigsten Erkenntnisse der empirischen Wörterbuchbenutzungsforschung zusammen. Das Manko all dieser Arbeiten ist es, dass sie sich fast ausschließlich auf Printwörterbücher beziehen. Engelberg/Lemnitzer nennen einige wenige Ausnahmen (ebd., S. 72). Es ist offensichtlich, dass gerade hinsichtlich der Modi ihrer Benutzung oder Konsultation ein deutlicher Unterschied zwischen Printwörterbüchern auf der einen Seite und elektronischen und webbasierten lexikalischen Ressourcen auf der anderen Seite besteht. Was Letztere angeht, so bilden sich Benutzungskonventionen als Teil einer kulturellen Praxis, die wahrscheinlich stark von webbasierten Suchmaschinen beeinflusst ist, erst heraus, und auch über Erwartungen an Webwörterbücher lässt sich noch sehr wenig sagen (vgl. aber hierzu die recht neue Arbeit von Christa Knapp 2006). Die Entwickler von Wörterbuchportalen müssen also mit Konzepten der Benutzerführung arbeiten, ohne ausreichend genau über die Erwartungen und das Verhalten der Benutzer Bescheid zu wissen.

In dieser Situation scheint es gerechtfertigt, mit einer explorativen Studie zu beginnen. Diese Studie, deren Ergebnis hier dargestellt werden soll, genügt nicht den in der Wörterbuchbenutzungsforschung etablierten Standards. Der Autor dieses Abschnitts ist die einzige Testperson dieser Studie und in mehrfacher Hinsicht kein durchschnittlicher Benutzer. Diese Studie kann und soll deshalb nicht eine Untersuchung mit einer ausreichend großen Zahl von Probanden ersetzen. Sie hat vielmehr den Status einer formativen Evaluation, die Designentscheidungen der beiden Portale auf den Prüfstand stellt und Hinweise zur weiteren Entwicklung der Portale geben kann.

In diesem Sinn soll nun zunächst die Benutzerführung von OWID und im Anschluss die des WBP unter die Lupe genommen werden.

3.2 OWID

3.2.1 Der Zugang

Die Einstiegsseite, auf die man z.B. bei Eingabe der URL www.owid.de geführt wird, ist aus Benutzersicht unzureichend (vgl. Abbildung 1). Zum einen genügt die nur einen Satz lange Präsentation des Portals („Dieses Portal dokumentiert korpusbasiert erarbeitete lexikografische Forschungsergebnisse verschiedener Projekte des IDS in Form elektronisch aufbereiteter lexikografischer Produkte.“) nicht den Erfordernissen der Benutzeransprache, die solche Fragen wie: „Was soll ich hier? Welche Informationen erhalte ich hier?“ beantworten sollte. Zum anderen wird der neue oder gelegentliche Benutzer mit wörterbuchspezifischen Konzepten konfrontiert – „*lexiko*-Gesamtstichwortliste“, „lexikografisch ausgearbeitete Artikel“ –, die aus der Werkstattsprache der Projekte stammen und zumindest im Stil einer kontextsensitiven Hilfe erläutert werden müssten.

Auch die Schaltflächen, die zu den einzelnen lexikografischen Projekten bzw. Produkten führen, sind nicht selbsterklärend. Immerhin wird man aufgefordert, eine „Stichwortsuche“ zu starten. Im nächsten Abschnitt sollen jedoch zunächst Rahmenseiten des Portals betrachtet werden.

3.2.2 Aufbau der Portalseiten

Die Portalseiten lassen einen konsistenten Strukturplan erkennen (vgl. z.B. Abbildung 2). Dies ist ein für die Benutzerführung außerordentlich wichtiges Merkmal. Es lassen sich klar portalübergreifende Seitenelemente von lokalen Elementen unterscheiden, also von Navigationselementen, die zum Kontext der aktuellen Seite gehören. Auf diese Weise wird dem Benutzer die Orientierung auf der globalen Ebene des Portals sowie auf der lokalen Ebene des jeweiligen Informationsangebots erleichtert. Leider ist dieses Schema nicht bei allen lokalen Informationsangeboten strikt durchgeführt worden, in einigen lokalen Kontexten fehlen Seitenelemente. Besonders der inkonsequente Einsatz von Hintergrundfarben ist hier zu kritisieren. Es handelt sich aber um Mängel, die leicht zu beheben wären, man könnte sie als Schönheitsfehler an einem insgesamt gelungenen und sehr benutzerfreundlichen Gesamtkonzept bezeichnen.

3.2.3 Suche

Bei der Suche sind drei Fälle zu unterscheiden: a) die Suche führt zu keinem Ergebnis (also zu keinem Stichwort in der Lemmaliste); b) die Suche führt zu genau einem Ergebnis (Stichwort); c) die Suche führt zu mehreren Ergebnissen/Stichwörtern. Für den ersten Fall wählen wir zwei Stichwörter: i) *wrxln* und ii) *Rythmus*. Beide Suchanfragen führen auf eine Seite, auf der recht lapidar mitgeteilt wird, dass a) die Suchanfrage 0 Treffer ergab und b) die Suche nun beendet sei. Es ist bemerkenswert, dass beide Suchwörter zum gleichen Ergebnis führen. Zum zweiten Suchwort gibt es mindestens ein Stichwort (nämlich *Rhythmus*), das, technisch gesprochen, lediglich eine Levenshtein-Distanz von 1 vom Suchwort entfernt ist. Dies gibt es bei dem ersten der beiden gewählten Suchwörter nicht. Viele Informationssysteme sind hier weiter, indem sie auf ähnlich geschriebene Wörter hinweisen. Aus Sicht der Benutzerführung kann man sagen, dass dem Benutzer in einem der beiden Fälle eine Information vorenthalten wird, die mit den heutigen technischen Möglichkeiten erzeugt werden kann. Es wäre mindestens wünschenswert, dass auf dieser Seite auf weitere Möglichkeiten der Suchwortgestaltung, etwa auf die Verwendung des Kleene-Star, hingewiesen wird. Insgesamt ist die Benutzeransprache in dem Fall, dass die Suche kein treffendes Stichwort erreicht, als misslungen zu betrachten (vgl. Abbildung 8).

Das Suchergebnis in dem Fall, dass das Suchwort genau ein Stichwort trifft, ist der entsprechende Artikel. Dies ist sicher die Antwort, die der Benutzer erwartet. Auf die Präsentation der Artikel gehen wir später ausführlicher ein.



Abbildung 8: Anzeige zu einer nicht erfolgreichen Suche in OWID

In dem Fall, dass ein Suchwort auf Stichwörter in zwei oder mehr Ressourcen zutrifft (wir haben hierfür das Stichwort *Globalisierung* gewählt), wird man auf eine Art Schaltseite geführt, auf der die beiden Treffer mit einem Hinweis auf die Quellen, denen diese Treffer entstammen, angezeigt werden. Die beiden Treffer sind angemessen als Hyperlinks realisiert, die beim Anklicken zu den jeweiligen Artikeln führen. Die Hinweise, dass die Suche zwei Treffer ergab und nun beendet ist, sind überflüssig und ein weiteres Beispiel für eine misslungene Benutzeransprache.

Es muss weiterhin mit Bedauern festgestellt werden, dass der Benutzer nicht von einer flektierten Form als Suchwort zur kanonischen Zitierform (oder Lemma) geführt wird – wir haben als Beispiele die Suchwörter *gäbe*, *trüge* und *Indizes* gewählt. Auch dies wäre beim Stand der Technik heute möglich. Die Beispiel-Suchwörter zeigen, dass die Ermittlung der kanonischen Zitierform für Nichtmuttersprachler keine einfache Aufgabe ist. Bei Adjektiven, die ausschließlich attributiv gebraucht werden (*oberer*) oder bei deadjektivischen Substantiven (*Kranke*) dürften selbst viele Muttersprachler dazu nicht in der Lage sein. Der Mangel wiegt umso schwerer, als orthografische Angaben eine wichtige Säule im Informationsprogramm von *lexiko* sind.

3.2.4 Die Lemmaliste

Die Lemmaliste ist die Zugriffsstruktur, in der die vier Ressourcen von OWID integriert sind. Die Lemmata der vier Ressourcen sind farblich unterschieden, sodass auch in dem Fall, dass zwei oder mehr Ressourcen einen Artikel zum gleichen Stichwort haben, die-

ses Stichwort zweimal aufgeführt wird, aber eben in unterschiedlichen Farben (Beispiele hierfür sind: [Globalisierung](#) [2 Stichwörter] und [Angst](#) [3 Stichwörter]).

Die Lemmaliste von OWID ist ungewöhnlich lang, sie umfasst mehr als 300.000 Einträge. Zu den meisten Artikeln wird aber nur ein lexikografisches Minimalprogramm angeboten, das korpusbasiert und mit relativ geringem Aufwand erstellt werden konnte. Insgesamt halten wir die Idee einer integrierten Lemmaliste wie auch deren Realisierung für gelungen. Hier wird eine der Stärken eines Wörterbuchportals als einer integrierten Informationsstruktur sichtbar (vgl. Abbildung 9).

The screenshot shows the OWID website interface. At the top, there is a search bar with the text 'elexiko' and the OWID logo. Below the search bar, there are navigation tabs: 'Startseite OWID', 'Projekt OWID', 'Startseite elexiko', 'Wortartikel', 'Stichwortliste', 'Projekt', 'Benutzungshinweise', and 'Erweiterte Suche'. A horizontal menu contains letters from A to Z, with 'Auswahl:' and a dropdown arrow. The main content area is titled 'frei' and includes sections for 'Lesartenübergreifende Angaben' (Orthografie, Normgerechte Schreibung, Worttrennung), 'Herkunft und Wandel', and 'Lesartenbezogene Angaben'. The 'Lesartenbezogene Angaben' section lists various word variants with their specific grammatical or semantic specifications, such as 'Lesart 'unabhängig'' with 'Spezifizierung 'nicht fest angestellt'' and 'Spezifizierung 'ungebunden''. A text box on the right provides additional information about semantic and grammatical details. At the bottom, there is a link 'Zum Zusammenhang der Lesarten'.

Abbildung 9: Beispiel für die Anzeige der Stichwortliste mit Einträgen aus verschiedenen Wörterbüchern in OWID

3.2.5 Die Artikelstruktur

Bevor wir die Struktur der Wörterbuchartikel in *elexiko* unter dem Aspekt der Benutzerführung betrachten – alle anderen Aspekte müssen hier außen vor bleiben, viele von ihnen werden in anderen Artikeln dieser Publikation behandelt –, müssen wir die Unterscheidung zwischen den lexikografisch ausgebauten und den nicht ausgebauten Artikeln aufgreifen.

Die nicht ausgebauten Artikel, etwa 98-99 % aller Artikel, bieten dem Benutzer ein Minimalprogramm an Informationen, vor allem formbezogene Angaben zur Orthografie und Worttrennung. Der interessanteste Aspekt dieser Artikel ist der Link zu CANOO, einem morphologischen Wörterbuch und Analysesystem. Die URL-Struktur der Website von CANOO (<http://www.canoo.net/>) ermöglicht einen direkten Link zur morphologischen Analyse des jeweiligen Stichworts. Diese zielgenaue Verlinkung zu einer externen, aber für verlässlich befundenen Ressource ist eine der interessanteren Möglichkeiten, die das World Wide Web als umfassendes Hypertextsystem bietet. Aus Sicht der Benutzerführung ist diese Angabeklasse in Form eines Verweises als gelungen zu bezeichnen. Leider gilt dies nicht für den ebenfalls vorhandenen Verweis auf die Kookkurrenzdatenbank des Instituts für Deutsche Sprache. Hier ist kein direkter Link möglich und der sachgerechte Umgang mit diesem Service erfordert zunächst eine längere Einarbeitung. Diese Diskrepanz erstaunt umso mehr, als diese Datenbank ebenfalls ein Service des Instituts für Deutsche Sprache ist. Dieser Teil der ressourcenübergreifenden Integration muss also als weniger gelungen bezeichnet werden. Mit Bedauern nehmen wir die semantische Blindheit dieser Artikel zur Kenntnis. Zu gerne hätten wir erfahren, was ein [Beamtenbagger](#) ist und was eine [Serverin](#). Ein Link zu Belegen aus den eigenen Korpora oder auch zu einer Suchmaschine mit dem entsprechenden Suchwort wäre hier durchaus medienadäquat und würde solche bedeutungsbezogenen Fragen, die sicher viele Benutzer haben, beantworten helfen. Dem Benutzer darf hier ruhig etwas Eigeninitiative bei der Interpretation der Ergebnisse zugetraut werden.

Bei den ausgebauten Artikeln von *ellexiko* (vgl. Abbildung 2 und Abbildung 10) werden zunächst ein Überblick über die lesartenübergreifenden Angaben und eine kurze Charakterisierung der einzelnen Lesarten präsentiert. Der Artikel kann an verschiedenen Stellen „aufgeklappt“ und so ein detaillierteres Informationsprogramm erschlossen werden. Die Methode eines auffaltbaren Artikelgerüsts ist aus Benutzersicht prinzipiell zu begrüßen. Er erlaubt sowohl einen schnellen Überblick als auch ein vertieftes Studium der Angaben zum Artikel, und die sind bei den ausgebauten Artikeln sehr reichhaltig. Ob die Anordnung von Angaben im Gerüst und in den auffaltbaren Mikrostrukturteilen glücklich verteilt ist, also den tatsächlichen Nachschlagebedürfnissen der Benutzer entspricht, dies kann nur eine umfangreiche Benutzerstudie beantworten. Der hier gewählte Ansatz zur Ausgestaltung der Mikrostruktur erscheint uns jedenfalls angemessen und mediengerecht.

3.3 Wörterbuch-Portal

Für die Benutzerführung sind folgende Eigenheiten des Portals wichtig: a) Das Portal fasst recht unterschiedliche und etablierte Wörterbuchprojekte der Akademien zusammen. Der Einfluss der Portalbetreiber auf die Struktur der einzubindenden Ressourcen ist also gering; b) die überwiegende Zahl der Wörterbücher (ausgenommen das WDG und *ellexiko*) wendet sich an Akademiker oder ein interessiertes und gebildetes Laienpublikum. Die Nutzung des Portals für die Beantwortung einer spontanen lexikografischen Frage oder für die Deckung einer lexikalischen Lücke dürfte also eher die Ausnahme sein.

3.3.1 Zugang

Der Zugang zum WBP der Akademien ist denkbar einfach. Wenn man die URL eingibt oder über eine Suchmaschine auf die Startseite geführt wird, sieht man im Webbrowser

eine vertikal dreigeteilte Webseite (vgl. Abbildung 4). Die für den Erstbesucher wichtige Information, vor allem welche Information auf diesem Portal erschlossen und präsentiert wird, befindet sich etwas versteckt am unteren Ende der Seite, ist recht kurz und nicht immer glücklich formuliert. Für ein lexikografisches Fachpublikum mag diese Information aber hinreichen. Ein Verweis auf für ein Laienpublikum besser geeignete Ressourcen wäre wünschenswert.



Abbildung 10: Lesartenansicht der *lexiko*-Wortartikel (am Beispiel von [Monat](#))

3.3.2 Aufbau der Portalseiten

Man verlässt die Startseite des Portals nur, wenn man in eines der integrierten Wörterbücher wechselt. Dies ist eine der Stärken und Schwächen des gewählten Konzepts. Bleibt man „innerhalb“ des Portals, dann bewegt man sich in einer virtuellen Umgebung, die konsistent aufgebaut ist und in der man sich nicht verirren kann. Die Lernkurve für die Nutzung dieser Seite ist sehr gering.

Die Portalseite ist vertikal dreigeteilt (vgl. Abbildung 4). Die linke Säule erlaubt den Zugriff auf die im Portal integrierten Wörterbücher. Die rechte Säule ermöglicht den Zugang zu diesen und weiteren lexikalischen Ressourcen, die nicht unmittelbar in das Portal eingebunden sind. Die mittlere Säule ist der aktuellen, vom Benutzer angeforderten Information vorbehalten. Auf der Startseite werden hier (leider nicht mehr ganz neue) Meldungen der Portalbetreiber sowie einige recht kurze Informationen über das Portal selber präsentiert.

Wenn man zu einem der eingebundenen lexikografischen Produkte wechselt, dann wird man zu der Startseite des jeweiligen Projekts bzw. der lexikalischen Ressource geführt und muss sich in der Binnenstruktur dieser Website zurechtfinden, die dem WBP in keiner Weise ähnelt. Insofern ist das Portal am ehesten für geisteswissenschaftliche Forscher geeignet, die wortbezogene Informationen aus einer Vielzahl verschiedener lexikalischer Quellen benötigen. Gelegentliche Benutzer etwa des Wörterbuchs der deutschen Gegenwartssprache oder des Deutschen Wörterbuchs der Gebrüder Grimm tun besser daran, einen direkten Link („Lesezeichen“) auf die Startseite dieser Ressource zu legen.

Eine Stärke des Portals ist es, dass man zunächst zu jeder lexikalischen Ressource eine knappe, aber ausreichend informative Beschreibung erhält, und dies auf der Portalseite selber. Die Portalseite eignet sich also ausgezeichnet zum Stöbern.

3.3.3 Suche

Das Eingabefenster für Suchwörter befindet sich im linken Bereich der Startseite (vgl. Abbildung 4). Darunter kann der Suchraum auf eine Auswahl von Wörterbüchern festgelegt werden. Der Benutzer kann die einzelnen lexikalischen Ressourcen an- bzw. abwählen. Nützlich für den gelegentlichen Benutzer ist es, dass er direkt Informationen zum Umfang der Lemmaliste der jeweiligen Ressource erhält, und indirekt, d.h. durch Anklicken der Wörterbuchsigle, weitere Informationen zu dieser lexikalischen Ressource. Etwas befremdlich wirken die mit drei Punkten („...“) gekennzeichneten Listeneinträge. Dies bedeutet wohl nicht, dass die Liste hier weitergeht, sondern vermutlich, dass weitere Einträge geplant sind.

Ebenso wie bei OWID sollen auch hier drei Fälle unterschieden werden: a) die Suche führt zu keinem Ergebnis (also zu keinem Stichwort in der Lemmaliste); b) die Suche führt zu genau einem Ergebnis (Stichwort); c) die Suche führt zu mehreren Ergebnissen/Stichwörtern. Es geht hier nicht darum, die Abdeckung der lexikalischen Ressourcen zu überprüfen, sondern lediglich darum, ob das System in allen drei Fällen in einer für den Benutzer nachvollziehbaren Weise reagiert. Darüber hinaus wird für jedes Suchwort die Auswahl der Wörterbücher modifiziert. Wir verwenden zum einen die Standardauswahl (vgl. Abschnitt 2.2.3), zum anderen wählen wir alle Ressourcen aus.

wrxln / Rythmus: Das Suchwort *wrxln* ist wie zu erwarten nicht als Stichwort in einer der lexikalischen Ressourcen vorhanden. Wie oben beschrieben wurde, erhält man die Wörter in der Nachbarschaft des nicht vorhandenen Stichwortes (\leq *Wringmaschine*, \geq *Wucher* bei der Standardeinstellung). Die Anzeige von Stichwörtern aus der unmittelbaren alphabetischen Nachbarschaft kann u.U. dazu verhelfen, dass man zu einem falsch geschriebenen Suchwort dennoch das richtige Stichwort erhält, nämlich immer dann, wenn sich der Schreibfehler relativ weit hinten im Suchwort befindet und das korrekt geschriebene Stichwort in einer der ausgewählten Ressourcen vorhanden ist. Bei unserem Verschreiberwort *Rythmus* funktioniert das nicht, da der Verschreiber zu weit vorne im Wort ist. Verwirrend ist lediglich, dass die alphabetisch recht weit entfernten Stichwörter [Aaterling](#) und [Aawasserbrief](#) aus der Lemmaliste des [Deutschen Rechtswörterbuchs](#) (DRW) erscheinen. Man muss daraus schließen, dass sie die letzten Stichwörter der Lemmaliste des DRW sind. Dass dies aber auch nicht zutrifft, merkt man, wenn man etwa nach dem Stichwort *blut* im DRW sucht. Als Ergebnis erhält man eine Kontextliste, die von [Blumenschnade](#) bis [Blutbannbrief](#) reicht.

Die Liste sieht nicht wesentlich anders aus, wenn man alle Ressourcen auswählt. Hier fällt in unserem Zusammenhang lediglich auf, dass es zu den Stichwörtern im Kontext keinen einzigen Eintrag in *ellexiko* gibt. Hier ist möglicherweise die Integration der Lemmaliste nicht gelungen (oder noch nicht abgeschlossen). Verwirrend ist es außerdem, dass als Kontext zu unserem Verschreiberwort *Rythmus* nicht etwa die benachbarten Stichwörter *rütteln* und *Saal* angezeigt werden, sondern das Ende der Q-Lemmastrecke und der Beginn der R-Lemmastrecke. In diesem Fall ist die Anzeige des Lemmalistenkontextes wenig hilfreich.

Meister: Das Suchergebnis in dem Fall, dass das Suchwort genau ein Stichwort trifft, ist das entsprechende Stichwort, grau hinterlegt, im Kontext der Lemmaliste, mit Angabe der Ressourcen, in denen es Artikel zu diesem Stichwort gibt.

Der Fall, dass ein Suchwort auf Stichwörter in zwei oder mehr Ressourcen zutrifft, ist im WBP in überzeugender Weise dadurch gelöst, dass die Lemmalisten der einzelnen Ressourcen nebeneinander präsentiert werden (vgl. Abbildung 7). So kann man auf einen Blick erfassen, in welchen Ressourcen man Artikel zum Suchwort finden kann.

Auch beim WBP muss das Fehlen sprachtechnologischer Unterstützung, vor allem bei der Lemmatisierung flektierter Suchwörter und bei der unscharfen bzw. fehlertoleranten Suche, aus Nutzersicht bemängelt werden. Die Sache wird nicht dadurch besser, dass die Portalentwickler wiederholt auf die mangelnde finanzielle Unterstützung seitens der Betreiber hinweisen. Ein überzeugenderes Argument wäre es, wenn man mit der Zielgruppe des Portals argumentierte. Der typische Benutzer dieses Portals wird vermutlich in der Lage sein, als Suchwort eine korrekt geschriebene kanonische Zitierform anzugeben.

3.3.4 Die Lemmaliste

Hierzu ist wenig anzumerken. Es handelt sich um eine kumulierte Lemmaliste aller Ressourcen (einzig der Bestand von *ellexiko* scheint zu fehlen). Der Umstand, dass bei einigen Ressourcen die Lemmaliste (noch) nicht die gesamte Alphabetstrecke abdeckt, führt zu den oben beschriebenen Kontextartefakten. Einige dieser Artefakte sind für mich als Testnutzer nicht erklärlich. Hier scheint es mir geraten, technische Abhilfe zu schaffen. Der Benutzer könnte sonst falsche Schlüsse ziehen.

3.3.5 Artikelaufbau

Da die Artikel der integrierten Ressourcen ausdrücklich nicht der Gegenstand der Integration in das Portal waren, ist zu diesem Thema aus Benutzersicht auch nichts weiter zu sagen. Alle Aspekte der Führung des Benutzers durch das lexikografische Informationsprogramm liegen in der Verantwortung derer, die die Webpräsenz der jeweiligen Ressource realisiert haben, nicht in der Verantwortung der Portalentwickler. Hier eine Standardisierung zu erreichen bzw. zu verlangen dürfte illusorisch sein.

3.4 Einbeziehung des Benutzers

Viele Portale haben eine communitybildende Funktion. Der Benutzer kann aktiv Einfluss auf die Gestaltung des Portals oder die Auswahl und Präsentation der Informationen nehmen. Wörterbuchportale im Web böten ebenfalls die Möglichkeit, Benutzer ein-

zubeziehen und aus der Rolle des passiven Konsumenten zu holen. Es muss ja nicht gerade das kollektive Erstellen von Artikeln sein wie beim [Wiktionary](#)-Projekt. Hinweise auf Fehler, auf fehlende, aber benötigte Informationen etc. sind als Feedback auch für die Wörterbuchredaktionen wichtig. Umso mehr verwundert, dass es bei beiden Portalen keinen Rückkanal gibt. Dies muss man als Designentscheidung der Portalbetreiber respektieren. Es stellt keinen konzeptuellen Fehler beim Design der Portale dar, ist aber dennoch bedauerlich. Das Fehlen eines Rückkanals ist beim WBP der Akademien insofern noch mehr zu bedauern, als die vermutliche Nutzergruppe überwiegend aus Experten der durch die Wörterbücher dargestellten Gegenstandsbereiche bestehen dürfte. Das Feedback dieser Benutzergruppe sollte den im Portal vereinigten Wörterbuchredaktionen bzw. den Portalbetreibern den Aufwand bei der Etablierung des Kontakts zum Benutzer wert sein.

4. Empfehlungen für die Weiterentwicklung der Portale (*Lothar Lemnitzer*)

Die hier gegebenen Empfehlungen ergeben sich aus der Kritik an der Benutzerführung in Abschnitt 3, die als konstruktive Kritik an zwei insgesamt überzeugenden, aber durchaus ausbaufähigen Wörterbuchportalen verstanden werden sollte.

4.1 OWID

Für die nähere Zukunft von OWID empfehle ich die folgenden Maßnahmen:

- Überarbeitung der Dialoge mit dem Benutzer; zurzeit sind diese Dialoge stark projektzentriert, viele Kommunikationshandlungen seitens der Portalbetreiber scheinen mir unangemessen zu sein, wenn man als übergreifendes kommunikatives Ziel annimmt, die lexikografischen Ressourcen auch für lexikografische Laien und gelegentliche Benutzer mit partiellen Informationsbedürfnissen nutzbar zu machen.
- Stärkere Integration externer Ressourcen und Dienste in das Informationsangebot; als Beispiel seien hier gezielte Anfragen bei Google und Wikipedia, aber auch die Bereitstellung unredigierter Belege aus den Korpora des IDS genannt. Die bisher geleistete Integration externer Ressourcen könnte verbessert werden (z.B. Einbindung der Kookkurrenzdatenbank).
- Stärkere Nutzung sprachtechnologischer Methoden und Werkzeuge, vor allem bei der Lemmatisierung und orthografischen Korrektur von Suchwörtern.
- Stärkere Einbindung des Benutzers durch den Aufbau von Rückkanälen bzw. Kommunikationsstrukturen.

4.2 Wörterbuch-Portal

Für die nähere Zukunft des WBP empfehle ich die folgenden Maßnahmen:

- Eine konsistentere und für den Benutzer weniger verwirrende Form der Präsentation der Lemmalisten-Ausschnitte; das Konzept ist an sich stimmig und überzeugend, es geht hier lediglich um einige Mängel bei der Durchführung.

- Stärkere Integration externer Ressourcen und Dienste in das Informationsangebot; externe Ressourcen bzw. Suchmaschinen sollten zumindest ins Angebot übernommen werden.
- Stärkere Nutzung sprachtechnologischer Methoden und Werkzeuge, vor allem bei der Lemmatisierung und orthografischen Korrektur von Suchwörtern.
- Stärkere Einbindung des Benutzers durch den Aufbau von Rückkanälen bzw. Kommunikationsstrukturen.

Prinzipiell scheint uns ein deutlicheres Bekenntnis der Portalbetreiber zu diesem Portal wichtig zu sein. Hierzu gehört auch die finanzielle Ausstattung der Portalentwicklung. Die Integration und Bereitstellung sprachlicher Ressourcen aller Art für die geisteswissenschaftliche Forschung ist ein Desiderat, das in letzter Zeit auch bei den Förderinstitutionen stärker ins Bewusstsein gerückt ist. Man sollte hier den Mut haben, die Gunst der Stunde zu nutzen.

5. Wo stehen und was leisten Wörterbuchportale? (*Lothar Lemnitzer*)

Eine Einordnung der Wörterbuchportale in die Landschaft elektronischer lexikalischer Ressourcen muss über den Aspekt der Benutzerführung hinausgehen. Wir wollen deshalb zum Abschluss dieses Artikels unseren Blickwinkel erweitern und über mögliche Perspektiven für Wörterbuchportale nachdenken.

Die hier dargestellten Wörterbuchportale müssen sich im folgenden Umfeld digital verfügbarer lexikalischer Ressourcen und Informationssysteme behaupten bzw. sich dem Vergleich stellen:

- **Suchmaschinen** wie Google bieten einige simple Nachschlagefunktionen quasi nebenbei an. Der Nutzer wird auf Suchwörter verwandter Schreibweisen hingewiesen, was in vielen Fällen auf eine orthografische Korrektur hinausläuft („Meinten Sie: ...“). Google bietet darüber hinaus eine „definition:“-Funktion, bei deren Verwendung bevorzugt Definitionen zum Suchwort präsentiert werden. Die Zahl der angezapften Quellen ist für das Deutsche noch gering und geht momentan kaum über die deutsche Wikipedia hinaus. Die Funktion ist aber ausbaufähig und wächst wie von selber mit der Zahl der angezapften Quellen. Nachteil dieser Methode ist es freilich, dass die präsentierten Definitionen nicht weiter geprüft werden. Man muss als Nutzer ein gewisses Misstrauen mitbringen oder von der Qualität solcher Quellen wie der Wikipedia überzeugt sein. OWID könnte mit einer ähnlichen Herangehensweise, die aber die genannten Risiken in sich birgt, seine semantische Blindheit überwinden. Die [Wortwarte](#), eine Sammlung von Neuwörtern des Deutschen, geht diesen Weg, indem zu jedem Stichwort als Teil des Artikels links zu Google und zu Wikipedia angeboten werden.
- **Wörterbuchsammlungen** für den PC, meist von kommerziellen Verlagen angeboten, haben ihre Produkte oftmals in einer Anwendung integriert. In diesem Zusammenhang wird aber meist die Metapher der „Bibliothek“, und nicht die eines „Portals“, bemüht. Dennoch sind beide Arten von Informationssystemen hinsichtlich ihrer Intention, verschiedene Quellen zu integrieren und gemeinsam zu präsentieren, vergleichbar. Moderne PC-Bibliotheken sind mittlerweile auch aus dem engen Raum des allein stehenden Rechners ausgebrochen und bieten Aktualisierungen und eine Erweiterung des Blickfelds über das Netz an. Die meisten PCs verfügen ja heutzutage

über den notwendigen Internetanschluss. Der Vorzug dieser lexikalischen Ressourcen ist, dass die Qualität geprüft ist und die Entwickler dieser Ressourcen für deren Qualität mit ihrem Namen bürgen. Freilich kommen all diese Vorzüge nicht kostenlos.

- **Produkte der Laienlexikografie** wie [Wikipedia](#) oder die zur Familie der Wiki-Anwendungen gehörenden [Wiktionaries](#). Die Qualität der deutschen [Wikipedia](#) ist für ein reines Laienprodukt erstaunlich gut, wird aber sicher nie die Qualität von professionell erstellten lexikalischen Ressourcen erreichen. Dies gilt in stärkerem Maße für die Wiktionaries, bei denen die Kunst sprachlexikografischer Beschreibungen deutlich überwiegt im Verhältnis zum Expertenwissen über die benannten Dinge und Sachverhalte, welches bei Wikipedia im Vordergrund steht.
- **Kommerzielle oder semikommerzielle elektronische Informationssysteme** wie [xipolis](#) (hier findet ein Abonnementssystem mit Mikropayment Anwendung) oder die Wörterbuchportale des [Klett](#)- und des [Langenscheidt](#)-Verlags; bei diesen wird das elektronische Nachschlagen in einigen Produkten als Mittel der Werbung für die Verlagsprodukte und als Element der Kundenbindung angeboten und ist ansonsten für den Benutzer kostenlos.

Die hier beschriebenen Typen von Ressourcen und Informationssystemen stellen eher eine Konkurrenz bzw. Vergleichsbasis für OWID dar. Diese Ressourcen und Dienste zielen ebenso wie OWID auf ein größeres, nur schwer weiter spezifizierbares Laienpublikum.

Anders ist die Situation beim WBP der Akademien. Das dort präsentierte Portfolio lexikalischer Ressourcen ist sozusagen konkurrenzlos, richtet es sich doch überwiegend an ein geisteswissenschaftlich forschendes Fachpublikum.

Dementsprechend unterschiedlich sind die Perspektiven, die wir für die beiden Wörterbuchportale sehen:

Die Zukunft von OWID kann nur darin liegen, konsequent dem [LEO](#)-Prinzip („link everything online“) zu folgen und selber eine Vielzahl akzeptabler Ressourcen zu verlinken bzw. von anderen Seiten verlinkt zu werden. Dies stellt eine Öffnung der eigenen Ressourcen nach außen dar. Jede Form des wissenschaftlichen Autismus verbietet sich da. Dieser soll OWID nicht unterstellt werden, wir wollen aber vor solchen Tendenzen warnen.

Die Zukunft des Portals der Akademien kann darin liegen, dass dieses Portal zum Motor bzw. Katalysator für eine verbesserte bzw. verstärkte Standardisierung, Dokumentation und letztlich auch Vernetzung der an dem Projekt beteiligten Ressourcen wird. Dies ist eine politische Aufgabe, die aber von einer technischen Umsetzung dieser Strategie begleitet werden muss. Dies könnten die Portalentwickler leisten, wenn sie finanziell entsprechend ausgestattet werden. Auch hier gibt es vermutlich die eine oder andere Form des wissenschaftlichen Autismus zu überwinden. Zum anderen könnte das Portal eine andere Qualität der Benutzerforschung und Bedarfsanalyse leisten, als dies in diesem Artikel und den diesem Artikel zugrunde liegenden Recherchen und Experimenten möglich war.

Insofern verstehen die Autoren diesen Artikel als einen Beitrag, der am Beginn einer längeren und hoffentlich fruchtbaren Diskussion über Wörterbuchportale steht.

6. Literaturverzeichnis

- Battenburg, John Douglas (1991): English monolingual learner's dictionaries. A user-oriented study. Tübingen.
- Engelberg, Stefan/Lemnitzer, Lothar (2001): Lexikographie und Wörterbuchbenutzung. Tübingen.
- Haß, Ulrike (2005): *exiko* – das Projekt. In: Haß (Hg.), S. 1-17.
- Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *exiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York.
- Herberg, Dieter (1985): Zur Funktion und Gestaltung von Wörterbucheinleitungen. In: Hyldgaard-Jensen, Karl/Zettersten, Arne (Hg.): Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography May 16-17, 1984 at the University of Copenhagen. Tübingen. S. 133-154.
- Knapp, Christa (2006): Vom World Wide Web zum World Wide Wörterbuch. Internetwörterbücher als Übersetzungshilfsmittel? Saarbrücken.
- Lemnitzer, Lothar (2001): Das Internet als Medium für die Wörterbuchbenutzungsforschung. In: Lemberg, Ingrid u.a. (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. (= Lexicographica, Series maior 107). Tübingen.
- Müller-Spitzer (2007): Das *exiko*-Portal – Ein neuer Zugang zu lexikografischen Arbeiten am Institut für Deutsche Sprache (IDS). In: Rehm, Georg/Witt, Andreas/Lemnitzer, Lothar (Hg.): Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of the Biennial GLDV Conference 2007 (April 11-13, 2007, Eberhard-Karls-Universität Tübingen). Tübingen. S. 179-188.
- Ripfel, Martha (1990): Wörterbuchbenutzung bei Muttersprachlern. Untersuchungsbericht über eine Befragung erwachsener muttersprachlicher Sprecher zur Wörterbuchbenutzung. In: Lexicographica 6, S. 237-251.
- Schnörch, Ulrich (2005): Die *exiko*-Stichwortliste. In: Haß (Hg.), S. 71-90.
- Wiegand, Herbert Ernst (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd. Berlin/New York.

7. Verzeichnis der Nachschlagewerke

(alle Online-Nachschlagewerke zuletzt eingesehen im November 2007)

Deutsches Rechtswörterbuch: <http://drw-www.adw.uni-heidelberg.de/drw/>

Deutsches Wörterbuch von Jacob und Wilhelm Grimm: <http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/dwb/woerterbuecher/suchmaschine/wbgui>

Deutsches Wörterbuch von Jacob und Wilhelm Grimm, Neubearbeitung: <http://pom.bbaw.de/dwb/>

Diskurswörterbuch 1945-55: <http://www.owid.de/Diskurs1945-55/index.html>

exiko: <http://www.owid.de/exiko/index.html>

Erb, Sabine/Harras, Gisela/Proost, Kristel/Winkler, Edeltraud (2004-2007): Handbuch deutscher Kommunikationsverben. (= Schriften des Instituts für Deutsche Sprache 10). Berlin/New York.

Feste Wortverbindungen: <http://www.owid.de/Wortverbindungen/index.html>

Goethe-Wörterbuch: <http://pom.bbaw.de/gwb>

Harras, Gisela/Proost, Kristel/Winkler, Edeltraud (2007): Handbuch deutscher Kommunikationsverben. Teil 2: Lexikalische Strukturen. (= Schriften des Instituts für Deutsche Sprache 10.2). Berlin/New York.

LEO: <http://www.leo.org>

Neologismenwörterbuch: <http://www.owid.de/Neologismen/index.html>

Wikipedia – die freie Enzyklopädie: <http://www.wikipedia.de/>

Wiktionary: <http://de.wiktionary.org>

Wörterbuch der deutschen Gegenwartssprache (WDG): <http://www.dwds.de>

Wörterbuchportal des Klett-Verlages: <http://www.woerterbuchportal.de>

Wörterbuchportal des Langenscheidt-Verlages: <http://services.langenscheidt.de/fremdwb/fremdwb.html>

Wortwarte: <http://www.wortwarte.de>

xipolis: <http://www.xipolis.net>

Wolfgang Bock

Technische Aspekte des OWID-Portals

Abstract

Der Beitrag behandelt die technischen Gesichtspunkte unter denen die Neustrukturierung und -gestaltung des OWID-Portals erfolgte. Ausgehend von den definierten Anforderungen werden die grundlegenden Strukturen des Portals dargelegt. Die verwendeten technischen Standards (Oracle-Datenbank, XML/XSL, HTML/CSS) werden beschrieben und ihr Zusammenspiel erläutert. Exemplarisch wird der Weg eines Artikels von der Entstehung über das Einchecken ins System bis hin zur Aushabe als HTML-Seite dargestellt.

This contribution covers the technical aspects of the restructuring and the new design of the OWID-portal. Based on the requirements the basic structures of the portal are explained. The technical standards used (Oracle database, XML/XSL, HTML/CSS) are described and their interaction is illustrated. The way of an article starting from the emergence, via the checking into the system to the output as an HTML page is treated.

Inhalt:

1. Anforderungen
 - 1.1 Umbau zum Portal
 - 1.2 Ansprechendes, modernes und flexibles Design der Oberfläche
 - 1.3 Einfache Pflege
2. Ausgangssituation
 - 2.1 Die Oberfläche
 - 2.2 Der Aufbau
3. Vom Online-Wörterbuch zum Portal
 - 3.1 XML als Standardformat
 - 3.2 Einbindung in die Datenbank
4. Von XML zur Webseite
 - 4.1 Einchecken eines Wortartikels
 - 4.2 Ausgabe eines Wortartikels

Die Neustrukturierung und Neugestaltung des OWID-Portals war und ist immer noch eine herausfordernde Aufgabe. Sie bestand nicht nur in der Entwicklung und Umsetzung eines neuen Designs für einen zeitgemäßen Auftritt, sondern auch in der Optimierung des technischen Aufbaus. Der zentrale Punkt war jedoch die Einbindung neuer Module in die bestehende Infrastruktur des *lexiko*-Wörterbuchs. Durch die Erweiterung der Präsentation eines einzelnen Projektes zur Darstellung verschiedener Projekte in einem gemeinsamen Kontext wurde das lexikografische Portal OWID geschaffen.

Die Herausforderung, ein Portal wie OWID aus einer bestehenden Infrastruktur heraus zu erschaffen, spielt sich auf drei unterschiedlichen technischen Ebenen ab: Die Datenbankebene beinhaltet alle Daten, die in der Darstellungsebene angezeigt werden. Die Verbindung zwischen diesen beiden Ebenen übernimmt die Kommunikationsebene. Sie regelt, was wie und wann angezeigt wird. Jede dieser drei Ebenen musste erweitert und gegebenenfalls umstrukturiert werden und für jede gab es spezifische Anforderungen.

1. Anforderungen

1.1 Umbau zum Portal

Die wichtigste Aufgabe bestand in der Erweiterung der Datenbank von der Verwaltung eines einzelnen Wörterbuchs zu derjenigen mehrerer Wörterbücher, von denen sich manche in der Zielsetzung und damit auch im internen Aufbau völlig unterscheiden. Auf diese Struktur und ihre Umsetzung in der Datenbank gehe ich im Verlauf des Artikels noch ausführlich ein.

1.2 Ansprechendes, modernes und flexibles Design der Oberfläche

In den Anfängen von *ellexiko* lag der Fokus bei der Gestaltung der Webseiten auf der klaren, einfachen und übersichtlichen Darstellung der Inhalte. Diese Anforderungen galten selbstverständlich auch für das neue Portal, denn der regelmäßige Besucher der Seiten musste sich auch im neuen Design sofort zurechtfinden. Darüber hinaus sollte die Site mit modernen Mitteln der HTML-Codierung aufgebaut werden, um so die Flexibilität zu erhöhen und eine möglichst weitgehende Barrierefreiheit zu erhalten.

1.3 Einfache Pflege

Ein Portal wie OWID ist kein starres Gebilde, das, einmal erstellt, für immer unverändert im Netz steht. Es soll sich verändern und entwickeln. Daher ist es von zentraler Bedeutung, dass Aufbau und Pflege des Portals so einfach wie möglich gestaltet werden. Um dieses Ziel zu erreichen, sollten, wo immer möglich, die einzelnen Bereiche so standardisiert aufgebaut werden, dass sie an einer zentralen Stelle des Systems erzeugt, gepflegt und aktualisiert werden können. Denn nur so können Änderungen zuverlässig und schnell portalweit umgesetzt werden. Zu diesem Zweck ist es wichtig, sich zuerst einmal die Voraussetzungen anzusehen, unter denen der Umbau in Angriff genommen wurde.

2. Ausgangssituation

2.1 Die Oberfläche

Die bereits bestehende Oberfläche des *ellexiko*-Wörterbuchs sollte als Ausgangspunkt für die neue Gestaltung des OWID-Portals dienen. Sie bestand im Wesentlichen aus drei Bereichen (vgl. Abbildung 1): der Navigationsleiste im Kopfbereich (1), einer Spalte zur Darstellung von Wortlisten (2) und dem eigentlichen Inhaltsbereich (3), in dem die Wortartikel und weitere Informationen dargestellt wurden. Dieser Aufbau hatte sich bewährt, und die Nutzer des OWID-Wörterbuchs hatten sich daran gewöhnt. So bestand weder die Notwendigkeit noch der Wunsch, diesen Aufbau grundlegend zu ändern. Er sollte erhalten, modernisiert und erweitert werden (vgl. Abbildung 2).

Die drei beschriebenen Bereiche der Seiten wurden in separaten HTML-Dateien erzeugt und mittels eines Framesets (ein HTML-Konstrukt, das es ermöglicht, mehrere HTML-Dateien in einem Browserfenster darzustellen) angezeigt. Der Vorteil und ursprüngliche Grund für diese Technik liegt darin, dass einzelne Bereiche einer Seite separat geladen werden können, was in Zeiten geringer Datenübertragungsraten einen großen Vorteil darstellte. Die heute übliche Bandbreite der Datennetze rechtfertigt einen Einsatz von Frames nicht mehr.

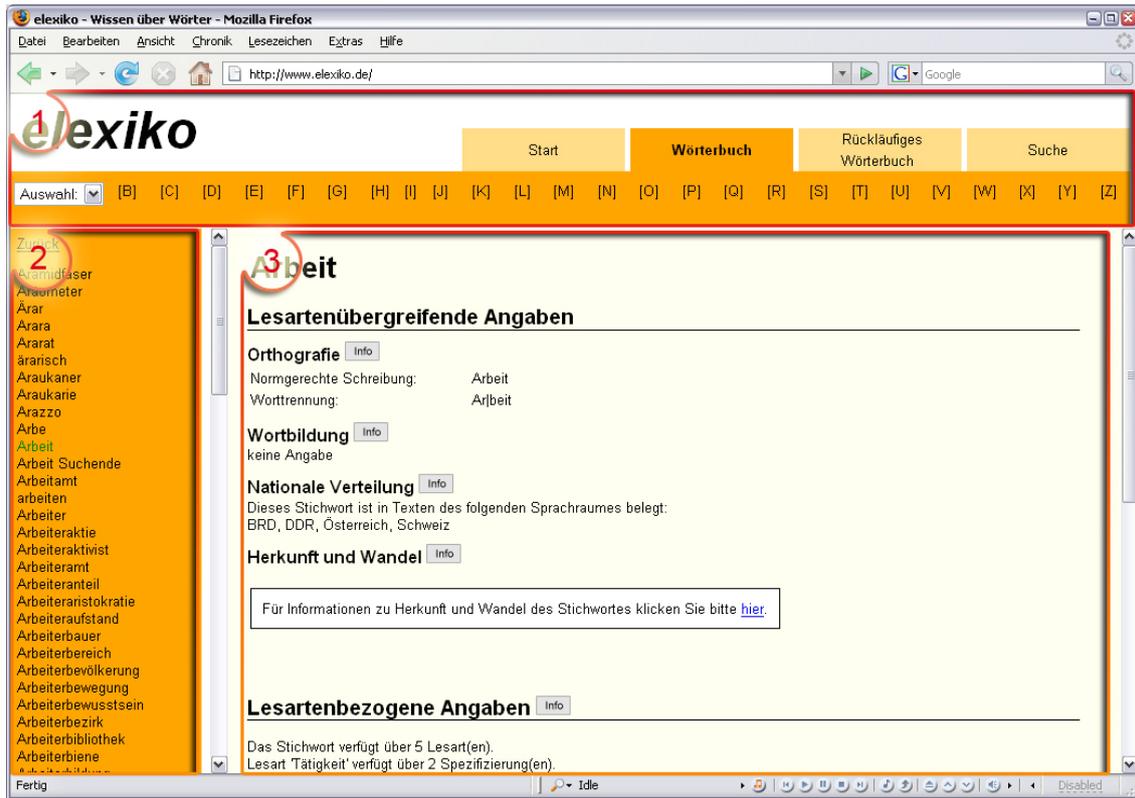


Abbildung 1: Aufbau der alten *elexiko*-Seiten

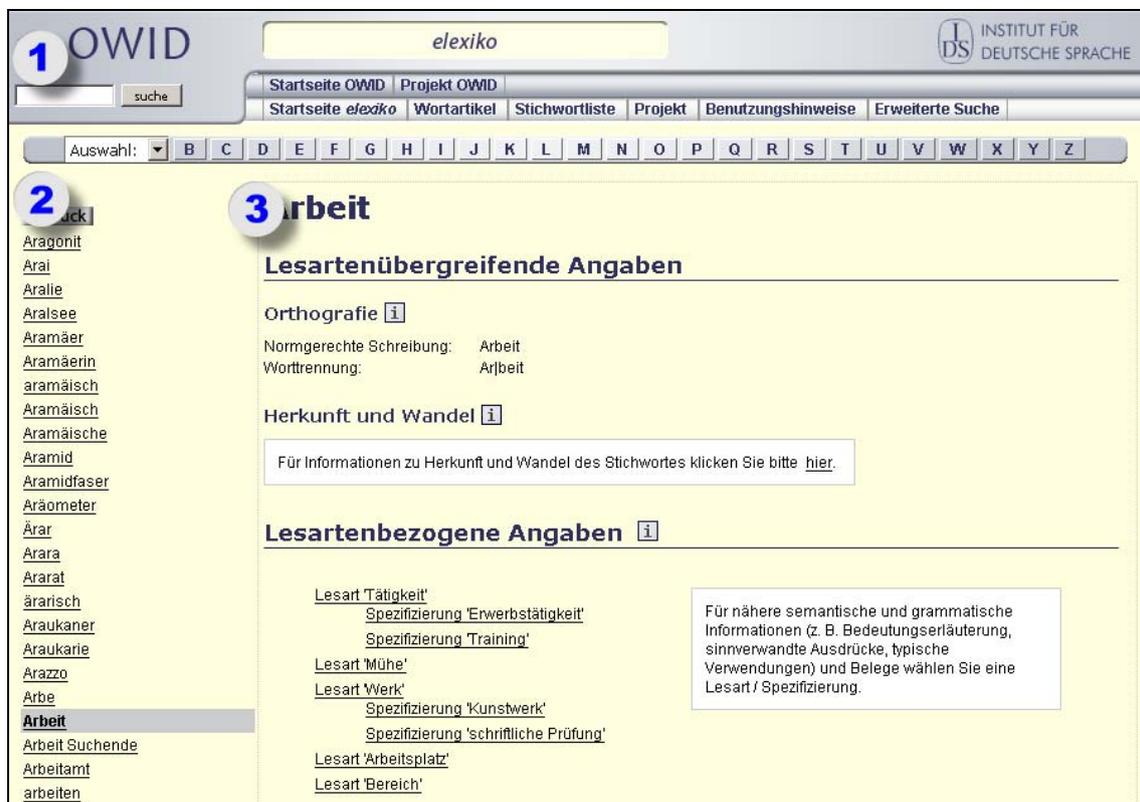


Abbildung 2: Der neue Aufbau der *elexiko*-Seiten

Die Nachteile von Frames liegen in der schlechteren Zugänglichkeit für Suchmaschinen und in der mangelnden Barrierefreiheit. So genannte Screenreader, die sehbehinderten Menschen Webinhalte vorlesen, können den Inhalt von Frames nicht erreichen. Daher sollte bei der technischen Neukonzeption des OWID-Portals auf Frames verzichtet werden.

Ebenso wie der strukturelle Aufbau der Site sollte auch der Aufbau der einzelnen HTML-Dokumente auf einer modernen, flexiblen Architektur basieren. Denn in der jüngeren, der ohnehin kurzen Geschichte des WWW setzt sich die Trennung von Inhalt und Formatierung immer mehr durch, und das aus gutem Grund.

Ursprünglich war das Internet zum schnellen Austausch strukturierter, meist wissenschaftlicher Daten gedacht. Diesen Zweck erfüllte es von Anfang an hervorragend. Nicht die Präsentation, sondern der Inhalt stand im Vordergrund. Mit der zunehmenden kommerziellen Nutzung des Webs änderten sich die Anforderungen, die an HTML gestellt wurden. Das Layout der Seiten, eine Funktion, die in HTML nur rudimentär integriert war, wurde immer wichtiger. Der HTML-Code einzelner Seiten wurde durch Formatierungsanweisungen extrem aufgebläht und gleichzeitig unübersichtlich, da Formatierungen für jedes Element auf jeder Seite einzeln eingesetzt werden mussten. Abhilfe schaffte die Einführung der *CascadingStyleSheets* (CSS), d.h. von Stilvorlagen, die flexibel auf bestimmte Bereiche angewendet werden können.

Um z.B. den Brotttext einer Seite, also den Text, der die eigentlichen Informationen enthält, zu formatieren, reicht es aus, die Formateigenschaften in einem Stylesheet zu definieren. Änderungen in diesem Stylesheet werden sofort auf alle Textpassagen angewendet, die als Brotttext ausgezeichnet sind. Dieses Konzept der Trennung von Formatierung und Inhalt ist nicht nur innerhalb einzelner Seiten, sondern auch über eine ganze Website wie z.B. das OWID-Portal anwendbar. In den HTML-Dateien findet sich also nur noch die reine Information in strukturierter Form. Die Formatierungsanweisungen finden sich in den CSS-Dateien, die auf die Strukturmerkmale der HTML-Dateien zugreifen.

2.2 Der Aufbau

Neben der Struktur der einzelnen Webseiten sollte auch der Aufbau der Funktionen innerhalb des Portals optimiert werden. Die einzelnen Seiten bestehen aus immer wiederkehrenden Elementen, die aber in verschiedenen Zusammenhängen unterschiedliche Inhalte haben können. Am einfachsten und besten wird das am Beispiel des Menüs deutlich.

Die Navigationsleiste, die ausnahmslos auf jeder Seite des Portals sichtbar ist, besteht aus zwei Zeilen, dem Hauptmenü mit den immer gleichbleibenden Punkten „Portal-Startseite“ und „Portal-Informationen“ und dem darunterliegenden Untermenü, das von Wörterbuch zu Wörterbuch unterschiedlich ist. Natürlich ist es möglich, das Menü für jede denkbare Situation vollständig und individuell in HTML zu erzeugen. Die Anzahl der verschiedenen Szenarien ist begrenzt. Sie liegt bei ca. drei verschiedenen Darstellungsformen für jedes Wörterbuch. Rechnet man noch drei Situationen für das Portal hinzu, kommt man bei vier Wörterbüchern auf 15 verschiedene Kombinationen. Da sich die einzelnen Szenarien nur minimal unterscheiden, wäre diese Arbeit per „Kopieren und Einfügen“ schnell erledigt. Dass diese Methode trotzdem nicht effizient ist, zeigt

sich spätestens bei nachträglichen Änderungen am Seitenaufbau, bei denen in diesem Fall 15 verschiedene Codeschnipsel berücksichtigt und gegebenenfalls geändert werden müssten. Diese Vorgehensweise ist nicht nur äußerst mühsam, sie ist darüber hinaus auch extrem fehleranfällig.

In der Erstprogrammierung zwar aufwendiger, aber in der Aktualisierung und Pflege einfacher ist die Erstellung eines eigenständigen Menümoduls, das von jeder Stelle des Portals aufgerufen wird und situationsabhängig auf die Anforderungen reagiert. Beim Aufruf des Menümoduls wird dem Programm ein Parameter übergeben, der beschreibt, für welches Wörterbuch das Menü erzeugt werden soll. Mithilfe dieses Parameters kann das Programm auf die gestellten Anforderungen reagieren. Durch die Übergabe weiterer Parameter können die möglichen Situationen weiter differenziert werden.

Zwar ist die Programmierung eines solchen Programmmoduls deutlich komplexer als die einfache Erstellung einzelner HTML-Szenarien, spätere Änderungen können dagegen viel einfacher und vor allem sicherer durchgeführt werden, da nur ein zentrales Modul für die Darstellung aller Menüs benutzt wird. Änderungen in diesem Modul werden sofort in allen Bereichen des Portals wirksam.

3. Vom Online-Wörterbuch zum Portal

Die dritte und größte Herausforderung lag darin, aus einem homogenen Wörterbuch ein Portal zu schaffen, dessen Inhalte aus ganz unterschiedlichen Quellen mit ebenso unterschiedlichen Zwecken geschaffen werden. Um diese teils gegensätzlichen Aufgaben unter einem Dach zu vereinen, bedurfte es einiger grundlegender Regeln.

3.1 XML als Standardformat

Die Auszeichnungssprache XML (*eXtensible Markup Language*) dient der „Darstellung hierarchisch strukturierter Daten in Form von Textdateien“ (Wikipedia, <http://de.wikipedia.org/wiki/XML>, Stand: 02.12.2007, 16:40 Uhr). Innerhalb der Datei werden also die Daten nicht formatiert, sondern lediglich strukturiert. Dabei unterliegt nur die Vorgehensweise der Strukturierung, die Struktur selbst unterliegt keinen strengen Regeln. Solange die Struktur formal richtig definiert ist, kann das XML-Dokument beliebige Strukturen enthalten. Der Vorteil von XML liegt in der verhältnismäßig einfachen maschinellen Verarbeitung.

Um jedoch eine große Gruppe gleichartiger Dokumente abarbeiten zu können, müssen diese auch zuverlässig auf derselben Struktur basieren. Minimale Abweichungen führen zwangsläufig zum Scheitern des gesamten Systems. Zur Definition und Prüfung von XML-Strukturen stehen *Document Type Definitions* (DTDs) zur Verfügung. Bei der Erstellung und Bearbeitung von XML-Dateien in dafür vorgesehenen Editoren wird die Struktur des Dokuments nach den in der DTD festgelegten Strukturen überprüft und bei Fehlern wird gewarnt. Selbst bei sehr komplexen XML-Konstrukten, wie z.B. den *lexiko*-Wortartikeln, sind dadurch einheitliche und fehlerfreie Dokumente garantiert, vorausgesetzt, die Wortartikel werden ausschließlich mit Editoren bearbeitet, die mit DTDs umgehen können.

Eine direkte Darstellung von XML-Dokumenten in Webbrowsern ist zwar möglich, in der Regel jedoch nicht ratsam, da Browser nicht nur die relevanten Informationen an-

zeigen, sondern auch die umgebenden Strukturelemente. Bei komplexeren Dateien geht die Übersichtlichkeit sehr schnell verloren. Abgesehen davon enthält XML per Definition keine Formatierungsanweisungen. Die Inhalte werden also listenartig und in immer gleicher Weise dargestellt, denn ein Browser kann nicht zwischen verschiedenen Gewichtungen der Elemente unterscheiden.

Um die Artikel korrekt in einem Browser darstellen zu können, müssen sie mit der *eXtensible Stylesheet Language* (XSL) in HTML transformiert werden. Mittels *XSL Transformation* (XSLT) lassen sich XML-Dateien in nahezu jedes andere Dateiformat überführen. Vor der Ausgabe auf den Bildschirm durchläuft das XML-Dokument die Anweisungen einer XSL-Datei und wird entsprechend den enthaltenen Anweisungen als HTML ausgegeben. Auf diese Weise können aus den streng strukturierten XML-Dateien beliebige HTML-Dateien generiert werden. Die Umwandlung von XML in HTML ist wahrscheinlich die häufigste Verwendung von XSLT. Andere Transformationen, wie diejenige ins PDF-Format oder in eine andere XML-Struktur, sind aber ebenfalls üblich.

Durch das Zusammenspiel von XML, DTD und XSLT steht eine hochflexible und gut verarbeitbare Dokumentenstruktur zur Verfügung, die universell in verschiedenste Umgebungen eingebunden werden kann.

3.2 Einbindung in die Datenbank

Zur Strukturierung beliebiger Datenarten ist XML hervorragend geeignet. Gleichzeitig ist es auf diese Aufgabe beschränkt. Es beinhaltet keinerlei Funktionen zur Verwaltung von Daten. Diese Aufgabe übernimmt für das Projekt OWID eine Datenbank aus dem Hause Oracle. In der Datenbank werden neben den verschiedenen Suchfunktionen innerhalb des Portals auch die gesamte Darstellung des Portals im Web und die Verwaltung der Daten auf dem Server definiert. Erst durch die logische Verbindung innerhalb der Wörterbücher und zusätzlich auch über die Wörterbuchgrenzen hinweg wird aus einer Ansammlung von Wörterbüchern ein zusammenhängendes Portal.

4. Von XML zur Webseite

Die Wortartikel werden mit einem Texteditor erstellt, der auf dem Rechner des jeweiligen Lexikografen installiert ist. Um einen Artikel dem Portal zur Verfügung zu stellen, muss er in das System eingecheckt werden.

4.1 Einchecken eines Wortartikels

Beim Einchecken durchläuft der Wortartikel mehrere Prozeduren. Nachdem der Autor die entsprechenden Daten auf seinem lokalen Rechner ausgewählt und hochgeladen hat, wird mittels der im System hinterlegten DTDs die Struktur des Artikels geprüft. Entspricht der Artikel den Vorgaben nicht, wird eine entsprechende Meldung abgegeben und eine Speicherung findet nicht statt.

Ist der Artikel fehlerfrei, wird er mittels PL/SQL, einer Oracle-eigenen Programmiersprache, und XPath, einer standardisierten XML-Abfragesprache analysiert. Bestimmte Elemente, wie z.B. die Lemmazeichengestaltung oder bestimmte Formangaben, werden aus dem Artikel extrahiert und in einzelne Tabellen in der Datenbank geschrie-

ben (vgl. Abbildung 3). Diese extrahierten Daten werden von der Datenbank indiziert und können so um ein Vielfaches schneller durchsucht werden, als direkte Abfragen des Portals auf den XML-Strukturen es ermöglichen würden. Darüber hinaus wird jeder Artikel mit einer eindeutigen ID versehen, anhand deren die in verschiedenen Tabellen abgelegten Informationen den jeweiligen Wortartikeln zugeordnet werden können.

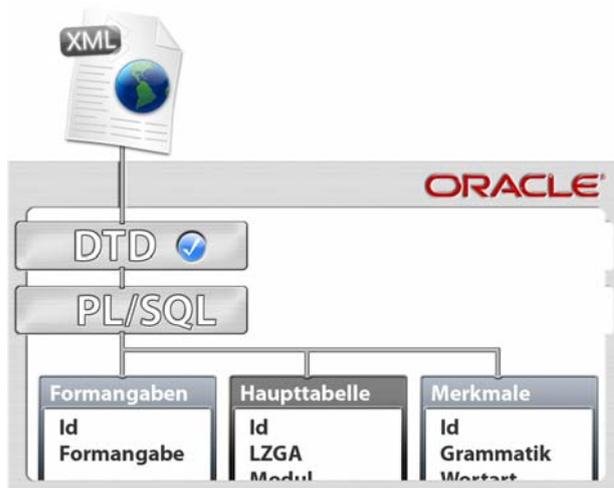


Abbildung 3: Speicherung der XML-Instanzen in die Datenbank

4.2 Ausgabe eines Wortartikels

Sobald der Artikel eingechekkt ist, steht er dem gesamten System zur Verfügung. Die Abfragemöglichkeiten im Portal sind so gestaltet, dass die Suchkriterien ausschließlich auf extrahierte Daten zugreifen. So ist trotz großer Datenmengen eine akzeptable Zugriffszeit gewährleistet. Wird anhand der gestellten Abfrage ein eindeutiger Treffer innerhalb aller relevanten Tabellen gefunden, identifiziert das System den Artikel anhand der ID und gibt ihn aus. Während der Ausgabe wandelt das im System abgelegte XSL die XML-Struktur des Artikels wie oben beschrieben in HTML um. Das generierte HTML enthält jedoch keine Formatierungsanweisungen. Diese werden durch ein externes Stylesheet gewährleistet (vgl. Abbildung 4).

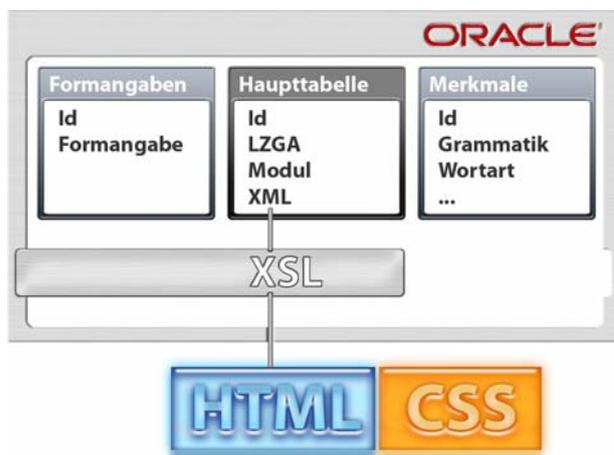


Abbildung 4: Ausgabe der Wortartikel aus der Datenbank

Durch die konsequente Verwendung offener Standards in Verbindung mit einem hochleistungsfähigen Datenbanksystem stellt das OWID-Portal ein flexibles und zukunftsfähiges System dar. Es kann weiter wachsen und bleibt für Veränderungen, Erweiterungen und Verbesserungen offen.

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 tiefergehend 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 granulareren, 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** (Info): Korrekte Schreibung: Video; Worttrennung: Vi|deo
 - Wortbildung** (Info): Wortbildungsart/-typ: unisegmentales Kurzwort Typ: Kopfwort; Gekürztes Wort: Videofilm; Wortart: Nomen
 - Nationale Verteilung** (Info): Dieses Stichwort ist in Texten des folgenden Sprachraumes belegt: BRD, Österreich, Schweiz
- Einzelbedeutungen** (Info): 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 einen Musiktitel gefilmt wird.

Additional elements include a 'Kommentar(e)' button and a 'Hinweis' box: '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.

Marion Hahn / Annette Klosa / Carolin Müller-Spitzer /
Ulrich Schnörch / Petra Storjohann

***elexiko* – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem**

Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche

Abstract

In diesem Beitrag werden wichtige Neukonzeptionen und umfangreiche Nachbearbeitungen einzelner Angabebereiche in *elexiko* erläutert. Die linguistische Konzeption dieser Angaben stellt eine Weiterentwicklung gegenüber der Konzeption dar, wie sie im Band „Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz“ (2005) vorgelegt wurde. Betroffen sind z.B. die Angabebereiche der typischen Verwendungen, der sinn- und sachverwandten Wörter und der Besonderheiten des Gebrauchs.

This paper presents extensive amendments to the original concept of *elexiko*, which has been published in the volume „Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz“ (2005). These complement, for example, information on habitual syntagmatic patterns, sense-related terms, and pragmatics.

Inhalt:

1. Einführung
2. Typische Verwendungen
 - 2.1 Typisierung und Gruppierung der Verwendungsmuster
 - 2.2 Semantische Kategorien und beispielhafte lexikalische Realisierungen
 - 2.3 Lexikografische Umsetzung der Neukonzeption
3. Sinn- und sachverwandte Wörter
 - 3.1 Neuerungen der Darstellung
 - 3.2 Variable Beziehungen
 - 3.3 Inhaltliche Neuerungen
 - 3.4 Methodologische Neuerungen
 - 3.5 Lexikografische Umsetzung der Veränderungen
 - 3.6 Ausblick
4. Weitere Überarbeitungen und Erweiterungen in *elexiko*
 - 4.1 Bedeutungsspektrum, Bedeutungserläuterungen und Kurzetikettierungen
 - 4.2 Besonderheiten des Gebrauchs
 - 4.3 Valenzangaben
5. Technische Aspekte der Umarbeitungen
6. Grundüberlegungen zu einem im Aufbau befindlichen Online-Nachschlagewerk
7. Literatur
8. Nachschlagewerke

1. Einführung (Annette Klosa)

Mit *elexiko* liegt innerhalb von OWID ein Online-Nachschlagewerk zur deutschen Gegenwartssprache vor, das insbesondere Bedeutung und Verwendung von Lexemen beschreibt. Seit 2004 ist *elexiko* mit einer umfangreichen [Stichwortliste](#) und einer ersten Gruppe ausführlich bearbeiteter Wortartikel (dem so genannten [Demonstrationswortschatz](#)) online, seitdem wird der Bestand an ausführlich bearbeiteten Wortartikeln kontinuierlich durch Ausbau des [Lexikons zum öffentlichen Sprachgebrauch](#) erweitert.¹ Daneben wird daran gearbeitet, die [niedrigfrequenten Stichwörter](#) um neue, durch (teil-)automatische Prozesse gewonnene Angaben zu bereichern.

Die linguistische Konzeption von *elexiko* ist im Band „Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz“ (Haß (Hg.) 2005) ausführlich erläutert. Die Konzeption wird daneben in zahlreichen weiteren projektbezogenen [Veröffentlichungen](#) dokumentiert.

Das Kolloquium zu OWID im Frühjahr 2007 hat die Projektgruppe von *elexiko* genutzt, um auf eine Reihe umfangreicher Neukonzeptionen und Nachbearbeitungen einzelner Angabebereiche hinzuweisen, da das Wörterbuch als solches schon in einem Kolloquium im Sommer 2004 der Öffentlichkeit präsentiert worden war. In den folgenden Abschnitten ist nachzulesen, wie die Typischen Verwendungen vollständig neu konzipiert und in allen ausführlichen Wortartikeln überarbeitet wurden und wie die Angabe von sinn- und sachverwandten Wörtern um neue Klassifikationen ergänzt und in der Online-Darstellung optimiert wurde. Hinzu kommt die Darstellung weiterer Ergänzungen und Umarbeitungen bestimmter Angabebereiche (z.B. der Besonderheiten des Gebrauchs), die insgesamt zu einer Verbesserung der Wortartikel geführt haben. Einige grundsätzliche Überlegungen zur Arbeit an einem im Aufbau befindlichen Online-Wörterbuch schließen diesen Abschnitt ab und stellen zugleich ein Fazit der ersten etwa vier Jahre praktischer Artikellarbeit in *elexiko* dar.

2. Typische Verwendungen (Petra Storjohann)

In der Wörterbuchrubrik „Typische Verwendungen“ finden Nutzer Informationen zu syntagmatischen Gebrauchsmustern, sprich mehrgliedrigen Wortverbindungen eines Stichwortes, die häufig im Sprachgebrauch vorkommen und sowohl semantische als auch syntaktische Kombinierbarkeit im unmittelbaren Kontext eines Stichwortes in einer bestimmten Lesart verdeutlichen. Diese werden konsequent aus dem [elexiko-Korpus](#) gewonnen, dort zunächst automatisch extrahiert, einer linguistischen Analyse und Interpretation unterzogen und anschließend von Lexikografen einer bestimmten Lesart bzw. Lesartenspezifizierung zugeordnet. Besonders für Nichtmuttersprachler ist diese Angabe von Interesse, da sie prototypische Verknüpfungsmöglichkeiten von Ausdrücken unterschiedlicher Wortarten und damit gleichzeitig Phänomene der Kollokation und Kolligation verdeutlicht.

2006 wurde dieser Teil der lesartenbezogenen Angaben aus *elexiko* einer vollständigen Neukonzeption unterworfen, mit dem Ziel, inhaltliche und die lexikografische Präsentation betreffende Mängel zu beseitigen. Die noch zahlreichen zu lösenden Aufgaben wurden bei der Beschreibung des alten Konzepts in Haß (Hg.) (2005) festgehalten (sie-

¹ Den aktuellen Bearbeitungsstand kann man [hier](#) abrufen.

he darin Storjohann 2005c) und bezogen sich vorwiegend auf die Suche nach einer konsequenten Systematisierung der angegebenen Muster, auf inhaltliche Ergänzungen und eine benutzerfreundlichere Darstellung. Um die festgehaltenen Ziele an dieser Stelle noch einmal nachvollziehbar zu machen, soll das folgende Artikelbeispiel *zustimmen* in der Lesart ‘sich einverstanden erklären’ die ehemalige Gestaltung dieser Angabe rekonstruieren.

Info

Typische Verwendungen

[Person] nickt zustimmend
 ablehnen oder zustimmen
 [...] müsste zugestimmt werden
 [...] wurde diskussionslos zugestimmt
 [Person] habe bereits zugestimmt
 [Person] hätten grundsätzlich zugestimmt
 dem vorliegenden Entwurf zustimmen
 der geplanten Fusion zustimmen
 einem entsprechenden Antrag zustimmen
 mit einem Kompromiss zustimmen
 mit großer Mehrheit zustimmen
 zustimmend zur Kenntnis nehmen
 sich weigern, einer Scheidung zuzustimmen
 [...] wurde in erster Lesung zugestimmt
 [Person] hat einem entsprechenden Vorschlag zugestimmt
 in der vorliegenden Form nicht zustimmen können

Abbildung 1: Ursprüngliche Anordnung Typischer Verwendungen des Stichwortes *zustimmen*, Lesart ‘sich einverstanden erklären’

Mit dem alten Auflistungsprinzip wurden Muster im Wesentlichen nach steigender Komplexität angeordnet. Das gleichzeitige lexikalische und syntaktische Zusammenspiel solcher Konstruktionen wurde weder inhaltlich systematisch dokumentiert, noch bot die alte Sortierweise für Fremdsprachenlernende nützliche Anleitungen für die eigene Textproduktion bzw. -rezeption. Sowohl semantisch zusammengehörige Verwendungen als auch syntaktisch-strukturell ähnliche Muster, die das Stichwort in seiner Funktion zeigen, wurden nicht gemeinsam gruppiert. Dieser Tatsache war sich das *lexiko*-Team bewusst und Storjohann (2005c) kritisiert selbst:

Das derzeitige Auflistungssystem der typischen Verwendungen vermag es nicht, semantisch zusammengehörige Verwendungen gemeinsam aufzulisten, wenn diese eine unterschiedliche Komplexität aufweisen. Auch syntaktisch-strukturell ähnliche Muster, die z.B. durch Attribuierung komplexer werden, erscheinen in der momentanen Anordnung auseinander gerissen. (Storjohann 2005c, S. 242)

Zwei Ziele ergaben sich aus den Mängeln dieser Anordnung und der lexikografischen Darstellung: Zum einen musste nach einem Kategorisierungssystem gesucht werden, das ähnliche Verwendungsmuster zusammen gruppiert. Die Einteilung der Syntagmen sollte für Nutzer möglichst nachvollziehbar, wenn nötig explizit gemacht werden. Darüber hinaus sollte eine bessere Lösung für das für bestimmte Syntagmenleerstellen genutzte so genannte Klammerelement gefunden werden (siehe Abbildung 1).

Diese leeren Klammern sind bisher unkommentiert, was eine Interpretation seitens des Benutzers oder der Benutzerin erschweren kann. Besonders für Nichtmuttersprachler(innen), die sich für diese Angabe interessieren könnten, ist kaum zu erschließen, was die Klammer repräsentieren soll. (Storjohann 2005c, S. 243)

Da die konkreten sprachlichen Realisierungen eines solchen Elements stark variieren können und sich häufig nicht einfach durch ein prototypisches, zum Muster gehöriges, einzelnes Lexem konkretisieren lassen, musste nach neuen Ideen gesucht werden, um Nachschlagenden eine bessere Vorstellung davon zu vermitteln, welche semantischen Inhalte in den durch das Klammerelement vorgegebenen Positionen charakteristisch wären. Diese Aufgaben standen im Mittelpunkt der Überlegungen im Zusammenhang mit der Neukonzeption dieses Angabetyps.

Dass die Neukonzeption dieser Wörterbuchangabe erst einige Jahre nach der Veröffentlichung des so genannten [Demonstrationswortschatzes](#) erfolgte, hatte gute Gründe. Für die Erstkonzeption lag keine ausreichende Materialgrundlage in Form fertig erarbeiteter Wörterbuchartikel vor, anhand deren typische Arten syntagmatischer Muster erkennbar gewesen wären. Für die z.T. hochfrequenten Stichwörter des Demonstrationswortschatzes und des späteren [Lexikons zum öffentlichen Sprachgebrauch](#) bot das *ellexiko*-Korpus z.T. eine Flut an Musterangaben, für die bezüglich ihrer Analyse, Systematisierung und lexikografischen Darstellung bisher nur unzureichende Erfahrungen zugrunde lagen. Auch andere Nachschlagewerke, besonders solche, die die Bedürfnisse von Nichtmuttersprachlern berücksichtigen und verstärkt mehrgliedrige syntaktische Verbindungen dokumentieren (z.B. WÖRTERBUCH DEUTSCH ALS FREMDSPRACHE, GROSSWÖRTERBUCH DEUTSCH ALS FREMDSPRACHE) boten kaum eine Orientierung, da dort häufig kein konsequentes Gruppieren erkennbar ist und die Anordnungsprinzipien nicht erläutert werden. Die Anzahl signifikanter syntagmatischer Gebrauchsmuster für ein Stichwort aus dem *ellexiko*-Korpus ist auch mit der z.T. vergleichsweise geringen Anzahl der gebuchten Syntagmen dieser Nachschlagewerke nicht vergleichbar, sodass dort kaum Indikatoren für ein für *ellexiko* notwendiges komplexeres Kategorisierungssystem zu finden sind.

Erst nachdem ca. 500 Stichwörter verschiedener Wortarten bearbeitet waren, existierte eine Datenbasis, die gezielt wortartenbezogen hinsichtlich der dort aufgeführten Muster untersucht werden konnte, um zu erkennen, welche Arten Muster für welche Stichwörter im Sprachgebrauch charakteristisch sind, ob diese bezüglich semantischer und/oder syntaktischer Funktion gruppierbar sind und wie diese lexikografisch präsentiert werden könnten. Die erste Phase des Artikelschreibens diente daher auch dem Sammeln von Erfahrung bei der Analyse umfangreicher korpusgestützt extrahierter Informationen und sie stellte ebenso die notwendige Voraussetzung zur Erstellung einer Datengrundlage in Form von Wörterbuchartikeln dar, die für die spätere Auswertung wichtige Einblicke lieferte.

Im Folgenden werden die wesentlichen Änderungen dieser Wörterbuchrubrik anhand des eingangs genutzten Verbbeispiels [zustimmen](#), Lesart 'sich einverstanden erklären' und zusätzlich anhand eines substantivischen Wortartikels, des Ausdrucks [Wissen](#), Lesart 'Kenntnisse', vorgestellt. Auf ein adjektivisches Wortbeispiel wird an dieser Stelle verzichtet, da ein solches bereits in Storjohann/Möhrs (2007) kurz illustriert wird.

2.1 Typisierung und Gruppierung der Verwendungsmuster

Nachdem eine ausreichend große Menge an Wortartikeln untersucht war, wurde deutlich, dass für die Anordnung der typischen Gebrauchsmuster kein wortartenübergreifendes Modell gefunden werden konnte, da die syntaktischen und semantischen Besonderheiten jeder einzelnen Wortart in usuellen Mustern sehr deutlich zum Vorschein kommen. Ergebnis der Auswertung war daher ein neuer umfangreicher wortartenbezogener Kriterienkatalog zur Klassifizierung und Subgruppierung von Verwendungsmustern, der eine formal bzw. inhaltlich sinnvolle Kategorisierung ermöglicht. Obwohl die Anordnungsprinzipien dabei stärker an syntaktischen Funktionen der Stichwörter im unmittelbaren Kontext ausgerichtet sind, ermöglicht es die neue Kategorisierung auch, semantisch zusammengehörige Strukturen gemeinsam zu präsentieren.

Verwendungsmuster von Verben lassen sich prinzipiell in folgende Typen unterscheiden: in Infinitivstrukturen, finite Strukturen und in Strukturen, in denen das Verb als Partizip in adjektivischer Verwendung auftritt. Diese grundlegende Unterteilung der Muster in Strukturtypen wird Nutzern jetzt in Form von Überschriften verdeutlicht (vgl. Abbildung 2).

alte Darstellung	neue Darstellung
<p>Typische Verwendungen Info</p> <p>[Person] nickt zustimmend ablehnen oder zustimmen [...] müsste zugestimmt werden [...] wurde diskussionslos zugestimmt [Person] habe bereits zugestimmt [Person] hätten grundsätzlich zugestimmt dem vorliegenden Entwurf zustimmen der geplanten Fusion zustimmen einem entsprechenden Antrag zustimmen mit einem Kompromiss zustimmen mit großer Mehrheit zustimmen zustimmend zur Kenntnis nehmen sich weigern, einer Scheidung zuzustimmen [...] wurde in erster Lesung zugestimmt [Person] hat einem entsprechenden Vorschlag zugestimmt in der vorliegenden Form nicht zustimmen können</p>	<p>Typische Verwendungen Info</p> <p><u>zustimmen in Infinitiv-Verwendungen</u> einem entsprechenden Antrag zustimmen dem vorliegenden Entwurf zustimmen der geplanten Fusion zustimmen in der vorliegenden Form nicht zustimmen können mit einem Kompromiss zustimmen mit großer Mehrheit zustimmen sich weigern, einer Scheidung zuzustimmen</p> <p><u>zustimmen in finiten Verwendungen</u> [Person] hat einem entsprechenden Vorschlag zugestimmt [z. B. der Regierungschef, das Bundeskartellamt] habe bereits zugestimmt [Person] hätte grundsätzlich zugestimmt [z. B. dem Antrag] wurde diskussionslos zugestimmt [z. B. der Rentenreform] wurde in erster Lesung zugestimmt</p> <p><u>zustimmen als Partizip in adjektivischer Verwendung</u> zustimmend zur Kenntnis nehmen [Person] nickt zustimmend</p> <p><u>Sonstige Verwendungen</u> ablehnen oder zustimmen</p>

Abbildung 2: Beispiel *zustimmen*, Lesart 'sich einverstanden erklären' in alter und neuer Darstellung

Neben der Unterteilung in Grundtypen erfolgt gegebenenfalls eine weitere Subgruppierung. Diese unterscheidet beispielsweise einfache Konstruktionen mit nominalem Bezugswort von Mustern mit Präpositionalphrasen und berücksichtigt u.a. Aspekte wie Genus Verbi (Aktiv vor Passiv), Modus (Indikativ vor Konjunktiv), Tempus (Gegenwartsform vor Vergangenheits- und Futurform) sowie das Vorkommen von Modalver-

ben (Muster ohne Modalverb vor solchen mit Modalverb), Negation (Form ohne Negation vor Form mit Negation) und Numerus (Singular vor Plural). Die Gruppierung der Muster, in denen das Verb als Partizip in adjektivischer Verwendung vorkommt, richtet sich nach dessen attributiver, prädikativer oder adverbialer Funktion. Unter der Überschrift „Sonstige Verwendungen“, die bei jeder Wortart vorkommen kann, werden Paarformeln (wie z.B. *ablehnen oder zustimmen*) und andere formelhafte Muster festgehalten.

Die Typisierung der Substantive orientiert sich an der Verwendung des Stichwortes mit Attribut, in Verbalphrasen und in Sätzen, oder sie orientiert sich daran, ob es selbst als Attribut vorkommt. Auch hier werden die Grundtypen explizit mit Überschriften versehen (vgl. Abbildung 3).

alte Darstellung	neue Darstellung
<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: right;">Info</p> <p>Typische Verwendungen</p> <p>über Wissen verfügen [.] vermittelt grundlegendes Wissen das nötige Wissen vermitteln die Vermittlung von Wissen großes Wissen und Erfahrung handwerkliches Wissen und Können nach bestem Wissen handeln Wissen unter Beweis stellen das Wissen um die Zusammenhänge praktisches und theoretisches Wissen vermitteln das Wissen in die Tat umsetzen mit Wissen und Billigung der [..] Wissen um die Bedeutung der [...] das erworbene Wissen in die Praxis umsetzen</p> </div>	<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: right;">Info</p> <p>Typische Verwendungen</p> <p><u>Wissen mit Attribut</u> großes Wissen und Erfahrung handwerkliches Wissen und Können</p> <p>Wissen um die Bedeutung der [z. B. <i>kulturellen Aufgaben</i>] das Wissen um die Zusammenhänge</p> <p><u>Wissen in Verbalphrasen und Sätzen</u> das nötige Wissen vermitteln praktisches und theoretisches Wissen vermitteln</p> <p>nach bestem Wissen handeln über Wissen verfügen</p> <p>das erworbene Wissen in die Praxis umsetzen das Wissen in die Tat umsetzen Wissen unter Beweis stellen</p> <p>[z. B. <i>der Kursus</i>] vermittelt grundlegendes Wissen</p> <p><u>Wissen als Attribut</u> mit Wissen und Billigung der [z. B. <i>Regierung</i>] die Vermittlung von Wissen</p> </div>

Abbildung 3: Beispiel *Wissen*, Lesart ‘Kenntnisse’ in alter und neuer Darstellung

Kriterien wie vorangestelltes oder nachgestelltes Attribut, Kasus- oder Präpositionsanschlüsse oder, ob das Nomen als Subjekt im Nominativ, als Objekt, in einer Präpositionalphrase mit einfachem Verb oder in satzförmigen Verbindungen mit erweitertem Verbalkomplex vorkommt, spielen u.a. für die weitere Subgruppierung eine Rolle. Auch für Substantive gibt es gegebenenfalls den Mustertyp „Sonstige Verwendungen“, der für Paarformeln und formelhafte Wortverbindungen vorgesehen ist.

Wie auch bei Nomen und Verben bietet sich für Adjektive eine syntaktisch ausgerichtete Anordnung an. Die attributive, prädikative oder adverbiale Funktion steht bei der Mustertypisierung der Adjektive im Vordergrund. Für weitere typeninterne Gruppen existieren Anordnungsprinzipien, die z.B. Muster nach ihrem Bezugsnomen, nach Prä-

position oder dem zugehörigen Verb sortieren sowie Infinitivformen vor flektierte Formen und Verbalphrasen vor Adjektivverbindungen positionieren (vgl. u.a. das Beispiel [unbegrenzt](#), Lesart 'ohne Einschränkung' in Storzjohann/Möhrs 2007).

2.2 Semantische Kategorien und beispielhafte lexikalische Realisierungen

Wie die beiden Beispiele in der alten Darstellung zeigen, befinden sich in einigen Mustern Klammerelemente, die für bestimmte Positionen eine Platzhalterfunktion innehaben. In einigen Fällen sind sie leer, wenn sie kein konkretes, für das Muster typisches Element enthalten, sondern an dieser Stelle verschiedene Lexeme variabel vorkommen können. In anderen Beispielen dagegen stehen darin Informationen semantisch-referenzieller Art, die das darin befindliche Element näher spezifizieren. Typischerweise stellen die Füllungen dieser Leerstellen im Muster Subjekte, Objekte oder Attribute dar, die z.T. den Kategorien [Person], [Zahl], [Land] zugeordnet wurden. In der Erstkonzeption war für diese Elemente keine Konvention bezüglich der Vergabe semantischer Kategorien oder einheitlicher Bezeichnungen getroffen worden, was ein inkonsequentes Ausfüllen dieser Angabe zur Folge hatte. Das Klammerelement erfüllt eine lexikogrammatistische Funktion, indem es eine semantische Füllung und syntaktische Anschlüsse veranschaulicht. Deshalb war es für die Neukonzeption wichtig, wenn möglich, diese beiden Informationen in die Angabe zu integrieren.

Die Auswertung der vorhandenen Wortartikel bedeutete daher eine Überprüfung bestehender semantischer Kategorien, also eine Bestandsaufnahme existierender Füllungen für die einzelnen Wortarten. Als Konsequenz wurde ein systematischer Kategorienkatalog als redaktionelle Richtlinie für die Lexikografen erstellt, der auch um zahlreiche neue Kategorien ergänzt wurde, z.B. [Eigenschaft], [Zeitangabe], [Einheit], [Ortsname].

Um syntaktische Merkmale des Klammerelements integrieren zu können, wurde beschlossen, wenn nötig, die entsprechende Leerstelle zusätzlich mit einem exemplarischen Korpusbeispiel zu illustrieren. Für die unterschiedlichen Lexeme, die keiner bestimmten Kategorie zugeordnet werden können, wurde entschieden, diese obligatorisch mit Beispielen für konkrete sprachliche Realisierungen zu versehen. Der exemplarische Charakter dieser sprachlichen Füllungen wird anhand der Einführung „z.B.“ verdeutlicht. Beispielhafte Leerstellenrealisierungen helfen, besonders Fremdsprachenlernenden eine Vorstellung davon zu vermitteln, welche Art Element die entsprechende Leerstelle einnimmt. Um dem Nutzer gleichzeitig syntaktische Anschlüsse zu demonstrieren, werden die Korpusbeispiele in der vom Muster geforderten grammatischen Form und nicht in ihrer Grundform angegeben. Für eine benutzerfreundliche Unterscheidung zwischen den Angaben des eigentlichen Musters, des Platzhalters und der konkreten sprachlichen Realisierung des Klammerelements wurden neue typografische Konventionen eingeführt. Die Beispiele in Abbildung 4 sind unterschiedlichen *lexiko-*Wortartikeln entnommen und illustrieren einige Fälle, bei denen die alte Form der Klammerelemente der neuen zum Vergleich gegenübergestellt wird.

<p><i>Urteil</i> 'juristischer Beschluss' alt: ein [...] Urteil fällen neu: ein [Eigenschaft: z. B. <i>abschließendes, endgültiges</i>] Urteil fällen</p> <p><i>Vorhaben</i> 'Projekt' alt: [Person] ließ sich von seinem Vorhaben nicht abbringen neu: [Person: z. B. <i>der Einbrecher</i>] ließ sich von seinem Vorhaben nicht abbringen</p> <p><i>Prozess</i> 'Rechtsstreit' alt: [Person] einen fairen Prozess machen neu: [Person: z. B. <i>dem Angeklagten</i>] einen fairen Prozess machen</p> <p><i>ereignen</i> 'geschehen' alt: [...] ereignete sich nach Angaben der Polizei am [Zeit] neu: [z. B. <i>der Mord</i>] ereignete sich nach Angaben der Polizei am [zeitangabe: z. B. <i>Dienstagvormittag</i>]</p> <p><i>anfordern</i> 'verlangen' alt: [...] kann kostenlos angefordert werden neu: [z. B. <i>die Broschüre</i>] kann kostenlos angefordert werden</p> <p><i>erheblich</i> 'beträchtlich' alt: erhebliche Vorbehalte gegen [...] neu: erhebliche Vorbehalte gegen [z. B. <i>den Vorschlag</i>]</p>
--

Abbildung 4: Beispiele für konkrete sprachliche Realisierungen des Klammerelements

Das einzige Klammerelement, das nach wie vor leer bleibt, steht ausschließlich für Satzerweiterungen, wie in Abbildung 5 z.B. im Stichwort [Zweifel](#), Lesart 'Skepsis' gezeigt wird.

<p>um alle Zweifel zu beseitigen, [...] es gibt begründete Zweifel, dass [...] es bestünden keine Zweifel, dass [...] [Person] hat große Zweifel, ob [...] [Person] meldete Zweifel an, ob [...] es bestehen erhebliche Zweifel daran, dass [...] [Person] ließ aber nie einen Zweifel daran, dass [...]</p>
--

Abbildung 5: Beispiele für Satzerweiterungen

Die Füllung des Klammerelements machte schnell deutlich, dass in mehreren Fällen ein Bedürfnis für weitere lexikografische Erläuterungen bestand. Das trifft vor allem für Verwendungseinschränkungen (z.B. regionaler Art beim Stichwort [Erfolg](#), Lesart 'Triumph' im Muster *mit gutem Erfolg maturieren*) oder grammatische Besonderheiten zu, die im eigentlichen kontextuellen Gebrauch zu beobachten sind. So kann es für ein Muster oder eine Gruppe an Mustern typisch sein, dass das Stichwort häufig pluralisch verwendet wird (z.B. [Bedürfnis](#), Lesart 'Wunsch'). Bei adjektivischen Stichwörtern dienen lexikografische Erläuterungen dazu, Wörterbuchbenutzer beispielsweise darauf aufmerksam zu machen, dass ein Stichwort in usuellen Verbindungen häufig in einer bestimmten Steigerungsform auftaucht (z.B. [billig](#), Lesart 'kostengünstig'). Um solche Beobachtungen festhalten zu können, stehen jetzt Lexikografen Kommentare und Hinweise zur Verfügung, die sich entweder auf ein spezielles Muster, eine Gruppe an Verwendungen, einen speziellen Mustertyp oder auf die gesamte Angabe beziehen können.

2.3 Lexikografische Umsetzung der Neukonzeption

Die Neukonzipierung dieser Angabe bedeutete eine Vielzahl von verschiedenen Arbeitsschritten, die zunächst diverse linguistische Analysen von Korpusdaten und existierenden *ellexiko*-Wortartikeln beinhalteten. Die Ergebnisse dieser Untersuchung mündeten in ein umfangreiches neues Konzept zur Typisierung und Gruppierung mehrgliedriger syntagmatischer Verbindungen für die einzelnen Wortarten. Die Erstellung eines neuen inhaltlichen Konzepts ging auch mit der Ausarbeitung neuer redaktioneller Richtlinien einher. Die Neukonzeption bedeutete auch, alle bereits bestehenden Wörterbuchartikel redaktionell nach den neuen Vorgaben nachzuarbeiten. Diese Artikel mussten auf die neue Struktur umgestellt werden (für technische Aspekte der Umarbeitung siehe Abschnitt 5). Diese Überführung machte eine nachträgliche Typisierung und Gruppierung der bereits aufgenommenen Muster für ca. 500 Stichwörter in all ihren Lesarten erforderlich, die Klammerelemente mussten konsequent einer Kategorie zugeordnet und gegebenenfalls mit einer beispielhaften Korpusfüllung ergänzt werden.

Der Wunsch nach einer benutzerfreundlichen Präsentation der inhaltlichen Veränderungen machte auch ein Konzept notwendig, in dem u.a. die Darstellung der Typen und Gruppen der typischen Verwendungsmuster sowie die Darstellung der Klammerelemente festgelegt wurden (zur technischen Umsetzung der Online-Präsentation siehe Abschnitt 5).

Für die Zukunft bleibt noch eine Aufgabe in diesem Bereich offen, die Analyse der Adverbien. Die Bearbeitung der *ellexiko*-Stichwörter im Wortschatzbereich des Demonstrationswortschatzes und des [Lexikons zum öffentlichen Sprachgebrauch](#) schloss überwiegend Nomen, Verben und Adjektive ein. Für die Analyse typischer Gebrauchsmuster von Adverbien lagen daher lange Zeit nicht genügend Beispielartikel zur näheren Untersuchung vor. Um auch für diese Wortart eine benutzerfreundlichere Beschreibung und Präsentation in *ellexiko* bieten zu können, wurden bereits Adverbien verstärkt als Stichwörter dem zu erarbeitenden Artikelbestand hinzugefügt, um sie künftig einer Auswertung unterziehen zu können.

3. Sinn- und sachverwandte Wörter (Petra Storjohann)

In der Wörterbuchrubrik „Sinn- und sachverwandte Wörter“ ist es seit der Online-Publikation des Demonstrationswortschatzes im Jahr 2004 ebenfalls zu wichtigen Veränderungen bezüglich der Darstellung und des Inhalts gekommen. Im Rahmen des Kolloquiums 2007 wurden nur einige der grundlegenden Neuerungen anhand von Wörterbuchbeispielen vorgestellt. Insgesamt hat aber genau diese Wörterbuchangabe von den korpuslinguistischen und -technologischen Entwicklungen der letzten Jahre am meisten profitiert. Diverse Korpusverfahren ermöglichen den gezielten Zugang zu semantisch verwandten Ausdrücken und versetzten *ellexiko* in die Lage, seinen Nutzern ein noch breiteres Spektrum an Informationen zu Sinnrelationen im Sprachgebrauch anbieten zu können. In den letzten Jahren konzentrierte sich *ellexiko* dabei auf eine systematischere und benutzerfreundlichere Präsentation der Ergebnisse und suchte verstärkt nach Möglichkeiten der Erläuterung von Gebrauchssituationen und Verwendungseinschränkungen.

3.1 Neuerungen der Darstellung

Bereits 2004 war *lexiko* das erste Nachschlagewerk zur deutschen Gegenwartssprache, das gebündelt Beziehungsarten der Synonymie, Antonymie, Inkompatibilität, Komplementarität, Konversonymie und Reversivität sowie auch Beziehungen der Über- und Unterordnung und Teil-von-Beziehungen für eine spezielle Lesart bzw. für eine Lesartenspezifizierung eines Ausdrucks beinhaltete. Diese wurden nach der traditionellen Einteilung in horizontale und vertikale Arten untergliedert und entsprechend angeordnet (vgl. Abbildung 6). Jedes sinnrelationale Partnerwort wurde, sofern es selbst Bestandteil der Stichwortliste war, als Link realisiert und mit einem Korpusbeleg veranschaulicht, um das gemeinsame Vorkommen im sprachlichen Kontext zu illustrieren.

Angst 'Furcht'

Sinnverwandte Wörter

<small>Info</small>	Synonym(e):	Bammel	<small>Beleg(e)</small>
		Bangigkeit	<small>Beleg(e)</small>
		Furcht	<small>Beleg(e)</small>
		...	
<small>Info</small>	Hyperonym(e):	Emotion	<small>Beleg(e)</small>
		Gefühl	<small>Beleg(e)</small>
<small>Info</small>	komplementär(e) Partner:	Hoffnung	<small>Beleg(e)</small>
		Traum	<small>Beleg(e)</small>
		...	
<small>Info</small>	inkompatible(r) Partner:	Aggression	<small>Beleg(e)</small>
		Depression	<small>Beleg(e)</small>
		...	

Abbildung 6: Alte Darstellung der semantischen Beziehungen am Beispiel *Angst*, Lesart 'Furcht'

Für Stichwörter, die in entsprechenden Lesarten/Lesartenspezifizierungen über zahlreiche Relationspartner verfügen und viele verschiedenartige Beziehungen eingehen, ergab sich das Problem, dass die vielen Beziehungsarten für die Nutzer nicht gut erfassbar waren oder ein Erkennen der Unterschiede und die Einordnung der Arten durch die Fachterminologie erschwert wurden. Daher wurde die Einteilung in horizontale und vertikale Strukturen aufgegeben und die einzelnen semantischen Beziehungen werden nun durch leicht verständliche Überschriften in grundlegende Arten unterteilt (vgl. Artikel [Angst](#), Lesart 'Furcht' in Abbildung 7).

Innerhalb einer konkreten Beziehungsart wurden die Relationsausdrücke bisher alphabetisch angeordnet. An diesem Prinzip wird immer noch festgehalten, da bei jedem aufgenommenen sinnverwandten Ausdruck davon ausgegangen werden kann, dass er im *lexiko*-Korpus mehrfach gemeinsam mit dem Stichwort in einem entsprechenden Kontext vorkommt und damit alle Partnerwörter über eine ausreichende Signifikanz verfügen. In Fällen, in denen unterschiedliche semantische Gruppierungen erkennbar sind, wird dieses Anordnungsprinzip aufgehoben und die Ausdrücke werden je nach semantischer oder thematischer Fokussierung in separate Gruppen unterteilt, dort aber wiederum alphabetisch aufgelistet. Auch in diesem Zusammenhang werden Überschriften zur

Markierung zusammengehöriger Gruppen genutzt, und zusätzliche Kommentare erläutern die vorgenommene Unterteilung (vgl. z.B. Artikel [Änderung](#), Lesart ‘das Modifizieren’ in Abbildung 8).

Angst 'Furcht'

Sinnverwandte Wörter

Beziehung(en) der Bedeutungsgleichheit/-äquivalenz

Info Synonym(e): [Bammel](#)
[Bangigkeit](#)
[Furcht](#)
 ...

Beziehung(en) des Bedeutungsgegensatzes

Info komplementär(e) [Hoffnung](#)
 Partner: [Traum](#)
 ...

Info inkompatible(r) [Aggression](#)
 Partner: [Depression](#)
 ...

Beziehung(en) der Überordnung

Info Hyperonym(e): [Emotion](#)
[Gefühl](#)

Abbildung 7: Darstellung der sinnrelationalen Beziehungsarten mit Überschriften am Beispiel *Angst*, Lesart ‘Furcht’

Änderung 'das Modifizieren'

Beziehung(en) der Unterordnung

Info Hyponym(e): **Betonung des Aspektes des Hinzufügens**

[Ergänzung](#)
[Präzisierung](#)

Kommentar:
 In dieser Gruppe sind spezielle Arten einer **Änderung** zusammengefasst, die den Aspekt des Hinzufügens von etwas betonen.

Betonung des Aspektes des Wegnehmens

[Aufhebung](#)
[Kürzung](#)
[Streichung](#)

Kommentar:
 In dieser Gruppe sind spezielle Arten einer **Änderung** zusammengefasst, die den Aspekt des Wegnehmens oder Reduzierens von etwas betonen.

Abbildung 8: Semantische Gruppierung innerhalb einer Beziehungsart am Beispiel *Änderung*, Lesart ‘das Modifizieren’

Solche Gruppierungen mit expliziten Erläuterungen sind besonders häufig für die Beziehung der Inkompatibilität anzutreffen, eine Beziehungsart, die kaum in anderen Nachschlagewerken dokumentiert ist (siehe u.a. [Freizeit](#), Lesart ‘Zeitraum’). Besonders für brisante Ausdrücke, wie sie im Bereich der hochfrequenten *lexiko*-Stichwörter häufig vorkommen, hat diese Relation eine wichtige diskursfokussierende Funktion (siehe z.B. die Wortartikel [Flexibilität](#), Lesart ‘Anpassungsfähigkeit’ und [Globalisierung](#), Lesart ‘Internationalisierung’).

3.2 Variable Beziehungen

Semantische Beziehungen existieren nicht zwischen Lexemen, sondern sind kontextuell gebunden, sprich sie können einer Lesart bzw. Lesartenspezifizierung zugeordnet werden. Alternativ gibt es erste kognitive (Croft/Cruse 2004) und konstruktionsgrammatische Ansätze (Murphy 2006), bei denen Sinnrelationen dynamisch als Konstruktionsstrukturen realisiert werden. Ungeachtet des theoretischen Ansatzes bestätigen Korpusdaten die starke Kontextabhängigkeit sowie die Dynamik und Flexibilität der sprachlichen Realisierung bestimmter Beziehungen. Variabilität zwischen Sinnrelationen, wie sie im Sprachgebrauch häufig und systematisch vorkommt, wurde bisher in keinem deutschsprachigen Nachschlagewerk, das Informationen zu Synonymen oder Gegensatzwörtern enthält, beschrieben. Nutzern ein besseres Verständnis von der Flexibilität der Sprache im tatsächlichen Gebrauch zu vermitteln, ist dabei eines der großen Anliegen von *lexiko*, auch wenn dies für die lexikografische Darstellung immer eine besondere Herausforderung bedeutet. Die folgenden Korpusbelege 1-4 demonstrieren ein derartiges systematisches Variieren von Beziehungsarten zwischen den Ausdrücken *Gefahr – Risiko* und *kritisch – ernst*.

1. Das Auswärtige Amt rät zurzeit noch von selbst organisierten Touren nach Kaschmir ab. Auf der Website heißt es: »Wegen der **Gefahr** terroristischer Gewalttaten wird vor Reisen nach Jammu und Kaschmir gewarnt.« Wer jetzt schon eine Reise nach Kaschmir unternehmen möchte, muss sich des **Risikos** bewusst sein, sollte aber auf jeden Fall Exkursionen in die Randgebiete der Provinz außerhalb von Srinagar vermeiden. (Die Zeit, 18.03.2004, Nr. 13, S. 79.)
2. Großes **Risiko** birgt eben große **Gefahr**. (taz, 25.11.1994, S. 25.)
3. Ein reales Bild über die **ernste** Situation in einigen Branchen gibt erst der Bezug zur Erwerbspersonenzahl, schreibt das IWW. **Kritisch** sei die Lage in der chemischen Industrie (63 Prozent Kurzarbeiter), Metallherzeugung (62 Prozent), [...]. (taz, 16.01.1991, S. 11.)
4. Ihr Gesundheitszustand sei „**ernst**, aber nicht **kritisch**“ erklärte ihr Arzt. (taz, 10.11.1989, S. 6.)

Während in Beleg 1 und 2 die Beziehung zwischen den Konzepten, die mit *Gefahr – Risiko* bezeichnet werden, zwischen Synonymie und konditionaler Implikation schwankt, variiert die Relation zwischen den mit *kritisch – ernst* bezeichneten Zuständen zwischen einer Bedeutungsgleichheit und einer kontrastiven Beziehung. Auf diese Variabilität wurde bisher in *lexiko* in Form von Kommentaren hingewiesen. Diese Kommentare werden jetzt durch das Symbol des Pfeils am entsprechenden Partnerwort ersetzt (vgl. Artikel [Gefahr](#), Lesart ‘Bedrohung’ in Abbildung 9).

Mit der Mouse-over-Funktion erscheint ein Kästchen mit entsprechender Information, die auf die kontextuelle Variabilität der Beziehung hinweist. Die Wahl einer neuen Präsentation dieser Information resultierte aus der Vielzahl an Kommentaren mit unterschiedlichen Inhalten in den bislang ausgearbeiteten Artikeln, die jetzt aufgrund einer anderen Darstellung besser voneinander getrennt werden.

Gefahr 'Bedrohung'

Beziehungen der Bedeutungsgleichheit/-äquivalenz

Info **Synonym(e):** [Bedrohung](#) Beleg(e)
[Gefährdung](#) Beleg(e)
[Risiko](#) ↑ Beleg(e)

Info **Sonstige Beziehung(en):** [Risiko](#) ↑ Beleg(e)

Vgl. auch **Risiko** als Synonym.

Kommentar:

Zwischen **Gefahr** und **Risiko** besteht ein konditionales Verhältnis. Ein **Risiko** existiert nur dann, wenn es die Möglichkeit einer **Gefahr** gibt. Daraus resultiert auch ein sich bedingendes, implizites Verhältnis.

Abbildung 9: Verweise auf Variabilität von Beziehungsarten am Beispiel *Gefahr*, Lesart 'Bedrohung'

3.3 Inhaltliche Neuerungen

Eine wesentliche inhaltliche Änderung ergab sich aus den Lücken bestehender traditioneller paradigmatischer Klassifizierungsmodelle,² die sich zu stark an einem idealistischen Sprachsystem orientierten und die verschiedene Beziehungsarten, wie sie im Sprachgebrauch realisiert werden, nicht abbilden konnten. Die Arbeit mit einem umfangreichen Korpus machte schnell deutlich, dass Ausdrücke eine Vielzahl von Sinnrelationen eingehen, die keiner herkömmlichen Kategorie zugeordnet werden können. Für alle diejenigen Beziehungsarten, die keiner klassischen Gruppe angehören, wurde daher zunächst eine Kategorie „Sonstige Beziehungen“ eingeführt. Betont werden muss, dass diese Kategorie vorerst den Charakter einer temporären Datensammlung hat. Erste kleinere Auswertungen bestehender Wortartikel ergaben aber bereits wichtige Einsichten, welche Arten hier typischerweise vorkommen.

Unter der Bezeichnung „Sonstige Beziehungen“ sind verschiedene Beziehungen zwischen Konzepten zusammengefasst, die mittels Lexikalisierungen gleicher Wortart realisiert werden und die kontextuell bestimmte konzeptuell-semantic Aspekte betonen. So setzen Sprecher beispielsweise Ausdrücke in Relation zueinander, die kausale (z.B. zwischen *Ursache*, Lesart 'Auslöser' und *Wirkung*, Lesart 'Effekt') oder konditionale Zusammenhänge lexikalisieren (z.B. *zusammenarbeiten*, Lesart 'kooperieren' – (*sich*) *austauschen* bzw. *vernetzen* (vgl. Abbildung 10), die eine Beziehung der semantischen Steigerung bzw. Graduierung akzentuieren (z.B. *Entlassung*, Lesart 'Kündigung' – *Suspendierung*) oder einen modalen Gegensatz zum Ausdruck bringen (z.B. *Bitte* – *Aufforderung*, Lesart 'Aufruf'). Darüber hinaus werden auch solche Relationen in dieser Rubrik dokumentiert, die verschiedenartige Folgebeziehungen verdeutlichen, z.B. resul-

² Auch wenn in *lexiko* die Typisierung der Sinnrelationen an traditionelle paradigmatische Kategorien angelehnt ist, wird nicht von einer strikten Trennung zwischen Syntagmatik und Paradigmatik ausgegangen. Sprachgebrauchsuntersuchungen semantischer Relationen verdeutlichen, dass u.a. Beziehungsarten der Synonymie und des Bedeutungsgegensatzes sprachlich häufig als syntagmatische Konstruktionen realisiert sind (vgl. Jones 2002, Murphy 2006, Storzjohann 2007).

tative Relationen (zwischen *Entdeckung* und *Erkenntnis*, Lesart ‘Wissen’) sowie responsive Beziehungen zwischen *fragen*, Lesart ‘sich erkundigen’ und *antworten*, Lesart ‘sich äußern’ oder zwischen *aufzwingen*, Lesart ‘aufdrängen’ und *hinnehmen*. Da unter dieser Gruppe verschiedene Beziehungen zusammengefasst werden, wird jede einzelne Relation obligatorisch mit einer Überschrift und/oder mit einem lexikografischen Kommentar näher erläutert (vgl. Abbildung 10).

zusammenarbeiten ‘kooperieren’

Info Sonstige Beziehung(en):

Handlung des Zusammenarbeitens als Bedingung
[\(sich\) austauschen](#) Beleg(e)

Kommentar:
 Zwischen *zusammenarbeiten* und *(sich) austauschen* besteht eine konditionale Beziehung, wenn die Handlung, die mit *zusammenarbeiten* bezeichnet wird, eine Voraussetzung dafür ist, die Handlung, die mit *(sich) austauschen* bezeichnet wird, ausüben zu können.

Bedingung für die Handlung des Zusammenarbeitens
[vernetzen](#) Beleg(e)

Kommentar:
 Zwischen *vernetzen* und *zusammenarbeiten* besteht ein konditionaler Zusammenhang, bei dem die Handlung, die mit *vernetzen* bezeichnet wird, eine notwendige Voraussetzung für die Handlung ist, die mit *zusammenarbeiten* bezeichnet wird.

Abbildung 10: Kategorie „Sonstige Beziehungen“ am Beispiel *zusammenarbeiten*, Lesart ‘kooperieren’

Liegen künftig umfangreichere empirische Untersuchungsergebnisse vor, könnte das klassische Klassifizierungsmodell um neue differenzierte Kategorien ergänzt werden. Die Dokumentation und Beschreibung solcher Beziehungen erfolgt in *lexiko* aus folgenden Gründen: Erstens werden in zahlreichen Synonymiken Ausdrücke als bedeutungsgleich gebucht, die im Sprachgebrauch kausale oder konditional-implizite Beziehungen zum Ausdruck bringen. Ein solches Beispiel ist u.a. die Beziehung zwischen *schützen*, Lesart ‘vor Gefahren bewahren’ und *decken*, die in zahlreichen Nachschlagewerken typischerweise als bedeutungsgleich gebucht ist (vgl. DUDEN 8), im *lexiko*-Korpus aber häufig durch die Struktur [PERSON *decken*, um PERSON zu *schützen*] kontextuell realisiert ist, oder die Relation zwischen *sinnlos* und *überflüssig*, die im *lexiko*-Korpus häufig in die Struktur [*sinnlos*, weil/da/daher *überflüssig*] eingebettet ist. Hier dient die Rubrik „Sonstige Beziehungen“ u.a. dem Aufdecken solcher Diskrepanzen. Ein weiterer Grund, warum solche Beziehungsarten dokumentiert und beschrieben werden, liegt darin, dass es sich bei zahlreichen Partnerwörtern um statistisch signifikante Kookkurrenzen handelt. Die Auswertung der Kookkurrenzen spielt bei der Bedeutungsanalyse der Stichwörter eine zentrale Rolle, und die entsprechenden semantischen Mitspieler werden in *lexiko* unterschiedlichen Angaben zugeordnet. Da auch die Relationspartner, die unter „Sonstige Beziehungen“ aufgenommen werden, sehr häufig in unmittelbarer Umgebung zum Stichwort auftauchen und in ihren gemeinsamen Kontexten dabei vielfältige, in Beziehung zueinander stehende Konzepte bezeichnen, ist die Beschreibung solcher Partnerwörter eine sinnvolle Ergänzung zur allgemeinen Beschreibung der Bedeutung und Verwendung eines lexikalischen Ausdrucks.

3.4 Methodologische Neuerungen

Dank der korpuslinguistischen und -technologischen Entwicklungen am IDS konnten für *lexiko* die Verfahren zur Gewinnung der semantischen Relationspartner weiter ausgebaut werden. Die beiden grundlegenden und für die Extrahierung der Sinnrelationen zueinander komplementären Verfahren, die corpus-driven und die corpus-based Methode,³ wurden ab 2006 durch eine Korpusanalysemethode ergänzt, die gezielt Ausdrücke mit ähnlichen Kookkurrenzprofilen ermittelt (vgl. Belica/Keibel 2007). Die Entwicklung dieser Methode (Modul *Related Collocation Profiles*) ist Bestandteil derzeitiger Forschungsarbeiten im Programmbereich Korpuslinguistik, der u.a. Korpusanalyse- und Erschließungsmethoden entwickelt und im Rahmen der korpuslinguistischen Denk- und Experimentierplattform (CCDB) testet und verfügbar macht.⁴ Die durch diese Analyseverfahren ermittelten Ausdrücke mit Kohäsionsbeziehungen zum Suchwort werden in *lexiko* zunächst als bedeutungsverwandte Kandidaten betrachtet, und die Beziehung zum Suchwort wird anschließend durch eine gezielte Korpusabfrage im *lexiko*-Korpus linguistisch untersucht und interpretiert, um eine relevante semantische Relation zu verifizieren. Diese zusätzliche Methode wirkt sich besonders im Zusammenhang mit Verbanalysen positiv auf die Ermittlung von sinn- und sachverwandten Ausdrücken aus. Da Verben vor allem durch ihre syntaktische Valenz charakterisiert sind und daher typischerweise in ihrer unmittelbaren Umgebung zusammen mit adverbialen Kookkurrenzen und nominalen Objektmitpielern auftauchen, werden durch die primär genutzte corpus-driven Methode (einer Kookkurrenzanalyse) sinnrelationale Ausdrücke gleicher Wortart, sprich Verben, kaum erfasst. Hier bietet die Analyse von Ausdrücken mit semantischen Ähnlichkeitsprofilen für die Ermittlung von Synonymen und Ausdrücken des Bedeutungsgegensatzes erhebliche Vorteile.⁵

3.5 Lexikografische Umsetzung der Veränderungen

Neben neuen inhaltlichen Überlegungen waren zahlreiche redaktionelle und technische Schritte bis zur vollständigen Online-Umsetzung der Neuerungen nötig. In den vergangenen drei Jahren wurden Änderungswünsche und konzeptionelle Neuüberlegungen, die sich aus der lexikografischen Praxis, der Auswertung sprachlicher Daten aus dem *lexiko*-Korpus sowie auf Anregungen von Kollegen hin ergaben, gesammelt. Zu Beginn des Jahres 2007 konnte daraufhin ein kleineres Konzept vorgelegt werden, das die wichtigsten Neuerungen vorstellte. Nach den notwendigen texttechnologischen Strukturanpassungen der Artikel (siehe dazu Abschnitt 5) konnten für die neu zu bearbeitenden Stichwörter die veränderten Vorgaben bereits eingearbeitet werden. Für ca. 500 bereits bearbeitete Stichwörter bedeutete das aber gleichzeitig eine redaktionelle Nacharbeitung. Für diese Stichwörter galt es, die Relationspartner hinsichtlich ihrer Anordnung und Eingruppierung zu überprüfen. Besonders für Stichwörter, für die bei der Suche nach Sinnrelationen bestimmte Korpusverfahren noch nicht zum Einsatz kamen, wurden existierende Angaben erneut mit Korpusdaten kritisch geprüft. Auf diese Weise

³ Die explizite Unterscheidung der Korpusverfahren geht auf Tognini-Bonelli (2001) zurück. Der Einsatz beider Methoden zur Extrahierung semantisch verwandter Ausdrücke wird u.a. in Storjohann (2005b; 2005d) beschrieben.

⁴ Siehe auch <http://www.ids-mannheim.de/kl/> und <http://www.ids-mannheim.de/kl/projekte/methoden/>. Für direkten Zugang zur CCDB-Datenbank siehe <http://corpora.ids-mannheim.de/ccdb/>.

⁵ Für ein konkretes Verbbeispiel siehe Storjohann (2006).

konnten für zahlreiche Stichwörter wichtige Ergänzungen vorgenommen werden. Zu den Aufgaben der redaktionellen Nachbearbeitung gehörte auch, semantisch zusammengehörige Partnerwörter aus der bisherigen alphabetischen Anordnung herauszunehmen und sie in Gruppen zu unterteilen sowie diese mit Überschriften und in der Regel mit lexikografischen Kommentaren zu versehen. Um die neuen Überlegungen auch online entsprechend präsentieren zu können, musste ebenfalls ein Konzept entwickelt und realisiert werden (siehe Abschnitt 5).

3.6 Ausblick

Die korpusgestützte Untersuchung semantischer Relationen bildet einen wichtigen Bestandteil der lexikalisch-lexikologischen Forschung in *ellexiko*. Ergebnisse und Erkenntnisse dieser Forschung werden praxisbezogen in die lexikografische Beschreibung und Präsentation der sinn- und sachverwandten Ausdrücke im *ellexiko*-Wörterbuch eingebunden. Daher ist in dieser Rubrik regelmäßig mit Neuerungen zu rechnen. In Storjohann (2005b) wurden künftige Ziele festgehalten, denen wir zwischenzeitlich bereits deutlich näher gekommen sind. Der Wunsch nach stärkerer Kommentierung bestimmter Verwendungssituationen und Gebrauchsnormen führte zu einer redaktionellen Nachbearbeitung sämtlicher Wörterbuchartikel. Der Forderung nach einer Gruppierung der Partnerwörter und der Nachvollziehbarkeit dieser Einteilung wird jetzt durch Überschriften und gezielte lexikografische Erläuterungen Rechnung getragen. Ebenso fiel auf,

dass in den Texten des *ellexiko*-Korpus regelhafte diskursrelevante Beziehungen auftauchten, die in der Literatur nur unzureichend oder gar nicht beschrieben sind. Für diese Relationen muss nach geeigneten Beschreibungen und Darstellungen gesucht werden. (Storjohann 2005b, S. 263).

Erste Ansätze werden u.a. mit den Beschreibungen der „Sonstigen Beziehungen“ gezeigt.

Besonders kritisch muss heute noch mit der Systematik der Verlinkung von Relationsausdrücken untereinander umgegangen werden. Viele korpusbasiert erarbeitete Wörterbücher müssen sich der Kritik aussetzen, eine sehr mangelhafte Rückverlinkung zwischen bidirektionalen Verweisstrukturen aufzuweisen (vgl. u.a. Paradis/Willners 2006).⁶

[...] we investigated how many of the antonyms are given in both directions. Of all the antonym relations in the dictionary, only 37% are given in both directions. [...] Again, it is not clear to us why these pairings should be helpful for the learner in the above directions only. (Paradis/Willners 2006, S. 217)

Obwohl mittels Software und Linkbanksystemen ein konsequentes Verlinken bidirektionaler Beziehungen erreicht werden kann, fehlt es *ellexiko* derzeit an technischen Verfahren zur systematischen Überprüfung von Vernetzungsstrukturen zwischen bedeutungsverwandten Ausdrücken, um mögliche Mängel bezüglich der Rückverlinkung zu vermeiden. Daher soll in den kommenden Jahren unter der Leitung von Carolin Müller-Spitzer ein detailliertes texttechnologisches Konzept zur Vernetzung erarbeitet werden, das neben vielen inhaltlich unterschiedlichen Vernetzungsstrukturen (z.B. im Bereich der Wortbildung) u.a. in Bezug auf semantische Relationen inhaltliche wie technische Lösungsvorschläge für ein konsequentes Verlinken und Rückverlinken und für die Do-

⁶ Die Untersuchung von Carita Paradis und Caroline Willners bezieht sich auf Verlinkungsstrukturen des Collins COBUILD Advanced Learner's English Dictionary (2003).

kumentation von Vernetzungs- und Verweisstrukturen enthalten soll. Deshalb stehen neben der fortlaufenden Erforschung semantischer Beziehungen im Sprachgebrauch die Erarbeitung und die technische Realisierung dieses Konzepts im Vordergrund der weiteren Arbeiten in diesem Bereich.

4. Weitere Überarbeitungen und Erweiterungen in *ellexiko*

Neben den in den vorhergehenden Abschnitten beschriebenen umfangreichen Neukonzeptionen und Ergänzungen ganzer Informationsrubriken sind weitere Angaben in *ellexiko* im Laufe der Artikelbearbeitung überdacht und, wo nötig, gegenüber der ursprünglichen Konzeption verbessert oder ergänzt worden. Diese werden im Folgenden vorgestellt, bevor abschließend aus diesem Vorgehen einige grundsätzliche Gedanken zur Arbeit an einem im Aufbau befindlichen Online-Wörterbuch abgeleitet und diskutiert werden.

4.1 Bedeutungsspektrum, Bedeutungserläuterungen und Kurzetikettierungen (Annette Klosa)

Da viele Wortartikel für *ellexiko* im Rahmen des so genannten Demonstrationswortschatzes im Jahr 2004 verfasst worden waren, entsprachen sie den inzwischen gewonnenen redaktionellen Erfahrungen nicht mehr völlig. Besonders auffällig geworden war dies im Hinblick auf das Spektrum der Lesarten insgesamt, die Formulierung der Bedeutungserläuterung und die Abstimmung zwischen Bedeutungserläuterung und so genannter lesartenidentifizierender Kurzetikettierung.

4.1.1 Zu Lesarten und Bedeutungsspektrum

Zu Beginn der Artikellarbeit galt der Grundsatz, dass nur solche Lesarten⁷ im Wortartikel beschrieben werden, die im *ellexiko*-Korpus belegt sind. Dieser Grundsatz gilt auch weiterhin, da *ellexiko* ein korpusgestütztes Online-Wörterbuch ist:

Das wohl wichtigste methodische Prinzip in *ellexiko* ist die Korpusbasiertheit. Es bedeutet, dass in *ellexiko* keine einzige Angabe enthalten ist, die allein aus der individuellen Sprecherkompetenz der LexikografInnen abgeleitet wurde. Die Angaben in *ellexiko* werden strikt aus dem 1,3 Milliarden Textwörter umfassenden und regelmäßig aktualisierten *ellexiko*-Korpus gewonnen. (Haß 2005b, S. 7)

Zusätzlich enthalten die Wortartikel jetzt einen Hinweis, wenn allgemeinsprachlich zu erwartende Lesarten in Sekundärquellen⁸ gebucht sind, diese sich aber im *ellexiko*-Korpus nicht zeigen (z.B. im Wortartikel [Arbeit](#)).

Auch in der ursprünglichen Angabe „Konzeptfamilie und Homonymie“, die online neu „Zum Zusammenhang der Lesarten“ heißt, wurde systematisch ergänzt. Hier galt zunächst die Bearbeitungsvorgabe, immer dann Verschiebungsrelationen wie z.B. Metaphorisierung oder Generalisierung zwischen Lesarten zu beschreiben, wenn diese mit einiger Sicherheit angenommen werden konnten (z.B. im Wortartikel [Information](#)). Da

⁷ Zur grundlegenden Konzeption des Bedeutungsspektrums in *ellexiko* vgl. Haß (2005a).

⁸ Zu den Sekundärquellen vgl. grundsätzlich Wiegand (1998, S. 140); Sekundärquellen für *ellexiko* sind z.B. folgende Wörterbücher: Duden (1999), Klappenbach/Steinitz (1964ff.).

in *ellexiko* aber unter einer Lemmazeichengestaltangabe auch Lesarten behandelt sind, die in anderen Wörterbüchern als Homonyme getrennt werden, ist es notwendig, auf die Nicht-Verwandtschaft zwischen einzelnen Lesarten explizit hinzuweisen (z.B. im Wortartikel [bereiten](#)). Für die Nachschlagehandlung selbst hat der Ansatz nur einer Lemmazeichengestaltangabe keine Nachteile:

Im elektronischen datenbankartigen Wörterbuch wie *ellexiko* beginnt die Nachschlagehandlung in der Regel mit der Eingabe einer gesuchten Wortform, nicht mehr mit dem Suchen in der alphabetischen Stichwortliste. Die Wortartikel selbst sind dabei zunächst unsichtbar. In *ellexiko* erhält ein Nutzer auf die Eingabe eines gesuchten Formativs zunächst eine mittels Signalwörtern oder Kurzetikettierungen erstellte Übersicht über die Lesarten, die mit diesem Formativ verbunden sind, ganz unabhängig von möglichen semantischen, wortbildungsmäßigen oder gar historischen Zusammenhängen dieser Lesarten untereinander. (Haß 2005a, S. 174f.)

Ausführlich erläutert werden daneben Fälle, in denen die Lesarten als semantisch in irgendeiner Weise zusammengehörig verstanden werden, aber nicht alle Lesarten einzelnen Verschiebungsrelationen zugeordnet werden können (z.B. im Wortartikel [Service](#)). Schließlich wird vermerkt, wenn einzelne Lesarten nicht völlig scharf voneinander zu trennen sind (z.B. im Wortartikel [normal](#)), was insbesondere bei Lexemen einer Wortklasse mit systematischer Polysemie auftritt (z.B. bei Institutionenbezeichnungen wie [Kindergarten](#)).

4.1.2 Zu Bedeutungserläuterungen und Kurzetikettierungen

Bei der redaktionellen Überprüfung aller bisher für die *ellexiko*-Wortartikel verfassten Bedeutungserläuterungen kam es vor allem darauf an, die Konsistenz innerhalb der Formulierung der Bedeutungserläuterung zu gewährleisten und zu kontrollieren, ob die Bedeutungserläuterung dem ursprünglichen Konzept entsprach.⁹ In *ellexiko* wird jede Lesart einer Wortklasse, genauer einer semantischen Wortklasse (z.B. Handlungsprädikator, Einstellungsprädikator) oder einer satzfunktionalen Klasse (z.B. Quantor, Junktor), zugeordnet.¹⁰ Die redaktionellen Richtlinien sehen vor, in Abhängigkeit hiervon die Bedeutungserläuterung zu formulieren. Die Handlungsprädikatoren kann man etwa daran erkennen, dass in der Bedeutungserläuterung der Handelnde und das Objekt, an dem gehandelt wird, benannt werden (z.B. im Wortartikel [planen](#), Lesart 'konzipieren').

Daneben wurden die Formulierungen der Bedeutungserläuterungen sprachlich geglättet und gegebenenfalls mit Beispielen angereichert. An der Bedeutungserläuterung für *planen* zeigt sich auch, wie aus dem *ellexiko*-Korpus gewonnene Konkretisierungen die Bedeutungserläuterung veranschaulichen; im Beispiel sind dies die Wörter *Architekt* und *Zeichner* bzw. *Projekt* und *Haus*:

Mit **planen** bezeichnet man eine Handlung, bei der eine Person(engruppe) (z.B. ein Architekt oder Zeichner) einen detaillierten Entwurf aufstellt, eine Skizze entwirft oder ein Konzept (z.B. für ein Projekt oder ein Haus) entwickelt.

⁹ Zur grundlegenden Konzeption von Bedeutungserläuterungen und Kurzetikettierungen in *ellexiko* vgl. Storjohann (2005a).

¹⁰ Eine Erläuterung der semantischen und satzfunktionalen Wortklassen, denen einzelne Lesarten in *ellexiko* zugewiesen werden, findet sich in Haß (2005a, S. 166ff.).

Insbesondere wurde bei diesem Arbeitsgang zwischen den so genannten Kurzetikettierungen und den Bedeutungserläuterungen eine Einheit hergestellt. Die Kurzetikettierungen kennzeichnen in *lexiko* einzelne Lesarten eines Stichwortes; es wird also auf eine sonst in Wörterbüchern übliche Nummerierung der Lesarten verzichtet. Daneben erfüllen die Kurzetikettierungen zwei Funktionen: Zum einen kann mit ihnen der begrifflich-referenzielle Bereich einer Einzelbedeutung schnell erfasst werden. Zum anderen wird durch das Anlegen von Links das Navigieren zwischen den verschiedenen Lesarten eines Stichwortes erleichtert.

Im Zuge der Überarbeitungen in diesem Bereich wurde deshalb versucht, die Etikettierungen so zu wählen, dass sie dem Nutzer einen möglichst guten Eindruck von den verschiedenen Lesarten verschaffen. Außerdem besteht jetzt zwischen Etikettierungen und Bedeutungserläuterungen eine Einheit, indem die Kurzetikettierung sprachlich, sofern möglich, in die Bedeutungserläuterung integriert wurde, wie im folgenden Beispiel in den Bedeutungserläuterungen zu den Lesarten 'Schuldigkeit', 'Aufgabe' und 'Wettkampfübung' des Wortartikels *Pflicht*:

Lesart 'Schuldigkeit': Mit **Pflicht** bezeichnet man eine Verantwortung, eine Schuldigkeit oder Normerwartung, die an das Handeln oder Verhalten einer Person(engruppe) gestellt wird.

Lesart 'Aufgabe': Mit **Pflicht** bezeichnet man eine Aufgabe oder Tätigkeit, die eine Person(engruppe) hat oder übernehmen bzw. ausüben soll.

Lesart 'Wettkampfübung': Mit **Pflicht** bezeichnet man im Sport den Teil einer Wettkampfübung, der in seiner Ausführung vorgeschrieben ist.

Das Konzept für die Benennung der Lesarten mithilfe von Kurzetikettierungen hat sich insgesamt bewährt. Ergänzend ist die Schaltfläche „Das Bedeutungsspektrum im Überblick“ in der Online-Ansicht der Artikel hinzugekommen. Nach Anklicken der Schaltfläche werden die jeweiligen Kurzetikettierungen mit Bedeutungserläuterungen untereinander angezeigt (z.B. im Wortartikel *Pflicht*), wodurch von jeder Lesartenansicht aus ein Einblick in das gesamte Bedeutungsspektrum gewährleistet ist.

4.2 Besonderheiten des Gebrauchs (*Marion Hahn und Ulrich Schnörch*)

Erfahrungsgemäß zeigt erst die fortlaufende lexikografische Praxis, d.h. die empirische Erprobung aller Kategorientypen, ob und welche Revisionen vorgenommen werden müssen. Während der Bearbeitung des 250 Wörter umfassenden Demonstrationswortschatzes bestätigte sich diese Erfahrung einmal mehr, weswegen Haß (2005c, S. 275) unter anderem im Zusammenhang mit den Gebrauchsbesonderheiten „Änderungen der Kategorien- und Artikelstruktur“ prognostizierte. Die mittlerweile mehrjährige Erfahrung in der Artikelproduktion und der Umgang mit dem zugrunde liegenden Korpus bildeten nun den Anlass und die Grundlage für die Modifizierung und Erweiterung der lexikografischen Konzeption in diesem Teilbereich. Hauptsächlich betraf das folgende Angebotsarten:

Die Verwendungsspezifik, die bisher in der Rubrik „Besonderheiten des Gebrauchs“ in vier Hauptkategorien behandelt wurde („Einstellung des Sprechers“, „Situationsbezug“, „Sachgebiet[e]“, „Text[sorten]bindung“), wurde um zwei neu geschaffene Angebotsarten erweitert, nämlich um die „Themengebundene(n) Verwendung(en)“ und die „Verwendung in mehrteiligen Eigennamen“.

Themengebundene Verwendung(en):

Bei der Arbeit mit dem *ellexiko*-Korpus wurde deutlich, dass viele Stichwörter bzw. Lesarten auffallend häufig in bestimmten thematischen Bereichen vorkommen. Einen ersten Hinweis darauf geben u.a. die semantischen Mitspieler,¹¹ die durch die Kookkurrenzanalyse ermittelt werden. Dabei geht es nicht um eine mögliche fachsprachliche Verwendung des Stichwortes – diese wird gesondert in der Sachgebietsangabe erläutert –, es handelt sich vielmehr um Thematisierungen, die erkennen lassen, in welchen im weitesten Sinn diskursiven Verwendungen ein Wort im öffentlichen Sprachgebrauch vorkommt: Der Gebrauch eines Wortes ist nämlich häufig mit Themen verknüpft, die öffentlich und kontrovers diskutiert werden.

In der Angabe „Themengebundene Verwendung(en)“ haben die Lexikografen nun die Möglichkeit, einen solchen auffälligen und statistisch signifikanten Gebrauch eines Stichwortes / einer Lesart in einem bestimmten Themenkomplex, wie er sich im Korpus darstellt, zu dokumentieren und auf derartige diskursähnliche Verwendungen hinzuweisen. Der Text hierfür ist frei formulierbar, im Umfang nicht festgelegt (aber für den Benutzer prägnant und überschaubar) und durch Überschriften gegliedert. Dabei liegt der Fokus auf der auffälligen, nicht erwartbaren Verwendung und weniger auf dem nach allgemeinem (Sprach- und Sach-)Wissen vorhersehbaren Vorkommen eines Stichwortes in einem bestimmten Themenbereich. So wird beispielsweise die signifikante Verwendung von *Bundeskanzler* im Kontext „Politik“ nicht beschrieben. Hingegen wird das Stichwort *Schriftsteller*, Lesart ‘Autor’ außer in der zu erwartenden Verwendung im Kontext von „Kultur“ im *ellexiko*-Korpus im Kontext von „Politik“ verwendet; dies wird als Gebrauchsbesonderheit festgehalten:

Themengebundene Verwendung(en):

Im Kontext von Literatur

Schriftsteller wird im *ellexiko*-Korpus im allgemeinen Kontext von Kunst und Kultur verwendet, insbesondere auch im Zusammenhang mit dem Literaturbetrieb. Hierbei wird häufig die aktuelle in- und ausländische Literaturszene beleuchtet, das Erscheinen von neuen oder neu aufgelegten Büchern kommentiert, und zwar in Form von kritischen bzw. lobenden Worten für Schriftsteller und ihr Werk (vgl. den Beleg).

Der Chef des Suhrkamp-Verlags, Siegfried Unseld, setzt in die jüngste Generation deutschsprachiger **Schriftsteller** große Hoffnungen. Als vielversprechende Autoren nannte er u.a. den Tiroler Norbert Gstrein, den ostdeutschen Lyriker Durs Grünbein und die in Baden bei Wien geborene Marlene Streeruwitz. (Tiroler Tageszeitung, 11.06.1997)

Im Kontext von Politik

Schriftsteller werden allgemein als (kritische) Personen des öffentlichen Lebens im *ellexiko*-Korpus im Bereich von Gesellschaft und Politik thematisiert. Als ‘unbequem’ denkende, frei ihre Meinung äuernde Menschen wurden und werden Schriftsteller in diktatorischen, totalitären usw. Systemen oder Regimen massiv bedroht und müssen bzw. mussten ihre öffentliche Kritik an Missständen und Missachtung von Menschenrechten im extremsten Fall auch mit dem Leben bezahlen (vgl. den Beleg).

An das Schicksal inhaftierter, getöteter, verschwundener und bedrohter Literaten erinnert heute der „Tag des inhaftierten **Schriftstellers**“ des internationalen PEN. Die Vereinigung der Dichter,

¹¹ Deren Ermittlung erfolgt – wie alle anderen Korpusanfragen – mithilfe von COSMAS II (vgl. <http://www.ids-mannheim.de/cosmas2/>) und des integrierten Tools „Statistische Kollokationsanalyse und Clustering“ (vgl. <http://www.ids-mannheim.de/kt/projekte/methoden/>).

Essayisten und Romanautoren will in diesem Jahr besonders auf sieben **Schriftsteller** aufmerksam machen. Der zum Tod verurteilte Regime-Kritiker Mansur Muhammad Ahmad Rajih ist seit 1983 in Jemen im Gefängnis. Sein Gesundheitszustand ist kritisch, ärztliche Behandlung wird ihm verweigert. Gegen Taslima Nasrin aus Bangladesh haben religiöse Extremisten einen Mordaufruf ausgesprochen. Nach Kritik an Extremismus und männlichem Chauvinismus wird sie der Blasphemie beschuldigt. Der türkische Dichter und Journalist Metin al-Tiok wurde bei einem Brandanschlag von Extremisten während eines Kulturfestivals in Sivas getötet. Der kubanische **Schriftsteller** und Führer einer sozialdemokratischen Organisation, Indamiro Restano, wurde zu zehn Jahren Gefängnis verurteilt und kürzlich in Einzelhaft isoliert. Der Südkoreaner Hwang Suk-young wurde wegen nicht genehmigter Besuche in Nordkorea zu acht Jahren verurteilt. (Frankfurter Allgemeine Zeitung, 00.00.1993)

In den themengebundenen Verwendungen spiegelt sich die starke gesellschaftsrelevante Prägung des Wortgebrauchs, wie sie sich in dem Zeitungskorpus, das *ellexiko* zugrunde liegt, darstellt. In *ellexiko* werden erstmals in einem allgemeinsprachlichen Wörterbuch diskursartige Verwendungen in dieser Form thematisiert und dokumentiert (vgl. hierzu auch [Globalisierung](#), Lesart 'Institutionalisierung').

Verwendung in mehrteiligen Eigennamen:

Im Rahmen der Einwortlexeme wurde im Zuge der konkreten Artikularbeit, genauer der Kookkurrenzanalyse, ein Phänomen augenfällig: Bestimmte Stichwörter/Lesarten finden in signifikanter Weise Eingang in mehrteilige Namen bzw. namenähnliche Fügungen. Insbesondere trifft das auf so genannte „Bereichs“-Lesarten zu, ja es ist geradezu ein Indikator für die Ansetzung einer solchen Lesart, vgl. z.B. den Wortartikel [Wirtschaft](#), Lesart 'Bereich':

Wirtschaft ist in dieser Lesart häufig Bestandteil von mehrteiligen Namen für Ministerien und andere Verwaltungseinrichtungen sowie von Forschungseinrichtungen, wie z.B.:

- Abteilung Wirtschaft und Finanzen,
- Bundesamt für Wirtschaft und Arbeit,
- Fachhochschule für Technik und Wirtschaft,
- Institut für die Deutsche Wirtschaft,
- Minister für Wirtschaft, Verkehr und Landesentwicklung,
- Ministerium für Wirtschaft und Finanzen,
- Ressort Wirtschaft.

Das Phänomen lässt sich nicht dem Bereich der usuellen Wortverbindungen zurechnen, es gehört vielmehr zur Besonderheit des Gebrauchs von bestimmten Lesarten. Allerdings war ein eigener Ort zur Beschreibung dieser Gebrauchsbesonderheit von Einwortlemmata in *ellexiko* nicht vorgesehen; also vermerkten Lexikografen bislang an zwei Orten, ob eine Lesart in mehrteiligen Namen verwendet wird, nämlich unter der Rubrik „Typische Verwendungen“ bzw. unter der Rubrik „Sachgebietsangabe“. Dies zeugt letztlich davon, dass bei den Lexikografen unabhängig voneinander das Bedürfnis erwuchs, die Information zu vermerken, dass diese aber in keine der bestehenden Rubriken wirklich passte; das Verfahren, die Angabe an zwei Orten zu platzieren, ist natürlich nicht nutzerfreundlich, müsste man doch die eine inhaltlich gleiche Angabe bzw. Information stets an zwei Orten in einem Wortartikel suchen.

Aus der Überlegung heraus, dass es sich um eine Besonderheit des Gebrauchs handelt, wenn bestimmte Lesarten häufig Eingang in mehrteilige Namen finden und dass dies mitunter einen lesartenkonstituierenden Faktor darstellt, der unbedingt an einem gesonderten Ort beschrieben werden sollte, wurde die Rubrik „Verwendung in mehrteiligen Eigennamen“ geschaffen.

4.3 Valenzangaben (Annette Klosa)

Teil der grammatischen Angaben zu Verben, Nomen und Adjektiven in *lexiko* sind Informationen zur Valenz des jeweiligen Stichwortes.¹² Dabei soll *lexiko* kein Valenzwörterbuch ersetzen oder gar ein neues schreiben, aber doch solche Satzbaupläne oder Anschlüsse zeigen, die im *lexiko*-Korpus sehr präsent sind. In den Online-Artikelansichten wird dies bezogen auf die jeweilige Lesart des Stichwortes in der Rubrik „Grammatik“ angegeben (z.B. im Wortartikel [Arbeit](#), Lesart ‘Tätigkeit’ bzw. im Wortartikel [behindern](#), Lesart ‘beeinträchtigen’).

Die Angaben zur Valenz wurden deshalb neu konzipiert und überarbeitet, weil die Unterscheidung nach fakultativen und obligatorischen Komplementen noch nicht richtig erfasst werden konnte, was sich im Laufe der Bearbeitung der ersten etwa 100 Verben gezeigt hatte. Auch die Online-Präsentation der Satzbaupläne genügte den Ansprüchen nicht mehr, weil die Markierung fakultativer Komplemente durch runde Klammern, die aus anderen Wörterbüchern und aus Grammatiken vertraut ist, auch in *lexiko* möglich sein sollte, z.B.:

JEMAND / ETWAS ermöglicht (JEMANDEM) ETWAS

JEMAND / ETWAS reagiert (AUF JEMANDEN / ETWAS)

Online werden auch weiterhin nur die Satzbaupläne (bei den Verben) bzw. die Anschlüsse (bei den Nomen) angezeigt. Zu prüfen ist, ob die im Wortartikel vorhandenen, detaillierten Informationen zu den Komplementen zukünftig als Rechercheangebot in *lexiko* angeboten werden können.

5. Technische Aspekte der Umarbeitungen (Carolin Müller-Spitzer)

Auch technisch stellten die in den vorangegangenen Abschnitten beschriebenen inhaltlichen Umarbeitungen eine besondere Herausforderung dar, da sie parallel zur Artikelproduktion laufen mussten und diese auch von technischer Seite so wenig wie möglich beeinträchtigen sollten.

Die lexikografischen Daten von *lexiko* liegen in einer granularen XML-Struktur vor, die jede Angabe einzeln maschinell zugreifbar macht und die so streng aufgebaut ist, dass ein Instruktionenbuch von formaler Seite für die Lexikografen nicht erstellt werden musste. D.h., in der XML-Struktur ist sehr genau abgebildet, in welcher Reihenfolge welche Angaben in die Artikelstruktur eingetragen werden sollen und wie diese intern strukturiert werden. (Festlegungen hinsichtlich der Beschreibungssprache einzelner Angaben und andere inhaltliche Richtlinien können naturgemäß nicht in der XML-Struktur abgebildet werden.) Die Vorteile einer solchen granularen Modellierung sind vor allem die flexible Darstellbarkeit der Daten und die gezielten Zugriffsmöglichkeiten auf die

¹² Vgl. hierzu im Einzelnen Klosa (2005a, S. 286, S. 289 und S. 292f.)

lexikografischen Daten (vgl. Storrer 2001 und Müller-Spitzer 2005). Die XML-strukturierten lexikografischen Daten von *ellexiko* werden dabei durch XSLT-Stylesheets für die Darstellung im Internet nach HTML transformiert (vgl. Abbildung 11).

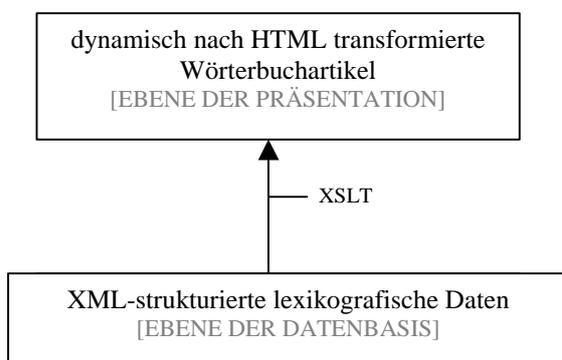


Abbildung 11: Generierung der Online-Artikelansichten in *ellexiko*

Jede Änderung in der inhaltlichen Gliederung einzelner Angabebereiche oder in der Binnenstrukturierung einzelner Angabegruppen erfordert daher viele kleine technische wie inhaltliche Schritte, die eng miteinander verzahnt sind und in deren Umsetzung die einzelnen Projektbeteiligten sich genau absprechen müssen. Im Einzelnen können diese Schritte folgendermaßen skizziert werden:

1. Die inhaltlichen Desiderata für die Umstrukturierung einzelner Angabebereiche werden konzipiert.
2. Anhand der bestehenden Modellierung muss daraufhin geprüft werden, welche Auswirkungen diese inhaltlichen Änderungswünsche auf die DTDs haben. Welche Elemente müssen geändert, welche gelöscht bzw. welche neu hinzugefügt werden? Insbesondere muss geklärt werden, welche Änderungen die bestehenden Artikel nicht valide, d.h. nicht konform zur bestehenden Modellierung werden lassen.
3. Es muss ein Konzept dafür erstellt werden, wie die Umarbeitung im Bereich der XML-Struktur ablaufen soll. Dabei sind Fragen zu klären wie: Welche Änderungen können durch Konvertierungen automatisch in den XML-Instanzen umgesetzt werden, welche müssen manuell vorgenommen werden? Müssen temporär zwei „Strukturäste“ für die alte und neue Struktur angelegt werden, damit die nach dem alten Muster strukturierten Daten nach und nach in die neue Struktur überführt werden können? (Beispielsweise musste für die Umarbeitung der Typischen Verwendungen die alte Struktur bestehen bleiben, damit die Typischen Verwendungen in der alten, unsortierten Form nach wie vor online gezeigt werden konnten und die neue Sortierung der Typischen Verwendungen gleichzeitig begonnen werden konnte.)
4. Die DTDs müssen geändert werden. Im Redaktionshandbuch müssen gegebenenfalls neue Schreibrichtlinien o.Ä. festgehalten werden.
5. Die inhaltlichen Umarbeitungen können beginnen.
6. Parallel zu den oben genannten Schritten müssen neue Vorgaben für die Darstellung der geänderten Daten erstellt und die XSLT-Stylesheets angepasst werden. Auch hier kann es notwendig sein, temporär für einen Teil der Daten die alte Darstellung bestehen zu lassen und für schon umgearbeitete Artikel die neue Darstellung zu aktivieren.

7. Nachdem die inhaltlichen Umarbeitungen abgeschlossen sind, müssen evtl. noch vorhandene alte Strukturteile aus den DTDs und Stylesheets gelöscht werden.

Wie an dieser Ablaufskizze zu erkennen sein wird, erfordern inhaltliche Umarbeitungen in den *lexiko*-Artikeln von allen beteiligten Mitarbeitern ein hohes Maß an Genauigkeit und Sorgfalt in der Arbeit und eine intensive Kommunikation untereinander. Da diese Änderungen außerdem in ihrem Aufwand nicht unterschätzt werden dürfen, werden sie in *lexiko* auch nur in den Fällen vorgenommen, in denen die neue Strukturierung einen echten Gewinn für die Benutzer darstellt. Außerdem werden sie immer nur in Blöcken durchgeführt. Auch wenn die granulare Strukturierung der lexikografischen Daten solche Umarbeitungen in gewissem Sinne erschwert, überwiegen jedoch insgesamt die Vorteile, die eine solche Modellierung bietet (vgl. Schlaps 2007).

6. Grundüberlegungen zu einem im Aufbau befindlichen Online-Nachschlagewerk (Annette Klosa)

lexiko ist, wie die vorangegangenen Abschnitte deutlich gezeigt haben, ein im Aufbau befindliches Online-Wörterbuch (vgl. Storrer/Freese 1996, S. 117). Der lexikografische Arbeitsprozess ist (anders als bei abgeschlossenen Online-Wörterbüchern wie etwa der retrodigitalisierten ersten Auflage des [Deutschen Wörterbuchs](#) von Jacob und Wilhelm Grimm) noch nicht abgeschlossen.

Neben den oben ausgeführten umfangreichen Neukonzeptionen und kleineren Umarbeitungen umfasst die Arbeit an *lexiko* als Daueraufgabe die Pflege der Stichwortliste,¹³ den Ausbau der Wortartikel um weitere Angaben und die Pflege ausgearbeiteter Wortartikel durch Konsistenzprüfungen. Im Einzelnen hieß dies etwa, die Stichwortliste an die revidierte neue Rechtschreibung anzupassen.¹⁴ Daneben wurden in den letzten Jahren zahlreiche Artikel beispielsweise um Angaben zu morphologischen Varianten angereichert. Verschiedene Korrekturgänge in den ausgearbeiteten Artikeln erhöhten die Konsistenz innerhalb der Wortartikel, z.B. im Bereich der bibliografischen Angaben. Verstärkt muss außerdem zukünftig die Konsistenz innerhalb der Verlinkungen geprüft werden.

Laufende Korrekturen und Aktualisierungen wie die genannten sind in der Erarbeitungsphase eines jeden Wörterbuches normal, denn:

Erst bei der praktischen Arbeit [...] stellen die Wörterbuchmacher fest, daß ihre Richtlinien Lücken haben, daß immer wieder Sonderfälle auftreten, an die man bei den theoretischen Überlegungen nicht gedacht hat, so daß die Richtlinien immer wieder ergänzt werden müssen. (Müller 1984, S. 363)

Diese Erfahrung lässt sich mit den üblicherweise angenommenen Phasen, in denen ein Wörterbuch entsteht, schwer vereinbaren. Für gedruckte Wörterbücher etwa benennt Landau (1984, S. 226-271) drei Hauptphasen, die hintereinander ablaufen sollten: „(I) planning, (II) writing, (III) producing“. Zugleich vermerkt er, dass Vorarbeiten für eine revidierte Fassung durchaus begonnen werden sollten, noch ehe das betreffende Wörterbuch erschienen ist (vgl. Landau 1984, S. 267). Ausführlicher legen etwa Riedel/Wille (1979) dar, welche einzelnen Arbeitsschritte für gedruckte Wörterbücher vorzusehen sind:

¹³ Zur grundlegenden Konzeption der *lexiko*-Stichwortliste vgl. Schnörch (2005).

¹⁴ Zur Orthografie der Stichwörter und zu den orthografischen Angaben in *lexiko* vgl. Klosa (2005b).

(I) Konzeption: inhaltlich-wissenschaftliche Konzeption – Umfangskonzeption – Textkonzeption – Bildkonzeption – organisatorisch-technische Konzeption

(II) Erarbeitung: Zusammenstellung der Stichwortliste – Zusammenstellung der Materialbasis – Ausarbeitung der Wörterbuchartikel – Auswahl von Illustrationen – Erstellung der lexikografischen Hilfsmittel – Verfassung der Benutzungsanleitung – laufende Korrekturen und Aktualisierungen – Endredaktion

(III) Druckvorbereitung: Festlegung der äußeren Gestaltung – Bestimmung der typografischen Gestaltung

Besonders interessant ist hier in der Phase der Erarbeitung der Punkt „laufende Korrekturen und Aktualisierungen“. Diese werden vermutlich hauptsächlich vorgenommen, um einem für Wörterbücher sehr wichtigen Aspekt zu genügen, und zwar dem „Aspekt der angestrebten Einheitlichkeit und Geschlossenheit, des Zusammenhangs der Teile als Teile eines Ganzen“ (Schaeder 1987, S. 130).

Ein striktes Nacheinander der drei Hauptphasen der Wörterbucharbeitung ist für gedruckte Wörterbücher nachzuvollziehen, wobei sich allerdings in der Praxis das Problem ergibt, dass die Konzeptions- und die Erarbeitungsphase zumindest eine Zeit lang Hand in Hand gehen dürften. Die Produktionsphase aber setzt tatsächlich erst ein, wenn die Erarbeitung abgeschlossen ist. Erst, wenn das gedruckte Wörterbuch erschienen ist, können mögliche Inkonsistenzen und (Druck-)Fehler auffallen und gesammelt werden. Revisionen in der Konzeption und Korrekturen müssen bis zu einer möglichen Neuauflage warten. Dann können auch Kritikpunkte der Rezensenten, Hinweise von Benutzern sowie eigene Vorstellungen der Lexikografen, die zunächst nicht realisiert werden konnten, berücksichtigt werden.

Für Online-Wörterbücher stimmt dies so nicht: Anders als gedruckte Wörterbücher oder auf CD-ROM publizierte Wörterbücher muss ihre Wörterbuchsubstanz nicht zu einem Tag X fertig sein, um in die Phase der Produktion überzugehen. Online-Wörterbücher können fortlaufend publiziert werden, wenn die redaktionelle Entscheidung für eine solche fortlaufende Publikation getroffen wurde.

Grundsätzlich kann man natürlich auch bei einem Online-Wörterbuch mit der Freischaltung bis zu dem Zeitpunkt warten, an dem das Wörterbuch komplett fertiggestellt ist. Bei Online-Wörterbüchern, die aus gedruckten Wörterbüchern retrodigitalisiert werden, ist dies vermutlich immer der Fall.¹⁵ Bei neu erarbeiteten Online-Wörterbüchern wie *ellexiko* oder Online-Wörterbüchern, die gedruckte Fassungen fortsetzen bzw. ergänzen (z.B. das [Deutsche Rechtswörterbuch](#)), wird häufiger ein anderes Vorgehen gewählt. Hier wird eher der Weg des sukzessiven Publizierens gewählt, denn „das Publikationsmedium Internet bietet die Chance, von der Statik des Wörterbuchs auf eine Dynamik zu wechseln“ (Lemberg 2001, S. 82). Die Phasen der Erarbeitung und der Produktion stehen damit nicht mehr in einem zeitlichen Nacheinander, sondern in einem zeitlichen Nebeneinander.

Lemberg (2001, S. 82) führt weiter aus, dass Dynamik „in diesem Fall fortgesetzte Korrektur- und Ergänzungsmöglichkeiten des Wörterbuchs auch in seinen bereits geschriebenen Teilen“ meint, und zwar „in allen einzelnen Informationspositionen des Wörterbuchartikels wie im mediostrukturellen und makrostrukturellen Bereich durch die

¹⁵ So wurde z.B. das [Neologismenwörterbuch](#) nach seinem Erscheinen in Buchform (Herberg/Kinne/Steffens 2004) in eine elektronische Fassung gebracht, die vollständig 2006 online freigeschaltet wurde.

permanent weitergeführte lexikalische Vernetzung oder das Einfügen neuer Wörterbuchartikel aufgrund neuer Belege“. Sind aber Mikro-, Medio- und Makrostruktur eines Online-Wörterbuchs von dynamischen Veränderungen betroffen, so bedeutet dies, dass auch die erste Phase des lexikografischen Arbeitsprozesses, nämlich die Phase des Konzipierens, nicht unberührt bleibt.

Bei solchen Veränderungen kann man im Übrigen unterscheiden nach völliger Neukonzeption (z.B. eines Angabetyps, wie im oben erläuterten Fall der Typischen Verwendungsmuster in *ellexiko*), nach Systematisierungen (z.B. indem bestimmte ursprüngliche redaktionelle Richtlinien ausgedehnt und stringenter angewendet werden, wie oben für die Formulierung der Bedeutungserläuterungen in *ellexiko* gezeigt wurde) oder nach Umgruppierungen (in *ellexiko* z.B., wie oben erläutert, indem paradigmatische Partner nach neuen Ordnungskriterien sortiert werden). Bei allen diesen Änderungen wird die ursprüngliche Konzeption z.B. aufgrund der inzwischen größeren lexikografischen Erfahrung überdacht und revidiert – man kehrt also zeitweise in die Phase des Konzipierens zurück.

Insofern gehen nicht nur die Phasen von Erarbeitung und Produktion Hand in Hand, sondern auch die Phasen von Konzeption und Erarbeitung können (zumindest zeitweise) für ein dynamisches Online-Wörterbuch parallel laufen. Auch deshalb unterscheiden Storrer und Freese zwischen im Aufbau befindlichen Online-Wörterbüchern (wie *ellexiko*) und abgeschlossenen Online-Wörterbüchern, bei denen „der lexikographische Arbeitsprozeß bereits abgeschlossen ist und das lexikographische Endprodukt nicht weiter zur Diskussion steht“ (Storrer/Freese 1996, S. 117).

Ein striktes zeitliches Nacheinander von Konzeptions-, Erarbeitungs- und Produktionsphase wie bei einem gedruckten Wörterbuch gibt es für *ellexiko* also nicht. Und *ellexiko* muss nicht an einem Tag X fertig sein, um in die Phase der Produktion überzugehen: *ellexiko* wird fortlaufend und tatsächlich täglich publiziert. Diese Entscheidung hat Vorteile wie Nachteile, und zwar für die Bearbeiter wie für die Nutzer.

Für die Lexikografen in *ellexiko* ist es zwar befriedigend, Erfahrungen aus der praktischen Wörterbucharbeit in Ergänzungen und Verbesserungen einfließen lassen zu können, aber dies kostet natürlich Zeit, sodass sich die Neuproduktion von Artikeln verlangsamen kann. Für die Nutzer ist gerade das möglicherweise ein großer Nachteil. Auf der anderen Seite erreichen Verbesserungen und Ergänzungen in *ellexiko* den Nutzer sehr schnell, da er nicht bis zu einer Neuauflage warten muss.

In Hinblick auf die Nutzer ist daneben grundsätzlich die Frage aufzuwerfen, ob man ihnen fortwährende Revisionen in einem im Aufbau befindlichen Online-Wörterbuch überhaupt zumuten darf bzw. soll. So kann man argumentieren, dass dieses Vorgehen dem (schnellen, sich ständig verändernden) Medium Internet zwar gerecht wird, die Dynamik aber zulasten der Verlässlichkeit geht:

Das Internet mit seiner dezentralen Organisationsform und seinem schnellen Wachstum ist bislang kein Ort der Verbindlichkeit und Verlässlichkeit. Wer in Eile etwas nachschlagen möchte, kann nicht wissen, ob das vertraute WWW-Wörterbuch nicht plötzlich mit neuen Zugriffsangeboten aufwartet, deren Funktionsweise man sich erst aneignen muß [...]. (Storrer/Freese 1996, S. 129)

Ein Vorteil von im Aufbau befindlichen Online-Wörterbüchern ist die Tatsache, dass sie aktuell gehalten werden können: In *ellexiko* können beispielsweise aktuelle Stichwörter

ter ergänzt werden, Belege durch aktuellere ausgetauscht werden, neue Typische Verwendungsmuster aufgenommen werden usw. Aber auch dieses Vorgehen hat Nachteile:

Was bei CD-Rom-Produkten [sic!] noch möglich ist, die Unterscheidung verschiedener Auflagen oder datierter Ausgaben, droht bei häufig aktualisierten online-Informationssystemen [sic!] verloren zu gehen. Welche Veränderungen wann vorgenommen wurden, welche Elemente herausgenommen und welche eingefügt wurden – alles dies ist durch Nutzer nicht mehr nachvollziehbar. (Haß-Zumkehr 2001, S. 377f.)

Haß-Zumkehr fordert deshalb, dass nicht nur jede Bildschirmseite mit Datum und Verfasserangabe gekennzeichnet werden müsste, sondern auch, dass „die jeweilige Version eines Nachschlagewerks zu jedem beliebigem [sic!] Zeitpunkt zu speichern und auch für Nutzer zugänglich zu archivieren“ wäre, wenn man zu den „seriösen Informationsanbietern“ gehören wolle (ebd.).

Doch bleiben diese Überlegungen zunächst noch theoretisch. Gerade in diesem Bereich wäre es wünschenswert, durch Benutzerbefragung und -beobachtung die tatsächlichen Erwartungen an ein im Aufbau befindliches Online-Wörterbuch festzustellen. So könnte man u.a. auch abfragen, ob Neubearbeitungen/Umarbeitungen besser sukzessive oder en bloc zur Verfügung gestellt werden sollten oder ob zukünftig alle Veränderungen markiert, jede Artikelversion datiert und ältere Versionen in einem Archiv angeboten werden müssen. Auch mit solchen Fragen und ihrer Beantwortung wird sich das Projekt *elexiko* zukünftig beschäftigen.

7. Literatur

- Belica, Cyril (2001-2007): Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. Internet: <http://corpora.ids-mannheim.de/ccdb/> (Stand: Februar 2008).
- Croft, William/Cruise, Alan (2004): *Cognitive linguistics*. Cambridge.
- Haß, Ulrike (2005a): Das Bedeutungsspektrum. In: Haß (Hg.), S. 163-181.
- Haß, Ulrike (2005b): *elexiko* – Das Projekt. In: Haß (Hg.), S. 1-17.
- Haß, Ulrike (2005c): Besonderheiten des Gebrauchs. In: Haß (Hg.), S. 265-276.
- 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.
- Haß-Zumkehr, Ulrike (2001): *Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte*. Berlin/New York.
- Hausmann, Franz Josef (1989): Die Markierung im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilbd. Berlin/New York, S. 649-657.
- Jones, Steven (2002): *Antonymy: A corpus-based perspective*. London.
- Keibel, Holger/Belica, Cyril (2007): CCDB: A corpus-linguistic research and development workbench. In: *Corpus Linguistics Conference proceedings*. Birmingham. Internet: http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf (Stand: Oktober 2007).
- Klosa, Annette (2005a): Grammatik. In: Haß (Hg.), S. 277-298.
- Klosa, Annette (2005b): Orthografie und morphologische Varianten. In: Haß (Hg.), S. 133-140.

- Landau, Sidney I. (1984): *Dictionaries: The art and craft of lexicography*. New York.
- Lemberg, Ingrid (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: Lemberg, Ingrid et al. (Hg.): *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher.* (= *Lexicographica, Series maior 107*). Tübingen. S. 71-91.
- Marello, Carla et al. (Hg.) (2006): *Proceedings of the 12th EURALEX International Congress (Atti del XIIth Congresso Internazionale di Lessicografia), EURALEX 2006, Turin, Italy, September 6th-9th, 2006. Vol. 2.* Turin.
- Müller, Wolfgang (1984): Zur Praxis der Bedeutungserklärung in (einsprachigen) deutschen Wörterbüchern und die semantische Umkehrprobe. In: Wiegand, Herbert Ernst (Hg.): *Studien zur neuhochdeutschen Lexikographie V.* (= *Germanistische Linguistik 3-6/84*). Hildesheim/New York. S. 359-461.
- Müller-Spitzer, Carolin (2005): Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess. In: Haß (Hg.), S. 20-54.
- Murphy, Lynne (2006): Antonyms as lexical constructions: or, why paradigmatic construction is not an oxymoron. In: *Constructions, Special Volume 1, 8.* Internet: <http://www.constructions-online.de/articles/specvol1/685> (Stand: Februar 2008).
- Paradis, Carita/Willners, Caroline (2006): What a corpus-based dictionary tells us about antonymy. In: Marello et al. (Hg.), S. 213-220.
- Riedel, Hans/Wille, Margit (1979): Über die Erarbeitung von Lexika. Grundsätze und Kriterien. (= Beiheft 91 zum Zentralblatt für Bibliothekswesen). Leipzig.
- Schaeder, Burkhard (1987): *Germanistische Lexikographie.* (= *Lexicographica, Series Maior 24*). Tübingen.
- 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.
- Schnörch, Ulrich (2005): Die *elexiko*-Stichwortliste. In: Haß (Hg.), S. 71-90.
- Schnörch, Ulrich (2007): Nicht 'gehoben', 'verhüllend', 'fachsprachlich': Ein neues Konzept zur lexikografischen Beschreibung von Kultur- und Diskursabhängigkeit des Wortgebrauchs in *elexiko*. In: Valentin, Jean-Marie unt. Mitarb. von Hélène Vinckel (Hg.) (2008): *Akten des XI. Internationalen Germanistenkongresses Paris 2005 „Germanistik im Konflikt der Kulturen“.* Bd. 4. (= *Jahrbuch für Internationale Germanistik, Reihe A, Kongressberichte 80*). Bern/Berlin u.a. S. 337-344.
- Storjohann, Petra (2005a): Semantische Paraphrasen und Kurzetikettierungen. In: Haß (Hg.), S. 182-203.
- Storjohann, Petra (2005b): Paradigmatische Relationen. In: Haß (Hg.), S. 249-264.
- Storjohann, Petra (2005c): Typische Verwendungen. In: Haß (Hg.), S. 235-248.
- Storjohann, Petra (2005d): Corpus-driven vs. corpus-based approach to the study of relational patterns. In: *Corpus Linguistics Conference proceedings. Birmingham.* Internet: <http://www.corpus.bham.ac.uk/PCLC/paper%20045PetraStorjohann.doc> (Stand: Februar 2008).
- Storjohann, Petra (2006): New Lexicographic Approaches to the Description of Sense Relations. In: Marello et al. (Hg.), S. 1201-1212.
- Storjohann, Petra (2007): Incompatibility: A no-sense relation? In: *Corpus Linguistics Conference proceedings. Birmingham.* Internet: http://corpus.bham.ac.uk/corplingproceedings07/paper/36_Paper.pdf (Stand: Februar 2008).
- Storjohann, Petra/Möhrs, Christine (2007). Das *elexiko*-Wörterbuch im Wandel – Typische Verwendungen neu beleuchtet. In: *Sprachreport 3/2007*, S. 12-16.

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

Storrer, Angelika/Freese, Kathrin (1996): Wörterbücher im Internet. In: Deutsche Sprache 24 (1996), S. 97-136.

Tognini-Bonelli, Elena (2001): Corpus linguistics at work. Amsterdam/Philadelphia.

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

8. Nachschlagewerke

Duden (1999): Duden – Das große Wörterbuch der deutschen Sprache. 3., völl. neu bearb. und erw. Aufl. Hrsg. vom Wissenschaftlichen Rat und der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich.

Duden 8 (2004): Duden 8– Das Synonymwörterbuch. 3., völl. neu erarb. Aufl. Mannheim/Leipzig/Wien/Zürich.

Götz, Dieter (Hg.) (1999): Langenscheidts Großwörterbuch Deutsch als Fremdsprache. Berlin u.a.

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

Kempcke, Günter (Hg.) (2000): Wörterbuch Deutsch als Fremdsprache. Berlin u.a.

Klappenbach, Ruth/Steinitz, Wolfgang (Hg.) (1964ff.): Wörterbuch der deutschen Gegenwartssprache. Berlin.

Sinclair, John (Hg.) (2003): Collins COBUILD Advanced Learner's English Dictionary. 4. Ausg. Glasgow.

Heidrun Kämper / Andreas Rothenhöfer

Das Online-Wörterbuch zum Schulddiskurs 1945-1955 in OWID

Abstract

Der folgende Beitrag beschreibt das OWID-Modul zum Schulddiskurs 1945-1955, das aus einem von der DFG geförderten Projekt hervorgegangen ist und dessen Ergebnisse in Kämper (2005) ausführlich dargestellt sind. Das OWID-Modul „Diskurswörterbuch“ ist die Online-Version des Wörterbuchs zum Schulddiskurs (vgl. Kämper 2006).

Erläutert wird der dem Wörterbuchtyp zugrunde gelegte Diskursbegriff und es wird dargestellt, inwiefern lexikalische Serialität Gegenstand des Wörterbuchs ist. Anschließend wird die besondere Struktur eines Diskurswortschatzes problematisiert, aus der als Aufgabe eines Diskurswörterbuchs die Darstellung begrifflich-semanticischer Bezugsrelationen des Wortbestands resultiert. Die Darstellung der äußeren Zugriffsstrukturen, der Artikelstrukturen, der Datendistribution und Verweise sowie Hinweise zur Benutzung werden dann exemplifiziert.

The following contribution describes the OWID-module ‘Schulddiskurs 1945-1955’. This module contains the results of a research project which has been supported by the DFG. The print version describes these findings in detail (cf. Kämper 2005). The OWID-module “Diskurswörterbuch” is the online version of the ‘Schulddiskurs’ (discourse of guilt) dictionary (cf. Kämper 2006).

The contribution explains the concept of discourse on which the dictionary is based, and demonstrates the role of lexical seriality within the dictionary. Next, the special structure of a discourse vocabulary is discussed. This structure determines the representation of the conceptual or semantic relations of the vocabulary. Finally, the explanation of different possibilities of access, the structure of the articles, the distribution of data and references are exemplified along with information on the use of the dictionary.

Inhalt:

1. Wörterbuchtyp
2. Wörterbuchgegenstand
3. Äußere Zugriffsstrukturen des Diskurswörterbuchs
4. Artikelstrukturen
5. Datendistribution und Verweise
6. Fazit
7. Literatur

Der folgende Beitrag beschreibt das OWID-Modul zum Schulddiskurs 1945-1955, das aus einem von der DFG geförderten Projekt hervorgegangen ist und dessen Ergebnisse in Kämper (2005) ausführlich dargestellt sind. Das OWID-Modul „Schulddiskurs im ersten Nachkriegsjahrzehnt“ ist die Online-Version des Wörterbuchs zum Schulddiskurs (vgl. Kämper 2006).

1. Wörterbuchtyp

Der Beschreibungsgegenstand eines diskursorientierten Wörterbuchs ist abhängig von dem Diskursbegriff, der seinem Konzept zugrunde liegt. Diskurs wird hier verstanden als eine Serie von Texten unterschiedlicher Sprecher bzw. Schreiber (Diskursbeteiligte) zu einem bestimmten Gegenstand innerhalb eines bestimmten Zeitraums. Da es sich um

Serien kollektiver kommunikativer Akte handelt, ist ein Diskurs, als ein gesellschaftliches und historisches Phänomen, komplex und heterogen hinsichtlich der Beteiligten und ihrer den Diskurs konstituierenden sprachlichen Äußerungen. Aus einem solchen Diskurszusammenhang rekonstruierte Semantik erhält durch die diskursive Darstellung insofern eine neue Dimension, als die lexikalische Serie es erlaubt, begriffliche bzw. semantische Einheiten nicht nur im thematischen, sondern auch im kommunikativen Kontext, d.h. in Bezug auf Funktionen und auf Absichten von Sprechergruppen, zu rekonstruieren. Der Wortschatz, der einen Diskurs konstituiert, besteht aus, den komplexen semantischen Gehalt dieses Diskurses komprimierenden, lexikalischen Einheiten. Ihre lexikografische Darstellung ist Gegenstand eines Diskurswörterbuchs: Ein Diskurswörterbuch stellt dar und beschreibt das relevante, einen Diskurs konstituierende und realisierende Vokabular nach spezifischen lexikografischen Prinzipien.

Die lexikalischen Elemente eines Diskurses stehen zueinander in Bedeutungsbeziehungen. Insofern sind Diskurse semantische Netze, die diese Bedeutungsbeziehungen ihres lexikalischen Inventars konstituieren. Das Netz des Diskursvokabulars gibt dem Diskurs eine begrifflich-semantische Ordnungsstruktur, die in Relation steht zum Diskursthema, zu den Diskursbeteiligten, zu den Texten, zu dem Zeitraum, über den sich ein Diskurs erstreckt, zu der Funktion / den Funktionen des Diskurses. Die Struktur dieses Diskurswortschatzes besteht aus der Summe des begrifflich-semantischen Potenzials der lexikalischen Diskurselemente und der Summe der regelmäßigen Beziehungen, die die lexikalischen Diskurselemente untereinander eingehen. Damit wird die onomasiologische Beschreibungsperspektive in das semasiologische Darstellungsprinzip dieses Wörterbuchs integriert. Ein Diskurswörterbuch, das die lexikalisch-semantischen Diskursrepräsentanten und ihre Relationen zueinander lexikografisch beschreibt, macht also die thematische, begriffliche und semantische Ordnung des Diskurses (die seine Kohärenz erzeugt) auf der Wortebene sowohl semasiologisch als auch onomasiologisch sichtbar.

Dargestellt wird dieser Wortschatz als begrifflich-semantisches Netz, dessen Knoten die lexikalischen Einheiten eines Diskurses bilden, auf die verwiesen wird. Damit bildet das Diskurswörterbuch die Struktur eines Diskurses auf lexikalisch-semantischer Ebene mit einem Verweissystem ab. Das Verweissystem ist die Darstellung eines durch die lexikologisch-semantische Analyse eines Diskurses erlangten Ergebnisses.

Diskurs ist ein Sprachgebrauchsphänomen und insofern empirisch aus einem konkreten Korpus (vgl. Busse/Teubert 1994) rekonstruierbar. Dieses Korpus bildet die Basis einer Diskursanalyse und damit auch eines Diskurswörterbuchs. Das Korpus des Schuldiskurses ist im [Quellenverzeichnis](#) des Online-Wörterbuchs nachgewiesen (zum Typ Diskurswörterbuch vgl. Kämper 2007).

2. Wörterbuchgegenstand

Gegenstand eines Diskurswörterbuchs ist die lexikografische Darstellung des thematisch strukturierten Beziehungssystems, m.a.W. des Netzes topikalisch aufeinander bezogener lexikalischer Diskursrepräsentanten, die in unterschiedlich beschaffenen begrifflich-semantischen Bezugsrelationen zueinander stehen. Insofern macht Diskurslexikografie die topikalische, begriffliche und semantische Ordnung des Diskurses (die seine Kohärenz erzeugt) mikrostrukturell auf der Wortebene sowohl semasiologisch als auch onomasiologisch sichtbar, und zwar im Kontext mit der funktionalen Diskursstruktur. Lexikografisch zu repräsentieren ist insofern der die Komplexität und Kohärenz do-

kumentierende Wortschatz eines Diskurses, dessen Elemente je unterschiedliche Funktionen in seinem Gesamtgefüge haben.

Der Gegenstand dieses Wörterbuchs ist der deutsche Schulddiskurs der frühen Nachkriegszeit. Das OWID-Modul ist die Online-Version der Printfassung (vgl. Kämper 2007). Das Wörterbuch zum Schulddiskurs besteht aus dem diesen Diskurs konstituierenden Wortschatz. Es umfasst 85 für besonders relevant gehaltene Hauptlemmata von *Abendland* bis *Zukunft* und 207 Sublemmata und wird nach Diskursbeteiligten unterschieden. Diese sind die Opfer des Nationalsozialismus, die nationalsozialistischen Täter sowie diejenigen, die nach 1945 nicht belastete gesellschaftliche, politische und künstlerische Funktionsträger waren. Der Wortschatz entspricht auf dieser Ebene der Makrostruktur der alphabetischen Liste der Hauptlemmata, der abgeleiteten Sublemmata sowie, vereinzelt, der in einer Bedeutungsbeziehung zu den Hauptlemmata stehenden, aber nicht zur Wortfamilie gehörenden Sublemmata. Diese [Lemmaliste](#) gibt Auskunft über die Bezeichnungsstruktur des in dem Wörterbuch dargestellten Wortschatzes zum nachkriegsdeutschen Schulddiskurs.

3. Äußere Zugriffsstrukturen des Diskurswörterbuchs

Das Online-Wörterbuch enthält dieselben Datenangebote, die sich auch in der Printversion finden. Durch das elektronische Medium werden aber andere Formen der Benutzung, der Suchanfrage und des Zugriffs auf das Datenangebot ermöglicht und notwendig gemacht. Zunächst kann eine Suchanfrage über das Suchfeld im OWID-Portal erfolgen (vgl. Abbildung 1).

The screenshot shows the OWID website interface. At the top left is the 'OWID' logo, and at the top right is the logo for the 'INSTITUT FÜR DEUTSCHE SPRACHE'. Below the logo is a navigation bar with 'Startseite OWID' and 'Projekt OWID'. The main content area is titled 'Willkommen in OWID, dem Online-Wortschatz-Informationssystem Deutsch des Instituts für Deutsche Sprache, Mannheim'. It includes a search bar with a 'suche' button, a checkbox for 'Groß-/Kleinschreibung beachten', and a 'Suche in ...' section with checkboxes for 'ellexiko', 'Neologismen', 'Feste Wortverbindungen', and 'Diskurswörterbuch 1945-55'. There are also 'Optionen' for 'in Gesamtstichwortliste' and 'in allen lexikografisch ausgearbeiteten Artikeln'. At the bottom, there are four buttons: 'ellexiko', 'Neologismenwörterbuch', 'Feste Wortverbindungen', and 'Diskurswörterbuch 1945-55'. The footer contains the copyright information: '© Institut für Deutsche Sprache (http://www.ids-mannheim.de) Startseite OWID | Projekt OWID' and some icons.

Abbildung 1: OWID-Portalseite mit Button „Diskurswörterbuch 1945-55“

Die Ergebnisse einer solchen Suchanfrage zu allen OWID-Modulen werden in Form einer unterstrichenen Lemmazeichenangabe zu jedem Treffer (pro gefundenem Wörterbuchartikel und Modul) angezeigt, die zugleich als Link zum betreffenden Artikel fungiert. Die Treffer aus der Lemmaliste des Moduls „Schulddiskurs im ersten Nachkriegsjahrzehnt“ sind dann jeweils in Grün hervorgehoben (vgl. Abbildung 2). Die andere Zugriffsmöglichkeit von der Portalseite aus besteht im Aufrufen der Moduleiten über den grün unterlegten Button „Diskurswörterbuch 1945-55“ (vgl. Abbildung 1). Damit gelangt der Benutzer / die Benutzerin auf die Startseite des OWID-Moduls zum Schulddiskurs. Dort erhält man Informationen zur Wörterbuchkonzeption und Datenbasis und kann auf die alphabetische Artikelliste des Schulddiskurs-Moduls zugreifen.

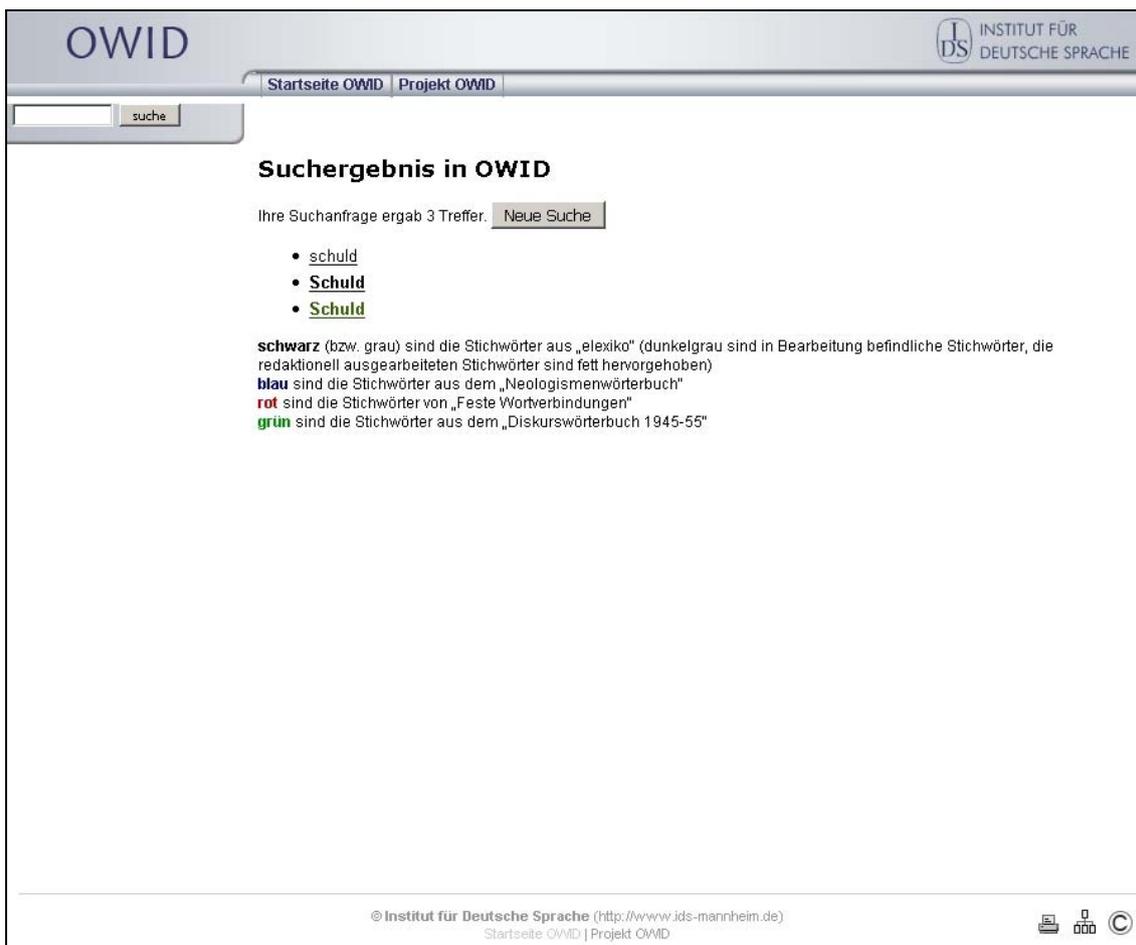


Abbildung. 2: Suchergebnisanzeige mit farblicher Markierung des Suchworts *Schuld*, das zugleich mit den entsprechenden Artikeln verlinkt ist

Innerhalb des Schulddiskurs-Moduls erfolgt der Zugriff auf die Einzelartikel über den Menüpunkt „Wortartikel“. Die Hauptlemmata sind in dieser Artikelliste alphabetisch angeordnet und jeweils halbfett in Grün und mit einer Unterstreichung hervorgehoben. Sie dienen zugleich als Hyperlinks zu den Wörterbuchartikeln. Ableitungen, Zusammensetzungen und semantisch verwandte Sublemmata, die in dem Artikel erwähnt oder mitbehandelt werden, erscheinen in der Artikelliste eingerückt unter den jeweiligen Hauptlemmata (vgl. Abbildung 3).



Abbildung 3: Verlinkte alphabetische Artikelliste des Diskurswörterbuchs 1945-55 mit eingerückten Sublemmata



Abbildung 4: Kopf des zentralen Wörterbuchartikels Schuld im Diskurswörterbuch 1945-55 mit Sublemmatblock, drei diskursperspektivischen Suchbereichen und Verweisangaben

Von einem Einzelartikel des Moduls aus (hier als Beispiel der Artikel *Schuld*) kann man nicht nur über vielfältige Verweisstrukturen zu anderen Artikeln wechseln, sondern die Liste aller Hauptlemmata im linken Bildschirmbereich ermöglicht ebenfalls das Aufrufen anderer Einträge (vgl. Abbildung 4).

4. Artikelstrukturen

Jeder Wörterbuchartikel hat als obligatorischen Bestandteil ein Lemma, als fakultativen Bestandteil einen Sublemmablock, in dem Zusammensetzungen und Ableitungen (zu *frei* z.B. *Freiheit* / *Befreiung* / *Selbstbefreiung*) sowie solche semantisch verwandten Wörter angegeben sind, die keine eigenständigen Artikel haben (zu *apokalyptisch* z.B. *eschatologisch*, zu *Finsternis* z.B. *Verfinsterung* / *dunkel* / *Nacht*, zu *Krematorium* z.B. *Krematoriumskamin* / *Krematoriumsofen* / *Kamin* / *Ofen* / *Schornstein*, zu *nominell* z.B. *Mitläufer*). Der semantische Kommentar enthält narrative Angabetexte zu semantischen Wortschatzbeziehungen und zu den Besonderheiten des diskursperspektivischen Wortgebrauchs. Die Artikel werden auf der höchsten Gliederungsebene nach den drei benannten Beteiligungsperspektiven ‘Opfer’, ‘Täter’ und ‘Nichttäter’ unterteilt. Darüber hinaus kann es in einzelnen Artikeln in Bezug auf die Beteiligtenperspektive ‘Nichttäter’ noch die Unterscheidung nach der politischen Perspektive ‘West’ und ‘Ost’ geben (vgl. z.B. die Artikel zu *Demokratie* und *Freiheit*). Die Artikelgliederung in einen oder mehrere perspektivische Subkommentare wird jeweils nach dem Sublemmablock in einer Perspektivengliederung in oranger Schriftfarbe dargestellt (vgl. Abbildung 5). Gibt es perspektivenübergreifende Gebrauchsmerkmale, so werden diese unmittelbar im Anschluss an die Perspektivengliederung zu Beginn des Artikels dargestellt.

The screenshot shows a Mozilla Firefox browser window displaying the article 'Schuld' from the 'Schulddiskurs' module. The address bar shows the URL: http://hypermedia.ids-mannheim.de/pls/lexiko/pl4_anzeige.artikel?v_id=309249&v_modul=Schuld. The left sidebar contains a list of sublemmata, including: *Einverständnis*, *Dämon*, *Demokratie*, *deutsch*, *Dienst*, *Draht*, *Faschismus*, *Finsternis*, *frei*, *Freiheit*, *Frieden*, *Gas*, *Gegenwart*, *gehört*, *Gehorsam*, *Geist*, *Geschichte*, *Gesundheit*, *Glaube*, *Glück*, *Haftung*, *Haltung*, *hart*, *heute*, *Hilfer*, *Hilflos*, *Hoffnung*, *Humanismus*, *Idealismus*, *irrtum*, *Kämpfer*, *Katharsis*, *Kollektivschuld*, *krank*, *Krematorium*, *Kultur*, *Leben*, *Liebe*, *Masse*, *Militarismus*, *Musemann*, *Nationalsozialismus*, *neut*, *nie*, *nominell*, *Nummer*, *Opfer*, *Pflicht*, *Politische Unreife*, *Reinigung*, *Revier*, *Revision*, *Sadist*, *Scham*, *Schuld*, *Selektion*, *SS*, *Steinbruch*, *Strafe*, *Stunde*. The main text area is divided into sections: **Opferdiskurs**, **Nichttäterdiskurs**, and **Täterdiskurs**. The text under 'Opferdiskurs' discusses the use of 'Schuld' in the early Nazi period, while 'Nichttäterdiskurs' and 'Täterdiskurs' discuss the perspectives of non-perpetrators and perpetrators respectively. The text is highlighted in orange to indicate perspective-based subcommentaries.

Abbildung 5: Orangefarbene Perspektivengliederung des Artikels *Schuld*

Während onomasiologische Bezüge zu nicht lemmatisierten Zusammensetzungen und Ableitungen des Schlüsselworts im Sublemmablock angegeben werden, erscheinen blau hervorgehobene Verweise in Form von Hyperlinks zu lemmatisierten Synonymen, Kollokationen und Antonymen.

Darüber hinaus gibt es zu jedem Artikel eine umfangreiche Quellendokumentation, die nicht nur der wissenschaftlichen Absicherung von sprachlichen Gebrauchsmustern dient, sondern weitere Informationsmöglichkeiten zum Sprachgebrauch in Form von authentischen Belegbeispielen bereitstellt. Diese Belege werden beim Überfahren mit dem Mauszeiger im Artikelfenster als gelbe Felder eingeblendet (vgl. Abbildung 6). Somit wird zum einen das Informationsangebot in den Artikeln übersichtlich gehalten, zum anderen aber werden auch vielfältige Möglichkeiten zur systematischen Konsultation zusätzlicher Angaben bezüglich des Wortgebrauchs eröffnet.

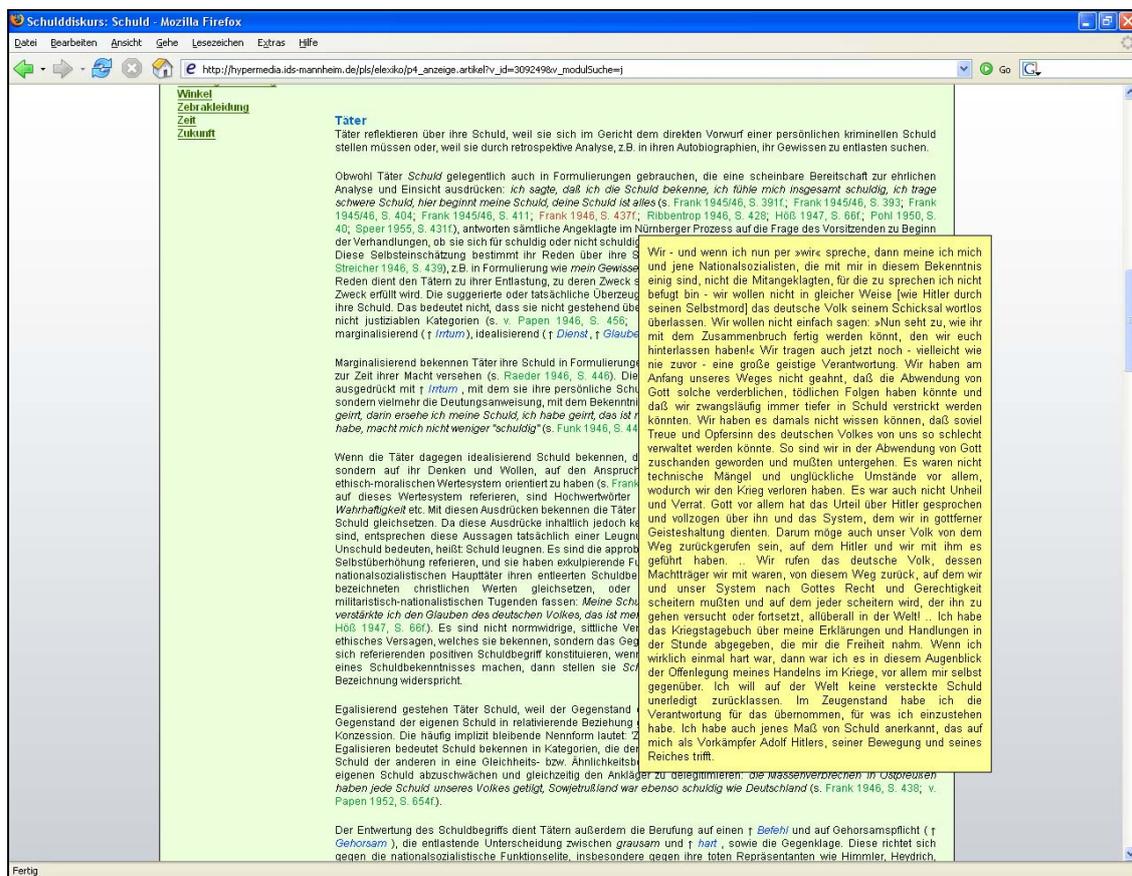


Abbildung 6: Belegangaben mit Quelleneinblendungen

5. Datendistribution und Verweise

Der Artikelkopf enthält die semantischen, enzyklopädischen und Gebrauchsinformationen. Diese Informationen werden in narrativem Stil mitgeteilt, denn das Wörterbuch soll auch ein Lesebuch sein, und in der Regel gestützt mit Beispielen, die aus dem Belegteil stammen, also authentisch sind. Sie wurden jedoch für die Verwendung im Artikelkopf u.U. gekürzt bzw. vereinfacht. Im Anschluss wird durch die Angabe der Kurztitel auf die Belege im Dokumentationsenteil verwiesen. Es wird nicht die gesamte Bedeutungsgeschichte eines Lemmas dargestellt und beschrieben, sondern nur der in

den Jahren 1945 bis 1955 belegte und aus den Diskurstexten rekonstruierte Gebrauch. Am Ende des Artikelkopfs wird gegebenenfalls auf Sekundärliteratur verwiesen. Ein wesentlicher Bestandteil jedes Artikelkopfes ist die onomasiologische Vernetzung des Diskurswortschatzes durch Verweise. Diese Vernetzung bezieht sich auf begriffliche Bedeutungsrelationen, so wird z.B. im Artikel [Abendland](#) auf *Kultur*, *Humanismus* und *Christentum* verwiesen. Außerdem werden antonymische und hyperonymische Bedeutungsrelationen durch Verweise kenntlich gemacht. Im Artikel [Antifaschist](#) wird z.B. auf *Kämpfer* und *Opfer*, im Artikel [Angst](#) wird auf *Glück* verwiesen. Vernetzt werden darüber hinaus argumentativ relevante Lexemkomplexe. Im Artikel [Befehl](#) wird z.B. auf *treu* und auf *Gehorsam* verwiesen, weil diese Lemmata in denselben argumentativen Kontexten gebraucht werden. Synonymische Bedeutungsbeziehungen werden darüber hinaus auch ohne Verweise hergestellt, und zwar dann, wenn es sich um Lexeme handelt, die keinen eigenen Artikel haben, die aber dennoch, wegen ihrer semantischen Serialität im Sinn von Ausdrucksalternativen, relevante Diskurselemente sind. So wird z.B. im Artikel [Apathie](#) die aus dem Korpus belegte Synonymik *inneres Absterben*, *Abstumpfung des Gemüts*, *innere Wurstigkeit*, *Gleichgültigwerden*, *Lethargie*, *primitive Stufe* dargestellt. Sie geht auch über die Wortartengrenze hinaus mit *seelisch abriegeln*, *stumpf geworden*, *primitiviert*.

Das Wesen eines Diskurswortschatzes besteht darin, dass er einen durch unterschiedliche Bedeutungsrelationen aufeinander bezogenen begrifflichen bzw. semantischen kohärenten Komplex bildet. Vorzustellen ist er als Netz, dessen Knoten, also die Lemmata als die lexikalischen Repräsentanten des Diskurses, in unterschiedlicher Relation zueinander stehen. Das begrifflich-semantisch komplexeste Zentrum des Schulddiskurses ist der Artikel zu dem Lemma *Schuld*. Er bildet den Zentralartikel des Wörterbuchs zum Schulddiskurs. Dieser Artikel ist gleichsam das semantische Kondensat des Schulddiskurses, in dem sozusagen die Schnüre seines topikalischen Netzes zusammenlaufen. Umgekehrt wird von diesem Zentralartikel aus auf die lexikalischen Repräsentanten des Diskurses, unterschieden nach den drei Sprecherperspektiven, verwiesen. Mit anderen Worten: In diesem Artikel ist die Ordnungsstruktur des Wortschatzes abgebildet, der den Schulddiskurs realisiert, und zwar pragmatisch unterschieden nach den drei Sprecherperspektiven und onomasiologisch durch die Vernetzung des lexikalischen Bestands des Schulddiskurses.

Entsprechend dem Zentralartikel des Schulddiskurses lassen sich darüber hinaus Schlüsselartikel jeweils für die drei Subdiskurse der Opfer, der Täter und der Nichttäter ausweisen. Sie kondensieren die semantische und die Argumentationsstruktur des jeweiligen Subdiskurses. Schlüsselartikel sind für den Opferdiskurs z.B. [Auschwitz](#), für den Täterdiskurs z.B. [Hitler](#), für den Nichttäterdiskurs z.B. [Verantwortung](#).

6. Fazit

Aufgrund seiner Konzeption als semasiologisches Wörterbuch mit einer alphabetischen Makrostruktur, komplementären onomasiologischen Sublemmareihen, einer sternförmig organisierten Mediostruktur mit einem synoptischen Kernwortartikel und verschiedenen Peripherieartikeln ergeben sich vielfältige Möglichkeiten der gezielten Konsultation und der strukturierten sprach- und kulturhistorischen Informationserschließung und -erweiterung.

Die Implementierung in OWID ermöglicht es dem Portalbenutzer / der Portalbenutzerin, zusätzliche diskurspezifische Informationen zu brisanten Wörtern des öffentlichen Sprachgebrauchs, wie etwa [Auschwitz](#), [Kollektivschuld](#) etc., zu erschließen sowie, vice versa, die Diskurslexik an allgemeinsprachliche lexikalische Zusammenhänge anzubinden.

7. Literatur

- Busse, Dietrich/Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hg.): Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik. Opladen. S. 10-28.
- Kämper, Heidrun (2005): Der Schuldiskurs in der frühen Nachkriegszeit. Ein Beitrag zur Geschichte des sprachlichen Umbruchs nach 1945. Berlin/New York.
- Kämper, Heidrun (2006): Opfer – Täter – Nichttäter. Ein Wörterbuch zum Schuldiskurs 1945-1955. Berlin/New York.
- Kämper, Heidrun (2007): Diskurs und Diskurslexikographie. Zur Konzeption eines Wörterbuchs des Nachkriegsdiskurses. In: Deutsche Sprache 4/06, S. 334–353.

Doris Steffens

Zu den Verlinkungen in den Wörterbüchern des Internetportals OWID – ein Erfahrungsbericht aus Sicht des Neologismenwörterbuches (NWB)

Abstract

Der Beitrag zeigt, ausgehend von der Darstellung der neuen Benutzungsmöglichkeiten der Onlinefassung des Neologismenwörterbuches gegenüber seiner Printfassung, welche Links derzeit innerhalb des Neologismenwörterbuches sowie von diesem auf die Wörterbücher des Portals OWID und auf andere elektronische Wörterbücher gesetzt werden. Am Beispiel der Wortartikel *Adresse* und *Klammeraffe*, die sowohl im Neologismen- als auch im *ellexiko*-Wörterbuch ausgearbeitet vorliegen, werden Überlegungen hinsichtlich der geplanten Verknüpfung zwischen Wortartikeln mit gleichlautenden Stichwörtern angestellt. Sie betreffen insbesondere die Lesarten und ihre Etikettierung sowie die Verlinkung von sinnverwandten Wörtern. Eine Verständigung darüber kann dazu beitragen, dem jeweiligen Projektkonzept besser gerecht zu werden und die Darstellung eindeutiger und damit letztlich auch benutzerfreundlicher zu machen.

Comparing the print version of the Dictionary of Neologisms with the online version of this dictionary, this paper shows which novel possibilities of usage are allowed by the online version and which links it includes to allow references to other OWID dictionaries as well as to dictionaries which are not part of OWID. On the basis of the dictionary entries for *Adresse* and *Klammeraffe* from the Dictionary of Neologisms and the OWID-dictionary *ellexiko*, this paper also demonstrates how entries with similar key words may be linked. Considerations about the linking possibilities planned focus on the representation of the senses of a head word, on how they may be labelled, and on the crosslinking of synonyms. A basic agreement about these issues is likely to contribute to an unambiguous representation of the OWID dictionary projects and to a user-friendly realisation of the dictionaries concerned.

Inhalt:

1. Einführung
2. Das NWB in der Online- und der Printversion – ein Vergleich auch in Hinblick auf die Verknüpfungsmöglichkeiten
3. Zur Handhabung der Verknüpfungen von Wortartikeln vonseiten des NWB
4. Ein spezieller Fall: Verlinkungen bei Wortartikeln mit gleichlautenden Stichwörtern in NWB und *ellexiko*-Wörterbuch (eWB)
 - 4.1 Beispiel *Adresse*
 - 4.2 Beispiel *Klammeraffe*
5. Schlussbemerkung
6. Literatur

1. Einführung

Im Internetportal OWID für die lexikografischen Publikationen der Abteilung Lexik des IDS ist auch das im Projekt Neologismen erarbeitete Neologismenwörterbuch vertreten. Es präsentiert korpusbasiert den neuen Wortschatz, der in den 90er Jahren des 20. Jahrhunderts in der deutschen Allgemeinsprache aufgekommen ist.¹

¹ Ursprünglich war das Neologismenprojekt allein auf die Veröffentlichung in einem Printwörterbuch ausgerichtet. Als aber in der Abteilung Lexik das Konzept für das lexikalisch-lexikologische korpusbasierte Informationssystem *ellexiko* (jetzt: OWID) entwickelt wurde, bot sich die Möglichkeit, auch den Wortschatzbereich der Neologismen, im ersten Arbeitsgang für die 90er Jahre, entsprechend zu

Das Portal strebt nicht das bloße Nebeneinander von unterschiedlichen lexikografischen Arbeitsergebnissen dieser Abteilung an, sondern verfolgt das Ziel, die Inhalte auf sinnvolle Weise so miteinander zu verzahnen, dass es gleichermaßen von Vorteil für die Wörterbücher wie für die Nutzer ist. Die Voraussetzungen dafür sind geschaffen: Die Angaben der im Portal vertretenen Wörterbücher sind nach einheitlichen Prinzipien modelliert, und somit ist auch die Möglichkeit gegeben, flexibel auf sie zuzugreifen.²

2. Das NWB in der Online- und der Printversion – ein Vergleich auch in Hinblick auf die Verknüpfungsmöglichkeiten

Die neuen Organisationsmöglichkeiten durch Hypertext werden im Vergleich des 2004 erschienenen gedruckten NWB³ mit der 2006 ins Netz gestellten Onlineversion deutlich. Das gedruckte NWB hat insofern eine besondere Vorgeschichte, als es von Beginn an in der XML-Artikelstruktur⁴ erarbeitet wurde, ihm also die gleiche XML-basierte Datenmodellierung zugrunde liegt wie der Onlineversion. Deshalb unterschied sich bereits der gedruckte Band von einem herkömmlichen allgemeinen Gesamtwörterbuch dadurch, dass die traditionellen lexikografischen Standards, die auch von platzmangelbedingter Formalisierung und Beschränkung geprägt sind, nicht mehr eingehalten werden mussten. Dies war möglich, weil das NWB als Spezialwörterbuch mit einer relativ kleinen Zahl von ca. 700 Stichwörtern Platz für eine dem erhöhten Informationsbedarf bei neuem Wortschatz gerecht werdende, vergleichsweise ausführliche Darstellung bot. Dem entsprach ihrerseits die XML-basierte Datenmodellierung, die ein vielfältiges, bei Bedarf erweiterbares Angebot an Datentypen bereithält, was nicht zuletzt auch eine breite Präsentation von sprachlichen und außersprachlichen Zusammenhängen ermöglicht. Vorteil der XML-basierten Datenmodellierung ist es auch, dass nicht nur die Darstellung der Onlineartikel, sondern auch das Layout für einen gedruckten Band⁵ umgesetzt werden kann.

Trotz dieses unbestreitbaren Fortschritts bleiben natürlich Beschränkungen, die der Linearität des Printmediums geschuldet sind und hier nur angedeutet werden können:

bearbeiten und online zugänglich zu machen. Somit ergab sich die für das Neologismenprojekt interessante Ausgangsposition, die ursprünglich auf das Printwörterbuch zugeschnittene Konzeption im Rahmen der neuen Möglichkeiten eines hypertextuellen Informationssystems zu erproben.

² „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 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.“ (Müller-Spitzer 2008)

³ Herberg, Dieter/Kinne, Michael/Steffens, Doris (2004): Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi. XXXIX/393 S. (= Schriften des Instituts für Deutsche Sprache 11). Berlin/New York.

⁴ XML (*eXtensible Markup Language*) = international standardisierte Metasprache zur Auszeichnung von Daten

⁵ Geplant ist ein weiterer Band des NWB für die Neologismen des ersten Jahrzehnts des 21. Jahrhunderts.

Im gedruckten NWB ist eine – gerade für neuen Wortschatz wichtige – kontinuierliche Aktualisierung, Ergänzung und Korrektur von Daten ebenso wenig möglich wie die Darstellung sämtlicher im Onlinewörterbuch präsentierter Daten⁶ und die Verknüpfung mit Wortartikeln anderer Wörterbücher und Nachschlagewerke.

Die in Wörterbüchern generell bevorzugte selektive Rezeption von Informationen funktioniert in einem gedruckten Wörterbuch nur bedingt. Das auch in der Printfassung des NWB praktizierte, durchaus verschiedene Kriterien berücksichtigende Verweissystem ist mit (zeitaufwendigem) Blättern verbunden und stößt spätestens dann an seine Grenzen, wenn bestimmte Informationen aus den Wortartikeln quer zu der alphabetischen Ordnung im Wörterbuch gesucht werden. Ein elektronisches Wörterbuch mit entsprechender Datenmodellierung bietet sich dagegen geradezu an, einzelne Angaben eines Wortartikels selektiv zu rezipieren. So kann beispielsweise auf eine einzelne Lesart oder auf Belege zugegriffen werden und es lassen sich – als eigentlicher Mehrwert des Internetwörterbuches gegenüber dem gedruckten Wörterbuch – beliebig viele Suchkriterien kombinieren und so die Stichwörter, auf die die gesuchten Angabeklassen zutreffen, ermitteln.

Die elektronischen Verweise, die Links, ermöglichen die vernetzte Präsentation von Daten, in unserem Fall die Verknüpfung von Textdaten⁷ innerhalb eines Wortartikels, zwischen den Wortartikeln eines Wörterbuches oder zwischen den Wortartikeln verschiedener Wörterbücher. Sie beschleunigen das „Nachschlagen“ und regen zum weiteren Stöbern an. Auch wenn die Links den zeitlichen Aufwand, zu den gesuchten Daten zu gelangen, wesentlich verringern helfen, sollten sie mit Bedacht gesetzt sein, um zum einen beim Nutzer Irritationen zu vermeiden, zum anderen, und dies zeigt sich bei den Wörterbüchern im Internetportal OWID, weil Wortschatzbereiche anderer Wörterbücher im Portal OWID betroffen sein können.

3. Zur Handhabung der Verknüpfungen von Wortartikeln vonseiten des NWB

Erwartungsgemäß am häufigsten sind Links gesetzt, die die Wortartikel des NWB systematisch miteinander verknüpfen. In erster Linie erfolgt die Verknüpfung in Hinblick auf die sinnverwandten Ausdrücke, die mit dem Stichwort gebildeten Wortbildungsprodukte und die Wortbildung des Stichwortes.

Sowohl innerhalb der sinnverwandten Ausdrücke als auch innerhalb des allein im NWB präsentierten Datentyps „Wortbildungsproduktivität“ wird auf die Stichwörter, die im NWB einen eigenen Eintrag haben, ein Link gesetzt, z.B.:

⁶ Im Printwörterbuch mussten bei aller Großzügigkeit in der Datenpräsentation trotzdem in Hinblick auf die Papierkosten bestimmte Passagen gekürzt werden: Das betraf einen Großteil der Belege, die Angabe der Wörterbücher mit der Stichwortbuchung sowie die nicht normgerechten Schreibungen des Stichwortes.

⁷ Hier sind nur Textdaten der Wortartikel selbst gemeint, nicht z.B. die der Wörterbuchbenutzung dienenden Umtexte. Bilddaten setzen wir häufig bei Stichwörtern, die Konkreta bezeichnen, zur Unterstützung von Textdaten ein, Audiodaten sind zur Ergänzung der herkömmlichen phonetischen Umschriften geplant.

[Onlinebanking](#)**Sinnverwandte Ausdrücke**Synonym(e): [E-Banking](#), [Electronic Banking](#), [Internetbanking](#)Kohyponym(e): [Direktbanking](#), Handybanking, [Homebanking](#), PC-Banking, [Telebanking](#), [Telefonbanking](#)Hyperonym(e): [Banking](#)[Euro](#) Lesart 'Einheitswährung'**Wortbildungsproduktivität**Zusammensetzung: als Grundwort, z.B. in *Buchgeldeuro*, *Weicheuro*sehr häufig als Bestimmungswort, z.B. in [Eurocent](#), [Eurokurs](#), [Euroland](#),[Euromünze](#), [Euroraum](#), [Eurozone](#)

Bei den Angaben zur Wortbildung wird auf Bestandteile bzw. auf die Basis von im Deutschen gebildeten Neologismen verlinkt, die ebenfalls Neologismen sind und einen eigenen Eintrag haben, z.B.:

[Gegenfinanzierung](#)**Wortbildung**

Wortbildungsart/-typ: Ableitung (explizite Derivation)

Basis: [gegenfinanzieren](#) (Verb)

Suffix: -ung

Innerhalb des Datentyps „Wortbildung“ wird auch bei speziellen Erscheinungen ein Link gesetzt, so von der Übersetzung eines Lehnwortes auf das Lehnwort selbst – ungeachtet einer Verlinkung im Bereich der sinnverwandten Ausdrücke (z.B. [Prepaidkarte](#) vs. [Prepaidcard](#)).

Die Verknüpfungen zwischen den Wörterbüchern im Internetportal OWID sind unterschiedlich stark ausgeprägt. Da im *lexiko*-Wörterbuch (eWB) und im NWB die Übereinstimmung hinsichtlich der Artikelstrukturen im Vergleich zu den anderen Wörterbüchern im Portal am größten ist, steckt aus Sicht des NWB hier auch das größte Verlinkungspotenzial.

Innerhalb des Datentyps „Wortbildung“ im NWB bietet die ca. 300.000 Einträge umfassende Stichwortliste des eWB für das NWB dann obligatorisch die Datenbasis, wenn die Bestandteile bzw. die Basis eines im Deutschen gebildeten Neologismus Nichtneologismen und gleichzeitig Stichwörter im eWB sind, z.B.:

[Hörbuch](#)**Wortbildung**

Wortbildungsart/-typ: Zusammensetzung (Determinativkompositum)

Bestandteil: [hören](#) (Verb)Bestandteil: [Buch](#) (Nomen)

Innerhalb des Datentyps „Wortbildung“ wird auch bei speziellen Erscheinungen ein Link auf das entsprechende Stichwort im eWB gelegt, z. B.:

[entschleunigen](#)

entschleunigen ist zu [beschleunigen](#) in Analogie zu Gegensatzpaaren wie *belüften* / *entlüften* gebildet, wobei das Präfix *be-* von *beschleunigen* durch *ent-* ersetzt wird.

Ein Link wird auch dann gesetzt, wenn zu den entsprechenden Stichwörtern keine lexikografisch ausgearbeiteten Artikel, sondern nur so genannte „Rumpf“-Artikel mit derzeit noch wenigen automatisch erstellten Angaben vorliegen. Für diese Vorgehensweise spricht, dass in Hinblick auf einen später auszuarbeitenden Wortartikel des eWB ein Link nicht nachträglich gesetzt werden muss, auch wenn er jetzt nicht zu den Informationen führt, die wahrscheinlich erwartet werden.

Nach Möglichkeit werden auch Links zu den anderen Wörterbüchern im Portal hergestellt. So sollen die beiden Wortartikel [Blind Date](#) im NWB und [Blind Date](#) in „Feste Wortverbindungen“ miteinander verlinkt werden, wobei Kommentare jeweils auf die Projektspezifika Bezug nehmen:

Im NWB wird die Erklärung, die sinngemäß bereits in den Benutzerhinweisen steht, lauten: „Als Wörter gelten auch solche Anglizismen wie *Blind Date*, denen in der Herkunftssprache eine Wortgruppe entspricht.“

In „Feste Wortverbindungen“ heißt es: „*Blind Date* wurde auf Grund des signifikanten Ko-Vorkommens als Stichwort angesetzt: *Date* ist ein statistisch auffälliger Kookkurrenzpartner von *blind*.“

Zum „Diskurswörterbuch 1945-1955“ ist bereits ein Link gesetzt: Er führt vom Neologismen-Wortartikel [Tätervolk](#) zum Wortartikel [Kollektivschuld](#).

Auf andere elektronische Nachschlagewerke verlinken wir, um dem Nutzer die Möglichkeit zu geben, auf weiterführende Informationen (z.B. die ältere Bedeutung eines Neologismus) zugreifen zu können. Bei sprachlichen Informationen kommen vor allem die folgenden elektronischen Nachschlagewerke infrage: [DWDS](#) (z.B. beim Stichwort [Klammeraffe](#)) und [wissen.de](#) (z.B. beim Stichwort [toppen](#)), das Internetportal mit Zugang zu Inhalten von zum Bertelsmann-Konzern gehörenden Verlagen, oder auch [Meyers Lexikon online](#) (z.B. beim Stichwort [online](#)). Für vertiefende Informationen zu dem vom Stichwort bezeichneten Sachverhalt, der im NWB besonders im Datentyp „Enzyklopädisches“ kommentiert wird, setzen wir gegebenenfalls einen Link auf [Wikipedia](#) bzw. auf [Meyers Lexikon online](#).

4. Ein spezieller Fall: Verlinkungen bei Wortartikeln mit gleichlautenden Stichwörtern in NWB und eWB

Dass die Entscheidung, wo ein Link gesetzt wird und wohin er führt, nicht formal zu handhaben ist, sondern gerade bei einem Portal wie dem unseren, das für sich in Anspruch nimmt, mehr als das reine Nebeneinander von Wörterbüchern zu sein, gut überlegt sein will, soll dieser Abschnitt zeigen. Die beiden folgenden Beispiele betreffen Stichwörter, die sowohl im NWB als auch im eWB einen ausgearbeiteten Wortartikel haben.

Es wird dargestellt, wie die Verlinkungen gehandhabt werden und welche Überlegungen sich daraus ergeben.

4.1 Beispiel *Adresse*

Im Mittelpunkt steht hier die Verknüpfung von Informationen bei Wortartikeln, deren Stichwort eine Neubedeutung der 90er Jahre repräsentiert, d.h. die in den 90er Jahren aufgekommene(n) Bedeutung(en) eines im deutschen Wortschatz bereits etablierten Lexems (z.B. [Lichterkette](#) Lesart ‘Demonstration’, [herunterladen](#) Lesart ‘Daten kopieren’).

Da im NWB der Begriffsbestimmung gemäß allein diese neue(n) Bedeutung(en) exhaustiv beschrieben werden, musste die Frage beantwortet werden, wie dem legitimen Bedürfnis des Nutzers nach dem Zugriff auf das gesamte Bedeutungsspektrum des Lexems am besten entsprochen werden kann.

Bisher erfolgt der Zugang auf die älteren Bedeutungen vom Wortartikel der Neubedeutung aus auf keines der Wörterbücher im Internetportal OWID. Entweder waren die betreffenden Stichwörter dort bisher nicht ausgearbeitet (eWB) oder sie kommen dort nicht vor (Feste Wortverbindungen, Diskurswörterbuch 1945-1955). Infolgedessen wird in erster Linie auf das WDG⁸ verlinkt, das in digitalisierter Form auf den Seiten des Wörterbuches der deutschen Sprache des 20. Jahrhunderts (DWDS) präsent ist. Das WDG bietet sich an, da es nach Prinzipien der wissenschaftlichen Lexikografie erarbeitet wurde und das Erscheinen so lange zurückliegt, dass Neubedeutungen in unserem Verständnis noch gar nicht gebucht sein können, so dass Unklarheiten in Bezug auf die gemeinte Bedeutung nicht auftreten können.

Hat das Stichwort im WDG keinen eigenen Wortartikel (meist bei Anglizismen wie [Button](#) oder Präverbfügungen wie [herausschreiben](#), die ihrer Basis, hier *schreiben*, zugeordnet sind), werden die elektronischen Nachschlagewerke [Wikipedia](#), [wissen.de](#) sowie [Meyers Lexikon online](#) für eine Verlinkung herangezogen. Wenn auch diese Quellen nichts erbringen, entfällt der Link.

Da im eWB nunmehr auch Stichwörter, die im NWB mit einer oder mehreren Neubedeutungen gebucht sind, ausgearbeitet werden,⁹ ist perspektivisch geplant, statt auf die oben genannten elektronischen Nachschlagewerke einen Link auf den entsprechenden Wortartikel des eWB zu setzen – und umgekehrt von dort auf den des NWB.

Wie stellt sich dies nun konkret am Beispiel der beiden Wortartikel mit dem Stichwort *Adresse* dar?

Im NWB hat das Stichwort [Adresse](#) zwei neue Bedeutungen. Die eine lässt sich (anhand einer so genannten Lesartenetikettierung, von der aus sich durch Klick die jeweiligen lesartenbezogenen Angaben öffnen) kurz mit ‘E-Mail-Adresse’ beschreiben, die andere mit ‘Internetadresse’ (siehe Abb. 1).

⁸ WDG = Wörterbuch der deutschen Gegenwartssprache (1964–1977). Hrsg. v. Ruth Klappenbach und Wolfgang Steinitz. 6 Bde. Berlin.

⁹ In der Gesamtstichwortliste der im Portal OWID vereinten Wörterbücher gibt es bei einer Neubedeutung, die Stichwort im NWB ist, jeweils noch einen weiteren gleichlautenden Eintrag für einen (potenziellen) Wortartikel im eWB.

The screenshot shows the OWID website interface. At the top, there is a search bar and navigation links. The main content area is titled 'Adresse' and includes a 'Benutzerhinweise' button and a 'NEOL' label. Below the title, there are sections for 'Lesartenübergreifende Angaben' and 'Lesartenbezogene Angaben'. The 'Neologismtyp:' is listed as 'Neubedeutung'. Under 'Lesartenbezogene Angaben', it shows 'Lesart: 'E-Mail-Adresse'' and 'Lesart: 'Internetadresse'' with a 'weiter' button. A paragraph explains that these two variants stand to the original 'Anschrift' in a metaphorical relationship. A 'Druckversion des gesamten Artikels' button is located in the bottom right corner.

Abbildung 1: Bildschirmansicht der Eingangsseite des Wortartikels *Adresse*

This screenshot shows the detailed entry for 'Adresse' (variant 'E-Mail-Adresse'). It includes a 'Benutzerhinweise' button and a 'NEOL' label. The 'Lesartenübergreifende Angaben' section is present. The 'Lesart 'E-Mail-Adresse'' section shows 'Lesart: 'Internetadresse'' and an 'Aufkommen:' note: 'seit Mitte der 90er Jahre des 20. Jahrhunderts in Gebrauch'. The 'Wortbildung' section lists 'Wortbildungsart/-typ:' as 'Kurzwort (unisegmental; Endwort)' and 'Basis:' as 'E-Mail-Adresse (Nomen) oder Mailadresse (Nomen)'. Below this is a table with three columns: 'Bedeutung und Verwendung', 'Grammatik', and 'Weitere Informationen'. The 'Bedeutungsangabe' section contains a blue box with the text: 'aus dem Benutzernamen, dem Zeichen @ und dem Rechnernamen bestehende Adresse, unter der jemand im Internet per E-Mail erreichbar ist'. At the bottom, there is a 'Hinweis(e)' button, a text box with 'Ältere Bedeutung(en): hier klicken', a 'Schließen' button, and a 'Beleg(e)' button.

Abbildung 2: Bildschirmansicht des Wortartikels *Adresse* (Lesart 'E-Mail-Adresse')

Der Link bei [E-Mail-Adresse](#) führt auf die zu dieser Lesart gehörende Seite mit dem geöffneten Reiter „Bedeutung und Verwendung“, in dem unterhalb der Bedeutungsangabe der Link auf die ältere(n) Bedeutung(en) positioniert ist. Hier kann (wie in der Lesart [‘Internetadresse’](#) auch) die ältere Bedeutung ‘Adresse, Wohnungsangabe’ im WDG angeklickt werden (siehe Abb. 2).

Ersetzt man nun diesen Link auf das WDG durch den auf das eWB, entfaltet sich im dortigen Wortartikel [Adresse](#) das Bedeutungsspektrum der folgenden fünf Lesarten [‘Anschrift’](#), [‘Anlaufstelle’](#), [‘E-Mail-Anschrift’](#), [‘Internetlokalisierung’](#), [‘IP-Nummer’](#). Die beiden erstgenannten, [‘Anschrift’](#), [‘Anlaufstelle’](#), lassen sich vom Rezipienten, sicher auch von einem Deutschlerner, als die im NWB gemeinten älteren Bedeutungen identifizieren. Bei der nächsten Lesart [‘E-Mail-Anschrift’](#) kann der Nutzer sicher ebenfalls den Bezug zur Lesart [‘E-Mail-Adresse’](#) im NWB herstellen.

Dagegen dürften die Lesarten [‘Internetlokalisierung’](#) und [‘IP-Nummer’](#) im eWB bzw. [‘Internetadresse’](#) im NWB Verwirrung stiften. Es wird nicht auf Anhieb klar, dass [‘Internetlokalisierung’](#) im eWB und [‘Internetadresse’](#) im NWB denselben Referenzbezug haben. Und wenn der Nutzer Klarheit hat, wird er sich zu Recht fragen, warum die Lesart [‘IP-Nummer’](#), zweifellos neu, im NWB nicht gebucht ist. Die Erklärung lautet, dass diese Lesart fachsprachlich ist und somit nicht das vom Neologismenprojekt zugrunde gelegte Kriterium der Allgemeinsprachlichkeit erfüllt. Diese Information bleibt dem Nutzer bisher vorenthalten.

Unterschiedliche Bezeichnungen von Lesarten können miteinander verzahnt werden, indem ein Link von Seiten des eWB auf die Neubedeutung im NWB gleich auf der Eingangsseite des *lexiko*-Wortartikels neben der betreffenden Lesart platziert wird. Auf jeden Fall wäre eine Abstimmung hinsichtlich der Bezeichnung für die Lesartenetikettierung – nicht zuletzt auch in Hinblick auf die Benutzerfreundlichkeit – wünschenswert.¹⁰

Da in einem Wörterbuch, das ausschließlich neuen Wortschatz präsentiert, etablierte Lexeme dementsprechend nur in Bezug auf ihre neue(n) Bedeutung(en) dargestellt werden, entfielen bisher im NWB bei Neubedeutungen die lesartenübergreifenden Angaben wie Schreibung und Aussprache, Wortbildung, Aussprache, da sie der Genese entsprechend bei den älteren Lesarten sinnvoller aufgehoben sind und zudem die Möglichkeit bestand, dass sie im eWB in einem entsprechenden Wortartikel Platz finden. Ausnahmen gibt es dort, wo ein erwartbares Benutzerinteresse es erfordert. So hat im NWB der Anglizismus [Tag](#) eine Ausspracheangabe, um von vornherein eine Verwechslung mit dem indigenen Lexem *Tag* auszuschließen.

Diese mit der Aussparung von Angaben doch sehr an die Vorgehensweise im herkömmlichen Wörterbuch erinnernde Darstellung ist zu überdenken. Zum einen gibt es keinen Grund, z.B. aus Platzmangel auf eine Angabe von Daten an zwei Stellen zu verzichten, zum anderen existiert ohnehin nur dann für eine Neubedeutung im NWB ein entspre-

¹⁰ Inzwischen wurde ein Anfang hinsichtlich einer Übereinstimmung in der Bezeichnung für die Lesart gemacht: Im NWB wurde bei der Neubedeutung *Seite* im Sinne von ‘Internetseite’ die Lesartenetikettierung, dem Eintrag im eWB folgend, in ‘Internetpräsenz’ geändert.

In diesem Zusammenhang wäre auch die Frage zu stellen, ob es bei nur wenigen Unterschieden hinsichtlich der Datentypen wirklich sinnvoll ist, dass zwei Projekte die gleichen Lesarten von Stichwörtern bearbeiten.

chender ausgearbeiteter Artikel im eWB, wenn das Wort den dortigen Auswahlkriterien entspricht. Auf jeden Fall bedeutet die explizite Darstellung der lesartenübergreifenden Angaben im Wortartikel einer Neubedeutung einen Informationszuwachs.

4.2 Beispiel *Klammeraffe*

Ein Vergleich in der Darstellung der Synonymiebeziehungen in den beiden Wortartikeln *Klammeraffe* für die Neubedeutung im NWB und das Lexem im eWB zeigt eine andere Facette des gleichen Problems.

<p>NWB Klammeraffe Sinnverwandte Ausdrücke Synonym(e): Affenschwanz, at, at-Zeichen</p>	<p>eWB Klammeraffe Sinnverwandte Wörter Beziehung(en) der Bedeutungsgleichheit/ -äquivalenz Synonym(e): Affenschwanz, at</p>
--	--

Im NWB wird ein Link bei den sinnverwandten Ausdrücken nur auf Neologismen mit ausgearbeiteten Wortartikeln (hier: [at](#)) gesetzt. Die anderen aufgeführten Synonyme (hier: *Affenschwanz*, *at-Zeichen*) haben keinen ausgearbeiteten Wortartikel, also auch keinen Link.

Im eWB wird bei *at* auf den gleichlautenden Neologismenwortartikel verlinkt, bei *Affenschwanz* auf einen Rumpffartikel des eWB mit derzeit noch wenigen automatisch erstellten Angaben.

Wenn es für einen Neologismus keinen ausgearbeiteten Artikel gibt, liegt der Grund meist in der – aus Sicht des NWB – unzureichenden Beleglage im elektronischen IDS-Textkorpus. Das NWB hat relativ strenge Kriterien für die Aufnahme eines Kandidaten in die Neologismen-Stichwortliste angelegt, die auch der korpusbasierten Arbeitsweise geschuldet sind, da die Aufnahme immer mit der Ausarbeitung eines Wortartikels verbunden ist. Bekäme ein nicht den Kriterien entsprechendes Lexem trotzdem einen Link, würde der auf die ca. 300.000 Stichwörter umfassende Stichwortliste des eWB führen¹¹ und damit auf einen Rumpffartikel, der im hier gemeinten Zusammenhang unseres Erachtens wenig hilfreich ist und nur die Frage heraufbeschwören würde, warum dieses zweifellos neue Wort im NWB kein Stichwort ist.

Eine Recherche (am 29.10.2007) im IDS-Textkorpus (Gesamtkorpus) erbringt z.B. bei *Affenschwanz* insgesamt 20 Belege, von denen sich fünf auf die hier relevante Lesart ‘Gliederungszeichen’ beziehen. Insofern haben wir keine Handhabe, einen Wortartikel *Affenschwanz* auszuarbeiten und damit den Link dorthin zu lenken.

¹¹ In dieser überaus umfangreichen Liste können durchaus auch Lexeme Eingang gefunden haben, die im hier relevanten Erfassungszeitraum aufgekommen sind, also neu sind. Sie gehen dann in die Stichwortliste des NWB über, wenn sie den Auswahlkriterien des Neologismenprojektes entsprechend eigene ausgearbeitete Wortartikel bekommen.

Gegebenenfalls könnte man in einem Rumpfartikel für einen Neologismus, der keinen Eintrag im NWB hat, einen Hinweis geben, dass das Stichwort zu wenig belegt ist.

Für Wörter, die als Synonyme zu einem Stichwort fungieren, aber wegen schwacher Belegung keinen eigenen Eintrag haben, geben wir beim Stichwort einen Standardsatz. Im konkreten Fall steht im Wortartikel *Klammeraffe* unter „Weitere Informationen“ bei „Sprachreflexives“: „Das Lexem *Affenschwanz* hat sich als seit Mitte der 90er Jahre des 20. Jahrhunderts belegtes Synonym von *Klammeraffe* nicht durchsetzen können.“

Um auch Neologismen wie *Affenschwanz* auffinden zu können, die mindestens an einer Stelle innerhalb eines Neologismen-Wortartikels vorkommen, aber nicht selbst Stichwort sind und sich somit weder alphabetisch noch systematisch ermitteln lassen, hat das Neologismenprojekt konzeptionelle Überlegungen angestellt, wie der Zugriff auf diesen Wortschatzbereich erfolgen kann. Es sollen dann Wörter aufgespürt werden können, die beispielsweise in einem Wortartikel als Kürzung des Stichwortes aufgeführt sind (z.B. **BMI** siehe Body-Mass-Index), die als mit dem Stichwort gebildet unter dem Datentyp „Wortbildungsaktivität“ aufgelistet sind (z.B. **Eurozone** siehe Euro Lesart ‘Einheitswährung’) oder die als mit dem Stichwort in semantischer Beziehung stehend unter dem Datentyp „Sprachreflexives“ erwähnt sind (z.B. **Affenschwanz** siehe Klammeraffe).

Natürlich steht es jedem Wörterbuch grundsätzlich frei, Links innerhalb des eigenen Wörterbuches nach der eigenen Systematik zu setzen. Da aber die Stichwortliste des eWB so umfangreich ist, dass sie zwangsläufig auch in den 90er Jahren aufgetauchten neuen Wortschatz enthält, der aus Gründen, die nur das NWB zu vertreten hat, in diesem keine ausgearbeiteten Wortartikel bekommen hat, wäre es empfehlenswert, Links in diesem Fall nur nach gegenseitiger Absprache zu setzen.

5. Schlussbemerkung

Wie die dargestellten Sachverhalte zeigen, wächst mit dem Fortschreiten der Arbeiten an diesem Wörterbuchportal auch das Problembewusstsein. Es sind noch viele Überlegungen für eine optimale Gestaltung, verbunden mit der Bereitschaft, sie auch umzusetzen, erforderlich.

Links an der richtigen Stelle mit den richtigen Inhalten tragen dazu bei, das den Onlinewörterbüchern innewohnende Potenzial, sich mit der über das eigene Wörterbuch hinausführenden vernetzten Präsentation von Daten als durchlässig und somit besonders nutzerfreundlich zu erweisen, möglichst umfassend auszuschöpfen.

6. Literatur

Müller-Spitzer, Carolin (2008): Der texttechnologische Aufbau von OWID. [In dieser Publikation.]

Kathrin Steyer / Meike Lauer / Annelen Brunner

Usuelle Wortverbindungen in OWID. Lexikografische Aspekte unter besonderer Berücksichtigung der Pragmatik

Abstract

Dieser Artikel fasst wichtige Aspekte der vom Projekt 'Usuelle Wortverbindungen' (UWV) erarbeiteten Konzeption für die korpusbasierte lexikografische Beschreibung von Wortverbindungen in OWID zusammen. Der Schwerpunkt in diesem Teilprojekt liegt auf der lexikografischen Beschreibung des typischen Gebrauchs von usuellen Wortverbindungen auf der Basis eines sehr großen Korpus des Deutschen. Zur differenzierten Untersuchung des Sprachgebrauchs werden korpusanalytische Methoden herangezogen und die Ergebnisse in einem nutzerfreundlichen Hypertextformat präsentiert. Zudem ist es ein Ziel, die sprachliche Vielfalt, die in den Korpora gerade auch in Bezug auf Wortverbindungen zu finden ist, durch eine große Menge authentischer Korpusbelege angemessen darzustellen.

This article presents important aspects of the corpus-based description of multi-word expressions in OWID, as developed by the project "Usuelle Wortverbindungen" (UWV). The main focus lies on the lexicographical description of the typical usage of multi-word expressions based on the evidence of a very large corpus of written German. Corpus analytical methods are employed to study language use and the results are presented in a user-friendly hypertext format. The diversity of language reflected in multi-word expressions is documented by a large number of authentic corpus samples.

Inhalt:

1. Die Erarbeitung der Musterartikel
2. Zur Mikrostruktur der Mehrwortartikel
 - 2.1 Lesartenübergreifende Angaben
 - 2.2 Lesartenspezifische Angaben
 - 2.3 Kommentare
3. Perspektiven
4. Literatur
5. Referenzwörterbücher

1. Die Erarbeitung der Musterartikel

In der ersten Arbeitsphase dieses Teilprojektes ging es um die Erarbeitung von Mehrwort-Musterartikeln, wobei insbesondere getestet werden sollte, welche Artikelstruktur sich in welchem Umfang für eine Mehrwortbeschreibung eignet. Des Weiteren sollte überprüft werden, mit welchem Arbeitsaufwand systematisches korpusbasiertes Vorgehen verbunden ist. Es wurden unterschiedliche Auswahlkriterien für die Lemmakandidaten in Abhängigkeit von der jeweiligen Zielsetzung zugrunde gelegt. Als Ergebnis entstanden zwei Artikeltypen: [25 Mehrwort-Musterartikel](#) und [100 Wortverbindungsartikel](#). Beide werden im Folgenden näher beschrieben.

Zur Ermittlung der Lemmakandidaten. Die Auswahl der Lemmakandidaten erfolgte durchweg auf der Basis der am IDS entwickelten [statistischen Kookkurrenzanalyse](#) (Bella 1995) mit unterschiedlichen Parametereinstellungen, derzeit u.a. implementiert in [COSMAS II](#).¹ Den Ausgangspunkt bildeten Kookkurrenzprofile von Einwortlemmata

¹ Die Artikelarbeit basiert auf dem Deutschen Referenzkorpus DeReKo, der mit über 3,2 Milliarden Wörtern weltweit größten linguistisch motivierten Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit (<http://www.ids-mannheim.de/kl/projekte/korpora/>, Stand: 25.6.2008).

des [*ellexiko-Demonstrationswortschatzes*](#) (alle Kookkurrenzpartner zum Nomen *Mobilität*), die vom [UWV-Projekt](#) linguistisch geprüft wurden.² Das obligatorische Kriterium für alle Lemmakandidaten stellte die statistisch auffällige Affinität zwischen den Basis-komponenten einer Wortverbindung dar. Diese Herangehensweise ist neuartig, da in der lexikografischen Praxis die Entscheidung für die Aufnahme in ein Wörterbuch stattdessen zumeist auf der Basis bereits vorhandener Idiomsammlungen und Nachschlagewerke getroffen wird. Dies führt in vielen Fällen zu einer Tradierung bestimmter phraseologischer Kernbestände, die nicht immer dem aktuellen Usus entsprechen. Mithilfe der statistischen Kookkurrenzanalyse kann man zwar nicht alle usuellen Wortverbindungen finden, die potenziell zum Bestand einer Sprachgemeinschaft gehören, da es viele gibt, die zu selten belegt und daher statistisch nicht auffällig sind. Allerdings lässt sich in einem Umkehrschluss sagen, dass diejenigen Wortverbindungen, die mithilfe der automatischen Methoden selektiert und als usuelle Wortverbindungen linguistisch qualifiziert wurden, eine gesicherte Festigkeit im Gebrauch aufweisen. Über die statistische Affinität hinaus mussten die ermittelten Kookkurrenzen eine Übersummativität aufweisen. Diese übersummativ Qualität ist jedoch nicht in einem engen Sinne als semantische Idiomatizität zu verstehen, sondern in einem weiten Sinne als ein „Mehr an Gebrauchsbedeutung“ des gesamten Mehrwortausdrucks im Vergleich zu den Bedeutungen ihrer Komponenten. Diese Wortverbindungen fungieren damit als kommunikative Einheiten.

25 Mehrwort-Musterartikel zu Demonstrationszwecken. Die Erarbeitung der 25 *Mehrwort-Musterartikel* (auf den OWID-Seiten eingeordnet unter [Mehrwortartikel vermischt](#)) sollte vor allem dazu dienen, die vom Projekt UWV entwickelte Artikelstruktur für die korpusbasierte Erarbeitung elektronischer Mehrwortartikel in der Praxis zu testen. Es wurden Wortverbindungen ausgewählt, die möglichst alle relevanten Mehrwortklassen, so wie sie die Mehrwortforschung bisher entwickelt hat (vgl. Burger 2007, Fleischer 1997), und alle in der Mehrwortartikelstruktur enthaltenen Angaben abdecken, z.B. verschiedene Idiomatizitätsgrade, Wortverbindungen mit mehreren Lesarten, verschiedene Mehrwortklassen wie Zwillingsformel oder Sprichwort, verschiedene Phrasentypen wie Präpositionalphrase oder satzwertige Wortverbindung oder verschiedene Realisierungen kommunikativer Funktionen wie Warnung oder Kritik (vgl. [verwendete Attributlisten](#)).³

100 Wortverbindungsartikel zu den Bezugsadjektiven *blind, geistig, gesund, normal, sanft*. In einem nächsten Arbeitsschritt wurden Wortverbindungen beschrieben, die sich aus der Auswertung der Kookkurrenzprofile ausgewählter Bezugswörter ergaben (auf den OWID-Portalseiten eingeordnet unter [Wortverbindungen zu](#)). Hier wurden die relevanten Informationen zum typischen Gebrauch im Korpus auf einige wenige Angaben konzentriert. Die Entscheidung für adjektivische Bezugswörter resultierte aus der Randständigkeit, die für diese Wortklasse in der bisherigen Mehrwortforschung noch festzustellen ist. Die Kookkurrenzanalyse ermöglicht die systematische Erfassung von usuellen Umgebungsmustern eines Adjektivs, die auf der Basis von Intuition nur schwer

² Zum Konzept der usuellen Wortverbindungen vgl. Steyer (2000), Steyer (i. Dr.). Zur linguistischen Interpretation und lexikografischen Anwendung von Korpusanalysemethoden vgl. u.a. Steyer (2004a; b), Steyer/Lauer (2007), Belica/Steyer (i.Dr.), Brunner/Steyer (2007). Zu aktuellen Tendenzen der phraseologischen Forschung und Phraseografie auch in Bezug auf korpusbasierte Empirie vgl. auch Hallsteinsdóttir/Farø (Hg.) (2006), Burger/ Dobrovolskij/Kühn/Norrick (Hg.) (2007).

³ Zu einzelnen Aspekten der spezifischen Mikrostruktur dieses Artikeltyps verweisen wir auf den online verfügbaren [Informationstext der OWID-Portalseite 'Feste Wortverbindungen'](#).

generierbar sind. Ein Ergebnis dieses Teilprojektes war die Erkenntnis, dass auch viele transparent und zunächst nicht idiomatisch scheinende Wortverbindungen, vor allem bei Adjektiv-Nomen-Verbindungen (z.B. im Feld von *geistig* und *normal*), eine Festigkeit in der Verwendungsspezifik aufweisen und demzufolge als zusammenhängende Einheiten zu beschreiben sind. Dies führte zu einer primär pragmatisch motivierten Form von Gebrauchsbeschreibungen, die nicht immer identisch mit dem traditionellen Konzept von 'Bedeutungsparaphrasen' ist. Sie rückt die typischen kommunikativen Aspekte der jeweiligen Wortverbindung ins Zentrum.

2. Zur Mikrostruktur der Mehrwortartikel

2.1 Lesartenübergreifende Angaben

Artikelstichwort. Nach der Identifizierung einer Wortverbindung stellt die Entscheidung für das Artikelstichwort in einem Mehrwortartikel eine weitere schwierige lexikografische Aufgabe dar. Die vielfältigen Formen des Artikelstichworts für usuelle Wortverbindungen erfordern eine vorgelagerte Reflexion und vor allem systematische Korpusuntersuchungen (insbesondere KWIC-Analysen), um den festen Kern der Wortverbindung zu bestimmen. Dazu sind in den meisten Fällen mehrere Recherchevorgänge mit verschiedenen Parametereinstellungen notwendig, um wirklich alle Fälle in den Blick zu bekommen. Ein Syntagma kann dann als fester Kern gelten, wenn kein weiteres Element getilgt oder ersetzt werden kann, ohne damit die Bedeutung der Wortverbindung zu verändern. Zugleich sollte aber dieses Muster ohne weitere Ergänzungen verstehbar sein. Beispiele für solche Lemmazeichengestaltungsangaben sind:

- [billiger Jakob](#)
- [Angst vor der eigenen Courage](#)
- [jenseits von Gut und Böse](#)
- [hoch und heilig](#)
- [an die große Glocke hängen](#)
- [im eigenen Saft schmoren](#)
- [auf Dauer](#)
- [Dienst nach Vorschrift](#)

In vielen Fällen ist jedoch eine solche korrekte und zugleich leicht verständliche Benennung nicht realisierbar, da auch der semantische Kern vielfachen Variationen unterliegt. Hier müsste man eigentlich kategoriale Angaben verwenden. So ist etwa bei der Wortverbindung *da liegt der Hund begraben* (im Sinne von: 'dies ist der (tatsächliche) Grund / der Kern / die Ursache des Problems') nur das Partizip *begraben* fix (z.B. *da / hier / im Verwaltungshaushalt / in Nordhausen liegt / ist der Hund / Hase / die Krux begraben*). Die invariante Form lautet also: [DEIKT. ELEMENT / LOC + VERB + NOMEN + *begraben*]. Der Lexikograf steht vor einem Dilemma: Wenn das Artikelstichwort nur den möglichst unveränderbaren (obligatorischen) Kern einer Wortverbindung abbilden soll, kann das zu sehr abstrakten Darstellungen führen (siehe Beispiel). Dies steht aber im Widerspruch zu einer nutzerfreundlichen Form, die in einem lexikografischen Artikel unbedingt anzustreben ist.⁴ Deshalb wurde in den Artikeln in solchen Fällen das im Korpus häufigste Verwendungsmuster als Artikelstichwort angesetzt. Aus

⁴ Dobrovol'skij spricht in diesem Zusammenhang vom Widerspruch zwischen der lexikografischen Universalität der Nennform, die aber oft befremdlich wirke, und der psychologisch realen und damit benutzerfreundlicheren Aussageinformation, die aber nie alle Realisationsformen abdecken könne (Dobrovol'skij 1993, S. 59).

diesem Grund wurde etwa bei der Wortverbindung [eigen / eigenes Fleisch und Blut](#) die eigentlich ältere Form *eigen Fleisch und Blut* als Artikelstichwort angesetzt und bei der Wortverbindung [reden / sprechen, wie einem der Schnabel gewachsen ist](#) nur das Verb *reden* ins Artikelstichwort aufgenommen, da dieses am häufigsten in der Wortverbindung vorkommt. Derartige lexikografische Entscheidungen zu einem Artikelstichwort wurden jedoch immer mit einem Kommentar versehen, um deutlich zu machen, dass es sich nur um die typischste Form handelt. Darüber hinaus werden im Korpus vorgefundene Varianten der Wortverbindung, wenn sie eine auffällige Signifikanz aufweisen, gleichfalls angegeben.

Frequenzangabe (25 Mehrwort-Musterartikel). Die Angabe zur Vorkommenshäufigkeit basiert auf einer konkreten Suchanfrage, die die Wortverbindung in ihrem lexikalischen Kern abdeckt. Es ist zu beachten, dass die angegebene Zahl nur etwas über die Häufigkeit der Wortverbindungen auf der Basis dieser Suchanfrage und des zum Analysezeitpunkt zugrunde gelegten Korpus aussagt. Bei einer anderen Suchanfrage oder einer anderen Korpuszusammenstellung kann sich die Frequenz verändern. Ein Beispiel für diese Angabe findet sich im Artikel [an die große Glocke hängen](#): „Die Wortverbindung *an die große Glocke hängen* hat über 850 Vorkommen in allen IDS-Korpora der geschriebenen Sprache. (Stand: Oktober 2005)“.

Suchanfrage für Recherche im Korpus (100 Wortverbindungsartikel). Diese Angabe dient dazu, dem Nutzer den Zugang zu allen Vorkommen einer Wortverbindung auf dem jeweils aktuellsten Stand der öffentlich zugänglichen IDS-Korpora der geschriebenen Sprache zu ermöglichen, indem er die vorformulierten Suchanfragen nach dem Aufrufen von COSMAS II in das dortige Suchfenster kopiert. Die Formulierung von Suchanfragen für Wortverbindungen ist keineswegs trivial, da sie bereits auf Hypothesenbildungen basiert, das Beherrschen der Suchanfragesprache voraussetzt und oft auch ein gewisses Vorwissen in Bezug auf die Aussagekraft von Korpusanalysemethoden verlangt. Daher stellen Wortverbindungsanfragen an das Korpus nicht selten eine Barriere für Nutzer dar. In der Regel sind zwei Möglichkeiten – entsprechend kommentiert – angegeben: erstens eine Suchanfrage, mit der man relativ gesicherte Resultate erhält, die aber aufgrund ihres hohen Fokussierungsgrades nicht das ganze Spektrum der Variationen und Modifikationen erfassen muss, und zweitens eine Suchanfrage, die sehr viel weiter gefasst ist und daher diese produktiven Abwandlungsphänomene besser erfasst. Hier muss der Nutzer allerdings damit rechnen, dass er auch Resultate erhält, die überhaupt nicht die gesuchte Wortverbindung abbilden. Durch die dynamische Suche erhält der Nutzer darüber hinaus die Vorkommenshäufigkeit einer Wortverbindung auf dem aktuellsten Stand der öffentlich zugänglichen IDS-Korpora der geschriebenen Sprache. Ein Beispiel für diese Angabe findet sich im Artikel [im Vollbesitz seiner geistigen Kräfte](#): „(&Vollbesitz /+w2 &geistig) /+w1 &Kraft“.

Basiskomponente(n). Diese Angabe beinhaltet die Auflistung aller autosemantischen Komponenten einer Wortverbindung. Sie dient dazu, die Komponenten der Mehrwortartikel mit den jeweiligen Einwortlemmata im *lexiko*-Wörterbuch zu verbinden. So verweisen z.B. die Basiselemente im Artikel [kapitaler Hirsch](#) auf die *lexiko*-Artikel [kapital](#) und [Hirsch](#). Des Weiteren ist so ein Zugriff auf die beschriebenen Wortverbindungen über alle ihre Basiskomponenten möglich. So werden z.B. bei einer [Suche im gesamten OWID-Portal nach gesund](#) auch alle Wortverbindungsartikel, die die Basiskomponente *gesund* enthalten, geliefert.

Eine interne Annotation stellt die Angabe zur Unikalität dar, also zu solchen sprachlichen Einheiten, die keine freie Bedeutung außerhalb dieser Wortverbindung besitzen (z.B. das Element *frank* in der Wortverbindung [frank und frei](#)).

Variation(en) und Modifikation(en). Die Unterscheidung zwischen Variation und Modifikation basiert auf dem Konzept von Burger (2007). Unter Variation werden usuell Varianten einer Grundform verstanden, die das konzeptuelle Schema einer Wortverbindung nicht antasten (vgl. z.B. die Artikel [eigen Fleisch und Blut](#), [reden, wie einem der Schnabel gewachsen ist](#)). Im Gegensatz dazu sind Modifikationen okkasionelle Abwandlungen einer Wortverbindung, die im Korpus sehr zahlreich zu finden sind. Diese Modifikationen wurden in einige Artikel testweise aufgenommen (vgl. z.B. [Tue Gutes und rede darüber](#)). Sie veranschaulichen den sprachlichen Reichtum, wie man ihn in den Korpora vorfindet, sehr gut. Es wird aber künftig noch zu überprüfen sein, welche Kriterien anzusetzen sind, um die lexikografische Relevanz von Variationen zu bewerten. Eine gewisse Musterhaftigkeit muss auch in den freieren Modifikationen erkennbar sein.

2.1 Lesartenspezifische Angaben

Gebrauchsbeschreibung. Eine korpusbasierte Beschreibung von Wortverbindungen ist in besonderer Weise an den usuellen Gebrauch gebunden. Wortverbindungen lassen sich als Mikrotexte mit Oberflächenstrukturen, propositionalem Gehalt und illokutiven Funktionen auffassen. Eine pragmatisch ausgerichtete Gebrauchsbeschreibung darf sich deshalb nicht auf eine eng semantische Paraphrase beschränken, sondern muss die Erfassung und Darstellung des kommunikativen Sinns einer Wortverbindung und ihrer Funktionen in der Kommunikation ebenso berücksichtigen. Wenn man beispielsweise nur die reine Bedeutung von *Dienst nach Vorschrift* mit ‘seinen Beruf streng nach den Vorgaben ausüben’ umschreiben würde, hätte man den kommunikativen Sinn und den typischen Gebrauch dieser Wortverbindung nicht erfasst. Erst durch den folgenden Kommentar wird der eigentliche Verwendungskontext deutlich: „Dabei kann sowohl gemeint sein, dass die Ausführung der Tätigkeit sehr genau geschieht, als auch, dass über die vereinbarten, vertraglich geregelten zu erbringenden Leistungen hinaus nicht mehr (zeitlich oder inhaltlich) gearbeitet wird. In der Regel wird *Dienst nach Vorschrift* als Druckmittel eingesetzt, oft anstelle eines Streiks.“ (vgl. [Dienst nach Vorschrift](#)).⁵ Das bedeutet, dass in sehr vielen Paraphrasen solche Formulierungen wie *Mit ... drücken Sprecher aus, dass / bewerten Sprecher ...* bevorzugt werden, wie etwa die Gebrauchsparaphrasen in den Artikeln [Kapital verspielen](#) und [recht und billig](#) zeigen.

Eine pragmatische Perspektive einzunehmen, bedeutet, folgende Fragen zu stellen:⁶

- Wer verwendet eine Wortverbindung wem gegenüber in welcher Situation mit welcher Intention?

⁵ Auch Burger diskutiert die unzureichende Aussagekraft einer Paraphrase, die auf den so genannten Bedeutungskern reduziert wird und nicht das damit einhergehende illokutive Potenzial berücksichtigt, u. a. anhand des usuellen Aufforderungsmodus des Phraseologismus „vor seiner eigenen Tür kehren“ (2007, S. 202f).

⁶ Es geht hier jedoch weniger um die in der phraseologischen Forschung relativ gut beschriebenen kommunikativen Formeln, Routineformeln usw., sondern vielmehr um sprachliche Einheiten, die nicht durch ein illokutives Potenzial direkt bestimmt sind. Einen guten Überblick über pragmatische Beschreibungsansätze in der Phraseologie bietet Filatkina (2007).

- Welche Verwendungsbedingungen und -beschränkungen gibt es für den Gebrauch einer Wortverbindung?
- Was geben Sprecher mit dem Gebrauch einer Wortverbindung über die eng an die Bedeutung gebundene Funktion hinaus noch zu verstehen?
- Worin besteht also der kommunikative Sinn einer Wortverbindung?

Die wichtigsten methodischen Zugänge sind heuristische Kookkurrenzfelder, KWIC- und Beleganalysen, wie am Ausschnitt aus der Umgebungsanalyse von *billiger Jakob* gut sichtbar wird:

Total	Anzahl	Autofokus	LLR	Kookkurrenzen	syntagmatische Muster
		von bis			
5	5	1	1	524 Jakobs Image	100% Image des billigen Jakobs
27	22	1	1	Jakobs	54% des die billigen Jakobs
35	8	-2	-2	188 nicht spielen	75% nicht ... den billigen Jakob ... SPÖ spielen
46	11	-2	-2	nicht sind	54% Wir sind [...] nicht der billige Jakob aber
55	9	-2	-2	nicht werden	44% nicht zum billigen Jakob zu werden
93	38	-2	-2	nicht	28% das Sojus nicht als den billigen Jakob an
96	3	-2	-2	96 Image	33% Wir ... Image ... Billigen
101	5	2	5	76 spielen	80% den billigen Jakob zu spielen

In den entsprechenden KWICs werden Muster deutlich, die Aufschluss über eine mit dieser Wortverbindung einhergehende negative Konnotation geben:

Jakobs Image :

L99 erwartet BM/dpa/StM Frankfurt/M. - Praktiker will das Image des "**billigen Jakobs**" unter den Baumarkt-Ketten loswerden. Der neue
 E97 Um vom auf die Dauer ohnehin unvorteilhaften Image des **billigen Jakobs** wegzukommen, besinnt sich das Fürstentum auf seine
 E98 Sortiment, die anderswo schwer erhältlich sind. Das Image eines **billigen Jakobs** für Wald-und-Wiesen-Bands oder zweitrangige Sänger und
 R99 radikal umkrempeln. Vor allem soll die Metro-Tochter das Image des "**billigen Jakobs**" loswerden. "Dauerhaft kann die Kundenbindung nur
 P98 führt er Titel, die schwer erhältlich sind. Das Image des **billigen Jakobs** für Wald-und-Wiesen-Bands oder zweitrangige Sänger lässt

nicht spielen :

B99 Vorstandschef des Fernsprechriesen verkündet, künftig nicht den "**billigen Jakob**" spielen zu wollen. Davon ist die Telekom nun abgerückt.
 T91 so groß geworden zu sein, daß sie sich auch nicht mehr scheut, den **billigen Jakob** zu spielen. So wird die soziale Fallstudie als ein »Stück
 I99 gereizt. Er werde nach der Nationalratswahl sicher nicht den "**billigen Jakob**" spielen, vertraute Schüssel dem Wiener Standard an: Es
 A98 Gebührenerhebung beim Postkonto. «Wir können nicht länger den **billigen Jakob** spielen und Top-Leistungen anbieten, ohne etwas dabei zu
 I99 gestellt würden. Schüssel wolle zwar nicht den "**billigen Jakob**" gegenüber der SPÖ spielen, wohl aber "der billige
 R99 Wettbewerber schlichtweg unterschätzt. "Wir wollen nicht den **billigen Jakob** spielen", lautete sein Credo. Mittlerweile ist das
 M98 mit insgesamt 20 000 Mitarbeitern dürften nicht länger den "**billigen Jakob**" spielen, warnte Hariegel. Sie müßten mehr auf Qualität,
 K99 Koalition. Der VP-Obmann hatte gemeint, man wolle nicht weiter den "**billigen Jakob** für die SPÖ" spielen. Ursula Stenzel billigte Rudas in

nicht sind :

P94 Angebot, kommen aber mit keinen Schleuderpreisen. Der **billige Jakob** sind wir nicht, sondern wollen Kunden möglichst viel Auto
 HMP05 nicht allein in punkto Preis absetzen: "Wir sind nicht der **billige Jakob**, sondern bieten die beste Mischung aus Preis und
 T93 Kröte zu schlucken. Und fügt trotz hinzu: "Wir sind nicht der **billige Jakob.**" Genau den glauben manche Sozialdemokraten in Markus
 T89 der SV Waldhof Mannheim, fühlte sich unwohl: "Wir sind nicht der **billige Jakob**, aber wir können wohl nichts dagegen machen", erklärte
 T98 müßten auch unangenehme Wahrheiten offen gesagt werden. "**Billiger Jakob** sind die anderen, nicht wir." Die Union habe die bessere

T01 des Angebots seines Angestellten reiben: "Wir sind zwar nicht der **billige Jakob**, aber Qualität macht sich eben bezahlt."
 B98 sein, es ist nicht so bequem mit der Union, nicht so billig. **Billiger Jakob** sind die anderen, nicht wir. So wird man bei jedem Thema
 I98 keine "Wegelagerer" am Brenner seien "wir sind aber auch nicht der **billige Jakob** im Transitverkehr". Der Tiroler SPÖ-Mandatar Erwin
 N98 wo er die Linie "Habsburg" mitentwickelte. "Wir sind nicht der **billige Jakob**, aber wir haben eben andere Kalkulationsvorstellungen als
 P93 Kasse." Denn, darüber sind sich Mayr und Seitlinger einig: "Der **billige Jakob** sind wir nicht. "
 B00 Vorstellungen zum Kaufpreis". "Wir sind nicht der **billige Jakob**", sagt er. Schließlich habe sich die Mega an die WoGeHe

nicht werden:

T02 Die Stiftung warnt, Engagement dürfe nicht als „**billiger Jakob** des Sozialstaats“ missbraucht werden. Was für
 I97 der Tiroler VP-NR Dieter Lukesch. Denn Tirol dürfe nicht zum **billigen Jakob** des Transitverkehrs werden. Konkret verlangte er
 R98 und wettbewerbsgerechte Preise" bieten, dabei aber nicht zum "**billigen Jakob**" werden. In anderen Ländern habe sich gezeigt, daß die
 Z98 Telekom-Chef Ron Sommer sich fest vorgenommen hatte, nicht zum "**billigen Jakob**" zu werden, ändert er nun abrupt seinen Kurs: Mit einer
 Z02 Auslandsgesprächen. Und obwohl Sommer angekündigt hatte, nicht zum "**billigen Jakob**" zu werden, blieb ihm schließlich doch nichts anderes,
 DPA06 sinken. Ricke beteuerte zugleich, dass der Konzern nicht der «**billige Jakob**» werden wolle. Wichtig seien auch Qualität und Service.
 DPA06 Ricke an. Zugleich beteuerte er, dass der Konzern nicht der «**billige Jakob**» werden wolle. Wichtig seien auch Qualität und Service.
 DPA06 Trotz der Preisabschläge wolle der Bonner Konzern nicht der «**billige Jakob**» werden, betont Ricke. «Den Kampf um die Kunden werden
 F97 Die Werkstätten für Behinderte dürften nicht als der "**billige Jakob** des Arbeitsmarkts" betrachtet werden. Sie stünden unter

Diese Erkenntnisse können lexikografisch umgesetzt werden – z.B. in Gebrauchsbeschreibungen wie im Artikel [billiger Jakob](#):

Mit *billiger Jakob* wird jemand bezeichnet, der ein Produkt oder eine Leistung zu niedrigen Preisen, oft unter Verzicht auf Anspruch und Qualität, anbietet. *Billiger Jakob* wird auch im Sinne eines 'Sich-unter-Wert-Verkaufens' oder eines 'nützlichen Idioten' verwendet. In den meisten Fällen ist damit eine negative Bewertung verbunden.

Wertungsaspekt: Mit *billiger Jakob* verbinden Sprecher eine negative Bewertung, die sich zum Teil darin ausdrückt, dass eine solche Zuschreibung zurückgewiesen wird.

Gerade in Bezug auf den pragmatischen Mehrwert, der mit einer sprachlichen Einheit verbunden sein kann, ermöglicht die Korpusanalyse gesicherte Aussagen, die über die individuelle Einzelwahrnehmung eines Sprechers hinausgehen. In der Lexikologie und Lexikografie wird u.E. bisher zu wenig deutlich gemacht, dass auch funktionale Aspekte einer abgestuften Festigkeit unterliegen, was unter Umständen auch zu Überinterpretationen und unzulässigen Generalisierungen führen kann. Deshalb unterscheiden wir zunächst prinzipiell zwischen zwei Perspektiven:

- a) verwendungsspezifische Komponenten, die konstitutiv für die Bedeutung einer Wortverbindung sind (**bedeutungsinhärente Gebrauchsspezifik**);
- b) verwendungsspezifische Komponenten, die eine Wortverbindung kontextuell und situativ spezifizieren bzw. ausdifferenzieren (**situative Gebrauchsspezifik**).⁷

⁷ Zentrale Aspekte unseres Ansatzes für pragmatisch motivierte lexikografische Angaben sind auch in die Artikelstruktur des *ellexiko*-Wörterbuchs für die Einzelwortbeschreibung eingeflossen (vgl. dazu auch Haß (Hg.) 2005, S. 265).

Bedeutungsinhärente Gebrauchsspezifika meint die Gebundenheit der Bedeutung einer Wortverbindung an eine oder mehrere ganz bestimmte Gebrauchsbedingungen, also ihre funktionale Festigkeit. Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf diese Gebrauchsbedingungen einzugehen. Bedeutungsinhärente Gebrauchsspezifika kann alle Aspekte der Verwendung betreffen und stellt ein obligatorisches Merkmal dar, wie folgende Beispiele zeigen:

Kein Grund zur Panik!: Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf den damit einhergehenden Sprechakt der Beruhigung einzugehen.

geistiger Brandstifter: Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf die negative Konnotation einzugehen.

hoch und heilig: Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf die emotionalisierende Verstärkungsfunktion der Zwillingsformel in Bezug auf ein Verb wie *versprechen* einzugehen.

eingeschränktes Gesichtsfeld: Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf die Domänengebundenheit bzw. die Fachsprachlichkeit dieses Ausdrucks (Augenheilkunde) einzugehen.

sanft entschlafen: Man kann die Bedeutung der Wortverbindung nicht erklären, ohne auf die Textsortengebundenheit (Todesanzeigen, Grabsteininschriften) einzugehen.

Somit können potenziell alle Aspekte der Verwendung einer Wortverbindung eine zentrale Stellung für die Erklärung der Bedeutung einnehmen. In unseren Artikeln betrifft dies vor allem: den Wertungsaspekt, die Funktion in der Kommunikation, das Sachgebiet, die Gruppensprache, die Textspezifika, die Medialität.

Die kontextuell-situative Gebrauchsspezifika betrifft demgegenüber typische Einstellungen, Situationen usw., die zur Bedeutung einer Wortverbindung hinzutreten und die durch die Korpusanalyse empirisch identifiziert werden können. Diese sind aber nicht obligatorisch für den Bedeutungskern. Bei diesen situativen Faktoren, die den Gebrauch einer Wortverbindung kontextuell spezifizieren, geht es nicht um das Sprachverhalten individueller Sprecher (etwa in einer aktuellen Face-to-Face-Situation) oder um konkrete Interaktionen (Gespräch, Einzeltext usw.). Es geht vielmehr um die Verwendungsmuster kollektiver Sprecher, deren Sprachverhalten sich in den Korpora manifestiert. Damit sind für uns auch hier nur jene Phänomene von Interesse, denen man aufgrund ihrer Rekurrenz im Korpus Musterhaftigkeit zusprechen kann.

In den 25 Musterartikeln wurden diese zwei verschiedenen verwendungsspezifischen Aspekte intern noch unterschiedlich ausgezeichnet und in verschiedenen lexikografischen Angaben verarbeitet. In den 100 Wortverbindungsartikeln floss diese Unterscheidung in die entsprechende Formulierung der Gebrauchsbeschreibung ein. Situative Aspekte der Verwendungsspezifika werden in den Artikeln durch einschränkende Adverbien wie *häufig*, *auch*, *überwiegend* oder *zumeist* oder auch typografisch durch einen neuen Absatz markiert. Diese Herangehensweise illustrieren die folgenden drei Beispiele:

Angst vor der eigenen Courage

Mit *Angst vor der eigenen Courage* bewerten Sprecher Absichten, Handlungen bzw. Folgen von Handlungen **zumeist** negativ.

geistige Höhenflüge

In den IDS-Korpora der geschriebenen Sprache werden *geistige Höhenflüge* auf **überwiegend** ironische Weise thematisiert, entweder um geistige Leistungen anderer gerade als unterdurchschnitt-

lich zu bewerten oder fremde Erwartungshaltungen gegenüber der eigenen Leistung zu bremsen und dem Adressaten zu vermitteln, dass er keine übertriebenen Erwartungshaltungen hegen dürfe.

der ganz normale Wahnsinn

Mit *der ganz normale Wahnsinn* werden Zustände oder Situationen, die extrem oder zugespitzt zu sein scheinen, doch als alltäglich bzw. völlig normal bewertet.

Häufig soll mit dieser Bewertung auch eine gewisse Gelassenheit ausgedrückt werden, verbunden mit der Ermunterung, die Dinge soweit möglich hinzunehmen, wie sie eben mal sind.

Auch die Beschreibung der situativen Faktoren des typischen Gebrauchs einer Wortverbindung kann sich auf sehr verschiedene Aspekte der Verwendungsspezifik beziehen.

Zur Erfassung und Beschreibung der situativen Gebrauchsspezifik:

Wertungsaspekt.⁸ Hier geht es um die Erfassung und Beschreibung typischer Einstellungsmuster, die Sprecher mit der Wortverbindung noch transportieren und die durch die Korpusanalyse als rekurrente Muster erkennbar werden, z.B.

geistiger Vater: Die Wortverbindung ist im Kern ohne auffällige Konnotation. Erst durch bestimmte Verwendungskontexte wird ihr eine negative Bewertung situativ zugeschrieben, z.B. wenn die Idee / Bewegung / Richtung, für die jemand als Urheber gilt, im aktuellen Diskurs negativ bewertet wird. Dies lässt sich anhand von typischen Kontextmustern wie *geistiger Vater des Fundamentalismus, des Krieges, von Auschwitz, der Terroristen* aus den Korpusbefunden identifizieren.

geistige Frische: Dem an sich schon positiv bewerteten Sachverhalt, dass jemand noch gute geistige Fähigkeiten aufweist, wird in bestimmten Verwendungskontexten eine weitere Bewertung im Sinne von Erstaunen oder Bewundern hinzugefügt: *in erstaunlicher / beneidenswerter geistiger Frische*.

gesundes Selbstbewusstsein: Mit der Wortverbindung verbinden Sprecher häufig eine Bewertung, die aber in der Regel vage gelassen wird. Die Korpusbefunde legen sowohl negative als auch positive Bewertungen des Sachverhalts, dass jemand ausgeprägt selbstbewusst ist, nahe. Die Interpretation hängt dann letztlich vom jeweiligen Wertekanon des Rezipienten ab.

Funktion in der Kommunikation. Sprecher vollziehen mit Wortverbindungen oft gerichtete Sprachhandlungen, z.B. einen Befehl, eine Bitte, eine Drohung, eine Warnung. Beispiele für solche in Wortverbindungen realisierten kommunikativen Funktionen sind: *Kapital verspielen* (wird oft als Warnung formuliert: *darf nicht verspielt werden*); *Tue Gutes und rede darüber* (ironische Zurückweisung / Kritik).

Sachgebiet, Gruppensprache, Textspezifik, Medialität. Diese Informationen sind bezogen auf auffällige Gebundenheiten im Gebrauch der Wortverbindung in Bezug auf die kommunikative Situation im weitesten Sinne. Sie können aus den Korpusbefunden natürlich nur vermittelt generiert werden, da alle diese Merkmale über schriftliche, in den IDS-Korpora vor allem massenmedial verbreitete Texte transportiert werden. Indikatoren für derartige Gebrauchsspezifika sind jedoch eine auffällig häufige metasprachliche Kommentierung, z.B. zum fachsprachlichen Status oder zur eigentlichen fach-

⁸ Zu linguistischen Aspekten von 'Bewertung' und entsprechenden sprachlichen Realisierungen vgl. Fries (1991).

sprachlichen Herkunft einer Wortverbindung, oder auffällig häufige Markierungen von Mündlichkeit. Informationen zum Sachgebiet beziehen sich auf ein auffälliges Vorkommen in einer oder mehreren Domänen im Korpus, wie beispielsweise Wirtschaft, Sport, Finanzwesen oder Politik, z.B. [blinder Fleck](#) (Augenheilkunde); [gesunde Härte](#) (Sport).

Die gruppensprachliche Information bezieht sich auf eine im Korpus auffällige Thematisierung von Vorkommen einer Wortverbindung im Sprachgebrauch bestimmter Sprechergruppen, z.B. [kapitaler Hirsch](#) (Jägersprache).

Die textsortenbezogenen Informationen dienen zur Markierung der typischen textsorten- bzw. typengebundenen Verwendung einer Wortverbindung. Eine Wortverbindung kommt im Korpus auffällig häufig in einer oder mehreren Textsorten vor, z.B. [geistige Frische](#) (Geburtstagsanzeigen / Glückwünsche).

Medialität lässt sich in schriftlichen Korpora nur insoweit erfassen, als dass die Vorkommen im Korpus auf auffallend häufige Markierungen von Mündlichkeit überprüft werden können (z.B. Anführungszeichen, Interjektionen) oder die Zugehörigkeit zu einer bestimmten medialen Kommunikationsform sehr explizit und wiederholt metasprachlich kommentiert wird, z.B. [noch ganz gesund sein](#); [Kein Grund zur Panik!](#).

Typische Kontextmuster. Die Kontextmusterangabe dokumentiert die typischen, durch automatische Umgebungsanalysen der jeweiligen Phrase gewonnenen, syntagmatischen Einbettungen der Wortverbindung. Die Kontextmuster enthalten keine systematischen Aussagen zur Bildbarkeit von Wortverbindungen, also präskriptive Angaben von Phrasenstrukturen oder Restriktionen (z.B. in Form von regelhaften Beispielphrasen). Das Spektrum typischer kontextueller Muster, so wie sie in der natürlichsprachigen Umgebung verwendet werden, dient vielmehr als Illustration des typischen Gebrauchs der Wortverbindung im Korpus. Die einzige Abstraktion wird dadurch getroffen, dass im Regelfall die übliche Nennform (z.B. bei Nomen: Nominativ Singular) angesetzt wird, es sein denn, es gibt Restriktionen innerhalb der Flexionsparadigmen oder Satzpositionen (also z.B. nur Pluralverwendungen, nur Genitiv, nur partizipial, nur prädikativ, nur nebengeordnet usw.). Diese Orientierung an den Resultaten der automatischen Umgebungsanalyse bedeutet in manchen Fällen, dass die Kontextmuster gemessen an regelhaften grammatischen Phrasen oder Sätzen unvollständig erscheinen. Damit soll dokumentiert werden, dass genau diese sprachlichen Komponenten als eine feste sprachliche Einheit erkannt und selektiert wurden. Zu jedem Kontextmuster werden Belege zur Illustration angegeben, sodass man die Kontextmuster quasi als Label für die eigentlichen Volltextstellen verstehen kann. Die Artikelstruktur bietet die Möglichkeit, Kontextmuster in Gruppen zusammenzufassen. Auch hier können die Kriterien für eine solche Kontextmustergruppe alle Ebenen des Sprachsystems betreffen; sie können aber auch andere Kriterien des Gebrauchs folgen, z.B. thematisch-diskursiven Zusammenhängen.

[Kapital verspielen](#)

ein / ihr / ... Kapital verspielen

viel Kapital verspielen

Kapital darf nicht verspielt werden

Gefahr, Kapital zu verspielen

politisches Kapital verspielen

wichtiges Kapital verspielen

wertvolles Kapital verspielen

Kapital leichtsinnig verspielen

Kapital leichtfertig verspielen

Kapital leicht verspielen

Beziehungen zu anderen Wörtern und Wortverbindungen. In dieser Angabe werden Verwendungszusammenhänge zwischen Wortverbindungen oder Beziehungen im Hinblick auf Wortklassen thematisiert, die durch vergleichende Kookkurrenzanalysen hervorgetreten sind. Diese Verwendungszusammenhänge lassen sich in den meisten Fällen nicht mit tradierten Kategorien wie etwa den paradigmatischen Relationen fassen. So gibt es aus korpusbasierter Sicht eigentlich nie eine echte Synonymie. Durch die Möglichkeit der feinen Kontextspezifikation treten auch distinktive Eigenschaften und Gebrauchsbesonderheiten einer Wortverbindung im Kontrast zu einer anderen deutlich zutage.

Folgende Beispielartikel enthalten Verweise zu anderen Wortverbindungen, die das Spektrum solcher Wortverbindungsfelder wenigstens andeuten:

[blind gegenüber](#) – [auf einem Auge blind sein](#) – [sich blind und taub stellen](#) – [...] [macht blind](#) – [Liebe macht blind](#)

[geistiger Brandstifter](#) – [geistiger Vater](#) – [geistiger Wegbereiter](#)

2.3 Kommentare

Neben den einzelnen lexikografischen Angaben nehmen auch frei formulierte Kommentare einen wichtigen Platz ein. Diese („Hintergrund“-)Kommentare können sowohl spezifische Aspekte des Gebrauchs betreffen als auch Fachkommentare darstellen, etwa zu einer speziellen lexikografischen Entscheidung, zum Gebrauch einzelner Basiskomponenten oder zu Zusammenhängen mit Inhalten anderer Wörterbücher (in OWID oder extern). Eine feinere Unterteilung, beispielsweise in Fachkommentare und Kommentare zum Gebrauch, wie sie noch bei den 25 Musterartikeln vorgenommen wurde, hat sich nicht bewährt und wurde für die 100 Artikel verworfen. Beispiele für Kommentierungen finden sich in allen Teilen der Mikrostruktur:

Kommentare zu „Artikelstichwort“

Mehrwortstatus:

[Blind Date](#): *Blind Date* wurde auf Grund des signifikanten Kovorkommens als Stichwort angesetzt: *Date* ist ein statistisch auffälliger Kookkurrenzpartner von *blind*. Im [Neologismenwörterbuch \(NWB\)](#) wurde ‘[Blind Date](#)’ dagegen als Einwortlemma angesetzt, da dort auch Anglizismen in Form von Wortgruppen als Einzelwörter behandelt werden.

Schreibung:

[Blind Date](#): *Blind Date* ist mit ca. 500 Treffern die häufigste Schreibvariante in den IDS-Korpora der geschriebenen Sprache. Weniger häufig sind die Schreibvarianten *blind date* (35 Treffer) und *Blind date* (22 Treffer).

Flexionsform:

[die normalste Sache der Welt](#): Die Wortverbindung ist in den IDS-Korpora der geschriebenen Sprache ausschließlich im Superlativ belegt.

Lexikalische Festigkeit:

[die sanfte Tour](#): Das Artikelstichwort wurde mit dem bestimmten Artikel *die* angesetzt, da der bestimmte Artikel usuell ist.

Typik:

[im Vollbesitz seiner geistigen Kräfte](#): Die Wortverbindung *im Vollbesitz der / seiner geistigen Kräfte* ist die signifikanteste Realisierung des Grundmusters *im Vollbesitz von etw. sein*. Andere Ergänzungen sind beispielsweise *seiner Fähigkeiten, ihrer Macht, der Wahrheiten, ihrer Stimme*.

Kommentare zu „Variation / Modifikation“

[auf Dauer](#): Die Wortverbindung *auf Dauer* weist ein Vorkommen in den IDS-Korpora geschriebener Sprache von über 20 000 Treffern, die Nebenform *auf die Dauer* von über 4000 Treffern auf.

[Gut gebrüllt, Löwe](#): Die Wortverbindung *Gut gebrüllt, Löwe* wird gelegentlich modifiziert, indem statt des Lexems *Löwe* Namen oder Bezeichnungen von Personen, Vereinen, Institutionen oder Ähnlichem verwendet werden.

[sanfte Revolution](#): Variation: *samtene Revolution*: Die Wortverbindung *samtene Revolution* ist in den IDS-Korpora der geschriebenen Sprache häufiger belegt. Sie kann allerdings nur als teilsynonym gelten, da sie im Korpus ausschließlich als Bezeichnung für den politischen Umbruch in der Tschechoslowakei 1989/1990 verwendet wird.

Kommentare zu „Suchanfrage für Recherche im Korpus“

[In einem gesunden Körper wohnt ein gesunder Geist](#): (&gesund /w1 &Körper) /s0 &Geist: Diese Suchanfrage bietet das genaueste Ergebnis. Wenn Sie an einem möglichst großen Modifikationsspektrum interessiert sind, sollten Sie die lexikalischen Einheiten innerhalb eines Satzes mit folgender Anfrage suchen: (&gesund /s0 &Körper) /s0 &Geist. Sie müssen dann aber damit rechnen, dass in bestimmten Fällen auch Kontextstellen angezeigt werden, die nicht zum Verwendungsspektrum dieser Wortverbindung gehören.

Kommentare zu „Gebrauchsbeschreibung“

[auf Dauer](#) (Lesart ‘auf lange Sicht’): *Auf Dauer* wird in dieser Lesart sehr häufig in Negationskontexten verwendet. Damit wird ausgedrückt, dass einer weiteren Zukunft eines Sachverhalts oder eines Zustands keine Chancen eingeräumt werden bzw. diese Zukunftsperspektive nicht für sinnvoll erachtet wird.

[blinder Aktionismus](#): (Mit *blinder Aktionismus* werden Entscheidungen oder Handlungen als übertrieben, voreilig, wenig durchdacht und damit als wenig hilfreich, wenn nicht gar als kontraproduktiv bewertet ...)

Kommentar: Kookkurrenzpartner von *Aktionismus* wie *hektisch, wild, blank, sinnlos, blindwütig* oder *kurzfristig* stützen diese Gebrauchsbeschreibung.

[geistiger Brandstifter](#): Die Wortverbindung *geistiger Brandstifter* ist in den IDS-Korpora der geschriebenen Sprache seit Beginn der 90er Jahre belegt.

Kommentare zu „Kontextmuster“

[blinder Alarm](#): Kontextmuster *sich als blinder Alarm herausstellen*: Die ergänzenden Verben dieser Gruppe sind an die Wortverbindung *blinder Alarm* gebunden und kommen nie oder nur selten allein mit dem Nomen *Alarm* kombiniert vor.

Kommentare zu „Beziehungen zu anderen Wörtern und Wortverbindungen“

[geistiger Vater](#): Die Wortverbindung steht in einem Verwendungszusammenhang mit [geistiger Wegbereiter](#); [geistiger Brandstifter](#).

Kommentar: Die Wortverbindungen *geistiger Wegbereiter* und *geistiger Brandstifter* werden in den IDS-Korpora der geschriebenen Sprache jedoch überwiegend (*Wegbereiter*) oder immer (*Brandstifter*) negativ konnotiert verwendet, was an signifikanten Verwendungsmustern wie *geistiger Wegbereiter des Nationalsozialismus*; *geistiger Wegbereiter des Faschismus*; *geistiger Wegbereiter des Terrorismus* erkennbar wird.

Wer zu spät kommt, den bestraft das Leben: Bei der Wortverbindung *Wer zu spät kommt, den bestraft das Leben* handelt es sich um eine Wortverbindung, die die strukturellen und funktionalen Eigenschaften eines Sprichworts bereits erfüllt, obwohl auch heute noch die Autorschaft und der ursprüngliche Äußerungszusammenhang präsent sind. Daher ist sie sowohl als Sprichwort als auch als Zitat einzuordnen.

Kommentare in Verbindung zu anderen lexikografischen Ergebnissen / externen Wörterbüchern / Internetquellen⁹

jenseits von Gut und Böse: Die Wortverbindung *jenseits von Gut und Böse* geht auf Friedrich Nietzsches philosophisches Werk „Jenseits von Gut und Böse. Vorspiel einer Philosophie der Zukunft“ von 1886 zurück. Die Autorschaft dieser Wortverbindung ist in den aktuellen Korpusbelegen nach wie vor sehr präsent, sodass man diese Wortverbindung sowohl als Phraseologismus im engeren Sinne als auch als Zitat einordnen kann.

3. Perspektiven

Die lexikografische Beschreibung usueller Wortverbindungen wird in den nächsten Jahren mit der Erarbeitung von Mehrwortartikeln zu ausgewählten Feldern weitergeführt. Die entsprechenden Annotationen sollen der späteren Recherchierbarkeit nach linguistischen Merkmalen dienen und, wo es möglich ist, analog zu den entsprechenden Attributen im [*ellexiko-Wörterbuch*](#) angesetzt werden. Der Zugang zu den Artikeln wird stets auf zweifache Weise möglich sein: zum einen über die projektspezifische Stichwortliste auf unserer [OWID-Portalseite](#), zum anderen über die Stichwortsuche im gesamten OWID-Portal. Die [UWV-Stichwortliste](#) unterscheidet sich von einer herkömmlichen, z.B. der [300 000er Stichwortliste](#) des [*ellexiko-Wörterbuchs*](#), dadurch, dass es keine im Voraus definierte Lemmaliste geben wird, sondern dass sie erst durch die Auswertung systematischer Kookkurrenzanalysen während der laufenden Forschungsarbeit des Projektes ‘Usuelle Wortverbindungen’ sukzessive entsteht. Vorgesehen ist, die OWID-Artikel mit anderen Projektergebnissen, die in elektronischer Form präsentiert werden, zu verlinken. Vorstellbar ist darüber hinaus, mit externen Partnern erarbeitete lexikografische Daten (Artikel, korpusvalidierte Stichwortlisten zu einzelnen Mehrwortklassen) zu integrieren.

⁹ Auch diese Kommentare beziehen sich auf Beobachtungen des typischen Gebrauchs im Korpus. Es werden also nicht systematisch externe Quellen analysiert oder entsprechende Einträge mit unseren Informationen verglichen. Diese Verweise kommen nur dann ins Spiel, wenn sich in den Korpora auffällig häufige Indizien z.B. für ein Zitat oder eine spezifische Herkunft oder auch für enzyklopädische Informationen finden. Im Beispiel [jenseits von Gut und Böse](#) wird die Quelle (F. Nietzsche) häufig in den Korpusbelegen angeführt. Sie ist quasi in den Köpfen der Sprecher noch präsent (vgl. auch [Wer zu spät kommt, den bestraft das Leben](#)).

4. Literatur

- Belica, Cyril (1995): Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethode. Mannheim. Internet: <https://cosmas2.ids-mannheim.de/cosmas2-web/> (Stand: Mai 2008).
- Belica, Cyril/Steyer, Kathrin (i. Dr.): Korpusanalytische Zugänge zu sprachlichem Usus. In: AUC (Acta Universitatis Carolinae), Germanistica Pragensia XX. Prag 2006. [Vorabdruck im Internet unter http://www.ids-mannheim.de/ll/uvw/dateien/Belica_Steyer_2005.pdf (Stand: Mai 2008)].
- Brunner, Annelen/Steyer, Kathrin (2007): Phraseologische und phraseographische Aspekte korpusgesteuerter Empirie. In: Jesenšek, Vida/Fabčič, Melanija (Hg.): Phraseologie kontrastiv und didaktisch. Neue Ansätze in der Fremdsprachenvermittlung. (= Zora 47). Maribor, S.181-194.
- Burger, Harald (2007): Phraseologie. Eine Einführung am Beispiel des Deutschen. 3., neu bearb. Aufl. (= Grundlagen der Germanistik 36). Berlin.
- Burger, Harald/Dobrovol'skij, Dmitrij/Kühn, Peter/Norrick, Neal R. (Hg.) (2007): Phraseologie / Phraseology. Ein internationales Handbuch zeitgenössischer Forschung / An international handbook of contemporary research. 2 Halbbde. (= HSK 28.1/2). Berlin/New York.
- Dobrovol'skij, Dmitrij (1993): Datenbank deutscher Idiome. Aufbauprinzipien und Einsatzmöglichkeiten. In: Földes, Czaba (Hg.): Germanistik und Deutschlehrerausbildung. Szeged/Wien, S. 51-67.
- Fleischer, Wolfgang (1997): Phraseologie der deutschen Gegenwartssprache. 2., durchges. u. erg. Aufl. Tübingen.
- Filatkina, Natalia (2007): Pragmatische Beschreibungsansätze. In: Burger/Dobrovol'skij/Kühn/Norrick (Hg.), Teilbd. 1, S. 132-158.
- Fries, Norbert (1991): Bewertung. Linguistische und konzeptuelle Aspekte des Phänomens. In: Sprache und Pragmatik. (= Arbeitsberichte 23). Lund, S.1-31.
- Hallsteinsdóttir, Erla/Ken Farø (Hg.) (2006): Neue theoretische und methodische Ansätze in der Phraseologieforschung / New theoretical and methodological approaches to phraseology. (= Linguistik online 2/06). Internet: http://www.linguistik-online.de/27_06/ (Stand: Mai 2008).
- Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz. Berlin/New York.
- Projekt Methoden der Korpusanalyse und -erschließung. Institut für Deutsche Sprache. Mannheim. Homepage. Internet: <http://www.ids-mannheim.de/kl/projekte/methoden> (Stand: Juli 2008).
- Steyer, Kathrin (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten. In: Deutsche Sprache 2/2000, S. 101-125.
- Steyer, Kathrin (2004a): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Steyer, Kathrin (Hg.): Wortverbindungen – mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003. Berlin/New York, S. 87-116.
- Steyer, Kathrin (2004b): Idiomatik hypermedial. Zur Repräsentation von Wortverbindungen im Informationssystem Wissen über Wörter. In: Palm-Meister, Christine (Hg.): EUROPHRAS 2000. Internationale Tagung zur Phraseologie vom 15. bis 18. Juni 2000 in Aske/Schweden. (= Stauffenburg Linguistik 25). Tübingen, S. 465-479.
- Steyer, Kathrin (i. Dr.): Zwischen theoretischer Modellierung und praxisnaher Anwendung. Zur korpusgesteuerten Beschreibung usueller Wortverbindungen. In: Mellado Blanco, Carmen (Hg.): Theorie und Praxis der Erstellung von idiomatischen Wörterbüchern. Tübingen.
- Steyer, Kathrin/Lauer, Meike (2006): Elektronische Mehrwortartikel und lexikografische Konzeption. Mannheim. Internet: <http://hypermedia.ids-mannheim.de/elexiko/ModulMV/Konzeption.html> (Stand: Mai 2008).
- Steyer, Kathrin/Lauer, Meike (2007): „Corpus-Driven“: Linguistische Interpretation von Kookkurrenzbeziehungen. In: Kämper, Heidrun/Eichinger, Ludwig M. (Hg.): Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache. (= Studien zur deutschen Sprache 40). Tübingen, S. 493-509.

5. Referenzwörterbücher in den Mehrwort-Artikeln

Duden 11 (2002) = Duden – Redewendungen. Wörterbuch der deutschen Idiomatik. Bearbeitet von Brigitte Alsleben und Werner Scholze-Stubenrecht. 2., neu bearb. u. aktual. Aufl. Mannheim/Leipzig/Wien/Zürich.

Duden 12 (2002) = Duden – Zitate und Aussprüche. Bearbeitet von Maria Grazia Chiaro und Werner Scholze-Stubenrecht. 2., neu bearb. u. aktual. Aufl. Mannheim/Leipzig/Wien/Zürich.

Röhrich, Lutz (2000): Lexikon der sprichwörtlichen Redensarten. (= Digitale Bibliothek 42). Berlin.

Werner Scholze-Stubenrecht

Zu den neuen lexikografischen Inhalten in OWID

Abstract

Ausgehend von den drei Einträgen „Gegenwart“, „blind“ und „Globalisierung“ und ihren jeweiligen Verknüpfungen mit einem der drei OWID-Produkte, versucht dieser Beitrag zu zeigen, ob und wie das *ellexiko*-Wörterbuch durch die Einbindung in OWID an Substanz und lexikografischem Informationsgehalt gewonnen hat.

Based on the *ellexiko* entries “Gegenwart”, “blind”, and “Globalisierung” and evaluating their connection to the other three OWID products, this article tries to show how the *ellexiko* dictionary has been improved by additional lexicographical substances in OWID.

Inhalt:

1. Der Wörterbucheintrag „Gegenwart“
2. Der Wörterbucheintrag „blind“
3. Der Wörterbucheintrag „Globalisierung“
4. Fazit
5. Nachschlagewerke

OWID ist die Verbindung des Online-Wörterbuches *ellexiko* mit den drei lexikografischen Produkten „Schulddiskurs im ersten Nachkriegsjahrzehnt“, „Wortverbindungen online“ und „Neologismenwörterbuch“ durch einen gemeinsamen Zugang per Stichwortsuche. Die Frage nach den neuen lexikografischen Inhalten stellt sich deshalb auch und vor allem als eine Frage nach dem lexikografischen Gesamtnutzen der in einem Portal zusammengefassten Einheit „*ellexiko* + drei weitere Wörterbuchprodukte“. Der Blickwinkel dieses Beitrages ist dabei ein lexikografischer: Es werden im Folgenden die lexikografischen Strukturen und Inhalte aus einer von der „kommerziellen“ Lexikografie geprägten Sicht beurteilt.

Anhand von drei Einträgen aus *ellexiko*, jeweils in Verbindung mit einem korrespondierenden Eintrag aus einem der drei anderen Produkte, soll exemplarisch gezeigt werden, inwiefern das *ellexiko*-Wörterbuch durch die Einbindung in OWID inhaltlich-lexikografisch gewonnen hat. Dabei erheben diese Betrachtungen keinen Anspruch auf eine vollständige Würdigung all dessen, was in den letzten Monaten und Jahren lexikografisch ausgebaut und optimiert wurde. Es werden lediglich eine Reihe von Beobachtungen weitergegeben, die zur einen oder anderen kleinen Nachbesserung anregen können. Schließlich soll der Versuch unternommen werden, daraus auch eine Aussage über die grundsätzliche Ausrichtung und Gewichtung von OWID abzuleiten.

1. Der Wörterbucheintrag „Gegenwart“

Der Wörterbucheintrag [Gegenwart](#) in *ellexiko* beginnt im Bereich der lesartenübergreifenden Angaben mit den orthografischen Angaben zur Normschreibung und Worttrennung. Das ist bei vielen Wörterbüchern so und dürfte der Erwartungshaltung geübter Wörterbuchbenutzer entsprechen. Was an dieser Stelle in vielen anderen Wörterbüchern geboten wird, nämlich eine Ausspracheangabe oder -hilfe, sei es durch eine phonetische Umschrift oder durch Betonungs- und Vokallängenkennzeichnung, gehört derzeit nicht zum Informationsangebot in *ellexiko*, wird aber teilweise durch die Zusammenführung mit dem Neolo-

gismenwörterbuch eingebracht, wo zumindest die Fremdwörter mit fremdsprachlichen Ausspracheelementen (z.B. *mailen*) eine phonetische Umschrift haben.

Es folgt ein Abschnitt zur Wortbildung, der bei diesem Eintrag nicht gefüllt ist, danach kommen die etymologischen Angaben, die allerdings nur aus Verlinkungen zum „Deutschen Wörterbuch“ von Jacob und Wilhelm Grimm und weiteren Literaturverweisen bestehen. Dieser Zugewinn an Informationsvermittlung durch Verknüpfung mit externen Quellen entspricht dem modularen Konzept von OWID durchaus, auch wenn hier der Zugang zu einem weiteren Wörterbuch nicht auf der Ebene des Portals selbst angesiedelt ist.

Auch in diesem Bereich verbreitert die Verbindung mit dem Neologismenwörterbuch das lexikografische Informationsangebot, denn dort werden kurze Herkunftserläuterungen in bestimmten Fällen (z.B. bei *Blind Date*) unmittelbar, also ohne externe Verlinkung gegeben. Es dürfte sich dabei um Einträge handeln, die in der Regel ohnehin nicht mit dem Grimmschen Wörterbuch zu verlinken wären, weil es sie dort nicht gibt.

Die nun folgenden Abschnitte „Wandel zwischen 1700-1945“¹ und „Wandel seit 1945“ sind mit Hinweisen darauf gefüllt, dass noch keine Analysen durchgeführt werden konnten und keine entsprechenden Beschreibungen vorliegen.

Hier kann man stutzen, denn das Wörterbuch „Schulddiskurs im ersten Nachkriegsjahrzehnt“ enthält einen Eintrag *Gegenwart* und es beschreibt ja zumindest einen spezifischen Aspekt der Wortverwendung zwischen 1945 und 1955. Auch wenn das nicht dasselbe ist wie eine Darstellung eines semantischen Wandels seit 1945, liegt es für den Benutzer nahe, sich den dortigen Eintrag anzusehen. Er ist von der Struktur her deutlich anders, als es die Wörterbucheinträge in *ellexiko* sowie im Wörterbuch „Wortverbindungen online“ und im Neologismenwörterbuch sind. Er bietet eine spezielle Art der lexikografischen Sprachgebrauchsbeschreibung, gibt das Ergebnis sehr ausgefeilter Interpretationen in einem entsprechend elaborierten, fachsprachlich oder mindestens bildungssprachlich geprägten Stil wieder, beruht auf der Auswertung von gezielt ausgewählter Literatur, geht mit Belegzitate ganz anders um als die anderen Wörterbücher in OWID. Vermutlich würden durchschnittlich gebildete Benutzer den Text als „schwierig“ einstufen. (Was nicht für alle Artikel dieses Wörterbuches gilt.)

Die Bedeutungserläuterung der entsprechenden Lesart in *ellexiko* ist demgegenüber natürlich sehr viel unspezifischer und dadurch sicher auch ein Stück alltagstauglicher; es geht ja hier auch nicht um den mikroskopischen Blick auf einen historisch eingeschränkten Diskurszeitraum, sondern um die gegenwärtige allgemeinsprachliche Verwendung des Wortes.

Was Wörterbücher häufig an die Bedeutungserläuterung anschließen, sind sogenannte „normierte Beispiele“, also ausgewählte Kontexte des Stichworts, deren Herkunft nicht näher belegt ist. Wenn etwa im Duden-Stilwörterbuch die Kollokationen *die Technik der Gegenwart* oder *die Gegenwart genießen* aufgeführt werden, ist nicht nachprüfbar, ob diese Verwendungen im tatsächlichen Sprachgebrauch häufiger oder überhaupt vorkommen. Theoretisch könnte der Lexikograf sie sich auch einfach ausgedacht haben. (Es gibt beide Kontexte im Dudenkorpus,² aber sie waren in diesem Wörterbuch schon zu lesen, bevor es das Dudenkorpus gab.)

¹ Als Überschrift wäre „Wandel 1700-1945“ überzeugender.

² Für Informationen zum Dudenkorpus vgl.

http://www.duden.de/deutsche_sprache/wort_in_den_duden.php

Dieses Risiko der Unangemessenheit im Hinblick auf den tatsächlichen Sprachgebrauch wird in *ellexiko* ebenso vermieden wie im Neologismenwörterbuch und im Wörterbuch „Wortverbindungen online“; hier werden alle Verwendungen entweder in zum Teil sehr ausführlichen Belegzitate vorgeführt oder aus Kollokationsanalysen der jeweils zugrunde liegenden Korpora gewonnen und als semantische Umgebungen oder im Rahmen typischer Verwendungen dargestellt.

Und hier liegt zweifellos die besondere Stärke der Wörterbücher in OWID (besonders von *ellexiko*), zumal der Bereich Sinnverwandtschaft (einschließlich der Gegenwörter) in die Belegauswahl einbezogen ist und die Wortschatzbeziehungen des Stichworts so umfassend dokumentiert und kommentiert sind, wie man das sonst wohl kaum findet.

Ungewohnt aus dem Blickfeld der traditionellen Lexikografie ist sicher der Abschnitt „Die semantische Umgebung und ihre Mitspieler“ in *ellexiko*. Die in den Benutzungshinweisen zu *ellexiko* gegebenen Beispiele für [einschränken](#) und [Mobilität](#) sind zwar durchaus plausibel, weil man sich aus Stichwort, W-Frage und Mitspielern einleuchtende Syntagmen bilden kann.

Bei der ersten Lesart von [Gegenwart](#) dagegen ist schon die Unterscheidung von „Was gibt es in der Gegenwart?“ und „Was hat die Gegenwart?“ nicht auf Anhieb transparent (vgl. unten), und warum der „Streifzug“ als Mitspieler bei „Was macht man mit der Gegenwart?“ und nicht bei „Was macht man in der Gegenwart?“ erscheint, wird auch nicht gleich deutlich.

Was gibt es in der Gegenwart?

Erfordernisse
Herausforderung
Kommunikation
Konflikt
Kriege
Probleme
Realität
Tendenzen

Was hat die Gegenwart?

Architektur
Film
Kultur
Literatur
Musik
Theater

Wen gibt es in der Gegenwart?

Autor
Komponisten
Künstler
Maler

Was macht man mit der Gegenwart?

Analyse
auseinandersetzen
beleuchten

bewältigen
Bewältigung
gestalten
(sich) beschäftigen
Streifzug
verstehen

Was macht man in der Gegenwart?

erhalten
erkennen
herüberretten
leben
retten
(sich) widmen
zurückholen

Was macht die Gegenwart?

bieten
einholen
(sich) widerspiegeln
zeigen

Wie ist die Gegenwart?

gelebt
real
rosig
trist
unmittelbar

Spätestens dann, wenn der Wörterbucheintrag den Nutzer wie in diesen Fällen zum längeren Nachdenken zwingt, fragt er sich natürlich, was er überhaupt damit gewonnen habe, wenn ihm dann letztlich vielleicht doch das eine oder andere klar geworden ist. Die *Erfordernisse der Gegenwart* zeigen wohl schon eine andere semantische Beziehung als die *Architektur der Gegenwart*, aber was fängt man mit dieser Erkenntnis an? Das Ganze ist, wie gesagt, zunächst einmal einfach ungewohnt, und es wäre daher sehr interessant, welche Benutzer von *ellexiko* auf diese Daten häufiger zugreifen und warum sie es tun.

Zurück zu den gewohnten lexikografischen Inhalten. *ellexiko* gibt, wie andere Wörterbücher auch, grammatische Informationen an. Während man sie dort aber meistens als lesartenübergreifend am Beginn eines Wörterbucheintrags findet, sind sie hier lesartenbezogen. Das ist gut begründbar, denn einzelne Lesarten können mit unterschiedlichen Wortartangaben verbunden sein (z.B. bei *normal*, wo drei Lesarten mit der Wortart Adjektiv verbunden sind, eine Lesart aber das Adverb betrifft) oder sie können unterschiedlichen Restriktionen etwa bei der Pluralbildung unterliegen. Beim Eintrag *Gegenwart* wird bei allen drei Lesarten angegeben, dass das Wort keinen Plural habe. Das steht wohl auch in anderen Wörterbüchern so, derzeit auch noch in „Duden – Das große Wörterbuch der deutschen Sprache“, sollte aber zumindest für die erste Lesart überprüft werden. Das Dudenkorpus, das im Augenblick noch etwas kleiner ist als die IDS-Korpora, bietet immerhin 70 plausible Belege für *Gegenwart* (im Kontext dabei auch mehrere für *Vergangenheiten* und *Zukünfte*), sodass eine Angabe wie „Plural selten“ vielleicht nicht falsch wäre. Eventuell könnte man in *ellexiko* auch unter der Rubrik „Besonderheiten des Gebrauchs“ etwas dazu sagen.

Als erstes Zwischenergebnis wäre festzuhalten: Der Eintrag *Gegenwart* in *ellexiko* wird über OWID durch den entsprechenden Eintrag des Schulddiskurs-Wörterbuches um eine soziolinguistische und sprachgeschichtliche Dimension erweitert. Eine ausgewiesene *ellexiko*-Lücke (Wandel seit 1945) wird dadurch wenigstens zu einem Teil weniger lückenhaft. Das Wörterbuch wird im Bereich Wortgebrauch/Wortfunktion punktuell vertieft, und zwar in einer fachspezifischen Ausrichtung.

2. Der Wörterbucheintrag „blind“

Für *blind* unterscheidet *ellexiko* vier Lesarten und eine Spezifizierung der ersten Lesart. Diese Anzahl der berücksichtigten Lesarten zeigt die hochgradige Korpusorientiertheit der *ellexiko*-Informationen. Andere Wörterbücher bieten nämlich nicht nur fünf explizite Bedeutungsunterscheidungen (Lesarten und Spezifizierung der Vergleichbarkeit halber zusammengefasst). Im „Wörterbuch der deutschen Gegenwartssprache“ werden sechs angegeben, in „Duden – Das große Wörterbuch der deutschen Sprache“ sind es zehn. Das erklärt sich daraus, dass diese beiden Wörterbücher auch eher Fachsprachliches verzeichnen und etwa die *blinden Arkaden* in der Architektur oder die *blinde Naht* aus der Schneiderei mit abdecken. Und beim Dudenwörterbuch werden unter „blind“ auch noch zwei Verwendungen aus dem 18./19. Jahrhundert (*blinde Nacht* und *blinder Lärm*) angeführt. All das ist in einem Textkorpus der allgemeinen deutschen Gegenwartssprache nicht mit statistischer Signifikanz zu erwarten.

Als gewiefter *ellexiko*-Benutzer kann man natürlich – falls man die genannten Lesarten vermisst – den Etymologiebereich anklicken und sich im grimmschen Wörterbuch um-

sehen, wo neben den wortgeschichtlichen auch Lesarteninformationen zu finden sind, und dort trifft man auch beispielsweise auf die *blinde Nacht*.

Neben Lesartenauswahl weist auch die Darstellung der Grammatikangaben auf die strenge Korpusbasiertheit des Wörterbuches hin. Zur Lesart ‘[maßlos](#)’ werden die Komparativformen zwar paradigmatisch vollständig aufgeführt (bis auf den Dativ Singular des Maskulinums), es wird jedoch bei mehreren Deklinationsformen des Komparativs darauf hingewiesen, dass sie im Korpus nicht belegt sind. Etwas irritierend erscheint hier der angefügte Kommentar, nach dem „blind“ den semantischen Gradunterschied nicht zulässt, während das Korpus einen entsprechenden Gebrauch dokumentiert:

Kommentar:

Blind lässt den semantischen Gradunterschied nicht zu. Im *lexiko*-Korpus können jedoch einige Fälle des Komparativs belegt werden (vgl. den Beleg).

Lange liest er so dahin, der Dichter, Wunderbares und Langweiliges nacheinander; er kann es nicht unterscheiden, er konnte es nie. H. C. Artmann, der Oberromantiker unter den Avantgardisten, hat zur Sprache ein Verhältnis wie der röm.-kath. Sünder zu seiner Religion: eine ziemlich ausschweifende und umso **blindere** Liebe zum Experiment. (die tageszeitung, 06.03.1992, S. 27, Du bist die Rotweinspur.)

Die Lesart ‘[unkritisch](#)’ im Eintrag [blind](#) gibt nun wieder Gelegenheit, etwas zur Aufwertung von *lexiko* durch die Einbindung in OWID zu sagen. Vergleicht man nämlich die lexikalischen Mitspieler in der Lesart ‘unkritisch’ mit denen der vorhergehenden Lesart ‘maßlos’, so fällt auf, dass bei der Frage „[Wer oder] was ist blind?“ in beiden Fällen Fanatismus und Glauben aufgelistet sind. Dieselben lexikalischen Mitspieler können also in der semantischen Umgebung unterschiedlicher Lesarten auftreten. Nach dieser eigentlich nicht sehr überraschenden Erkenntnis kann man sich fragen, ob der Mitspieler *Aktionismus*, der nur in der Lesart ‘unkritisch’ ausgewiesen ist, nicht auch in der Lesart ‘maßlos’ hätte angesiedelt sein können. Zur Beantwortung dieser Frage könnte es hilfreich sein, den Eintrag [blinder Aktionismus](#) des Wortverbindungen-Wörterbuches zu konsultieren. Freundlicherweise zeigt einem die Suchergebnisliste in OWID nach der Sucheingabe „blind“ auch alle die Einträge des Wortverbindungen-Wörterbuches, bei denen „blind“ das so genannte Basiselement ist, sodass man dieses zusätzliche Informationsangebot beim Einstieg in die Wörterbuchbenutzung gleich vor Augen hat:

Suchergebnis in OWID

Ihre Suchanfrage ergab 21 Treffer.

[Suchergebnis speichern](#)

[Neue Suche](#)

- [blind](#)
- **blind** Basiselement zu [blinder Aktionismus](#)
- **blind** Basiselement zu [blinder Alarm](#)
- **blind** Basiselement zu [auf einem Auge blind sein](#)
- **blind** Basiselement zu [blind buchen](#)
- **blind** Basiselement zu [Blind Date](#)
- **blind** Basiselement zu [blinder Eifer](#)
- **blind** Basiselement zu [blinde Fenster](#)
- **blind** Basiselement zu [blind gegenüber](#)
- **blind** Basiselement zu [blindes Huhn](#)
- **blind** Basiselement zu [Ein blindes Huhn findet auch mal ein Korn](#)
- **blind** Basiselement zu [blinde Kuh](#)

- **blind** Basiselement zu [Liebe macht blind](#)
- **blind** Basiselement zu [\[...\] macht blind](#)
- **blind** Basiselement zu [blinder Passagier](#)
- **blind** Basiselement zu [blinde Scheiben](#)
- **blind** Basiselement zu [sich blind und taub stellen](#)
- **blind** Basiselement zu [blind ins Verderben](#)
- **blind** Basiselement zu [sich blind verstehen](#)
- **blind** Basiselement zu [blind vor](#)
- **blind** Basiselement zu [blinder Fleck](#)

Im Wortverbindungen-Wörterbuch gibt es statt der Bedeutungserläuterung eine Gebrauchsbeschreibung, und in *lexiko* finden sich bei [blinder Aktionismus](#) Synonyme für das Adjektiv, die sich auf die *lexiko*-Bedeutungserläuterung beider hier relevanten Lesarten verteilen ließen: *übertrieben* findet sich so in der Lesart ‘maßlos’, und *wenig durchdacht* entspricht dem *unreflektiert* der Lesart ‘unkritisch’. Die Vermutung bestätigt sich also, dass *blind* in Verbindung mit *Aktionismus* semantisch nicht so eindeutig festgelegt ist, wie das die Zuordnung in *lexiko* annehmen lässt.

Noch eine Anmerkung zu den lexikalischen Mitspielern der Lesart ‘maßlos’: Das Wort *Verständnis* scheint hier in die falsche Rubrik gerutscht zu sein. Ein *blindes Verständnis* ist wohl kein „maßloses Verständnis“, sondern gehört zu *sich blind verstehen* und damit im vorliegenden *lexiko*-Eintrag eher zur Spezifizierung ‘ohne hinzusehen’ oder zu einer noch zu ergänzenden Lesart ‘ohne Absprache, intuitiv’.

Als zweites Zwischenergebnis ist festzuhalten: *lexiko* wird durch die Verknüpfung mit dem Wörterbuch „Wortverbindungen online“ sowohl im Bereich der Kollokationen als auch in der semantischen Feindifferenzierung erweitert und vervollständigt. Es besteht eine relativ hohe Kompatibilität; am Beispiel [blind](#) gewinnt man den Eindruck, das Wortverbindungen-Wörterbuch sei gezielt auf *lexiko*-Lücken ausgerichtet. Das kann natürlich daran liegen, dass Einträge wie „Alarm“, „Eifer“, „Fleck“ oder „Passagier“ in *lexiko* derzeit noch nicht bearbeitet sind. Aber es macht zumindest schon einmal den Eintrag [blind](#) umfassender und ergänzt so das Informationsangebot von *lexiko*.

3. Der Wörterbucheintrag „Globalisierung“

Als drittes Beispiel soll der Eintrag [Globalisierung](#) betrachtet werden. Das Stichwort ist in *lexiko* extensiv behandelt worden. Obwohl es nur in einer Lesart dargestellt ist, umfasst der Ausdruck der Druckansicht insgesamt 22 Seiten. Wer diesen Eintrag aufruft, wird kaum das Gefühl haben, dass er nicht ausreichend informiert wird. Es gibt hier 5 Belegzitate zur Rubrik „Bedeutungserläuterung“, 20 im Bereich „Sinnverwandte Wörter“ und noch einmal 13 im Abschnitt „Besonderheiten des Gebrauchs“. Dazu kommen etwa ein Dutzend Kommentare und auch eine recht ausführliche Ausarbeitung der semantischen Umgebungen und deren lexikalische Mitspieler. Der Teil „Typische Verwendungen“ mit nicht einmal ganz einer Druckseite Umfang nimmt sich dagegen schon fast ärmlich aus.

Im Neologismenwörterbuch umfasst der Eintrag [Globalisierung](#) nur 4 Seiten, er enthält 11 Belegzitate. Wenn man noch in den Vergleich einbezieht, dass die 38 *lexiko*-Belege grob geschätzt im Schnitt zwei- bis dreimal so lang sind wie die 11 im Neologismenwörterbuch, dann wird wohl kaum ein Benutzer vermuten, dass hier noch wichtige In-

formationen zu finden sind, die *ellexiko* nicht bietet. Und trotzdem kann es sein, dass ihm die in OWID angebotene Alternative durchaus willkommen ist. Dann nämlich, wenn er keine Lust hat, sich durch 22 Seiten durchzuarbeiten, und darauf vertraut, dass das für ihn Wesentliche auch in den 4 Seiten des deutlich kürzeren und damit bei vergleichbarer Struktur auch übersichtlicheren Eintrags im Neologismenwörterbuch enthalten ist.

Eingangs wurde schon erwähnt, dass das Neologismenwörterbuch mit seinen Ausspracheangaben und seinen expliziten Angaben zur Wortherkunft an lexikografischen Inhalten etwas bietet, was über *ellexiko* hinausgeht. Aber auch eine alternativ angebotene knappere Darstellung (als erster Einstieg ins Thema oder als zeitsparender Ersatz der ausführlichen Ausarbeitung) kann als Bereicherung eines Wörterbuchs angesehen werden.

Natürlich ist das Hauptargument dafür, das Neologismenwörterbuch in OWID zu präsentieren, ein anderes: Die dort behandelten Stichwörter sind nur zu einem sehr kleinen Teil bereits in *ellexiko* ausgearbeitet; [Adresse](#) (*ellexiko*) / [Adresse](#) (Neologismen) oder [Globalisierung](#) (*ellexiko*) / [Globalisierung](#) (Neologismen) sind eher die Ausnahme. Und so wird die Zahl der „fertigen“ Einträge im Portal durch den Neologismenteil deutlich gesteigert, gegenüber *ellexiko* allein wohl mindestens verdoppelt.

4. Fazit

Durch die Verknüpfung von *ellexiko* mit den Wörterbüchern „Schulddiskurs im ersten Nachkriegsjahrzehnt“ und „Wortverbindungen online“ wie dem Neologismenwörterbuch in OWID sind seine lexikografischen Inhalte in quantitativer und in qualitativer Hinsicht vorangebracht worden. Alle hinzugekommenen Stichwörter erhöhen die Chance eines Benutzers, das zu finden, was er sucht. Die spezifischen Charakteristika von *ellexiko*, seine strikte Korpusbasiertheit zum einen, seine Konzentration auf Bedeutungsbeschreibung und Wortverwendung und die grundsätzliche Ausführlichkeit in der Darstellung der Wortverwendung zum anderen, werden durch das Wortverbindungen-Wörterbuch und das Neologismenwörterbuch nicht beeinträchtigt, da beide mit sehr ähnlichen oder denselben Prämissen arbeiten.

Eine Vertiefung des lexikografischen Inhalts, wenn auch in einer sehr spezifischen Interessenrichtung, liefert das Wörterbuch „Schulddiskurs im ersten Nachkriegsjahrzehnt“. Eine gewisse Verbreiterung entsteht durch für *ellexiko* neue lexikografische Beschreibungselemente wie Aussprache oder unmittelbare Herkunftsinformation, die im Neologismenwörterbuch zu finden sind. Sowohl das Neologismenwörterbuch als auch das Wortverbindungen-Wörterbuch erhöhen die Zahl der Korpusbelege beim einzelnen Eintrag, und beide schließen einen Teil der derzeit noch großen Lücken in der Abdeckung des Wortschatzes der deutschen Gegenwartssprache.

Dem Benutzer, der das Glück hat, nach lexikografischen Informationen zu einem Wort zu suchen, das über OWID schon als fertig bearbeiteter Eintrag zu erreichen ist, und der sich in besonderem Maße für den Bedeutungsgehalt und für die kontextuellen Verwendungen des Wortes interessiert, dem winkt ein Erfolgserlebnis, das derzeit wahrscheinlich von keinem anderen Wörterbuch vermittelt werden kann.

5. Nachschlagewerke

- Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm (1854-1960). 16 Bde. [in 32 Teilbänden]. Leipzig. [Retrodigitalisierte Fassung im Internet unter: <http://germazope.uni-trier.de/Projects/DWB> (Stand: Mai 2008)].
- Duden (1999): Duden – Das große Wörterbuch der deutschen Sprache. In zehn Bänden. 3., völlig neu bearb. u. erw. Aufl. Hrsg. vom Wissenschaftlichen Rat und der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich.
- Duden (2001): Duden – Das Stilwörterbuch. Die Verwendung der Wörter im Satz. 8., neu bearb. Aufl. Hrsg. von der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich.
- Wörterbuch der deutschen Gegenwartssprache (1964-1980). Herausgegeben von Ruth Klappenbach und Wolfgang Steinitz. Berlin. [Retrodigitalisierte Fassung im Internet unter: http://www.dwds.de/woerterbuch/1_1 (Stand: Mai 2008)].

Thomas Burch

Fragen der Vernetzung in OWID¹ und im Wörterbuchnetz²

Abstract

Nachschlagewerke sind aufgrund ihres primären Verwendungszweckes, Informationen schnell und gezielt zu finden, auf vielfältige Art aufeinander bezogen und damit bereits in der gedruckten Fassung in gewisser Weise sowohl explizit als auch implizit vernetzt angelegt. Im Falle elektronischer Nachschlagewerke lassen sich diese „Netzwerke“ zusätzlich ausweiten, indem auch Beziehungen zwischen Informationseinheiten etabliert werden können, die in der gedruckten Fassung, beispielsweise aufgrund der Entstehungsgeschichte der Einzelwerke, bisher gar nicht möglich waren. Diese Vernetzungen können weit über eine rein ausdrucksseitige Verknüpfung hinaus, indem sie philologische und informationswissenschaftliche Methoden verbinden. Im Folgenden werden die Wörterbuchverbünde *OWID* und *Wörterbuchnetz* vorgestellt und insbesondere auf das darin enthaltene Vernetzungspotential eingegangen. Neben den in beiden Ansätzen vorhandenen expliziten Verweise zwischen den Wortartikeln werden für das Trierer Wörterbuchnetz zusätzlich automatische Verfahren und Methoden aufgezeigt, mit deren Hilfe bisher nur implizit gegebene Beziehungen zwischen den Wortartikeln ermittelt und zur Überprüfung vorgeschlagen werden können.

Because of their primary usage, i.e. to retrieve information very fast and systematically, dictionaries are strongly cross-linked. Already in the printed edition this network is established in an explicit and an implicit way. But in the case of an electronic publication these interconnections can be additionally extended by integrating relations between information positions, which until now could not be part of the printed versions. The reason for these missing links results e.g. from the historical order of dictionary development, older ones can not link to newer ones. But these connections can not only be found by simply considering similarities within the text patterns. They can be calculated with the help of algorithms from the field of information retrieval and statistical analysis. In the following the two dictionary networks *OWID* and *Wörterbuchnetz* will be presented especially focussing on their capabilities of cross-linking. Beside the basic concepts of the two approaches we will demonstrate the underlying methods and algorithms for the *Wörterbuchnetz* by the help of which we are able to calculate and measure similarities between the different dictionaries.

Digitale Nachschlagewerke sind, wie auch ihre gedruckten Entsprechungen, auf vielfältige Art aufeinander bezogen und damit in gewisser Weise implizit „vernetzt“; eine übergreifende, integrierte Recherche ist jedoch wegen der Unterschiede in der Anlage, Anordnung und Struktur der einzelnen Werke auch im digitalen Medium nicht ohne Weiteres möglich.³ Nachschlagewerke, die durch inhaltlich-strukturelles Markup in standardisierte und damit vergleichbar gemachte Informationseinheiten gegliedert und durch Metadaten angereichert sind, machen jedoch die impliziten Vernetzungen explizit. Dadurch kann eine neue Qualität der Informationsgewinnung erreicht und die Lücke zwischen der schwerfälligen Benutzbarkeit und eingeschränkter Verfügbarkeit der Buchversionen einerseits und der fehlenden Systematik und Beliebigkeit der Informationen im Internet andererseits geschlossen werden.

¹ *OWID (Online-Wortschatz-Informationssystem Deutsch)*, das Portal für lexikografische Arbeiten am Institut für Deutsche Sprache (IDS), Mannheim, www.owid.de (zuletzt eingesehen am 27.12.2007). Im Folgenden wird OWID (bzw. *ellexiko*) im Stand von Mai 2007 präsentiert, da sich am hier Gezeigten inhaltlich nichts Grundsätzliches verändert hat.

² *Das Wörterbuchnetz*, Verbund von neun digitalisierten Wörterbüchern, Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier, www.woerterbuchnetz.de (zuletzt eingesehen am 27.12.2007).

³ Diese unterschiedlichen Strukturen erklären sich zum einen wissenschaftshistorisch aus der Entwicklung der lexikografisch-lexikologischen Methoden, zum anderen aus den spezifischen Zielsetzungen der verschiedenen Wörterbuchtypen.

Im Gegensatz zu einer bloßen, unverbundenen Bereitstellung verschiedener Nachschlagewerke bieten Wörterbuchverbände mit multidirektionalen Verlinkungen komplexe und gezielt spezifizierbare Zugänge zum Material. Diese Verlinkungen gehen weit über eine rein ausdrucksseitige Verknüpfung hinaus, indem sie philologische und informationswissenschaftliche Methoden verbinden. Die Entwicklung von Verfahren, Methoden und Technologien, die eine intelligente, dynamische Vernetzung von Nachschlagewerken unterschiedlichen Typs einschließlich ihrer Quellenverzeichnisse und Primärquellen ermöglichen und dadurch eine neue Qualität des Informations- und Wissensmanagements erreichen, erfordert neben informationstechnologischen Konzepten auch eine stärkere Einbeziehung philologischer und lexikografischer Verfahrensweisen. Im Folgenden werden die Wörterbuchverbände OWID und *Wörterbuchnetz* vorgestellt und insbesondere wird auf das darin enthaltene Vernetzungspotenzial eingegangen. Neben den in beiden Ansätzen vorhandenen expliziten Verweisen zwischen den Wortartikeln werden für das Trierer Wörterbuchnetz zusätzlich automatische Verfahren und Methoden aufgezeigt, mit deren Hilfe bisher nur implizit gegebene Beziehungen zwischen den Wortartikeln ermittelt und zur Überprüfung vorgeschlagen werden können (vgl. Rapp 2006).

In OWID, dem Informationssystem zum deutschen Wortschatz des Instituts für Deutsche Sprache (IDS) (vgl. Klosa et al. 2006), werden die Forschungsergebnisse verschiedener Projekte virtuell vereinigt. Die korpusbasiert erarbeiteten Nachschlagewerke werden unter einer einheitlichen grafischen Benutzeroberfläche elektronisch im Internet publiziert. Zurzeit umfasst es die Komponenten *lexiko*, *Neologismenwörterbuch*, *Wortverbindungen online* und ein Nachschlagewerk zum *Schulddiskurs im ersten Nachkriegsjahrzehnt*. Das *lexiko*-Wörterbuch ist auf über 300.000 Stichwörter ausgelegt, von denen zurzeit etwa 800 ausführlich lexikografisch beschrieben sind. Das *Neologismenwörterbuch* präsentiert in mehr als 750 umfangreichen Wortartikeln neue Wörter, neue feste Wortverbindungen sowie neue Bedeutungen von etablierten Wörtern, die in den 90er-Jahren des 20. Jahrhunderts in die Allgemeinsprache eingegangen sind. Im Modul *Wortverbindungen online* werden empirische Ergebnisse auf dem Gebiet der korpusgesteuerten Mehrwortforschung veröffentlicht, die sich für eine Online-Präsentation eignen. Die Präsentationsformen haben unterschiedliche linguistische Beschreibungstiefen und Darstellungsformate. Derzeit sind ca. 130 lexikografische Mehrwortartikel über die produktinterne Stichwortliste abrufbar. Das Wörterbuch zum *Schulddiskurs im ersten Nachkriegsjahrzehnt* verzeichnet 85 Haupt- und über 200 Unterstichwörter. Jeder Artikel ist mit einem umfangreichen Beleganhang versehen. Der hier dargestellte Wortschatzbereich ist erarbeitet worden aus einem breit angelegten Korpus von Texten, die in den Jahren 1945 bis 1955 erschienen sind.

Die inhaltlichen Bezüge zwischen den einzelnen Modulen bieten dabei ein großes Vernetzungspotenzial, welches durch das elektronische Medium in idealer Weise für den Benutzer aufbereitet werden kann. Aufgrund der dynamischen Weiterentwicklung des Artikelbestandes ist es jeder gedruckten Publikation weit überlegen.

Allerdings müssen dazu entsprechende Voraussetzungen in Form von wohldefinierten Konzepten geschaffen werden, damit ein elektronisches System einwandfrei funktionieren kann. Dies beginnt in der Regel mit einer standardisierten Datencodierung, die eine konzeptuelle Inhaltsmodellierung der Artikelstrukturen beinhaltet. In OWID wird hier eine einheitliche XML-Codierung zugrunde gelegt (vgl. Müller-Spitzer 2007a), die (zusammen mit dem geplanten Einsatz von Xlink/XPointer (vgl. Müller-Spitzer 2007b))

einen plattform- und systemneutralen Datenbestand garantiert, in dem sowohl die inhaltlichen wie auch die Vernetzungskonzepte enthalten sind. Aufgrund dieser konsequenten Datencodierung zeigt sich das Vernetzungspotenzial in OWID, für den Benutzer völlig transparent, auf unterschiedlichen Ebenen. Eine den Modulen übergeordnete Vernetzung ergibt sich durch die gemeinsame Stichwortliste, in der alle behandelten Artikel aufgeführt werden. Durch Auswahl eines Eintrags wird der Benutzer in das zugehörige Teilmodul geführt (vgl. Abbildung 1).

In dem gezeigten Beispiel werden aus der Stichwortliste von *exlexiko* auch Einträge in den Modulen zum *Schulddiskurs im ersten Nachkriegsjahrzehnt* bzw. in den *Wortverbindungen online* aufgerufen. Verbunden mit der übergeordneten Stichwortsuche kann hier also über eine gemeinsame Schnittstelle der gesamte Artikelbestand abgefragt werden.

Eine weitere Vernetzungsebene ergibt sich durch Bezüge im Innern der Wortartikel, indem Referenzen auf verwandte Stichwörter durch entsprechende Verweise realisiert werden. Dabei spielt es keine Rolle, ob diese Verweise zwischen den Modulen wie im Falle der Stichwortliste oder innerhalb eines einzelnen Moduls bestehen. Eine solche interne Modulvernetzung zeigt Abbildung 2.

Abbildung 1: Vernetzung zwischen Einzelmodulen in OWID

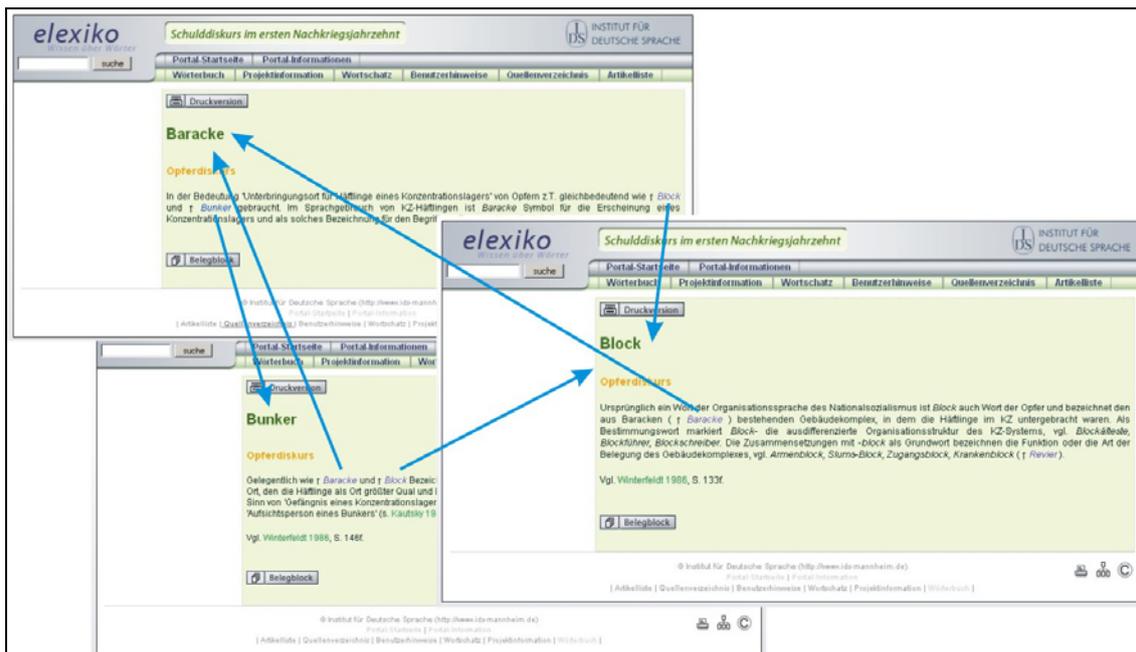


Abbildung 2: Vernetzung innerhalb eines Wörterbuches in OWID

In dem gezeigten Beispiel verweisen die Nennungen verwandter Stichwörter im Artikeltext auf die zugehörigen Einträge in OWID. Prinzipiell sind diese Verweise an jeder Stelle des Artikeltextes möglich. Es ergeben sich dadurch auch Verweise unterschiedlicher Qualität, indem beispielsweise synonyme oder antonyme Bezüge, aber auch Relationen zwischen Ober- und Unterbegriff dargestellt werden können. Wie im Beispiel zu sehen, können die Verweise bidirektional eingerichtet sein und durch sukzessive Verkettung Stichwörter zu Wortgruppen zusammengefasst werden. Die Begriffe *Baracke*, *Block* und *Bunker* sind durch die transitive Vernetzung implizit zu einem Wortcluster kombiniert. Eine Erweiterung der grafischen Oberfläche von OWID könnte darin bestehen, diese bisher nur intern gegebenen Strukturen zu visualisieren. Da das gesamte Material dynamisch erweitert wird, würde dieses Netzwerk mit jedem neu aufgenommenen Artikel dichter werden. Durch die redaktionell betreute Artikelarbeit kann hier eine sehr hohe Qualität der Vernetzungsinformation garantiert werden. Der geplante Einsatz von XLink/XPointer soll darüber hinaus zusätzliche Verbesserung bewirken, z.B. im Bereich der Typisierung der Links, der differenzierteren Visualisierbarkeit, aber auch in der Verwaltung der Vernetzungen für die Lexikografen.

Gegenüber dem dynamisch wachsenden Datenbestand in OWID stellt sich die Ausgangslage für das Trierer Wörterbuchnetz anders dar. Hier stehen unterschiedliche Typen tief annotierter Wörterbücher in digitaler Form zur Verfügung, die bereits in unterschiedlicher Dichte untereinander vernetzt, durch Symptomwertangaben und Metadaten intellektuell erschlossen und recherchierbar oder auch mit Primärquellen vernetzt sind:

- Typ Sprachstadienwörterbuch: Mittelhochdeutsches Wörterbuch, Mittelhochdeutsches Handwörterbuch (einschlägige Nachträge), Findebuch zum mittelhochdeutschen Wortschatz (als Verbund mit multidirektionalen Verweisen), Neues Mittelhochdeutsches Wörterbuch einschließlich digitalem Quellen- und Belegarchiv

- Typ sprachstadienübergreifendes Wörterbuch: Deutsches Wörterbuch von Jacob und Wilhelm Grimm (DWB)
- Typ Dialektwörterbuch: Pfälzisches Wörterbuch, Rheinisches Wörterbuch, Wörterbuch der elsässischen Mundarten, Wörterbuch der deutsch-lothringischen Mundarten, Luxemburgische Wörterbücher
- Typ Autorenwörterbuch: Goethe-Wörterbuch (GWB)
- Typ Sachwörterbuch: Ökonomische Enzyklopädie von J. G. Krünitz (Projekt der UB Trier)

Ziel der Überlegungen ist, anhand der verschiedenen Wörterbücher zum „Deutschen“ einen standardisierten Einstieg zu schaffen, von dem aus sich die nicht standardisierten, heterogen organisierten und verstreuten Informationen verzweigen – heterogen im Hinblick auf synchrone wie diachrone Varianz (vgl. Burch/Rapp 2007). Dieses Konzept lässt sich nicht nur auf Wörterbücher anwenden, sondern auch auf die Erschließung großer Korpora übertragen.

Geht man, wie im Falle des *Wörterbuchnetzes*, von retrospektiv digitalisierten Wörterbüchern aus, so ist der Aufbau einer entsprechenden Verweisstruktur ungleich schwieriger. Schon aufgrund der riesigen Datenmenge ist eine manuelle Erarbeitung und Prüfung der Verweise kaum in vertretbarem Zeitaufwand zu bewältigen. In der Regel beschränkt sich dies auf die Codierung der explizit im Druck genannten Verweise, die in vielen Fällen durch eindeutige Texteinleitungen (z.B. *s.*, *vgl.*, *s.u.*, *s.v.a.*) zumindest durch eine Programmroutine aufgefunden und dann durch einen Bearbeiter verifiziert werden können. Eine darüber hinausgehende inhaltliche Vernetzung der Wörterbucheinträge ist nur durch den Einsatz automatischer Verfahren denkbar. Benutzbar werden die von diesen Verfahren gelieferten Ergebnisse aber nur dann, wenn sie durch eine anschließende redaktionelle Bearbeitung überprüft und damit in ihrer Qualität kontrolliert werden.

Aus der Sicht der Informatik stellt sich somit zunächst die Frage, mit welchen Methoden sich der Aufbau eines Wörterbuchverbundes algorithmisch unterstützen lässt. Eine erste Ansatzmöglichkeit, die im Folgenden näher vorgestellt werden soll, stammt hier aus dem Bereich des Information-Retrieval. In einem ersten Schritt wird von der Realität abstrahiert und ein geeignetes Modell für einen Wörterbuchverbund erstellt, welches dann als Ausgangspunkt zur Implementierung der Algorithmen dient. Ganz allgemein entspricht einem beliebigen Verbund dabei die informationstheoretische Datenstruktur eines aus Knoten und Kanten bestehenden Graphen (vgl. Mehlhorn 1984). Die Knoten des Graphen repräsentieren zunächst die Wörterbücher und die Kanten beschreiben die Verbindungen, das heißt die Verweise, zwischen den Wörterbüchern. Diese grobe Struktur lässt sich weiter verfeinern und genauer modellieren, indem nicht das gesamte Wörterbuch als Knoten aufgefasst wird, sondern die einzelnen Wortartikel. Die Kanten verlaufen dann zwischen den Artikeln verschiedener Wörterbücher bzw. auch innerhalb eines einzelnen Wörterbuchs. Die Kanten sind gerichtet, das heißt, Ausgangs- und Zielpunkt sind eindeutig bestimmt.

Ausgehend von den gedruckten Wörterbüchern lassen sich die Verweise zunächst in drei Klassen einteilen: 1. Verweise, die explizit im Wörterbuch genannt sind, beispielsweise durch Angabe einer Seiten-, Spalten-, Zeilen- oder Artikelreferenz; 2. Verweise, die sich aufgrund von statistischen Berechnungen der Wörter in den Artikeln identifizie-

ren lassen; 3. Verweise, die sich durch Graphalgorithmen aus den Klassen 1 und 2 berechnen lassen, beispielsweise durch transitiven Abschluss. Mit den im Folgenden vorgestellten Verfahren sollen zunächst Verweise der Klasse 2 bestimmt werden.

Die Ausgangsbasis zur Implementierung der Algorithmen bildet ein Datenbanksystem (siehe Abbildung 3), in welchem die Wörterbücher in strukturierter Form abgespeichert sind. Die Architektur des Gesamtsystems basiert auf einer standardisierten SGML/XML-Codierung (vgl. Goldfarb 1990) der Wörterbuchdaten, die sich nach unterschiedlichen Document-Type-Definitionen (DTDs) richten kann. Zu jeder der vorgegebenen Codierungen existiert ein zugehöriger Importfilter, über den die Wörterbuchdaten in ein Datenbankmanagementsystem (DBMS) (vgl. Rob/Carlos 1993) übernommen werden können. In umgekehrter Richtung wird ein Exportfilter entwickelt, über den die durch die Bearbeitung innerhalb des Systems neu eingefügten Informationen in Form von zusätzlichen Metadaten zu den grundlegenden Wörterbuchdaten ebenfalls in standardisiertem XML-Format aus dem Datenbanksystem herausgezogen werden können. Auf diese Weise stehen sowohl die Wörterbuchdaten als auch die Beziehungen zwischen den Wörterbüchern in einer plattformunabhängigen und damit langfristig nutzbaren Form zur Verfügung.

Das Kernstück des gesamten Systems bildet ein relationales Datenbankmanagementsystem, in dem jeweils ein Wörterbuch durch eine zugehörige Datenbank verwaltet wird. Der Aufbau der einzelnen Datenbanken richtet sich dabei nach der Granularität der Auszeichnung. Als kleinste gemeinsame Schnittmenge wird nur festgelegt, dass separate Tabellen zur Abfrage der Stichwörter existieren. Weitere Konstituenten der Wörterbuchartikel wie beispielsweise Angaben zur Wortart, Bedeutungserläuterungen, Belege, Angaben zur Etymologie usw. werden dann in eigenen Datentabellen verwaltet, wenn sie durch die XML-Auszeichnung innerhalb des Gesamtartikels markiert wurden. Je feiner die Auszeichnung ist, desto differenzierter sind die Suchmöglichkeiten und desto genauer die Suchergebnisse.⁴

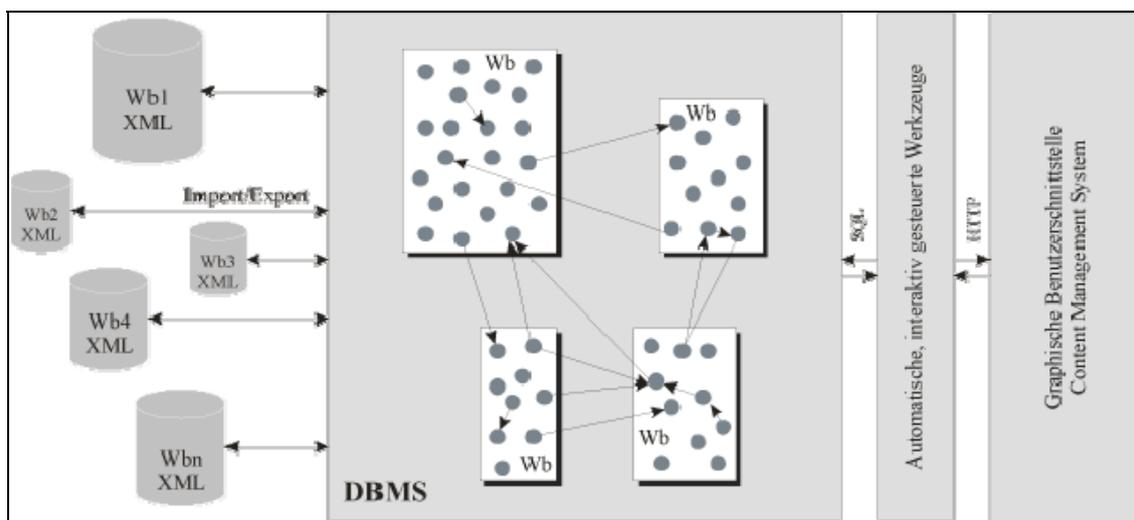


Abbildung 3: Architektur des Datenbanksystems zur Speicherung der Wörterbuchdaten

⁴ Bei Wörterbüchern, bei denen die Artikelmikrostrukturen kaum differenziert werden, bliebe im ungünstigsten Fall nur der Zugriff über eine herkömmliche Volltextrecherche.

Ausgehend von den in den Datenbanken abgespeicherten Wörterbuchartikeln können nun automatische Verfahren entwickelt werden, die zunächst zur Aufgabe haben, sämtliche Wortformen der Wörterbücher zu gewichten. Diese Gewichtung erfolgt durch quantitative Angaben über die Häufigkeit einer Wortform (= Term) innerhalb eines Wortartikels und innerhalb des gesamten Wörterbuchs. Berechnet werden die so genannte relative Häufigkeit und die inverse Dokumentfrequenz für jeden Term. Aus beiden Werten ergibt sich dann durch Multiplikation das sogenannte Termgewicht. Ein Beispiel für diese Berechnungsvorlage zeigt Abbildung 4.

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumentenhäufigkeit	Inverse Dokumentfrequenz	Gewicht
1507749	butterblume	butterblume	GB13619	1	31	21	14.791	-58.486
1507750	f.	f	GB13619	1	31	93291	2.674	-10.572
1507751	nnl.	nnl	GB13619	1	31	3829	7.280	-28.788
1507752	boterbloem,	boterbloem	GB13619	1	31	1	19.183	-75.854
1507753	gilt	gilt	GB13619	1	31	2896	7.683	-30.381
1507754	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507755	mehrere	mehrere	GB13619	1	31	1408	8.724	-34.495
1507756	kräuter,	kraeuter	GB13619	1	31	467	10.316	-40.791
1507757	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507758	caltha	caltha	GB13619	1	31	31	14.229	-56.264
1507759	palustris,	palustris	GB13619	1	31	163	11.834	-46.795
1507760	dotterblume,	dotterblume	GB13619	1	31	10	15.861	-62.718
1507761	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507762	leontodon	leontodon	GB13619	1	31	42	13.791	-54.531
1507763	taraxacum,	taraxacum	GB13619	1	31	53	13.455	-53.204
1507764	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507765	ranunculus	ranunculus	GB13619	1	31	128	12.183	-48.174
1507766	auricomus.	auricomus	GB13619	1	31	8	16.183	-63.991
1507767	das	das	GB13619	1	31	84491	2.817	-11.137
1507768	volk	volk	GB13619	1	31	4258	7.127	-28.182
TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumentenhäufigkeit	Inverse Dokumentfrequenz	Gewicht
1507769	wähnt,	waeht	GB13619	1	31	74	12.974	-51.300
1507770	davon,	davon	GB13619	1	31	5680	6.711	-26.538
1507771	dasz	dasz	GB13619	1	31	35412	4.071	-16.098
1507772	sie	sie	GB13619	1	31	45488	3.710	-14.670
1507773	die	die	GB13619	2	31	123292	2.271	-7.653
1507774	kuhe	kuehe	GB13619	1	31	374	10.636	-42.058

Abbildung 4: Berechnung der inversen Dokumentfrequenz für einen Wortartikel aus dem Deutschen Wörterbuch

Die Tabelle zeigt die durch den Algorithmus berechneten Werte für den Artikel *Butterblume* aus dem Deutschen Wörterbuch von Jacob und Wilhelm Grimm. Entscheidend zur Berechnung des Gewichtes (letzte Spalte) sind die Angaben über die Häufigkeit der Wortform, die Länge des Artikels (gemessen in Wortformen) sowie die Dokumentenhäufigkeit, die die Anzahl der Dokumente (= Wortartikel) repräsentiert, in denen der betreffende Term auftritt. Jedes Dokument wird dabei nur einmal gezählt, auch wenn der Term in ihm mehrfach vorkommt. Aus diesen Angaben lässt sich das Gewicht eines Terms bestimmen: Der Bezug zwischen Häufigkeit der Wortform und Artikellänge ergibt seine relative Häufigkeit im betrachteten Wortartikel; der Bezug zwischen Dokumentenhäufigkeit und der Gesamtanzahl an Dokumenten⁵ liefert die inverse Dokumentfrequenz; das Produkt aus beiden Werten bildet das Gewicht des Terms im vorliegenden Artikel. Sortiert man nun die Terme eines jeden Artikels nach den berechneten Gewichten, so erhält man einen so genannten Relevanzvektor der Terme. Für den betrachteten Beispielartikel ist diese Sortierung in Abbildung 5 gezeigt.

⁵ Beim Deutschen Wörterbuch sind dies 297.613 Dokumente.

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumenten-häufigkeit	Inverse Dokumentfrequenz	Gewicht
1507752	boterbloem,	boterbloem	GB13619	1	31	1	19.183	-75.854
1507766	auricomus.	auricomus	GB13619	1	31	8	16.183	-63.991
1507760	dotterblume,	dotterblume	GB13619	1	31	10	15.861	-62.718
1507749	butterblume	butterblume	GB13619	1	31	21	14.791	-58.486
1507758	caltha	caltha	GB13619	1	31	31	14.229	-56.264
1507762	leontodon	leontodon	GB13619	1	31	42	13.791	-54.531
1507763	taraxacum,	taraxacum	GB13619	1	31	53	13.455	-53.204
1507769	wähnt,	wæhnt	GB13619	1	31	74	12.974	-51.300
1507765	ranunculus	ranunculus	GB13619	1	31	128	12.183	-48.174
1507759	palustris,	palustris	GB13619	1	31	163	11.834	-46.795
1507774	kühe	kuehe	GB13619	1	31	374	10.636	-42.058
1507756	kräuter,	kraeuter	GB13619	1	31	467	10.316	-40.791
1507779	gelb.	gelb	GB13619	1	31	522	10.155	-40.156
1507778	butter	butter	GB13619	1	31	666	9.804	-38.766
1507775	fressen,	fressen	GB13619	1	31	967	9.266	-36.638
1507755	mehrere	mehrere	GB13619	1	31	1408	8.724	-34.495
1507753	gilt	gilt	GB13619	1	31	2896	7.683	-30.381
1507751	nnl.	nnl	GB13619	1	31	3829	7.280	-28.788
1507768	volk	volk	GB13619	1	31	4258	7.127	-28.182
1507776	werde	werde	GB13619	1	31	5146	6.854	-27.101
TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumenten-häufigkeit	Inverse Dokumentfrequenz	Gewicht
1507770	davon,	davon	GB13619	1	31	5680	6.711	-26.538
1507771	dasz	dasz	GB13619	1	31	35412	4.071	-16.098
1507772	sie	sie	GB13619	1	31	45488	3.710	-14.670
1507767	das	das	GB13619	1	31	84491	2.817	-11.137
1507750	f.	f	GB13619	1	31	93291	2.674	-10.572
1507754	für	fuer	GB13619	4	31	40359	3.882	-10.220

Abbildung 5: Sortierung der Wortformen nach den berechneten Termgewichten (= Relevanzvektor)

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumenten-häufigkeit	Inverse Dokumentfrequenz	Gewicht
1117899	Gackelblumm	gackelblumm	PG00045	1	43	1	17.251	-76.359
1117918	Gackelsblumm	gackelsblumm	PG00045	1	43	1	17.251	-76.359
1117927	Gakkelsblumm	gakkelsblumm	PG00045	1	43	1	17.251	-76.359
1117930	Kuhschelle.	kuhschelle	PG00045	1	43	3	15.666	-69.343
1117936	Gackel(s)-ei	gackelsei	PG00045	1	43	3	15.666	-69.343
1117893	Gackel(s)-blume	gackelsblume	PG00045	1	43	4	15.251	-67.506
1117903	Gackel1	gackel1	PG00045	1	43	4	15.251	-67.506
1117901	Godramst]	godramst	PG00045	1	43	5	14.929	-66.081
1117911	(Ranunculus)	ranunculus	PG00045	1	43	7	14.444	-63.933
1117897	(Caltha	caltha	PG00045	1	43	8	14.251	-63.080
1117910	'Hahnenfuß	hahnenfuss	PG00045	1	43	9	14.081	-62.328
1117926	pulsatilla)',	pulsatilla	PG00045	1	43	9	14.081	-62.328
1117924	'Kuchenschelle	kuechenschelle	PG00045	1	43	10	13.929	-61.655
1117914	Butterblume	butterblume	PG00045	1	43	11	13.792	-61.046
1117898	palustris)',	palustris	PG00045	1	43	12	13.666	-60.491

Abbildung 6: Relevanzvektor zum Artikel *Gackel(s)-blume* aus dem Pfälzischen Wörterbuch

Für jeden Wortartikel eines Wörterbuchs wird der Relevanzvektor nach einer festen Maximalanzahl von Elementen abgebrochen.⁶ Auf diesem Weg entsteht also für jedes in das Netz zu integrierende Wörterbuch eine Liste von Relevanzvektoren.

Im darauffolgenden Schritt werden nun je zwei dieser Vektorlisten elementweise miteinander verglichen, das heißt, pro Vektorvergleich werden die in ihnen enthaltenen Wortformen gegeneinander geprüft und es wird die Anzahl der Übereinstimmungen ermittelt. Hier wird zunächst auf exakte Gleichheit getestet, womit eine Basis für die möglichen Verweise zwischen den Wortartikeln geschaffen wird. In weiteren Verfeine-

⁶ Auf diese Weise werden die für die Vergleiche irrelevanten hochfrequenten Wortformen ausgeschlossen, da sie sich durch sehr niedrige Gewichtung auszeichnen.

rungen dieses Algorithmusschrittes könnten auch Informationen über die Wortformen berücksichtigt werden (zum Beispiel durch Lemmatisierung, Stemming, approximative Vergleiche etc.), um zusätzliche Übereinstimmungen zu ermitteln, die dann auch entsprechend bewertet werden können. Einen Vergleichsvektor zu obigem Beispiel zeigt Abbildung 6. Hierbei handelt es sich um den Artikel *Gackel(s)-blume* aus dem Pfälzischen Wörterbuch.

Vergleicht man die normierten Wortformen der Vektoren aus Abbildung 5 und Abbildung 6 miteinander, so findet man drei exakte Übereinstimmungen in den Formen *caltha*, *ranunculus* und *butterblume* innerhalb der ersten 20 gewichteten Terme. Über einen bestimmten Schwellenwert wird für den Gesamtalgorithmus vorgegeben, wann ein Vergleich positiv zu bewerten ist, das heißt, wann ein Verweis zwischen den verglichenen Wortartikeln etabliert werden soll. In obigem Beispiel bedeutet die Anzahl von drei Übereinstimmungen, dass eine Kante im Wortartikelgraphen vom Artikel *BUTTERBLUME* (Deutsches Wörterbuch) zum Artikel *Gackel(s)-blume* (Pfälzisches Wörterbuch) eingerichtet werden soll.

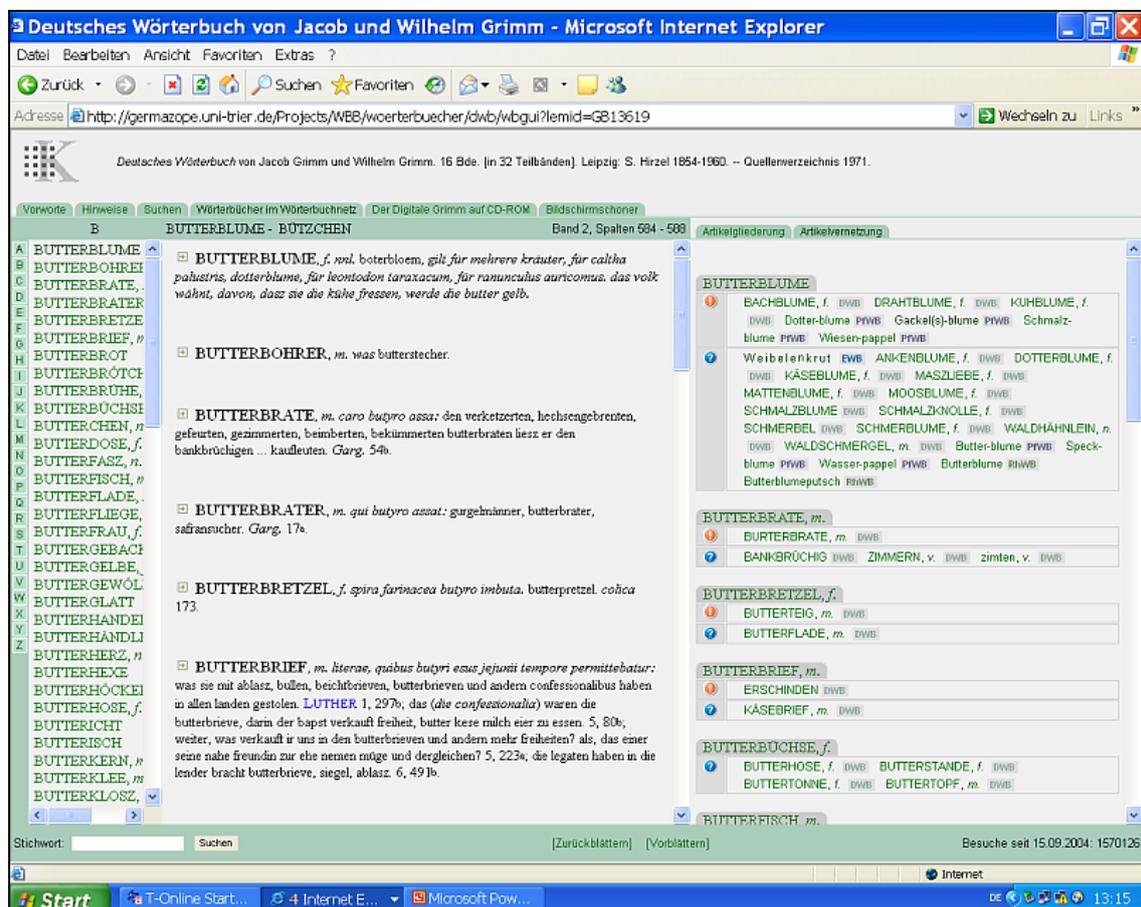


Abbildung 7: Visualisierung der Vernetzungsinformation zur Artikelstrecke ab *BUTTERBLUME* im Deutschen Wörterbuch von Jacob und Wilhelm Grimm

Jeder derart ermittelte Verweis wird in einer zusätzlichen Datenbank als Metainformation über den eigentlichen Wörterbuchdaten abgespeichert, die Basisdaten bleiben also unverändert. Beim Zugriff auf ein Wörterbuch kann aus dieser Datenbank die Vernetzungsinformation ausgelesen und ebenfalls in der grafischen Benutzeroberfläche visualisiert werden (siehe Abbildung 7).

Abbildung 7 zeigt die Vernetzung des Artikels *BUTTERBLUME* im Deutschen Wörterbuch mit weiteren Artikeln des DWB (zum Beispiel *BACHBLUME*, *DRAHTBLUME*, *KUHBLUME*, *KÄSEBLUME* etc.), aber auch mit Artikeln aus anderen Wörterbüchern (zum Beispiel *Dotterblume*, *Schmalzblume* im Pfälzischen Wörterbuch, *Weibelenkrut* im Elsässischen Wörterbuch oder *Butterblume* im Rheinischen Wörterbuch). Die Symbole vor den Verweisgruppen in Abbildung 7 kennzeichnen die Zahl der Übereinstimmungen während der Vektorenvergleiche. Eine höhere Anzahl von Übereinstimmungen führt zu einer Höhergruppierung des Verweises und bedeutet in der Regel eine höhere Zuverlässigkeit der Verknüpfung.

Das hier skizzierte ambitionierte „Idealziel“, ein vielfältig verknüpftes, intelligentes „Meta-Wörterbuch“ der „deutschen Sprache“, steht noch in seinen Anfängen und ist gewiss nur in einer gemeinsamen, koordinierten Forschungs- und Umsetzungsanstrengung realisierbar. Die dafür notwendigen Konzepte, Methoden und Verfahren müssen in einem interdisziplinären Methodenbündel aus den folgenden Bereichen zusammenwirken: Informatik (Ontologien, Graphalgorithmen, Information-Retrieval etc.), Computerphilologie/Computerlinguistik (automatisches Markup, Pattern-Matching, Konkordanzen, automatische Lemmatisierung etc.), systematische Linguistik (Abbildung von Lautgesetzen synchron und diachron, Erarbeitung von Hyper-Lemmalisten, Phraseologie etc.), Lexikografie/Lexikologie (Semasiologie, Onomasiologie, Umkehrlexikografie). Von einem solchen Metazugriff können alle an Sprachinformationen interessierten Disziplinen in höchstem Maße profitieren, darüber hinaus sind weitreichende Impulse für neue Fragestellungen und Lösungsansätze auf verschiedensten Forschungsfeldern zu erwarten.

Literatur

- Burch, Thomas/Rapp, Andrea (2007): Das Wörterbuch-Netz: Verfahren – Methoden – Perspektiven. In: Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung .hist 2006, hrsg. für Clio-online von Daniel Burckhardt, Rüdiger Hohls und Claudia Prinz, unter Mitwirkung von Sebastian Barteleit, Gudrun Gersmann, Peter Haber, Madeleine Herren, Patrick Sahle, Daniel Schlögl, Georg Vogeler, Claudia Wagner und Irmgard Zündorf. (= Historisches Forum 10/I). Berlin. S. 607-627. [Online-Version: <http://www.akademienunion.de/projektbeschreibungen/woerterbuch-netz.htm> (Stand: Mai 2008)].
- Goldfarb, Charles (1990): The SGML handbook. Oxford.
- Klosa, Annette/Schnörch, Ulrich/Storjohann, Petra (2006): *ELEXIKO* – A lexical and lexicological, corpus-based hypertext information system at the Institut für Deutsche Sprache, Mannheim. In: Marelllo, Carla et al. (Hg.): Proceedings of the 12th EURALEX International Congress (Atti del XII Congresso Internazionale di Lessicografia), EURALEX 2006, Turin, Italy, September 6th-9th, 2006. Bd. 1. Turin. S. 425-430.
- Mehlhorn, Kurt (1984): Data structures and algorithms 2: Graph-algorithms and NP-completeness. (= EATCS Monographs on Theoretical Computer Science). Berlin/Heidelberg.
- Müller-Spitzer, Carolin (2007a): Das *ellexiko*-Portal: Ein neuer Zugang zu lexikografischen Arbeiten am Institut für Deutsche Sprache. In: Rehm, Georg/Witt, Andreas/Lemnitzer, Lothar (Hg.): Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of the Biennial GLDV Conference 2007 (April 11-13, 2007, Eberhard-Karls-Universität Tübingen). Tübingen. S. 179-188.
- Müller-Spitzer, Carolin (2007b): Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung. In: Hermes 38/2007, S. 137-171.

Rapp, Andrea (2006): Das Wörterbuchnetz. Verfahren – Methoden – Perspektiven. Internet: <http://www.goethe.de/wis/med/dos/dig/mew/de1526620.htm> (Stand: Mai 2008).

Rob, Peter/Coronel, Carlos (1993): Database systems. Design, implementation and management. In: The Wadsworth Series in Management Information Systems. Belmont, California.

Wörterbücher

Mittelhochdeutsches Wörterbuch, Mittelhochdeutsches Handwörterbuch (einschlägige Nachträge), Findebuch zum mittelhochdeutschen Wortschatz (als Verbund mit multidirektionalen Verweisen):

BMZ (1990): Mittelhochdeutsches Wörterbuch. Mit Benutzung des Nachlasses von Georg Friedrich Benecke ausgearbeitet von Wilhelm Müller und Friedrich Zarncke. 4 Bde. Nachdruck der Ausgabe Leipzig 1854-1866. Mit einem Vorwort und einem zusammengefaßten Quellenverzeichnis von Eberhard Nellmann. Stuttgart 1990.

Burch, Thomas et al. (Hg.) (2002): Mittelhochdeutsche Wörterbücher im Verbund. CD-ROM und Begleitbuch. Stuttgart. [CD-ROM-Version auf dem CD-ROM-Server der UB Trier; Internetversion unter www.mwv.uni-trier.de].

Gärtner, Kurt et al. (1992): Findebuch zum mittelhochdeutschen Wortschatz. Datenverarbeitung: Gerhard Hanrieder. Mit einem rückläufigen Index. Stuttgart.

Lexer, Matthias (1872-1878 [1992]): Mittelhochdeutsches Handwörterbuch. 3 Bde. Nachdruck der Ausgabe Leipzig 1872-1878. Mit einer Einleitung von Kurt Gärtner. Stuttgart.

Neues Mittelhochdeutsches Wörterbuch einschließlich digitalem Quellen- und Belegarchiv:

Digitales Mittelhochdeutsches Textarchiv: <http://www.mhgta.uni-trier.de>.

Gärtner, Kurt et al. (Hg.) (2006): Mittelhochdeutsches Wörterbuch. Erster Bd., Doppellieferung 1/2; Lieferung 1: *a-amurschaft* bearb. in der Arbeitsstelle der Akademie der Wissenschaften und der Literatur Mainz an der Universität Trier von Ralf Plate und Jingning Tao, Lieferung 2: *an – balsieren* bearb. von der Arbeitsstelle der Akademie der Wissenschaften zu Göttingen von Susanne Baumgarte, Gerhard Diehl und Bernhard Schnell. Mit einer CD-ROM. Stuttgart. [vgl. <http://www.mhdwb.uni-trier.de>]

Deutsches Wörterbuch von Jacob und Wilhelm Grimm (DWB):

DWB (1984): Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Nachdr. der Erstbearbeitung. München.

Der Digitale Grimm (2004): Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung, bearb. von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge. Hrsg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs, Benutzerhandbuch, Begleitbuch. 1. Aufl. Frankfurt a.M. Internet: <http://www.dwb.uni-trier.de>.

Pfälzisches Wörterbuch, Rheinisches Wörterbuch, Wörterbuch der elsässischen Mundarten, Wörterbuch der deutsch-lothringischen Mundarten, Luxemburgische Wörterbücher (Internet: <http://www.dwv.uni-trier.de>):

Alff, S. (Hg.) (1906): Wörterbuch der luxemburgischen Mundart. Luxemburg.

Gangler, Jean-François (1847 [2002]): Lexicon der Luxemburger Umgangssprache. Luxemburg. [Unveränderter Nachdruck Vaduz 2002].

Luxemburger Wörterbuch. Bd. 1-4 und Nachtragsband. Luxemburg 1950-1977.

Pfälzisches Wörterbuch. Begründet von Ernst Christmann, fortgeführt von Julius Krämer, bearbeitet von Rudolf Post unter Mitarbeit von Josef Schwing und Sigrid Bingenheimer. 6 Bde. Wiesbaden/Stuttgart 1965-1997.

Rheinisches Wörterbuch. Im Auftrag der Preußischen Akademie der Wissenschaften, der Gesellschaft für Rheinische Geschichtskunde und des Provinzialverbandes der Rheinprovinz auf Grund der von Johannes Franck begonnenen, von allen Kreisen des Rheinischen Volkes unterstützten Sammlung bearbeitet und herausgegeben von Josef Müller, Heinrich Dittmaier, Rudolf Schützeichel und Matias Zender. 9 Bde. Bonn/Berlin 1928-1971.

Wörterbuch der deutsch-lothringischen Mundarten. Bearbeitet von Ferdinand Follmann. Leipzig 1909. [Nachdruck Hildesheim/New York 1971].

Wörterbuch der elsässischen Mundarten. Bearbeitet von Ernst Martin und Hans Lienhart. 2 Bände. Straßburg 1899-1907. [Nachdruck Berlin/New York 1974].

[Vgl. hierzu auch: Fournier, Johannes (2003): Vorüberlegungen zum Aufbau eines Verbundes von Dialektwörterbüchern. In: Zeitschrift für Dialektologie und Linguistik 70, S. 155-176.]

Goethe-Wörterbuch (GWB):

GWB (1978ff.): Goethe-Wörterbuch. Hg. v. der Berlin-Brandenburgischen Akademie der Wissenschaften [bis Bd. 1, 6. Lfg.: Deutsche Akademie der Wissenschaften zu Berlin; bis Bd. 3, 4. Lfg.: Akademie der Wissenschaften der DDR], der Akademie der Wissenschaften in Göttingen und der Heidelberger Akademie der Wissenschaften. Bd. 1 (*A-azurn*), 1978; Bd. 2 (*B-einweisen*), 1989; Bd. 3 (*einwenden-Gesäusel*), 1989; Bd. 4, Lieferung IV/1-10 (*Geschäft-hinzutreten*). Internet: <http://www.gwb.uni-trier.de>.

Ökonomische Enzyklopädie von J. G. Krünitz:

Krünitz, Johann Georg (1773-1858): Oeconomische Encyclopädie oder allgemeines System der Land-, Haus- und Staats-Wirthschaft: in alphabetischer Ordnung. Bd. 1-242. Berlin. Internet: <http://www.kruenitz.uni-trier.de>.

Fons Moerdijk

Das „Algemeen Nederlandse Woordenboek“ (ANW) und *ellexiko* – ein Vergleich

Abstract

ANW und *ellexiko* repräsentieren eine neue Generation von wissenschaftlichen elektronischen (Online-) Wörterbüchern: sie sind keine digitalisierten Klone von schon existierenden Printwörterbüchern, sondern werden inhaltlich neu und mit voller Berücksichtigung der Möglichkeiten des neuen Mediums realisiert. In diesem Beitrag werden zuerst pauschal einige wichtige Parallelen und Unterschiede zwischen dem ANW und *ellexiko* beleuchtet. Anschließend wird der substanzielle Unterschied in den Suchoptionen eingehend behandelt. *ellexiko* hantiert mit dem Unterschied „einfache Suche“ neben „Expertensuche“ – ein bekanntes System. Das ANW hat ein eigenes, neues System mit den folgenden Suchmöglichkeiten entwickelt: Suche nach Information zu einem Wort, Suche nach einem Wort (von der Bedeutung aus), Suche nach Wörtern (auf Grund eines oder mehrerer gemeinschaftlicher Merkmale), Suche nach Beispielen mit gemeinsamen Merkmalen und Suche nach Information über das Wörterbuch selbst. In den onomasiologischen Suchformen, die vom Inhalt zum Wort führen, spielt das „Semagramm“, die Darstellung von Kenntnis, die mit einem Wort zu verbinden ist, in einem Rahmen mit „Slots“ und „Fillern“ eine substanzielle Rolle. Das Semagramm ist eine weitere Erneuerung des ANW.

The ANW and *ellexiko* represent a new generation of scholarly online dictionaries: they are no digitalised copies of existing printed dictionaries, but they are, as far as content and technology is concerned, developed from scratch as brand-new products. First a short overview is given of the main similarities and differences between the two dictionaries. Then the differences in search options offered by *ellexiko* and the ANW are discussed in more depth. *ellexiko* follows the well-known system with the distinction ‘simple search’ and ‘advanced search’. The ANW has created a new system with the following search options: search for information about a word, search for a word (with the meaning as starting-point), search for words (on the basis of one or more common features), search for examples (with common features) and search for meta-information about the dictionary itself. In the onomasiological searches from content to word the ‘semagram’, the representation of conceptual knowledge connected with a word in a frame with ‘slots’ and ‘fillers’, plays a prominent role. The semagram is also an innovation of the ANW.

Inhalt:

1. Parallelen und Unterschiede zwischen ANW und *ellexiko*
 - 1.1 Parallelen
 - 1.2 Unterschiede
2. Suchformen im Allgemeinen
 - 2.1 Das Dichotomiesystem
 - 2.2 Einwände gegen das Dichotomiesystem
3. Spezielle Suchformen bei *ellexiko* und im ANW
 - 3.1 Die *ellexiko*-Suche
 - 3.2 Die ANW-Suche
4. Abschluss
5. Literatur
6. Nachschlagewerke

Das *elexiko*-Wörterbuch,¹ ein Produkt des Instituts für Deutsche Sprache, ist schon einige Jahre im Internet zu bewundern. Etwa 800 Wortartikel zu hochfrequenten Stichwörtern sind bereits redaktionell bearbeitet und publiziert. Es bedarf also keiner näheren Einführung. Das gilt nicht für das „Algemeen Nederlandse Woordenboek“ (ANW). Das ANW wird ein Online-Wörterbuch der niederländischen Gegenwartssprache und ist ein Projekt des Instituts für Niederländische Lexikologie (INL) in Leiden. Im Gegensatz zu *elexiko* ist es noch nicht an die Öffentlichkeit gebracht. Erst mit Anfang des Jahres 2009 sind die ersten Resultate im Internet zu erwarten.

Mit dem ANW wurde im Jahre 2001 begonnen. Der Zeitraum von 2001 bis 2006 stand im Zeichen zweier Hauptaufgaben: Korpusaufbau und Konzeptentwicklung. 2006 wurden Probeartikel verfasst. Sehr viel Zeit hat man der inhaltlichen Analyse von Definitionen aus anderen Wörterbüchern gewidmet, auf deren Grundlage der Wortschatz in Bedeutungsklassen unterteilt wurde. Aus diesen jeweiligen Klassen wurden Schablonen für die semantische Beschreibung von Wörtern aufgestellt.

In diesem Beitrag behandle ich zuerst, in sehr groben Zügen, einige wichtige Parallelen und Unterschiede zwischen dem ANW und *elexiko*. Anschließend werde ich näher auf einen sehr substanziellen Unterschied, der mit der Präsentation im Internet zusammenhängt, und zwar den Unterschied bei den Suchoptionen, eingehen. Ich habe mich dafür entschieden, weil das ANW sich in dieser Hinsicht nicht nur von *elexiko*, sondern meines Wissens auch von anderen wissenschaftlichen Online-Wörterbüchern unterscheiden wird.

1. Parallelen und Unterschiede zwischen ANW und *elexiko*

1.1 Parallelen

ANW und *elexiko* repräsentieren eine neue Generation von wissenschaftlichen elektronischen (Online-)Wörterbüchern. Ich spreche von einer neuen Generation; erstens, weil sie keine digitalisierten Klone von schon existierenden Printwörterbüchern sind, sondern inhaltlich einen ganz neuen Anfang machen; und zweitens, weil sie von Anfang an in ihrem Entwurf die speziellen neuen Möglichkeiten und Perspektiven des Wörterbuchs als Hypertext berücksichtigt haben (Moerdijk 2007a).

Beide Wörterbücher beschreiben die allgemeine (vorwiegend geschriebene) Gegenwartssprache (unter Berücksichtigung der nationalen Sprachvarietäten, der „Natiolekte“). Sie wollen dabei einem breiten Spektrum von Benutzern (von Laien bis zu Sprachwissenschaftlern) dienen. Wie *elexiko* widmet das ANW Neologismen spezielle Aufmerksamkeit.

Eine dritte Übereinstimmung ist, dass sie „korpusbasiert“ sind. Sie werden zusammengestellt mithilfe eines Korpus im modernen Sinne einer Sammlung elektronischer Texte, die auf repräsentative Weise die Sprachwirklichkeit, die man beschreiben will, widerspiegeln.

Weiterhin haben beide eine sehr umfassende, reichlich verzweigte Artikelstruktur. Das ist teils eine Konsequenz der vielgestaltigen Zielgruppen, teils bedingt durch Bedürfnis-

¹ *elexiko* wird im Folgenden im Stand von Mai 2007 präsentiert, da sich am hier Beschriebenen inhaltlich nichts Grundsätzliches geändert hat.

se der Wissenschaftler, die Material für ihre Untersuchungen brauchen. Sie stellen vor allem Fragen, die von bestimmten Merkmalen aus zu den Wörtern führen, die diese Merkmale besitzen. Es sind gerade solche Fragen, die eine Ausbreitung der Suchmöglichkeiten und deshalb der Mikrostruktur erfordern. Für die Festlegung dieser Struktur benutzen sie die Markierungssprache XML. Jeder Wörterbuchartikel ist ein XML-Dokument und wird als solches in einer Datenbank gespeichert.

1.2 Unterschiede

Obwohl beide Wörterbücher ein Korpus als empirische Basis haben, gibt es doch auch einen wichtigen Unterschied. *ellexiko* konnte sich eines bestehenden Korpus bedienen: des Deutschen Referenzkorpus des IDS Mannheim. Das *ellexiko*-Korpus umfasst derzeit ca. 1,3 Milliarden Textwörter und besteht aus Zeitungs- und Zeitschriftentexten. Das ANW hat sich sein Korpus selbst bauen müssen. Das hat einen großen Teil der Vorbereitungsperiode beansprucht. Das ANW-Korpus hat einen Umfang von 104 Millionen Textwörtern. Wir haben gezielt eine Diversität von Themen und Genres angestrebt. Für etwa 1500 verschiedene Gesellschaftsgebiete wurden repräsentative Texte aus dem Internet heruntergeladen. Auch ein Teilkorpus aus literarischen Texten wurde zusammengestellt. Nur ungefähr 40% der Korpus-Texte stammten aus Zeitungen und Zeitschriften.

Beide Wörterbücher benutzen, wie gesagt, die Markierungssprache XML, aber auch hier gibt es einen Unterschied. *ellexiko* gebraucht als XML-Editor *XMetal*, das ANW verwendet *Altova XMLSpy* und verfügt dabei über ein selbst entwickeltes, „lexikografenfreundliches“ Bearbeitungsformular.

Die Bedeutungsbeschreibung besteht bei *ellexiko* im Anschluss an die lexikografische Tradition aus einer Definition, aus einer erklärenden Paraphrase. Beim ANW ist sie zweiteilig und besteht aus Definition und so genanntem „Semagramm“. Das Semagramm ist eine Innovation des ANW. Ein Semagramm ist die Darstellung der Kenntnis, die mit einem Wort verbunden ist, in einem Rahmen mit „Slots“ und „Fillern“. „Slots“ sind abstrakte konzeptuelle Strukturelemente, die sich auf Merkmale und Relationen beziehen, die für die Bedeutungsklasse, wozu ein Wort gehört, gelten. Sie werden mit konkreten Daten, die für das betreffende Wort gelten (die „Filler“), ausgefüllt. Ich komme später noch darauf zu sprechen (siehe auch Moerdijk 2007b).

Beim ANW findet die Bearbeitung nicht alphabetisch und nach Frequenzgrad statt, sondern nach so genannten „lexikografischen Typen“, Wörtern, die wegen gemeinsamer syntaktischer, morphologischer und semantischer Eigenschaften eine eigenständige Gruppe ausmachen. Ein solches Verfahren sollte zu besseren, einheitlicheren und systematischeren lexikografischen Beschreibungen und Erklärungen führen.

2. Suchformen im Allgemeinen

2.1 Das Dichotomiesystem

Ein wesentlicher Vorteil des elektronischen Wörterbuchs ist, dass es verschiedene Zugriffswege zu den Wörterbuchdaten bietet, während das prototypische Wörterbuch nur den alphabetischen Weg über die Stichwörter zu bieten hat. Im alphabetischen Druckwörterbuch geht man von der Form zum Inhalt. Das elektronische Wörterbuch besitzt diese Möglichkeit auch, darüber hinaus bietet es mehrere Suchmöglichkeiten, die

von inhaltlichen Merkmalen zu den Wortformen führen. Mit einem bekannten terminologischen Unterschied aus der traditionellen Lexikologie kann man hier von semasiologischen (von Form zu Inhalt) und onomasiologischen Suchoptionen (von Inhalt zu Form) sprechen.

Im Allgemeinen werden die verschiedenen Suchmöglichkeiten von den Wörterbüchern selbst in Dichotomien wie „basiszoeken“ gegenüber „uitgebreid zoeken“ (i-WNT), „simple search“ gegenüber „advanced search“ (i-OED), „recherche d'un mot“ versus „recherche assistée“ und „recherche complexe“ (i-TLF), „Standardmodus“ gegenüber „Expertenmodus“ oder „erweiterte Suche“ (GWDS, auf CD-ROM) dargeboten. Wir benennen dieses System in der Folge der Einfachheit halber als „das Dichotomiesystem“.

Die einfache Suche ist die Stichwortsuche, die Suche, die zum Artikel führt. Wenn wir von den Unterschieden in der Zahl der gebotenen Möglichkeiten und in der Präsentationsweise absehen, dann können wir sagen, dass es sich im Allgemeinen bei der ausführlichen, fortgeschrittenen Suche um eine Ausweitung der Suchoptionen und um die Möglichkeit von komplexeren Suchanfragen handelt, wie die Suche mit mehreren Suchtermini, die Suche in bestimmten Kategorien oder Kombinationen von Kategorien, die Suche mit Operatoren, die Auswahl des Kontextbereiches und Ähnliches.

2.2 Einwände gegen das Dichotomiesystem

Ein großer Nachteil des Dichotomiesystems ist, dass es alles andere als benutzerfreundlich ist. Die Benutzer wissen gar nicht, welche Optionen sich hinter Begriffen wie „einfache Suche“, „basiszoeken“, „simple search“, „komplexe Suche“, „erweiterte Suche“ oder „Expertensuche“ verstecken, sondern finden dies erst mit zunehmender Erfahrung heraus.

Und selbst wenn man sich einmal auf den betreffenden Webseiten befindet, ist häufig überhaupt nicht klar, was man nun mit all den angebotenen Auswahl Fenstern, Menüleisten und Buttons anfangen soll, um die erwünschten Suchergebnisse zu bekommen. Das ist insbesondere der Fall bei komplexen Suchanfragen, z.B. der Expertensuche (advanced search), wo die Benutzer ihre Suchanfrage durch das Ausfüllen von Eingabefeldern und das Anklicken von Menüpunkten in mehreren Auswahl Fenstern selbst zusammensetzen müssen. In vielen Fällen ist der sprachliche und computertechnische Jargon dabei ein zusätzliches Problem. Selbstverständlich gibt es Hilfetexte. Wer das Internet konsultiert, will aber schnell, direkt und deutlich informiert werden und vermeidet Aktionen, die dem im Wege stehen.

Ein weiteres Problem ist, dass (auch in den Hilfetexten) die Formulierungen der Suchoptionen häufig nicht mit den Suchanfragen, wie die Benutzer sie tatsächlich im Kopf haben, übereinstimmen. Diese Unzulänglichkeit wird verursacht durch die Vermischung der Suche des Computers mit der Suche der Benutzer. Ich gebe ein Beispiel: Bei dem iWNT hat man eine Suchoption „zoek woord in betekenis“. Sie ist vergleichbar mit der Suchfunktion „simple search on definitions“ im OED. Im Hilfetext wird das Ziel dieser Suchfunktion mit dem Beispiel „to search for the word *generous* in definition text“ näher erklärt. Die Benutzer suchen aber gar nicht nach dem Wort *generous* in allen Definitionen oder nach allen Definitionen, die das Wort *generous* enthalten. Sie suchen jedoch nach allen Wörtern, die Synonyme von *generous* sind (oder die in einer anderen bestimmten semantischen Relation mit *generous* stehen). Es ist der Computer, der in allen

„Definitionen“ des vollständigen Datenbestandes suchen soll, um diesen Suchwunsch der Benutzer zu erfüllen. Auch Suchfunktionalitäten wie „Suche im Volltext“ oder „Suche in Teilen des Artikels“ sind Formulierungen, die auf das Suchgebiet des Computers gerichtet sind und nicht auf dasjenige, was die Benutzer tatsächlich finden wollen.

Schließlich möchte ich bemerken, dass im Dichotomiesystem bei Weitem nicht alle Möglichkeiten des „umgekehrten“ Suchens, des Suchens von Merkmalen zu Wortformen, ausgeschöpft worden sind. Man hat sich mehr oder weniger beschränkt (besser: beschränken müssen) auf Suchoptionen, die auf einer Auswahl von Angabeklassen basieren, die mit einer begrenzten Anzahl von festen Angabewerten zu der Mikrostruktur gehören (zum Beispiel: Wortart, mit den Werten „Adjektiv“, „Adverb“, „Nomen“, „Verb“ usw.). Das Angebot wird eher durch praktische Verfügbarkeit als durch theoretische Überlegungen bestimmt. Im Prinzip können jedoch alle Kategorien der Mikrostruktur Grundlage für solche umgekehrten Suchanfragen sein.

3. Spezielle Suchformen bei *ellexiko* und im ANW

3.1 Die *ellexiko*-Suche

ellexiko unterscheidet zwischen der einfachen Suche nach einem Stichwort, mit der man zum Artikel gelangt, und der „Expertensuche“ und gehört also zu den Wörterbüchern mit einem „Dichotomiesystem“. Bei dieser Expertensuche kann man nach allen Wörtern mit bestimmten Merkmalen in einem der folgenden Bereiche suchen: Orthografie, Wortart, Grammatik, Wortbildung, sinnverwandte Wörter und semantische Klasse. Bei diesen Angabeklassen kann man aus einer begrenzten, in einer Menüleiste präsentierten Reihe von möglichen Werten den zutreffenden Wert durch Anklicken auswählen. Bei Wortart werden „beliebig“, „Adjektiv“, „Adverb“, „Nomen“, „Verb“ als Werte angeboten, bei Wortbildung „beliebig“, „Komposita“, „explizite Derivate“, „Konversionen“, „Kurzwörter“ usw. „Beliebig“ ist immer der Defaultwert. Das ist der Wert, den die Benutzer anfangs im Auswahlfenster zu sehen bekommen. (Man würde den Benutzern meiner Meinung nach übrigens einen besseren Anhaltspunkt bieten, wenn man nicht „beliebig“, sondern einen der anderen, inhaltlich signifikanteren Wert als Anfangswert zeigen würde.)

Die Einwände, die wir im Allgemeinen gegen das Dichotomiesystem vorbrachten, treffen in gewissem Maße auch auf *ellexiko* zu. Benutzer werden sich auch hier fragen, was sie unter „Expertensuche“ zu „suchen haben“. Was den Umfang der dargestellten Suchoptionen betrifft, gilt, wie für die anderen Wörterbücher, dass nur ein sehr geringer Teil der potenziellen Funktionalitäten, wenigstens bis jetzt, realisiert ist. Die reiche Mikrostruktur des *ellexiko*-Wörterbuchs bietet sicherlich viel mehr Möglichkeiten für Suchanfragen, die auf Angabeklassen oder festen Angabewerten basieren.

Es gibt aber auch Verbesserungen. Die Präsentation ist unkompliziert und übersichtlich und die Kategorien und ihre Eingabefelder werden präsentiert mit der Überschrift: „Suche nach Stichwörtern mit bestimmten Merkmalen“. Diese Formulierung schließt ohne Zweifel schon besser an das an, was die Benutzer haben wollen (obwohl „Suche nach allen Wörtern mit bestimmten Merkmalen“ vielleicht noch besser gewesen wäre).

3.2 Die ANW-Suche

Das ANW wird sich nicht an diese offensichtlich schon etablierte Ordnung des Dichotomiesystems anschließen. Es will eine positive Lehre aus dessen Schwächen ziehen und strebt deshalb nach einem benutzerfreundlichen System, das folgende Bedingungen erfüllt:

- Die Benutzer sollten möglichst direkt und klar die Informationen bekommen, die sie sich wünschen.
- Die Suchformulare sollten in solcher Weise formuliert und präsentiert werden, dass die Benutzer gleich sehen, wo und wie sie ihre Suchanfragen stellen müssen.
- Die Formulierungen der Suchoptionen sollten nicht auf das abgestimmt sein, was der Computer suchen soll, sondern auf das, was die Benutzer suchen.
- Die Benutzer sollten möglichst wenig Fragen selbst konstruieren müssen.
- Die Suchmöglichkeiten sollten erheblich ausgeweitet werden, namentlich jene, die von Merkmalen zu Wörtern führen.

Dies hoffen wir zu erreichen, indem wir folgende Suchformen unterscheiden (die Begriffe „Wort“ und „Wörter“ sind hier großzügig zu betrachten; andere lexikalische Einheiten wie Phrasen, Sprichwörter und Namen sind mit eingeschlossen):

- 1) Suche nach Informationen über ein Wort (Wort → Artikel)
- 2) Suche nach einem Wort (von der Bedeutung aus) (Wort ← Bedeutung)
- 3) Suche nach Wörtern (aufgrund eines oder mehrerer gemeinschaftlicher Merkmale) (Wörter ← Merkmale)
- 4) Suche nach Beispielen/Beispielsätzen mit gemeinsamen Merkmalen (Beispielsätze ← Merkmale)
- 5) Suche nach Information über das Wörterbuch selbst.

Bei der **Suche nach Informationen über ein Wort** bekommt man durch Eingabe in ein Eingabefeld oder durch Anklicken eines Wortes in einer Stichwortliste Zugang zu dem Artikel, der dem entsprechenden Wort gewidmet ist. Es ist die Suchform, die mit der traditionellen Suche in dem alphabetischen, semasiologischen Printwörterbuch korrespondiert. Man bekommt hier Antwort auf Fragen vom Typ „Was bedeutet Wort x?“, „Wie schreibt man Wort x?“, „Wie spreche ich Wort x aus?“, „Kommt Wort x auch im Singular vor?“, „Was ist die Herkunft von Wort x?“ usw. Es sind *par excellence* Fragen, die das Publikum sich stellt. Von einer allgemeinen Anfangsseite aus kann man mit Hyperlinks die Webseiten der einzelnen Artikelsegmente erreichen. Daneben hat man die Möglichkeit, den Artikel als Ganzes anzufordern.

Die **Suche nach einem Wort** soll Antwort geben auf Fragen vom Typ „Wie heißt dies?“, „Wie nennt man dies?“, „Was ist der Name dafür?“. Man geht von der Bedeutung, von inhaltlichen Elementen, hin zum Wort. Es ist die onomasiologische Fragestellung im reinen, strengsten Sinne. Solche Fragen stellen sich im täglichen Leben, wenn Wörter uns nicht einfallen, und sind außerdem beliebte Quizfragen. Weil der Gang von

der Bedeutung zum Wort hin nur gut gelingt, wenn die Wortwahl in der Definition mit der Wortwahl der Benutzer zusammenfällt – d.h. in den bestehenden elektronischen Wörterbüchern bis jetzt fast nie –, haben wir in unserer Struktur auch einer neuen Art und Weise der Bedeutungsbeschreibung, dem „Semagramm“, einen Platz eingeräumt. Wie wir schon sagten, ist ein Semagramm die Darstellung von Kenntnissen, die mit einem Wort zu verbinden sind, in einem Rahmen mit „Slots“ und „Fillern“. „Slots“ sind abstrakte konzeptuelle Strukturelemente, die sich auf Merkmale und Relationen beziehen, die für die Bedeutungsklasse, zu der ein Wort gehört, gelten. Z.B. für die Klasse der Tiernamen, mit „Säugetier“, „Vogel“, „Fisch“, „Insekt“, „Reptil“, „Amphibie“ und „Weichtier“ als Subklassen, enthält die Klassenschablone folgende Slots:

SINNLICHE WAHRNEHMUNG

- ·Farbe
- ·Aussehen
- ·Umfang / Größe
- ·Form
- ·Geräusch, Schall, Laut

ZUSAMMENSETZUNG

- Teil
- Bau, Konstruktion

FUNKTION ODER ZIEL

- Funktion, Ziel, Bestimmung

ENTSTEHUNG

- Herkunft, Ursprung
- Fortpflanzung

RAUM, PLATZ

- Behausung
- Herkunft
- Platz, Ort (geografisch)

ZEIT

- Alter

ZUSTAND

- Zustand allgemein
- Stadium

EIGENSCHAFT / QUALITÄT

- Geschlecht
- unstoffliche Eigenschaft

HANDLUNG ODER WIRKUNG

- Bewegung
- Wirkung, Einwirkung
- Verwendung, Nutzung
- Behandlung oder Bearbeitung
- Verhalten

BETEILIGTE PERSONEN ODER SACHEN

- Produkt, Frucht

WERTSCHÄTZUNG UND BELANG

- Anerkennung

VERGLEICH

- Ähnlichkeit

BEZIEHUNG UND ZUSAMMENHANG

- Beziehung oder Relation

Nehmen wir an, dass jemand das Wort *koe* (‘Kuh’) nicht finden kann. Er weiß aber, dass es sich um ein Tier handelt, das auf dem Bauernhof lebt, öfters schwarz-weiß gescheckt ist (jedenfalls in Holland), auf der Wiese oder im Stall steht und Milch produziert. Der Benutzer kann dann solch eine Merkmalschablone durchgehen und bei jedem Merkmal sozusagen „gelenkt“ ausfüllen, was er weiß. Was er ausgefüllt hat, wird mit dem Semagramm gematcht, das der Lexikograph zusammengestellt hat und das wie folgt aussieht (hinter dem Zeichen # werden Schlüsselwörter aus den kurzen Sätzen wiederholt und mit möglichen Alternativen ergänzt, um das Retrieval zu verbessern; bei dem Artikel *koe*, hier im Beispiel als *Kuh* wiedergegeben, sehen die Benutzer nur die Sätzchen):

EINE KUH

OBERKATEGORIE:	ist ein Tier # Tier; Säugetier; Wiederkäuer
KATEGORIE:	ist ein Rind # Rind; Wiederkäuer
FARBE:	ist öfters schwarz-weiß gescheckt, aber auch braun-weiß, ganz schwarz, braun oder weiß # schwarz-weiß; schwarz-bunt; braun-weiß; rotbunt; bunt; schwarz; schwarz gescheckt; weiß; braun; gescheckt; rostbraun
UMFANG:	ist groß # groß

GERÄUSCH:	muht, brüllt, macht ein Geräusch, das wir mit einem tiefen, langgezogenen „Muh“ nachahmen # muht; brüllt; muhen; brüllen; muh
TEIL:	hat ein Euter, Hörner und vier Mägen: Pansen, Netzmagen, Blättermagen, Labmagen # Euter; Hörner; Mägen; Pansen; Netzmagen; Blättermagen; Labmagen; Käse-magen
BAU:	ist grobknochig, grobgliedrig, von derbem Gliederbau # grobknochig; grobgliedrig
FUNKTION:	gibt Milch und Fleisch # Milch; Fleisch; Rindfleisch; Milchproduktion; Fleischproduktion
PLATZ:	lebt auf einem Bauernhof; steht auf der Wiese und im Winter im Stall # Bauernhof; Wiese; Weide; Weideland; Stall
ALTER:	ist erwachsen, hat gewöhnlich schon einmal gekalbt # erwachsen; gekalbt
GESCHLECHT:	ist weiblich # weiblich; Weibchen; weibliches Rind
EIGENSCHAFT:	ist nützlich und zahm; wird betrachtet als ein freundliches, faules, träges, dummes, neugieriges, soziales Tier # nützlich; zahm; freundlich; träge; dumm; neugierig; sozial
VERHALTEN:	grast und kät wieder # grast; grasen; kät wieder; wiederkauen
BEHANDLUNG:	wird täglich gemolken; wird geschlachtet # melken; schlachten
PRODUKT:	gibt Milch und Fleisch # Milch; Fleisch
WERTSCHÄTZUNG:	ist nützlich # nützlich

Damit ist die Chance, dass das Wort *koe* auch tatsächlich vom Computer gefunden wird, beträchtlich größer geworden. In Wörterbüchern, die *Kuh* definieren als „erwachsenes weibliches Rind“, wäre dies nicht gelungen.

Die Suche nach Wörtern findet statt auf der Grundlage eines Merkmals oder einer Kombination mehrerer Merkmale, die die Wörter gemeinsam haben. Im Prinzip können alle Kategorien, Subkategorien und festen Angabewerte der Mikrostruktur als Ausgangspunkte für solche Fragen benutzt werden. Es handelt sich hier um Fragen oder Suchaufträge wie: Zeige mir alle Wörter mit *-iser-* oder *-izer-* in der Mitte (Orthografie/Wortform), zeige mir alle Wörter mit englischer Aussprache oder alle Wörter mit Endbetonung (Aussprache), zeige mir alle Substantive (Wortart), alle Komposita mit einem Imperativsatz als erstem Glied (Wortbildung), zeige mir alle Wörter für Fahrzeuge (Bedeutung/Begriffsklasse), alle Wörter aus dem Golfspiel (gesellschaftlicher Bereich), alle französischen Lehnwörter (Etymologie), alle Neologismen aus den Jahren 2000-2003 (Datierung/Etymologie) usw. Auch kombinierte Fragen sind möglich, zum Beispiel: Zeige mir alle Substantive, die in den Niederlanden einen Plural auf *-en*, aber in Belgien

einen Plural auf -s haben (Kategorien: Wortart, Substantiv Pluralform, Sprachvarietät). Diese Art von Fragen wird vor allem von Fachleuten, die Material für ihre Untersuchungen brauchen, gestellt. Onomasiologische Fragen sind hier die Fragen nach Benennungen für Begriffe oder nach Wörtern innerhalb bestimmter gesellschaftlicher Domänen. Insofern andere Fragen von Merkmalen zu Wortformen führen, könnte man auch von onomasiologischen Fragen sprechen, wobei dann onomasiologisch in weiterem (und nicht ganz exaktem) Sinn aufzufassen ist.

Die Suche nach Beispielsätzen bietet u.a. die Möglichkeit, Beispielsätze aus einer bestimmten Quelle zu sammeln, mit einem bestimmten Wort oder mit mehreren bestimmten Wörtern, von einem bestimmten Autor oder einer bestimmten Autorin, aus einem bestimmten Jahr oder einem bestimmten Zeitabschnitt. Auch hier sind Kombinationen möglich. Diese Fragen beziehen sich nicht auf Wörter und fallen deshalb aus der semasiologischen wie der onomasiologischen Ordnung. Es sind übrigens meiner Meinung nach Fragen, die man eher einem Korpus als einem Wörterbuch stellen sollte. Man kann aber das Wörterbuch als ein Korpus benutzen.

Bei der Suche nach Informationen über das Wörterbuch selbst geht es um Fragen wie: Was für ein Wörterbuch ist das ANW, welche literarischen Quellen hat es benutzt, wie viele lexikalisierte Wortgruppen hat es? Welche Artikel wurden vom Lexikografen X gemacht? Wie viele belgisch-niederländische (flämische) Wörter enthält es, welche Abbildungen sind im Wörterbuch aufgenommen? Das Suchgebiet des Computers ist hier nicht nur das Wörterbuch, sondern auch die Titelei und die verschiedenen Dokumentatexte und Hilfetexte des Wörterbuchs und des ANW-Korpus.

4. Abschluss

Das ANW ist zurzeit (Ende 2007) mit den Vorbereitungen für die Publikation der Resultate aus der ersten Bearbeitungsphase im Internet beschäftigt. Mit Informatikern der Abteilung „Automatisierung“ des Instituts für niederländische Lexikologie wird an dem so genannten funktionellen Entwurf gearbeitet, der im April 2008 fertig sein soll. Dann folgen der technische Entwurf, der grafische Entwurf und der Bau eines Prototyps der Anwendung. Im November beginnt das Stadium der Realisierung, Implementation und der Tests. Anfang Mai 2009 soll das ANW im Internet zugänglich sein – und somit eine weitere Übereinstimmung mit *ellexiko* haben.

5. Literatur

- Moerdijk, Fons (2007a): Een nieuwe generatie elektronische woordenboeken. In: *Philologia Frisica Anno 2005. Lezingen fan it sechtjinde Frysk filologekongres 14, 15 en 16 desimber 2005*. Leeuwarden. S. 195-217.
- Moerdijk, Fons (2007b): Definities, frames en semagrammen. In: Moerdijk, Fons/Tempelaars, Rob/Van Santen, Ariane (Hg.): *Leven met woorden*. Leiden. S. 63-75.

6. Nachschlagewerke

- GWDS = Duden – Das große Wörterbuch der deutschen Sprache (2000). 10 Bde. auf CD-ROM. Mannheim.
- ioED = Oxford English Dictionary Online: <http://dictionary.oed.com/>
- iTLF = Le Trésor de la Langue Française informatisé: <http://atilf.atilf.fr/tlf.htm>
- iWNT = Woordenboek der Nederlandsche Taal op internet: <http://www.wnt.inl.nl>

Jörg Asmussen

***ordnet.dk* und *elexiko* – ein Vergleich**

Abstract

Sowohl *OWID/elexiko* für das Deutsche als auch *ordnet.dk* für das Dänische gehören zu den Pionieren bei der Nutzung des Web für wörterbuchähnliche Wortschatz-Informationssysteme. Da sich im Web viele unterschiedliche herkömmliche mediale Funktionen modellieren lassen, z.B. Druckmedien oder audiovisuelle Medien, und darüber hinaus eine unmittelbare Interaktion mit dem Benutzer möglich ist, stellt die Konzeption eines Wortschatz-Informationssystems eine komplexe Herausforderung für die traditionelle buchorientierte Lexikografie dar. Auf dem Hintergrund der Frage, inwiefern das klassische Buchmedium – und hiermit verbundene Interaktionsmuster – in den beiden Wortschatz-Informationssystemen nicht nur einfach virtuell nachgebaut, sondern nutzbringend erneuert wird, soll ein Vergleich von *elexiko* und einem Prototyp der Wörterbuchkomponente von *ordnet.dk* durchgeführt werden. Hierbei wird deutlich, dass *elexiko* und *ordnet.dk* etwas unterschiedliche Schwerpunkte haben: Das Ausloten mikrostruktureller Möglichkeiten zur möglichst umfassenden lexikografischen Beschreibung bei *elexiko* bzw. die Erweiterung der Wörterbuchfunktionen bei *ordnet.dk*, z.B. bezüglich der Recherche.

Both *OWID/elexiko* for German and *ordnet.dk* for Danish may be considered as pioneers in the field of using the web as a base for dictionary-like lexical information systems. Because the web permits modeling of many different traditional media, e.g. printed or audio-visual, and because it facilitates immediate user interaction, the design of a lexical information system becomes a complex challenge for traditional book-oriented lexicography. Based on the question, to what extent the classical book medium and the typical ways of interacting with it are renewed fruitfully and not just virtually copied, we aim to compare *elexiko* and a prototype of the dictionary component of *ordnet.dk*. As a result of this, it becomes clear that *elexiko* and *ordnet.dk* are focusing on slightly differing issues: Whereas *elexiko* emphasizes exploring new micro-structural possibilities in order to achieve a comprehensive lexicographic description, *ordnet.dk* prioritizes functionality of the dictionary, e.g. new ways of looking up lexical information.

Inhalt:

1. Wörterbücher und Korpora der DSL
2. Das Projekt *ordnet.dk*
3. Charakteristika elektronischer Wörterbücher
4. *elexiko* vs. *ordnet.dk*
 - 4.1 Gesamtpräsentation im Web
 - 4.2 Lesarten
 - 4.3 Bedeutungserläuterungen
 - 4.4 Belege
 - 4.5 Semantische Umgebung
 - 4.6 Typische Verwendungen
 - 4.7 Sinnverwandte Wörter
 - 4.8 Grammatische Angaben
 - 4.9 Wortbildung
 - 4.10 Zugriffsmöglichkeiten
5. Vergleich
 - 5.1 Gelockerte strukturelle Bindungen
 - 5.2 Erweiterte Zugriffsmöglichkeiten
 - 5.3 Anpassbare Präsentation
 - 5.4 Vernetzung mit anderen Ressourcen
 - 5.5 Einbeziehung des Benutzers
6. Schlussfolgerungen
7. Literatur

1. Wörterbücher und Korpora der DSL

Das Projekt *ordnet.dk* wurde im Jahre 2004 von der Gesellschaft für dänische Sprache und Literatur (*Det Danske Sprog- og Litteraturselskab*, DSL) ins Leben gerufen. Die 1911 gegründete, akademieähnliche Gesellschaft erarbeitet editionsphilologische Textausgaben, sprachgeschichtliche Darstellungen, Bibliografien sowie Wörterbücher und Textkorpora, wobei IT-gestützte Erstellung und Publikation eine immer größere Rolle spielen. Das Projekt *ordnet.dk* ist im Bereich Wörterbücher und Textkorpora angesiedelt und wird im Wesentlichen aus Mitteln des dänischen Kulturministeriums sowie der Carlsberg-Stiftung finanziert.

Die Ziele des Projekts *ordnet.dk* sind die Verknüpfung und zum Teil die Erweiterung dreier bei der DSL erarbeiteter Wortschatz-Ressourcen, nämlich des Wörterbuchs der dänischen Sprache (*Ordbog over det danske Sprog*, ODS, Dahlerup u.a. 1956), des Dänischen Wörterbuchs (*Den Danske Ordbog*, DDO, Hjorth/Kristensen/Lorentzen/Trap-Jensen/Asmussen u.a. 2005) sowie des Textkorpus *Korpus 2000* (vgl. Andersen/Asmussen, H./Asmussen J. 2002). Das Ergebnis soll spätestens Ende 2009 einer sprachlich interessierten Öffentlichkeit als portalähnliches Wortschatz-Informationssystem für Dänisch unter der Bezeichnung *ordnet.dk* ('Wortnetz') im Internet präsentiert werden.

Die inhaltlichen Grundpfeiler dieses Projekts bilden somit zwei Wörterbücher und ein Korpus: Bei dem älteren der beiden Wörterbücher, dem ODS, handelt es sich um ein historisches Wörterbuch des Dänischen in 28 Bänden sowie fünf Ergänzungsbänden. Es beschreibt die Lexik des Zeitraumes 1700-1950 auf der Grundlage überwiegend literarischer Exzerpte. Die ersten 28 Bände dieses Wörterbuchs entstanden bzw. erschienen im Zeitraum 1915-1956. Von seiner Konzeption her ist es mit dem Grimmschen Wörterbuch (Grimm 1956) vergleichbar. Als Teilprojekt von *ordnet.dk* wurde das ODS im gleichen Verfahren wie das Grimmsche Wörterbuch retrodigitalisiert, vgl. Asmussen (2003) und ist seither in einer [vorläufigen Version im Web](http://ordnet.dk/ods) zugänglich [<http://ordnet.dk/ods>].

Das DDO versteht sich im Prinzip als die Weiterführung des ODS und beschreibt den Wortschatz von etwa 1950 bis 2000. Es ist das zurzeit umfangreichste und ausführlichste Wörterbuch zur dänischen Gegenwartssprache und umfasst sechs Bände mit etwa 100.000 Stichwörtern in 60.000 Artikeln. Beschrieben werden Wortart und Flexion, Aussprache, Lesarten, syntaktische und semantische Verwendung, feste Wendungen, semantische Relationen, Wortbildung und Etymologie. Es erhebt somit den Anspruch, universell von vielen unterschiedlichen Zielgruppen benutzt werden zu können. Die Grundlage dieses Wörterbuchs bildete ein eigens hierfür erstelltes Textkorpus, vgl. Norling-Christensen/Asmussen (1998). Das Wörterbuch wurde im Zeitraum 1991-2005 erarbeitet und entstand von Anfang an in der Gestalt eines SGML- bzw. XML-Dokuments. Ein [Prototyp der Webversion](http://dev.ordnet.dk/en) des DDO ist nach Anforderung eines Passworts zugänglich [<http://dev.ordnet.dk/en>]. Der in den folgenden Abschnitten 4 und 5 vorgenommene Vergleich von *lexiko* und *ordnet.dk* wird sich allein auf diesen Prototyp beziehen.

Das [Korpus 2000](http://korpus2000.dk) [<http://korpus2000.dk>] ist ein Referenzkorpus der dänischen Gegenwartssprache. Es umfasst etwa 56 Millionen Tokens laufenden Text aus dem Zeitraum 1982-2002. Ein Teil dieses Korpus bildete zusammen mit weiterem, urheberrechtlich geschütztem Textmaterial die Grundlage für das DDO. Das Korpus 2000 wurde in den Jahren 1991-2002 erstellt, ist auf Tokenebene mit Angaben zu Wortart und Flexion und

auf Textebene mit bibliografischen, kommunikativen und soziologisch orientierten Angaben annotiert.

2. Das Projekt *ordnet.dk*

Wesentliches Ziel des Projektes *ordnet.dk* ist es – wie bereits eingangs erwähnt – Wörterbücher und Korpusmaterial in einem webbasierten Wortschatz-Informationssystem miteinander vernetzt zugänglich zu machen. Die Aufgabenstellung ist dabei recht komplex, da bei der Lösung folgende Teilbereiche zu berücksichtigen sind:

- eine Beantwortung der Frage nach den sinnvollerweise vernetzbaren Elementen der vorhandenen Ressourcen;
- eine Harmonisierung der vorhandenen Ressourcen, die ihr Zusammenspiel ermöglichen bzw. verbessern soll;
- die Entwicklung neuer Möglichkeiten der Recherche in Wörterbuch- und Korpusdaten;
- die Entwicklung eines leicht zu bedienenden Interface, das auch dem Nichtexperten den Zugriff auf das Material gestattet;
- die fortwährende Erweiterung und Ergänzung des Wörterbuch- und Korpusmaterials.

Das Projekt ist also ziemlich umfangreich und eine Realisierung sämtlicher Teilbereiche ist eher langfristig und kaum innerhalb des gesteckten Zeitrahmens vollständig zu bewältigen. Das Projekt wurde daher in eine Reihe von eigenständigen Teilprojekten aufgeteilt, die bis zu einem gewissen Grad unabhängig voneinander gelöst werden können und die im Zuge des Gesamtprojektes eine Vorabveröffentlichung bestimmter Funktionen ermöglichen:

1. Entwicklung eines Prototyps einer DDO-Webversion, der internen Evaluierungs- und Studienzwecken dienen soll. Dieser Prototyp liegt dem folgenden Vergleich zugrunde;
2. Retrodigitalisierung des ODS und Etablierung einer vorläufigen Webversion hiervon;
3. Bearbeitung und Erweiterung des DDO, insbesondere auch seiner Dokumentenstruktur, im Hinblick auf die Webveröffentlichung. Ein Unterprojekt hierzu bildet ein WordNet-Projekt, vgl. Pedersen/Nimb/Asmussen/Sørensen u.a. (2006), durch das die onomasiologischen Recherchemöglichkeiten im Web-DDO verbessert werden sollen.
4. Erweiterung des Korpus 2000 sowie Entwicklung eines neuen Webinterface hierzu, das sich mit dem geplanten DDO-Interface verzahnen lässt.
5. Veröffentlichung einer DDO-Webversion
6. Verknüpfung des DDO mit dem Korpus, sodass eine kombinierte Ressource entsteht, in welcher der Benutzer von Wörterbucheinträgen zu Korpusbelegen und von Korpusbelegen zu Wörterbucheinträgen gelangen kann.

Ein wichtiger theoretischer Auftrag liegt außerdem darin, eine generelle konzeptuelle Beschreibung von Webwörterbüchern zu entwickeln. Hierbei stehen insbesondere Überlegungen zu möglichen funktionalen Erweiterungen und ihren Auswirkungen auf die Wörterbuchstruktur im Mittelpunkt. Es ist evident, dass ein Wörterbuch im Web auf

funktionaler Ebene die Möglichkeit potenziell stark erweiterter Interaktionsformen mit dem Benutzer bietet, die allerdings einer näheren Bestimmung und Beschreibung bedürfen. Hinsichtlich seiner Struktur wird es eher als eine generalisierte lexikalische Ressource aufzufassen sein als ein linear verlaufender Text: Eine generalisierte lexikalische Ressource ist zu einem gewissen Grad multifunktional und lässt sich nicht nur als herkömmliches Wörterbuch konfigurieren, sondern prinzipiell auch als sprachtechnologisches Hilfsmittel.

3. Charakteristika elektronischer Wörterbücher

Der in den folgenden beiden Abschnitten durchgeführte Vergleich wird sich ausschließlich auf die Komponente *ellexiko* im Online-Wortschatz-Informationssystem Deutsch (OWID) des IDS und auf den DDO-Prototyp des Wortschatz-Informationssystems für das Dänische, *ordnet.dk*, beschränken. Mit anderen Worten geht es hier nur um die Wörterbuchkomponenten der beiden Systeme. Bei beiden handelt es sich um webpublizierte Wörterbücher der jeweiligen Gegenwartssprache.

Ein *webpubliziertes Wörterbuch* wird hier nicht als ein im Web nachmodelliertes Buch verstanden, sondern medial gesehen wesentlich breiter als Wortschatz-Informationssystem. Gleiches gilt für die im Folgenden synonym verwendeten Bezeichnungen *elektronisches Wörterbuch* bzw. *digitales Wörterbuch*. *Buch* bezieht sich hier nicht auf das Medium, sondern auf den durch ein Medium vermittelten Inhalt, in diesem Fall Informationen zum Wortschatz einer Sprache.

Webpublizierte Wörterbücher müssen als besondere Variante elektronischer Wörterbücher angesehen werden, die sich in mancherlei Hinsicht von solchen unterscheiden, die sich z.B. von CD-ROM o.Ä. als eigenständige Anwendung auf einem Rechner installieren lassen. Das Web war im Unterschied zu selbstständigen Computeranwendungen lange dadurch gekennzeichnet, dass es weitgehend dieselbe lineare Textualität aufwies wie das gedruckte Medium, wohingegen die Textualität in Computeranwendungen oft nicht-linear und in rekombinierbare Einzelbestandteile zerlegt ist, z.B. in Gestalt relationaler Datenbanken.

Zwar unterliegt das Textualitätskonzept des Web zurzeit einem raschen Wandel, dennoch gilt immer noch überwiegend, dass die einsehbaren Inhalte als linear verlaufende Dokumente dargestellt werden, die allerdings im Gegensatz zu herkömmlichen Texten von der Möglichkeit anklickbarer Querverweise, so genannter Hyperlinks, Gebrauch machen können. Die Implikationen dieses Datenmodells sind im allgemeinen Sprachgebrauch mittlerweile fest verankert: So spricht man von 'Seiten im Web', ganz so als wären es Seiten in einem gedruckten Buch.

Ein weiteres Charakteristikum des Web war bisher das Verfahren der Publikation, das dadurch gekennzeichnet ist, dass von einem Herausgeber eine Webseite zur allgemeinen öffentlichen Rezeption ins Netz gestellt wurde, wiederum ganz parallel zum Buchmedium. Insofern ist das Web in seinem Ursprung gar nicht so weit vom Buchmedium entfernt: Hier wie dort werden einer interessierten Öffentlichkeit textuelle Informationen von einigen Spezialisten bereitgestellt, hierunter auch solche Informationen, die traditionell in Wörterbüchern zu Hause sind. Die typische Interaktion mit einem Webwörterbuch entspricht daher ebenfalls grundlegend der mit einem gedruckten Wörterbuch: Der Benutzer sucht nach einem Wörterbuchartikel über das zu diesem Artikel gehörige

Schlagwort und bekommt als Ergebnis die Seite(n) mit dem entsprechenden Artikel, die er nunmehr rezipieren kann. Lediglich durch zwei Merkmale scheint sich ein herkömmliches Webwörterbuch von seiner gedruckten Variante zu unterscheiden: eine potenziell schnellere Suche und mehr Platz.

Interessanterweise ist das Internet technologisch jedoch nicht auf die Rolle des Lieferanten von Seiteninhalten auf Anforderung des Benutzers festgelegt. Vielmehr lassen sich durchaus komplexe Anwendungen realisieren, bei denen die Einflussnahme des Benutzers auf das Geschehen viel größer sein kann als im herkömmlichen Web, wodurch die tatsächlichen Bedürfnisse und Anforderungen des jeweiligen Benutzers – und nicht die seitens eines Herausgebers angenommenen – im Mittelpunkt zu stehen kommen. Durch die schrittweise Realisierung solcher Möglichkeiten wurde eine Veränderung der Wahrnehmung und Nutzung des Internets eingeleitet, die gern mit *Web 2.0* bezeichnet wird, vgl. z.B. Kerres/Nattland (2007).

Der folgende Vergleich wird somit vor allem auf dem Hintergrund der Frage durchgeführt, inwiefern das klassische Buchmedium – und hiermit verbundene Interaktionsmuster – in den Webversionen der jeweiligen Wörterbücher nicht nur einfach virtuell nachgebaut, sondern nutzbringend erneuert wird. Somit steht der funktionale Aspekt der beiden Wörterbücher im Vordergrund, während strukturelle Fragestellungen zum internen Aufbau der Wörterbücher – obwohl von zentraler Bedeutung in diesem Zusammenhang – hier eher im Hintergrund bleiben werden. Der funktionale Aspekt bezieht sich also im vorliegenden Kontext im Wesentlichen auf die Zweckerfüllung eines Wörterbuches hinsichtlich der Bedürfnisse seiner Adressaten – so wie diese sie für sich selbst definieren – unter möglichst geschickter Zuhilfenahme vorhandener Technologien. Hierbei spielt die bereits in Abschnitt 2 angedeutete Tatsache, dass ein digitales Wörterbuch nicht als ein herkömmliches Buch aufzufassen ist und somit auch nicht den medialen Bindungen des Buches unterliegt, eine wesentliche Rolle. Ein digitales Wörterbuch ist im Gegensatz zum gedruckten Wörterbuch durch die folgenden funktionalen Charakteristika gekennzeichnet, wobei die ersten drei auf sämtliche digitalen Wörterbücher zutreffen, während die letzten beiden nur für digitale Wörterbücher im Internet gelten.

I. Gelockerte strukturelle Bindungen:

Gedruckte Wörterbücher unterliegen einer Reihe von strukturellen Bindungen, die durch das Buchmedium bedingt sind. Da die lexikografischen Informationen, die ein gedrucktes Wörterbuch enthält, für den Benutzer rasch auffindbar sein müssen, wird ein gedrucktes Wörterbuch in der Regel alphabetisch nach ausdrucksseitigen Repräsentationen, in aller Regel den Schlagwörtern, systematisch organisiert (Teil der Makrostruktur). Zu jedem Wort gibt es sodann einen ebenfalls systematisch aufgebauten Eintrag, aus Platzgründen in der Regel in stark kondensierter Form, die sich dem Benutzer nicht immer unmittelbar erschließt (Mikrostruktur). Zwischen einzelnen strukturellen Elementen können implizite sowie explizite Querverweise etabliert werden, die jedoch stets statischer Natur sind (Mediostruktur). Die Notwendigkeit einer fixierten Makro- und Mikrostruktur sowie der Kondensation fallen beim digitalen Wörterbuch weg. Querverweise lassen sich dynamisch, je nach Bedarf, gestalten. Die Textsorte Wörterbuchartikel als konstitutive Grundkomponente eines Wörterbuchs ist strukturell obsolet: An seine Stelle kann – in teilweiser Anlehnung an WordNets – z.B. ein komplexes Netzwerk aus lexi-

kografischen Beschreibungen und der durch sie beschriebenen ausdrucksseitigen Repräsentationen treten.

II. Erweiterte Zugriffsmöglichkeiten:

Da die Reihenfolge der Einträge im digitalen Wörterbuch nicht fixiert sein muss und sich eine Indexierung eines jeden Elements einer entsprechend ausgezeichneten Mikrostruktur vornehmen lässt, entsteht die Möglichkeit eines unmittelbaren Zugriffs auf sämtliche lexikografischen Informationseinheiten im Material. Durch geschickte Verknüpfung der einzelnen Informationseinheiten sind ungeahnte, vielfach kombinierbare Zugriffsmöglichkeiten gegeben. Zudem gestattet die Möglichkeit, sämtliche Bedienelemente grafischer Benutzerinterfaces – auch solcher, die bislang typisch nur in Desktop-Anwendungen zu finden sind – einsetzen zu können, neue Interaktionsmöglichkeiten mit dem lexikografischen Material.

III. Anpassbare Präsentation:

Da Platz bei einem digitalen Wörterbuch praktisch unbegrenzt zur Verfügung steht, fällt die Notwendigkeit einer kondensierten Darstellung weg, sodass diese wesentlich anwenderfreundlicher gestaltet werden kann. Darüber hinaus erlaubt die Lockerung der bereits angesprochenen strukturellen Bindungen aber auch eine auf die Bedürfnisse des jeweiligen Benutzers zugeschnittene, variierbare Anordnung des Materials.

Das Erscheinungsbild der Informationen muss nicht länger ein für alle Mal festgelegt sein, sondern lässt sich gezielt dem jeweiligen Benutzer und seinen Voraussetzungen, Bedürfnissen und Gebrauchskontexten von ihm selbst, automatisch oder von einer Redaktion anpassen.

IV. Möglichkeit der Vernetzung mit anderen Ressourcen:

Ein digitales Wörterbuch muss im Gegensatz zum gedruckten Wörterbuch nicht länger ein in sich geschlossenes Produkt darstellen, sondern lässt sich, sofern bestimmte strukturelle Voraussetzungen erfüllt sind, mit anderen digitalen Ressourcen vernetzen. Dies können andere lexikografische oder linguistische Ressourcen sein wie z.B. Korpora, Texte und WordNets, oder aber auch audiovisuelle Ressourcen (Aussprache, Bedeutungserklärungen durch Illustrationen, Animationen, Filme) bzw. andere Angebote im Internet wie z.B. fremdsprachige Wörterbücher oder *Wikipedia*.

V. Einbeziehung des Benutzers in den redaktionellen Prozess:

Der Benutzer muss nicht an die Rolle des Rezipienten gebunden sein, sondern kann durchaus in den redaktionellen Prozess mit einbezogen werden, dadurch dass er z.B. neue Lesarten ausfindig macht und belegt oder das Material durch neue Wörter und Verwendungen ergänzt. Benutzerbeiträge können z.B. als Kommentare oder ähnliches vom eigentlichen redaktionellen Inhalt abgegrenzt werden. Somit bekommt der Benutzer die Rolle eines sprachlichen Informanten und freiwilligen redaktionellen Mitarbeiters – ein Verfahren übrigens, das bereits bei der Erarbeitung des DDO verwendet wurde, wo eine Vielzahl freiwilliger *Spordhunde* ('Wort-Spürhunde') ihre Beobachtungen und Belege der Wörterbuchredaktion – damals noch per Post – zukommen ließen und somit einen wertvollen Beitrag zur Vollständigkeit des Wörterbuches lieferten.

4. *elexiko* vs. *ordnet.dk*

Im diesem Abschnitt werden nach einer Beschreibung der Webgesamtpräsentation der beiden Werke eine Reihe von Elementen der Mikrostruktur von *elexiko*¹ denen des DDO-Prototyps in *ordnet.dk* gegenübergestellt. Abschnitt 5 bringt einen zusammenfassenden Vergleich auf der Grundlage der oben beschriebenen funktionalen Charakteristika.

4.1 Gesamtpräsentation im Web

Abbildungen 1 bzw. 2 zeigen das unmittelbare Erscheinungsbild von *elexiko* und dem DDO-Prototyp. Bei den Abbildungen kommt es hier nicht auf die Details an, die noch in den folgenden Abschnitten beschrieben werden, sondern lediglich um die Gesamterscheinung (der Hauptseite) der beiden Webwörterbücher.



Abbildung 1: Präsentation von *elexiko* im Web

¹ *elexiko* wird im Folgenden im Stand von Mai 2007 präsentiert, da sich am hier Diskutierten inhaltlich nichts Grundsätzliches geändert hat.

The screenshot shows the DDO-Prototyp web interface. At the top, there is a navigation bar with the logo 'Ordnet.dk' and a search input field containing 'arbejd'. To the right of the search bar are several menu items: 'Ordnet' (Use wordfinder, Preferences, Help), 'Entertainment' (Guessing words, ???, ???), and 'Services' (New words, Korpus 2000, Download). Below the navigation bar, the main content area is divided into several sections. On the left, there is a 'Matching headwords' list (2) containing various forms of 'arbejde'. The central area displays the entry for 'arbejde' (3), including its etymology, meanings (1), and phrases (4). The right side of the interface features a 'More about this word' section with links to 'In the corpus' (6), 'In the dictionary', and 'Evaluation'. The interface is annotated with yellow circles and numbers 1 through 6, highlighting key features: 1 (search input), 2 (word list), 3 (main entry), 4 (meanings/phrases), 5 (example sentence), and 6 (corpus links).

Abbildung 2: Präsentation des DDO-Prototyps im Web

Bei *ellexiko* ist das Browserfenster in drei Hauptbereiche unterteilt: eine Navigationsleiste mit Schaltflächen im oberen Teil des Fensters, von denen der Benutzer zu weiteren Informationen und Diensten (z.B. einer „Expertensuche“ (6), vgl. Abschnitt 4.10) geführt wird. Die Navigationsleiste enthält außerdem ein Eingabefeld (1) für Suchanfragen sowie einen alphabetischen Index zum schnellen Auffinden von bestimmten Buchstabenstrecken innerhalb der Stichwortliste. Darunter befindet sich rechts die zum aktuellen Schlagwort gehörende lexikografische Beschreibung und links davon ein alphabetischer Ausschnitt der Stichwortliste um das aktuelle Schlagwort herum (2).

Beim DDO-Prototyp ist der Aufbau der Hauptseite auf den ersten Blick sehr ähnlich: Auch hier befindet sich im oberen Teil des Browserfensters eine Navigationsleiste mit einem Eingabefeld (1) und Hyperlinks zu weiteren Informationen und Diensten (z.B. einer onomasiologischen Suche, vgl. Abschnitt 4.10), darunter befindet sich in der Mitte die lexikografische Beschreibung und links davon eine Stichwortliste (2), allerdings im Gegensatz zu *ellexiko* nur mit den Treffern der jeweiligen Suchanfrage. Wurde weder eine spezielle (z.B. onomasiologische) Suche noch eine mit Wildcards durchgeführt, wird automatisch immer eine rechtstrunkierte Suchanfrage ausgeführt und die entsprechenden Treffer werden angezeigt. Die Trefferliste einer jeden Anfrage befindet sich im DDO-Prototyp stets in diesem Fensterbereich, wodurch eine gewisse funktionale Konsistenz des Interface erzielt wird. Bei *ellexiko* werden bei Wildcard-Anfragen gesonderte Webseiten mit den Treffern angezeigt, die hier dann als Hyperlinks zu den Beschreibungen gegeben werden. Rechts von der lexikografischen Beschreibung befinden sich im DDO-Prototyp Hyperlinks zu vernetzten Ressourcen (6), wo weitere Informationen zum aktuellen Schlagwort zu finden sind.

Die lexikografische Beschreibung nimmt in beiden Wörterbüchern die zentrale Fläche des Browserfensters ein. Dieser Bereich gliedert sich sowohl in *ellexiko* als auch im DDO-Prototyp in drei Teilbereiche, in denen bestimmte Gruppen von lexikografischen Angaben zu finden sind. So sind lesartenübergreifende Angaben (3) sowie eine Lesartenübersicht (4) in beiden Systemen unmittelbar auf der angezeigten Seite zu finden.

Die detaillierte lesartenbezogene Beschreibung (5) befindet sich im DDO-Prototyp weiter unten auf der angezeigten Seite, während sie in *ellexiko* auf gesonderte Seiten verteilt ist.

Die wichtigste Orientierungshilfe, um sich in der lexikografischen Beschreibung zu rechtzufinden, bildet im DDO-Prototyp die Lesartenübersicht (4), die unmittelbar sichtbar für den Benutzer dargestellt wird. In *ellexiko* befindet sich die Lesartenübersicht (4) – zumindest für den hier untersuchten Eintrag *Arbeit* – außerhalb des sichtbaren Bereichs der angezeigten Webseite. Dem Benutzer wird somit zugemutet, die Textseite zu durchsuchen, bis die Übersicht nach dem Weiterrollen des Fensterinhalts erscheint. Hierbei dienen ihm im Wesentlichen nur Überschriften als grafische Orientierungshilfen. Von der Lesartenübersicht kann der Benutzer in beiden Wörterbüchern per Hyperlink zu den detaillierten lesartenbezogenen Angaben gelangen (5). Bei *ellexiko* gelangt der Benutzer für jede Lesart auf eine gesonderte Seite, behält dabei aber die Übersicht im Auge.

Viel Raum nehmen in *ellexiko* Etikettierungen der Artikelstruktur ein wie die in Abbildung 1 zu sehenden Überschriften „Lesartenübergreifende Angaben“ und „Lesartenbezogene Angaben“. Zweifelhaft ist zum Einen, ob der Benutzer deren fachlich geprägten Inhalt versteht, und zum Anderen, ob überhaupt ein Text zur Kenntlichmachung der einzelnen lexikografischen Angaben hilfreich für den Benutzer ist: Denn dadurch wird der Eintrag sehr textreich und es bleibt dem Benutzer überlassen, sich in der Textmenge zurechtzufinden und zwischen Etikettierungen und eigentlichen lexikografischen Angaben zu unterscheiden.

Im DDO-Prototyp wurde auf solche Etikettierungen verzichtet. Es geht aus dem Aufbau der lexikografischen Beschreibung nur indirekt hervor, um welche Art von Angaben es sich im Einzelnen handelt: So muss der Benutzer selbst eruieren, dass die Angaben vor der Lesartenübersicht lesartenübergreifend zu verstehen sind und die weiteren sich auf die jeweiligen Lesarten beziehen. Dasselbe gilt auch für den Aufbau jeder Lesartenbeschreibung, wo der Benutzer nur wenige textliche Informationen darüber bekommt, um welche lexikografische Angabe es sich jeweils handelt. Der DDO-Prototyp setzt im Gegensatz zu *ellexiko* Farben als Hilfsmittel zur Kenntlichmachung unterschiedlicher Informationstypen ein. So wird Blau verwendet, um ausdrucksseitige Angaben (Schlagwort, feste Wendungen, semantisch relatierte Wörter, Wortbildung) hervorzuheben, Rot bezieht sich auf inhaltsseitige Angaben (Bedeutungserklärungen) und Grün soll Belege signalisieren.

Ein weiterer Unterschied zwischen den beiden Wörterbüchern besteht in der Verteilung der verschiedenen Angaben: *ellexiko* verteilt die lexikografische Beschreibung über viele Webseiten, die sich jedoch zu einer Seite (zum Ausdrucken) bündeln lassen. Die detaillierten lesartenbezogenen Angaben werden in *ellexiko* durch eine Indexordner-Metapher geliedert. Durch Anklicken bestimmter Reiter gelangt der Benutzer zu weiteren Seiten mit entsprechenden lexikografischen Angaben. Der DDO-Prototyp versucht, sämtliche Angaben auf einer Seite unterzubringen, wo die einzelnen Lesarten mittels Hyperlinks mit der Lesartenübersicht verknüpft sind, die der Benutzer allerdings leicht aus den Augen verliert. Eine Gliederung der lesartenbezogenen Angaben wird hier durch unterschiedlich eingefärbte, voneinander abgesetzte typografische Blöcke erzielt.

Insgesamt versucht der DDO-Prototyp das in Abbildung 2 wiedergegebene Erscheinungsbild konsequent beizubehalten, also einen statischen Rahmen zu verwenden, in

welchem die unterschiedlichen Inhalte stets auf dieselbe Weise übersichtlich präsentiert werden. Bei *elexiko* wechselt der Rahmen mit dem angezeigten Inhalt: Jeder Inhaltstyp hat eine eigene Seite.

elexiko nutzt zwar eine der grundlegenden Möglichkeiten des Web, Texte durch Hyperlinks miteinander zu verbinden, wesentlich ausgiebiger als der DDO-Prototyp, der die für das Buchmedium typische Linearität der Texte beibehält. Allerdings geht eine Verteilung des Textmaterials über viele Webseiten hier zu Lasten der Übersichtlichkeit, während die relativ knappen Beschreibungen im DDO-Prototyp zu Lasten der lexikografischen Ausführlichkeit gehen. Wirklich nutzen würde man die Potenziale des Web erst dann, wenn sowohl Übersichtlichkeit und schnelle Orientierung als auch die Vertiefung je nach Bedarf des Benutzers angeboten werden könnten.

Im Folgenden soll nun die jeweilige Präsentation einer Reihe von lexikografischen Angaben in den beiden Wörterbüchern näher untersucht werden.

4.2 Lesarten

Da eines der landläufigen Merkmale eines Webwörterbuchs im Gegensatz zum gedruckten Wörterbuch der schier unerschöpflich scheinende Platz ist, der zur Speicherung und Präsentation zur Verfügung steht, ist es naheliegend, insbesondere für polyseme Lemmata Lesartenübersichten zu präsentieren, die einen raschen Überblick über einen umfangreichen Wörterbuchartikel geben und die die Navigation im Artikel vereinfachen können. Abbildung 3 und 4 zeigen die Lesartenübersichten in *elexiko* und im DDO-Prototyp, beide für den Eintrag *Arbeit* (dän. *arbejde*).

Lesartenbezogene Angaben

Das Stichwort verfügt über 5 Lesart(en).
 Lesart 'Tätigkeit' verfügt über 2 Spezifizierung(en).
 Lesart 'Werk' verfügt über 2 Spezifizierung(en).

Lesart 'Tätigkeit'
 Spezifizierung 'Erwerbstätigkeit'
 Spezifizierung 'Training'

Lesart 'Mühe'

Lesart 'Werk'
 Spezifizierung 'Kunstwerk'
 Spezifizierung 'schriftliche Prüfung'

Lesart 'Arbeitsplatz'

Lesart 'Bereich'

Abbildung 3: Lesartenübersicht in *elexiko*



Abbildung 4: Lesartenübersicht im DDO-Prototyp

Auf den ersten Blick fällt auf, dass die *ellexiko*-Übersicht wesentlich textreicher ausfällt und sich darüber hinaus einiger Begriffe („Lesart“ und „Spezifizierung“) bedient, die wohl eher in der Fachsprache des Lexikografen als in der Allgemeinsprache anzusiedeln sind und die den Zugang für bestimmte Benutzergruppen erschweren könnten. Es stellt sich somit unwillkürlich die Frage nach der Zielgruppe dieses Systems und ihrer Voraussetzungen: Wendet sich *ellexiko* primär an sprachwissenschaftliche Laien oder Experten?

Die verwendeten Bezeichnungen ließen sich eventuell dadurch entschärfen, dass sie durch die geläufigeren Bezeichnungen „Bedeutung“ und „Unterbedeutung“ ersetzt würden; dies würde aber dennoch nicht das grundsätzliche Problem dieser Übersicht lösen, dass sie ebenso wie die Gesamtpräsentation der lexikografischen Angaben (vgl. Abschnitt 4.1) textuelle Etikettierungen mit lexikografischen Angaben mischt, wobei die Etikettierungen in dieser Übersicht wesentlich mehr Platz einnehmen als die eigentlich wichtigen Angaben zur Charakterisierung der jeweiligen Lesart. Es fragt sich, ob sich nicht gerade hier die grafischen Möglichkeiten der Webtechnologie stärker nutzen ließen, um die Inhalte klarer zu vermitteln.

Die Übersicht wird durch einen Textblock mit einigen quantitativen Angaben zu den Lesarten eingeleitet, vgl. Abbildung 3. Solche Angaben lassen sich zwar maschinell leicht erstellen, aber lediglich dadurch, dass eine Sache technisch machbar ist, wird sie nicht unbedingt benutzerrelevant. Angaben zur Häufigkeit eines Wortes oder gar seiner Lesarten im zugrunde liegenden Korpus wären – entsprechend grafisch aufbereitet und bezogen auf einen allgemein nachvollziehbaren Standardwert – hier vielleicht wesentlich informativer.

In der Lesartenübersicht des DDO-Prototyps wird eine Nummerierung der Lesarten vorgenommen, die übrigens dieselben quantitativen Informationen zwar implizit, aber vielleicht wesentlich unmittelbarer vermittelt. Zudem deutet eine nummerierte Übersicht eine geordnete Reihenfolge an, im DDO in der Regel nach Korpusfrequenz. Auf eine Nummerierung wurde in der *ellexiko*-Übersicht verzichtet und das Prinzip für die Reihenfolge der einzelnen Lesarten ist nicht unmittelbar auszumachen. Offenbar spielt Korpusfrequenz eine Rolle, denn die recht niederfrequente Lesart ‘Kraft’ aus dem Bereich der Physik fehlt in der *ellexiko*-Übersicht, ist jedoch an anderer Stelle als nicht im Korpus belegt erwähnt.

4.3 Bedeutungserläuterungen

Das Bedeutungsspektrum im Überblick

Lesart 'Tätigkeit'
 Spezifizierung 'Erwerbstätigkeit'
 Spezifizierung 'Training'

Lesart 'Mühe'
 Spezifizierung 'Kunstwerk'
 Spezifizierung 'schriftliche Prüfung'

Lesart 'Arbeitsplatz'
 Spezifizierung 'Bereich'

Bedeutungserläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinnverwandte Wörter	Besonderheiten des Gebrauchs	Grammatik
-----------------------	---	-----------------------	----------------------	------------------------------	-----------

Bedeutungserläuterung

Mit **Arbeit** wird eine Handlung bezeichnet, bei der eine Person(engruppe) zur Berufsausübung und zum Erwerb des Lebensunterhaltes in mehreren, aufeinander folgenden Schritten eine Tätigkeit ausübt oder sich mit einem Sachverhalt beschäftigt.

Beleg(e)

Abbildung 5: Bedeutungserläuterung in *ellexiko*

Die Präsentation der Bedeutungserläuterungen in *ellexiko* (vgl. Abbildung 5) zeichnet sich dadurch aus, dass die Lesartenübersicht zwar ganz oben auf der Seite der aktuellen Lesart steht, dass der Benutzer sie aber bei etwas umfangreicheren Bedeutungserklärungen – genauso wie im DDO-Prototyp auch – aus den Augen verliert: Auch hier wird das Prinzip der im Web weit verbreiteten linearen Textualität wieder deutlich – auch dann noch, wenn die Texte wie bei *ellexiko* auf mehrere Webseiten verteilt werden: Es wird ein Text angezeigt, durch den sich der Benutzer hindurchscrollen muss. Steuerelemente, die sich in Desktop-Anwendungen stets an derselben Stelle befinden, sind in den beiden Webwörterbüchern in der Regel Bestandteil der jeweils angezeigten Seite, wie in diesem Fall die Lesartenübersicht, die sich auch als ein unabhängig von anderen Informationen fest platziertes Steuerelement hätte realisieren lassen können.

Die Tatsache, dass das Web viel Platz bietet, auf dem man sich lexikografisch entfalten kann, wird von *ellexiko* weidlich genutzt: So sind die Bedeutungserklärungen sehr umfangreich, während sie im DDO-Prototyp eher knapp gehalten sind, was zeigt, dass der Ausgangspunkt für dieses elektronische Wörterbuch eben ein für das Buchmedium geschriebenes Manuskript war. Dennoch sind die Bedeutungsangaben im DDO-Prototyp mit einigen Hyperlinks zu weiteren semantischen Informationen versehen. Abbildung 6 zeigt den Aufbau einer Bedeutungserläuterung im DDO-Prototyp.

2. åndelig virksomhed [▶ [summary](#)]

pl. sj. uden for talesprog · **fysisk el. åndelig** → **virksomhed** som man til stadighed udøver for at forsørge sig selv (og sin familie) fx som ansat i en virksomhed

↔ **virksomhed**

↔ **associations**

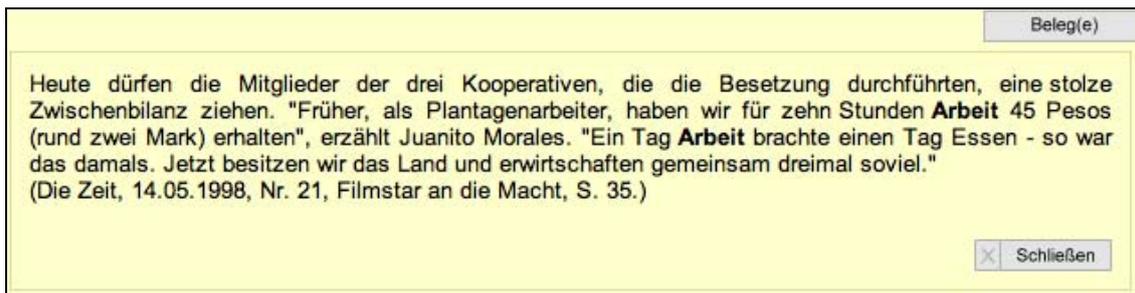
Abbildung 6: Bedeutungserläuterung im DDO-Prototyp

Eingeleitet wird eine Bedeutungserläuterung mit der jeweiligen Lesartennummer sowie durch die aus der Bedeutungsübersicht übernommene Kurzangabe der Bedeutung, in der Abbildung die Nummer 2 und *åndelig virksomhed* ('geistige Tätigkeit'). Diese Kurzangaben wurden automatisch aus den vollen Bedeutungsparaphrasen abgeleitet und sind nicht immer zuverlässig, da der strukturelle Aufbau der Bedeutungsparaphrasen im DDO gewissen Schwankungen unterliegt. So wäre in diesem Beispiel eher *fysisk el. åndelig virksomhed* ('körperliche oder geistige Tätigkeit') zu erwarten gewesen. Für eine endgültige Lösung wird wahrscheinlich eine manuelle Nachbereitung dieser Angaben erforderlich werden.

Während *ellexiko* seine Bedeutungserläuterungen in Gestalt einer umfangreichen Paraphrase gibt, die nicht in weitere mikrostrukturelle Einheiten gegliedert ist, sind die Bedeutungserläuterungen des DDO-Prototyps zwar knapp, aber dennoch aus mehreren Elementen zusammengesetzt. So besteht z.B. die Möglichkeit, die Bedeutungserklärung entweder semantisch oder morphologisch zu restringieren (im angegebenen Beispiel hellgrau ganz zu Beginn der Bedeutungserklärung wiedergegeben). Außerdem ist das in der Bedeutungsparaphrase verwendete *genus proximum* (genauer: der lexikalisierte Ausdruck für das *genus proximum*) gesondert hervorgehoben und fungiert als Hyperlink. Allen Hyperlinks in der lexikografischen Beschreibung sind verschiedene Arten von Pfeilen vorangestellt. Letztlich kann die Bedeutungsparaphrase durch eine erklärende Ergänzung (z.B. zur Konnotation) abgeschlossen werden (in der Abbildung etwas schwächer dargestellt). Die Bedeutungsangaben im DDO-Prototyp enthalten jeweils insgesamt vier Hyperlinks mit immer identischen Funktionen, wovon *summary* zur Bedeutungsübersicht zurückführt, während die drei übrigen erweiterte Zugriffsmöglichkeiten auf das Wörterbuch bieten: So lassen sich Kohyponyme und semantisch assoziierte Wörter ermitteln, außerdem lässt sich das *genus proximum*, dessen Bedeutung ja zum Verständnis der Bedeutungserklärung maßgeblich beiträgt, unmittelbar durch Anklicken nachschlagen.

4.4 Belege

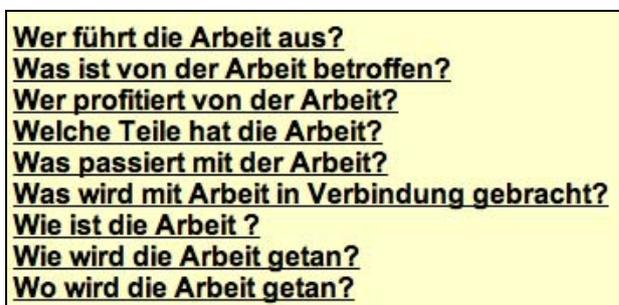
Bei der Angabe von Belegen für eine bestimmte Lesart nutzt *ellexiko* wieder die Möglichkeiten des scheinbar reichlichen Platzes im Web aus: Die Belege sind ziemlich umfangreich, was es dem Benutzer in der Regel erleichtern dürfte, die für die jeweilige Lesart spezifische Semantik des jeweiligen Lemmas zu erschließen. Sämtliche Belege sind mit einer präzisen, in der Regel seitengenauen bibliografischen Angabe versehen. Die Anzahl der für eine Lesart gegebenen Belege scheint allerdings eher zufällig zu sein. Obwohl es in einem Korpus für frequentere Lesarten normalerweise mehr Belege gibt, spiegelt sich dieser Sachverhalt nicht in der Anzahl der in *ellexiko* angegebenen Belege wider; so findet sich für *Arbeit* im Sinne von 'Erwerbstätigkeit' lediglich ein Beleg, der das Lemma in dieser Lesart allerdings gleich zweimal enthält, vgl. Abbildung 7.

Abbildung 7: Beleg in *ellexiko*

Im DDO-Prototyp fallen die Belege wesentlich knapper aus, was an der medialen Herkunft dieses Werks liegt. Allerdings besteht eine gewisse Korrespondenz zwischen der Frequenz einer Lesart und der Anzahl von Belegen, die für sie gegeben wird. So werden für die Lesart 'Erwerbstätigkeit' von *arbejde* zwei Belege angegeben. Die bibliografischen Angaben sind im DDO-Prototyp als Kürzel dargestellt, die als Hyperlinks dienen und in einer endgültigen Version des Systems nicht nur die volle bibliografische Information anzeigen, sondern direkt ins Korpus verweisen werden. In diesem Punkt verhält sich *ellexiko* sehr wie ein Buch mit klassischen bibliografischen Angaben, während der DDO-Prototyp die mögliche Vernetzung unterschiedlicher Ressourcen anstrebt.

4.5 Semantische Umgebung

Unter dem Reiter „Semantische Umgebung und lexikalische Mitspieler“ bietet *ellexiko* für jede Lesart einen Fragenkatalog, durch den semantische Rollen bezüglich der jeweils beschriebenen Lesart des Lemmas umrissen werden. Abbildung 8 zeigt den Katalog für die Lesart 'Tätigkeit' des Lemmas *Arbeit*.

Abbildung 8: Fragen zur semantischen Umgebung in *ellexiko*

Jede dieser Fragen führt durch Hyperlink zu einer Liste von passenden Lemmata, die durch Korpuskookkurrenzanalysen ermittelt wurden. Bei diesem Informationstyp wird eindeutig der Rahmen herkömmlicher gedruckter Wörterbücher gesprengt, die solche Informationen in diesem Umfang schon aus Platzgründen nicht bieten können. Dennoch bleibt offen, was die Grundlage dieses Fragenkataloges ist und für welche Zielgruppe er gedacht ist. Schade ist, dass die Ergebnislisten nur Lemmata ohne weitere Angaben, Verweise oder Belege enthalten. Im DDO-Prototyp gibt es für diesen Informationstyp keine Entsprechung.

4.6 Typische Verwendungen

Unter dem Reiter „Typische Verwendungen“ finden sich unterschiedliche phrasale Verwendungen der betreffenden Lesart. Abbildung 9 zeigt die typischen Verwendungen für die Lesart ‘Erwerbstätigkeit’ des Lemmas *Arbeit*.



Abbildung 9: Angaben in *elexiko* zur typischen Verwendung

Die typischen Verwendungen sind grammatisch gegliedert, wodurch zwar eine gewisse Systematik gewährt wird, was aber das Auffinden bestimmter fester Wendungen ziemlich erschwert. Wo findet der Benutzer beispielsweise *in Arbeit stehen* und was bedeutet dieser Ausdruck? Darüber hinaus sind Wendungen wie *ganze/gründliche Arbeit leisten/tun/machen*, *etwas in Arbeit haben/geben*, *in Arbeit stehen*, *von seiner Hände Arbeit leben*, *nur halbe Arbeit machen* sowie das Sprichwort *nach getaner Arbeit ist gut ruhn*, die allesamt im Duden Universalwörterbuch (DUW: Wermke u.a. 2006) verzeichnet sind, nicht (ohne weiteres) zu finden. Da es den engen Zusammenhang zwischen dem IDS-Korpus und *elexiko* gibt, ist es verwunderlich, dass Belege aus dem Korpus nur selten gegeben werden (typisch nur einer für jede Lesart), und dass eine enge funktionale Verknüpfung zwischen dem Korpus und dem Wörterbuch nicht besteht.

Die Angabe typischer Verwendungen findet sich zu jeder Lesart auch im DDO-Prototyp. Für die Lesart ‘Tätigkeit’ des Lemmas *Arbeit* finden sich die in Abbildung 10 wiedergegebenen Angaben.

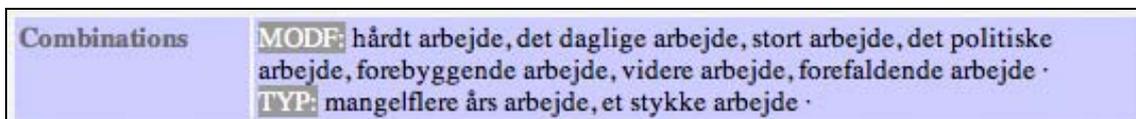


Abbildung 10: Angaben im DDO-Prototyp zur typischen Verwendung

Darüber hinaus werden parallel zur Bedeutungsübersicht auch feste Wendungen mit dem jeweiligen Lemma angezeigt; diese sind nicht nach Lesarten des Lemmas gegliedert (dies ist oft unmöglich). Sie sind als Hyperlinks repräsentiert und ermöglichen so den direkten Zugriff auf die jeweilige semantische Beschreibung mit weiteren Belegen aus dem Korpus, vgl. Abbildung 11.



Abbildung 11: Übersicht über feste Wendungen im DDO-Prototyp

4.7 Sinnverwandte Wörter

Auch bei der Angabe sinnverwandter Wörter zeigt sich die bereits festgestellte Tendenz nach Ausführlichkeit bei *ellexiko* und Platzökonomie beim DDO-Prototyp. Dadurch dass *ellexiko* ein ausschließlich für die Verwendung im Internet konzipiertes Werk ist, unterliegt es kaum Platzbeschränkungen. So bietet *ellexiko* z.B. auch wesentlich mehr Synonyme als der DDO-Prototyp. Beide Systeme erlauben es, durch Verlinkung zum Artikeleintrag des jeweiligen Synonyms zu gelangen. *ellexiko* bietet außerdem die Möglichkeit, an Ort und Stelle Belege für die Verwendung der angegebenen Synonyme einzusehen.

4.8 Grammatische Angaben

Die grammatischen Angaben beziehen sich in beiden Werken im Wesentlichen auf Angaben zur Wortart und zur Flexion sowie auf rudimentäre syntaktische Angaben, in *ellexiko* „Anschlüsse“, im DDO-Prototyp „Konstruktion“ genannt. Darüber hinaus bietet der DDO-Prototyp bei Verben in der Regel auch Angaben zu syntaktischen Komplementen. Die grammatischen Angaben in *ellexiko* sind nicht mit anderen Angaben oder Informationen vernetzt, während im Projekt *ordnet.dk* eine Vernetzung der grammatischen Angaben mit Belegen aus dem Korpus geplant ist.

4.9 Wortbildung

Angaben zur Derivation und Komposition auf der Basis des jeweiligen Lemmas bietet *ellexiko* nicht, während solche Angaben im DDO-Prototyp vorhanden sind. Wortbildungen, für die es außerdem eigene Einträge im DDO gibt, fungieren dabei als Hyperlinks.

4.10 Zugriffsmöglichkeiten

Grundsätzlich gilt für beide Systeme, dass unmittelbar nach Einträgen der Stichwortlisten (auch mit Wildcards) gesucht werden kann. Außer der Anfrage durch Eintippen eines Suchwortes ist in *ellexiko* ein alphabetischer Zugriff über Stichwortliste und alphabetischen Index möglich. Darüber hinaus bietet eine so genannte „Expertensuche“ die Möglichkeit, auf Wörter auch durch andere Parameter als ihre orthografische Repräsentation in der Stichwortliste zuzugreifen; beispielsweise lassen sich Wortart, Wortbildung oder semantische Klasse bei der Suche festlegen. Allerdings ist die Suchmaske der „Expertensuche“ verwirrend aufgebaut. So lassen sich unterschiedliche Suchräume (*ellexiko*-Wörterbuch, bearbeitete Artikel im *ellexiko*-Wörterbuch, *ellexiko*-Demos-

trationswortschatz) offenbar sowohl durch die entsprechende Einstellung von Radiobuttons als auch durch unterschiedliche Schaltflächen festlegen, vgl. Abbildung 12. So passiert es leicht, dass der Benutzer minutenlang auf eine Antwort warten muss, wenn er sich in der Wahl des Suchraumes versehentlich vertut: die Suche nach sämtlichen zehn Kurzwörtern in der *elexiko*-Stichwortliste dauert mehrere Minuten, falls sie überhaupt durchgeführt wird, während sie unter den bearbeiteten Artikeln nur wenige Sekunden beansprucht.



Abbildung 13: Onomasiologische Suche im DDO-Prototyp



Abbildung 12: Suchmaske der „Expertensuche“ in *elexiko*

Über die basale Recherchemöglichkeit einer Lemmasuche hinaus bietet der DDO-Prototyp eine bislang experimentelle onomasiologische Suche, mit der sich zu einem Begriff die ihn bezeichnenden Wörter finden lassen. Abbildung 13 zeigt einen Teil der Ergebnisliste für die Suche nach *beholder* ('Behälter'). Das Wortinventar der Liste ist unterteilt nach weiteren semantischen Merkmalen, die mithilfe eines korpusstatistischen Verfahrens (vgl. Asmussen 2007) den Bedeutungsparaphrasen der jeweiligen Lemmata entnommen sind; in diesem Fall die Merkmale *hank* 'Henkel', *cylinderformet* 'zylindrisch' und *opbevaring* 'Aufbewahrung'. Die Liste ist nicht fehlerfrei, vgl. *liggehøne* 'Legehenne'.

5. Vergleich

In Abschnitt 3 wurden fünf Charakteristika von Webwörterbüchern im Unterschied zu gedruckten Wörterbüchern hervorgehoben, nämlich (1) gelockerte strukturelle Bindungen, (2) erweiterte Zugriffsmöglichkeiten, (3) anpassbare Präsentation, (4) Vernetzbarkeit mit anderen Ressourcen sowie (5) die Möglichkeit der Einbeziehung des Benutzers in den redaktionellen Prozess. Im Folgenden soll zusammengefasst werden, wie *elexiko* und der DDO-Prototyp sich hinsichtlich dieser Charakteristika verhalten.

5.1 Gelockerte strukturelle Bindungen

Gelockerte strukturelle Bindungen bedeuten die Möglichkeit, von einer fixierten Makro- und Mikrostruktur abzuweichen, Querverweise dynamisch anzulegen und die Möglichkeit einer umfassenden, detaillierten lexikografischen Beschreibung.

ellexiko nutzt einige dieser Möglichkeiten insofern, als dass im Vergleich zum gedruckten Wörterbuch sowohl eine wesentlich umfangreichere Mikrostruktur mit einer großen Menge an unterschiedlichen lexikografischen Informationstypen etabliert wurde als auch eine wesentlich umfassendere, differenziertere Beschreibung mit umfangreichen Belegen gegeben wird. Beides scheint aber im Wesentlichen eine Konsequenz aus der Tatsache zu sein, dass im Web unmittelbar viel Platz zur Verfügung steht, nicht aber daraus, dass das Web auch flexible, wordnetartige Strukturen zuließe. Der DDO-Prototyp nutzt solche Möglichkeiten allerdings noch weniger als *ellexiko* und bleibt von seiner strukturellen Konzeption her dem Buchmedium, also seinem Ursprung, verhaftet – ein Zustand, der sich jedoch im Laufe der Entwicklung von *ordnet.dk* zunehmend ändern sollte.

Dynamische Querverweise, d.h. auch Querverweise dort, wo sie nicht von vornherein geplant waren, nutzt der DDO-Prototyp insofern, als dass z.B. vom *genus proximum* einer Bedeutungsparaphrase auf automatisch ermittelte Kohyponyme verwiesen wird, wodurch ein neuer Verweistyp nachträglich etabliert wurde, den es im gedruckten Wörterbuch nicht gibt. In *ellexiko* ist die Verweisstruktur etwas konservativer: es werden zwar zahlreiche Querverweise in Gestalt von Hyperlinks zu anderen Teilen des jeweiligen Artikels eingesetzt, nicht zuletzt bedingt durch die weitverzweigte Artikelstruktur; aber eben nur solche, die das Ergebnis einer klassischen Wörterbuchstrukturbeschreibung sind, nicht solche, die sich als dynamisch zu verwirklichende, durch die Materialstruktur implizit gegebene Möglichkeit anböten.

Bei allen potenziellen Vorteilen gelockerter struktureller Bindungen bergen sie allerdings auch die Gefahr in sich, dass die Lexikografen einen Detaillierungsgrad ihrer Beschreibungen anstreben könnten, der teils den Benutzer, teils die Lexikografen selbst überfordert. Da das Web es ermöglicht, auch unabgeschlossene Produkte zu publizieren und die Unabgeschlossenheit zum Teil auch kaschieren kann, ist eine mögliche Folge hiervon, dass das Wörterbuch von Artikel zu Artikel in seiner Beschreibungstiefe sehr unheitlich werden kann und dass überhaupt Wörterbücher publiziert werden, die zu wenig Einträge haben, um für den Benutzer wirklich von Nutzen zu sein. Das Problem der Unabgeschlossenheit bzw. stark schwankenden Beschreibungstiefe kennzeichnet *ellexiko* leider sehr. Dadurch dass er seinen Ursprung in einem bereits publizierten Wörterbuchwerk hat, ist es in diesem Punkt um den DDO-Prototyp wesentlich besser bestellt.

5.2 Erweiterte Zugriffsmöglichkeiten

In Webwörterbüchern besteht die Möglichkeit eines unmittelbaren Zugriffs auf sämtliche lexikografischen Informationseinheiten sowie die Möglichkeit, diese bei der Recherche miteinander zu kombinieren.

Bei *ellexiko* steht allerdings die klassische Wörterbuchsuche über einen alphabetischen Index bzw. durch Eintippen des Suchwortes im Mittelpunkt. Darüber hinaus gibt es die

„Expertensuche“, die es ermöglicht, gewisse andere Elemente der lexikografischen Beschreibung in die Suche mit einzubeziehen.

Der DDO-Prototyp unterscheidet sich hinsichtlich der Zugriffsmöglichkeiten zunächst nur unwesentlich von *ellexiko*: Auch hier bildet die Schlagwortsuche den unmittelbaren Zugang. Darüber hinaus wird jedoch mit weiteren Möglichkeiten experimentiert, allen voran mit der onomasiologischen Suche (vgl. Abschnitt 4.10). Weiter gibt es die bereits in Abschnitt 5.1 erwähnte Möglichkeit, durch Hyperlinks von bestimmten Informationstypen (Synonyme, Wortbildung, assoziierte Wörter, *genus proximum*) eines Artikels direkt die entsprechenden Artikel anzusteuern. Insgesamt scheint der DDO-Prototyp in diesem Punkt also etwas weiter fortgeschritten zu sein, indem er neue, wenngleich noch unausgereifte Recherchemöglichkeiten anbietet. Wahrscheinlich ist die Datengrundlage von *ellexiko* mit seinen nur etwa 600 Artikeln auch noch viel zu begrenzt für neuartige Suchanfragen.

5.3 Anpassbare Präsentation

Weniger strukturelle Bindungen, wie z.B. mehr Platz, bedeuten auch, dass sich hinsichtlich der Anordnung und Präsentation der lexikografischen Beschreibung im Vergleich zur sehr kondensierten, platzsparenden Darstellung im herkömmlichen Wörterbuch ganz neue und kaum erprobte Möglichkeiten ergeben.

ellexiko nutzt diese Möglichkeiten, indem die lexikografische Darstellung weit verzweigt über etliche Webseiten verteilt wird, was allerdings die Gefahr in sich birgt, dass der Benutzer sich in den vielen Informationen bzw. Seiten verirrt.

Der DDO-Prototyp versucht hingegen so viel wie möglich auf einer Webseite unterzubringen. Hier ist das übergeordnete Konzept, dass ein Wörterbuchartikel wenn möglich maximal eine Webseite beanspruchen darf. Der Artikel selbst ist in Informationsblöcke unterteilt, die je nach Art der Information farblich unterschiedlich hinterlegt sind.

Ein wesentliches Element der flexiblen Präsentation bildet jedoch auch die Möglichkeit, auf die Bedürfnisse des jeweiligen Benutzers spezifisch zugeschnittene Informationen anzubieten. Prinzipiell ließe sich ein Wörterbuchwerk durchaus multifunktional konzipieren, sodass es unterschiedliche Benutzerbedürfnisse befriedigen kann. Ein Großteil der hierzu notwendigen wörterbuchkonzeptionellen Arbeit betrifft dabei die jeweils adäquate Präsentation. So ließen sich letzten Endes auf der Grundlage ein und desselben Datenmaterials durch eine jeweils entsprechende Präsentation des Materials unterschiedliche Wörterbücher oder Wörterbuchfunktionen anbieten.

Weder *ellexiko* noch der DDO-Prototyp nutzen solche Möglichkeiten bislang. Der Gedanke einer „Expertensuche“ bei *ellexiko* ist ein erster Ansatz für eine solche Benutzerdifferenzierung, allerdings nur hinsichtlich der Recherchefunktionalität, nicht der Präsentation.

5.4 Vernetzung mit anderen Ressourcen

Die angebotenen lexikografischen Informationen lassen sich durch Informationen aus anderen sprachlichen Ressourcen ergänzen, vorausgesetzt, es besteht eine geeignete Schnittstelle. So ist *ellexiko* im Augenblick bereits an das morphologische Lexikon *canoonet*, vgl. Boland (2004), gekoppelt und Bestandteil des OWID-Portals (vgl. Ab-

schnitt 3), in dem u.a. auch Neologismen und Wortverbindungen über eine gemeinsame Webseite recherchiert werden können.

Ziel des *ordnet.dk*-Projekts ist es – wie bereits eingangs in Abschnitt 2 beschrieben – eine Vernetzung mehrerer Sprachressourcen zu realisieren: So soll das DDO in seiner endgültigen Webversion mit einem morphologischen Lexikon, einem WordNet und einem Korpus vernetzt werden, sodass ein kombiniertes Wortschatz-Informationssystem entsteht.

5.5 Einbeziehung des Benutzers

Webwörterbücher bieten prinzipiell die Möglichkeit, den Benutzer aktiv in den redaktionellen Prozess mit einzubeziehen, z.B. indem er bestehende Artikel kommentiert, neue Wörter oder neue Lesarten oder aber auch Fehler in der bestehenden Beschreibung mitteilt.

ellexiko bezieht den Benutzer nicht in den redaktionellen Prozess mit ein, der DDO-Prototyp tut es nur insofern, als dass es möglich ist, Verbesserungsvorschläge unmittelbar über eine von der Hauptseite her erreichbare Eingabemaske zu machen. In seiner endgültigen Fassung wird *ordnet.dk* dem Benutzer mindestens die Möglichkeit anbieten, sprachliche Beobachtungen sowie Fehler im Informationssystem allgemein mitzuteilen.

6. Schlussfolgerungen

Sowohl *ellexiko* als auch *ordnet.dk* sind Pioniere in der Nutzung des Webs für wörterbuchähnliche Wortschatz-Informationssysteme. Da sich im Web viele unterschiedliche herkömmliche mediale Funktionen modellieren lassen, z.B. Druckmedien oder audiovisuelle Medien, und darüber hinaus eine unmittelbare Interaktion mit dem Benutzer möglich ist, stellt die Konzeption eines Wortschatz-Informationssystems eine komplexe Herausforderung für die traditionelle buchorientierte Lexikografie dar. Einerseits lassen sich durchaus herkömmliche Wörterbücher im Web modellieren, die in ihrer Funktion ihrem Buchpendant völlig entsprechen und lediglich den Vorteil der relativ leichten Zugänglichkeit durch das Web bieten. Andererseits ließen sich in einem komplexen, gezielt für das Web entwickelten Wortschatz-Informationssystem sämtliche medialen Möglichkeiten des Webs ausnutzen. Sowohl *ellexiko* als auch der DDO-Prototyp befinden sich irgendwo zwischen diesen beiden Extremen.

So steht bei *ellexiko* eine neue, insbesondere auf das Platzangebot im Web zugeschnittene Mikrostruktur im Mittelpunkt, wodurch *ellexiko* einige der strukturellen Bindungen des herkömmlichen gedruckten Wörterbuchs hinter sich lässt und einen gewissen Grad der Hypertextualität einführt, jedoch in der Präsentation, den Zugriffsmöglichkeiten, der Vernetzung mit anderen Ressourcen sowie bei der Einbeziehung des Benutzers den Möglichkeiten des Buchmediums weitgehend verhaftet bleibt.

Beim DDO-Prototyp geht es zunächst nicht um eine neue Mikrostruktur, sondern darum, ein bereits existierendes gedrucktes Wörterbuch im Web zu präsentieren und mit neuen Funktionalitäten auszustatten, insbesondere geht es um die Entwicklung von Zugriffsmöglichkeiten, die das herkömmliche gedruckte Wörterbuch nicht unmittelbar bietet. Im weiteren Verlauf des *ordnet.dk*-Projekts wird es daneben auch um die Vernetzung

von Wörterbuch und Korpus gehen. Es ist ebenfalls geplant, den Benutzer vermehrt in den redaktionellen Prozess mit einzubeziehen, so wie es bereits bei der Entstehung des DDO durch die so genannten *Sporhdunde* der Fall war, vgl. Abschnitt 3.

elexiko und *ordnet.dk* haben also etwas unterschiedliche Ansätze bezüglich der in Abschnitt 3 gegebenen Charakteristika webbasierter Wörterbücher: Das Ausloten mikrostruktureller Möglichkeiten zur möglichst umfassenden lexikografischen Beschreibung bei *elexiko* – die Erweiterung der Wörterbuchfunktionen bei *ordnet.dk*. Ideal wäre wahrscheinlich die Kombination dieser beiden Ansätze.

Ein ganz wesentlicher Aspekt bei der Entwicklung webbasierter Wörterbücher, den sich die jeweiligen Redaktionen stets vor Augen halten sollten, ist der, dass die Zielgruppe nicht nur aus Benutzern herkömmlicher gedruckter Wörterbücher besteht, die auf das Web ‘umgestiegen’ sind, sondern vielleicht sogar vorwiegend aus Benutzern, die das Web überhaupt als ihre primäre Informationsquelle nutzen. Über diesen Benutzertypus weiß man bislang nur wenig; dennoch lässt sich wohl mutmaßen, dass ein Webwörterbuch ihm etwas bieten muss, was er nicht ebenso gut woanders im Web finden kann. Entscheidendes Erfolgskriterium für die Lexikografen wird dabei sein, ob das jeweilige Webwörterbuch dem User einen Mehrwert gegenüber dem gesamten restlichen Web bietet – eine große Herausforderung für die Weblexikografie!

7. Literatur

- Andersen, Mette S./Asmussen, Helle/Asmussen, Jørg (2002): The project of Korpus 2000 Going Public. In: Braasch, Anna/Povlsen, Claus (Hg.): Proceedings of the 10th EURALEX International Congress. Bd. 1. Copenhagen. S. 291-299. Internet: <http://korpus.dsl.dk/staff/ja/papers/euralex.pdf> (Seite zuletzt besucht am 25.01.2008).
- Asmussen, Jørg (2003): Zur geplanten Retrodigitalisierung des Ordbog over det danske Sprog. Konzeption, Vorgehensweise, Perspektiven. In: Burch, Thomas/Fournier, Johannes/Gärtner, Kurt/Rapp, Andrea (Hg.): Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001. (= Abhandlungen der Akademie der Wissenschaften und der Literatur, Mainz). Stuttgart. S. 161-176.
- Asmussen, Jørg (2007): Korpuslinguistische Verfahren zur Optimierung lexikalisch-semantischer Beschreibungen. In: Kallmeyer, Werner/Zifonun, Gisela (Hg.): Sprachkorpora – Datenmengen und Erkenntnisfortschritt. Jahrbuch 2006 des Instituts für Deutsche Sprache. Berlin/New York. S. 123-151.
- Boland, Jan H. (2004): Morphologie im Internet. In: Zeitschrift für Germanistische Linguistik 32, 3, S. 455-457.
- Dahlerup, Verner et al. (Hg.) (1956): Ordbog over det danske Sprog. Bde. 1-28. København.
- DUW = Wermke, Matthias et al. (Hg.) (2006): Duden Deutsches Universalwörterbuch. 6., überarb. u. erw. Aufl. Mannheim/Leipzig/Wien/Zürich.
- Grimm, Jacob und Wilhelm (Hg.) (1956): Deutsches Wörterbuch. Leipzig. Internet: <http://germazope.uni-trier.de/Projects/DWB> (Seite zuletzt besucht am 25.01.2008).
- Hjorth, Ebba/Kristensen, Kjeld/Lorentzen, Henrik/Trap-Jensen, Lars/Asmussen, Jørg et al. (Hg.) (2005): Den Danske Ordbog. Bd. 1-6. København.
- Kerres, Michael/Nattland, Axel (2007): Implikationen von Web 2.0 für das E-Learning. In: Gehrke, Gernot (Hg.): Web 2.0 – Schlagwort oder Megatrend? Fakten, Analysen, Prognosen (= Schriftenreihe Medienkompetenz des Landes Nordrhein-Westfalen 6). München/Düsseldorf.
- Norling-Christensen, Ole/Asmussen, Jørg (1998): The corpus of The Danish Dictionary. In: Lexikos. Afrilex Series 8, S. 223-242.

Pedersen, Bolette S./Nimb, Sanni/Asmussen, Jørg/Sørensen, Nicolai H. et al. (2006): DanNet – a WordNet for Danish. In: Proceedings of the Third International WordNet Conference. GWC 2006, Seogwipo, Korea, January 22-26, 2006. Jeju. S. 329-331.

Andrea Abel

ELDIT (Elektronisches Lernerwörterbuch Deutsch – Italienisch) und *ellexiko*: ein Vergleich

Abstract

Ziel des folgenden Beitrags ist eine vergleichende Gegenüberstellung der elektronischen Wörterbücher ELDIT („Elektronisches Lernerwörterbuch Deutsch-Italienisch“) und *ellexiko*. Im Mittelpunkt der Darstellung stehen ein allgemeiner Überblick und die Beschreibungen der beiden Benutzerschnittstellen mit entsprechenden Hintergrundinformationen sowie ein technischer Vergleich der beiden Systeme in tabellarischer Form.

Die Synopse zeigt, dass beide Wörterbücher einerseits eine Reihe von Gemeinsamkeiten aufweisen, die unter anderem in der modularen Struktur und im Zugang zu einzelnen Angabeklassen, aber auch in den eingesetzten Technologien und der Hypermedianutzung bestehen, andererseits unterscheiden sie sich wesentlich voneinander. Aus den möglichen Konstellationen verschiedener Parameter (Adressat, Fachgebiet, Zweck/Benutzungssituation, Medium) ergeben sich nämlich jeweils spezifische Anforderungen, die eine differenzierte Ausgestaltung nicht nur dieser, sondern ein- oder mehrsprachiger Benutzerschnittstellen im Allgemeinen auf mehreren Ebenen nötig machen.

The present contribution aims at comparing and contrasting the two electronic dictionaries ELDIT (Electronic Learner's Dictionary German-Italian) and *ellexiko*. The focus of the article lies on outlining and describing the main features of the two user interfaces as well as on providing a table overview over the characteristics of both systems.

The synopsis clearly shows the common features of the two dictionaries, such as their modular structure, the way of providing access to the different information categories, the technologies applied and the use of hypermedia facilities. Notwithstanding these similarities, the two tools differ widely from each other. The range of possible combinations of several parameters (e.g. target group, subject field, purpose/context of use, medium) results in specific requirements that necessarily lead to differences at different levels – not only between the tools discussed here but between monolingual or multi-lingual user interfaces in general.

Inhalt:

1. Einführung
2. Synopse
 - 2.1 Stand der Arbeiten
 - 2.2 Adressaten
 - 2.3 Fachgebiet und Umfang
 - 2.4 Ziel(e) und Funktion(en)
 - 2.5 Quellen und Datengewinnung
 - 2.6 Typus
 - 2.7 Medium
 - 2.8 Benutzungshinweise
 - 2.9 Zugriffsstrukturen
 - 2.10 Metasprache
 - 2.11 Angabeklassen – Informationsarten
 - 2.12 Benutzerschnittstelle
3. Schlussbemerkung
4. Daten zur technischen Umsetzung
5. Verwendete und weiterführende Literatur

1. Einführung

Der vorliegende Beitrag geht auf die Einladung zurück, eine vergleichende Gegenüberstellung von [ELDIT](#) („Elektronisches Lernerwörterbuch Deutsch – Italienisch“) und [ellexiko](#) im Rahmen eines 2007 am IDS Mannheim veranstalteten Kolloquiums zu OWID zu präsentieren.¹ Die Aufgabe, einen Vergleich zu erstellen, klingt zunächst einfach, doch stellt sich sofort die Frage: Was soll denn eigentlich analysiert und verglichen werden? Zwei Webauftritte zur Sprache? Zwei Wörterbücher? Dieser Zweifel ergibt sich durch ein Merkmal, das vielen elektronischen Sprachressourcen eigen ist, nämlich dass sie mehr als ein einfaches Wörterbuch, ein Korpus o.Ä. sind, was sich vielfach bereits an der Bezeichnung, im konkreten Fall *Portal* und *Plattform*, zeigt.

Unter *Portal* lässt sich allgemein eine Webseite verstehen, die den Einstieg in andere Webseiten ermöglicht. Ähnlich bildet eine *Plattform* eine Basis für den Zugang zu verschiedenen Ressourcen bzw. den Austausch zwischen Interessierten. ELDIT versteht sich demnach als ein Portal zu verschiedenen Anwendungen und wird von seinen Machern als Online-Plattform für Sprachlerner bezeichnet. OWID bietet Zugang zu den lexikografischen Produkten des IDS und bildet seinerseits „eine Plattform für die Präsentation“² dieser verschiedenen Produkte. Wegen der ursprünglichen Namensgleichheit von Portal und Wörterbuch (*ellexiko*) hatte man sich durch die Anregungen im Laufe des Kolloquiums in der Zwischenzeit am IDS dafür entschieden, das Portal in OWID („Online-Wortschatz-Informationssystem Deutsch“) umzubenennen. Bei ELDIT stößt man auf dasselbe Phänomen: Ursprünglich aus einem einfachen Wörterbuch geboren, hat sich der Name auf die gesamte Plattform übertragen. Auch hier wäre eine Namensänderung eine Überlegung wert.

Im Folgenden werden lediglich die Wörterbücher ELDIT und *ellexiko* synoptisch betrachtet.³ Im Grunde könnte man dabei eine Vielzahl an Aspekten gegenüberstellen, z.B. den lexikografischen Ansatz, die Datenmodellierung, die technische Umsetzung, die Benutzerschnittstellen, einzelne Angabeklassen, die Arbeitsumgebung etc. Im Mittelpunkt der knappen Darstellung stehen ein allgemeiner Überblick und die Beschreibungen der beiden Benutzerschnittstellen mit entsprechenden Hintergrundinformationen. Ein technischer Vergleich der beiden Systeme in tabellarischer Form⁴ sowie weiterführende Literatur über verschiedene Aspekte von ELDIT folgen am Ende.

2. Synopse

2.1 Stand der Arbeiten

Die Grundzüge von ELDIT wurden 1998/99 konzipiert und seit 1999 steht das Wörterbuch, zunächst nur mit einigen wenigen Einträgen, im Netz. Die Ausarbeitung der Wör-

¹ Die Ausführungen folgen im Wesentlichen dem Stand vom Mai 2007, da sich am hier Diskutierten inhaltlich nichts Grundsätzliches geändert hat.

² <http://www.owid.de/owid/stand.html>

³ Die Beschreibungen von ELDIT beschränken sich auf den deutschsprachigen Teil des Wörterbuchs. Als Quelle für die Beschreibungen von OWID bzw. *ellexiko* dienen vor allem die entsprechenden Webseiten, die Publikation über *ellexiko* von Haß (Hg.) (2005) sowie Feedback von OWID-/*ellexiko*-Mitarbeiterinnen, letzteres besonders in Bezug auf die technischen Details.

⁴ Für die Erstellung der Übersicht der technischen Details gilt der Dank Judith Knapp, für weiteres Feedback Stefano Bracco.

terbucheinträge ist im Wesentlichen abgeschlossen. Die künftigen Aufgaben liegen in der Verbesserung und Überarbeitung einzelner Wörterbuchmodule besonders mithilfe computerlinguistischer Methoden (z.B. Ermittlung und Darstellung semantischer Relationen,⁵ Erweiterung der Suchmöglichkeiten, Implementierung des Moduls zur Lesartendisambiguierung,⁶ Umsetzung des Web-Authoring-Tools) sowie im Ausbau der gesamten Plattform durch weitere Lehr- und Lernangebote in Kooperation mit anderen Institutionen.

ellexiko wurde im Jahr 2004 mit einer Stichwortliste und einem Demonstrationswortschatz online gestellt. Ein Ausbau mit weiteren Wortartikeln einerseits und hin zu einem lexikografischen Portal mit weiteren lexikografischen Produkten andererseits erfolgte in den Folgejahren. Nun sind eine stärkere Verknüpfung der Wörterbücher im Portal und eine Auseinandersetzung mit Fragen der Benutzerführung geplant.⁷

Demnach sind beide Produkte als „work in progress“, als Forschungsprototypen zu verstehen, die ständig weiterentwickelt werden, die jedoch zugleich den Benutzern als sprachliche Ressourcen kostenlos zur Verfügung gestellt werden.

2.2 Adressaten

Die anvisierten Benutzer von ELDIT (allgemeine Beschreibungen von ELDIT vgl. z.B. Abel/Weber 2000, Abel 2003) sind deutschsprachige Italienischler und italienischsprachige Deutschler, und zwar Anfänger bis leicht Fortgeschrittene (Niveaustufen A1–B1/B2 nach dem Gemeinsamen Europäischen Referenzrahmen). Außerdem richtet sich das Wörterbuch an linguistische Laien. Während ELDIT eine genau definierte und damit zahlenmäßig beschränkte Zielgruppe anspricht, wendet sich *ellexiko* an ein allgemeineres Publikum bzw. an „alle an Bedeutung und Verwendung Interessierten, sowohl linguistische Laien [...] als auch an linguistische Experten [...]“ (Haß (Hg.) 2005, S. 3). Deutschler gehören dabei nur am Rande zum Zielpublikum; als hilfreich wird *ellexiko* – allein aufgrund des zugrunde gelegten Korpus, des Beschreibungsvokabulars u.Ä. – für fortgeschrittene Lerner, nicht aber für Anfänger beschrieben (vgl. dazu Haß (Hg.) 2005, S. 6).

2.3 Fachgebiet und Umfang

In beiden Wörterbüchern steht die Beschreibung der Allgemeinsprache im Mittelpunkt. Dabei zielt *ellexiko* mit 300.000 Lemmata, die kontinuierlich ausgearbeitet werden, auf die Beschreibung des Gesamtwortschatzes der deutschen Gegenwartssprache. Derzeit sind rund 650 redaktionell ausgearbeitete Wortartikel einsehbar. Fachsprachliche Lesarten sind insofern enthalten, als sie für gesellschaftlich aktuelle Diskurse als wichtig erachtet werden.

ELDIT hingegen enthält einen Grundwortschatz von rund 3.500-4.000 Lemmata pro Sprache.⁸ Im Mittelpunkt der lexikografischen Arbeit stand die Ausarbeitung von Ver-

⁵ Dies geschieht derzeit im Rahmen eines Doktoratsstudiums in Zusammenarbeit zwischen der Europäischen Akademie Bozen und der Universität Trient.

⁶ Siehe dazu: Ignatova/Abel (demn.)

⁷ <http://www.owid.de/stand/stand.html>

⁸ Zur Auswahl der Lemmata wurden verschiedene Grundwortschätze und Lehrwerke der entsprechenden Niveaustufen herangezogen, wie z.B. Vorbereitungsmaterialien für das Zertifikat Deutsch des Goethe-Instituts; weitere Informationen dazu vgl. Abel (2003).

ben, Substantiven und Adjektiven/Adverbien;⁹ Erklärungen von Funktionswörtern sind über eine Kurzgrammatik zugänglich. Der Bereich der Mündlichkeit wurde – im Unterschied zu *ellexiko* – ansatzweise mit berücksichtigt, z.B. durch die Aufnahme kommunikativer Formeln wie „Gute Reise!“ o.Ä.

Wie es für Lernerwörterbücher charakteristisch ist, verfügt ELDIT bei selektiver Makrostruktur über eine ausgeprägte Mikrostruktur (vgl. dazu Zöfgen 1994). *ellexiko* hingegen weist eine ausgeprägte Makro- und Mikrostruktur auf.

2.4 Ziel(e) und Funktion(en)

ELDIT wie *ellexiko* zeichnen sich durch ihren instruktiven Charakter aus, der Wörterbüchern im Allgemeinen eigen ist. Dabei liegt der Schwerpunkt von ELDIT auf einer didaktischen Ebene, wie dies bereits die Spezifizierung „Lernerwörterbuch“ in der Bezeichnung nahelegt. *ellexiko* dient darüber hinaus auch der Dokumentation und versteht sich als Wortdatenbank und Instrument der Wortschatzforschung. Während sich ELDIT also allein der Wortschatzvermittlung verschrieben hat, widmet sich *ellexiko* auch der Wortschatzermittlung.

ELDIT eignet sich in erster Linie – und allein schon aufgrund der beschränkten Anzahl von Einträgen – als Hilfsmittel für die Textproduktion und erst in zweiter Linie für die Textrezeption, nicht aber für die professionelle Übersetzung. Daneben dient es der Festigung und Erweiterung des Wortschatzes. Die Funktionen von *ellexiko* sind breit gefächert (Textrezeption, -produktion und Wortschatzlernen, vgl. Haß (Hg.) 2005, S. 3) und werden im Wesentlichen vom Benutzer durch die Auswahl einzelner Datenelemente festgelegt, d.h., „dass der Wörterbuchtext erst in der Rezeptionssituation eine je spezifische Funktion erhält und dass nicht die Autoren, sondern die Nutzer über die je aktuelle Funktion entscheiden“ (ebd.).

Bei der Konzeption von ELDIT wurden einerseits die Bedürfnisse und Probleme des Lernalerners mit der L2 (z.B. Lernerbedürfnisse in ausgewählten kommunikativen Situationen, Aspekte des Wortschatzlernens, L2-Fehleranalyse, vgl. dazu Abel 2003, auch Bergenholz/Tarp 2002) berücksichtigt, andererseits die Bedürfnisse und Probleme des Lernalerners im Umgang mit dem Wörterbuch (z.B. Ergebnisse aus der Wörterbuchbenutzungsforschung, Wörterbuchkritik, vgl. dazu Abel 2003, und besonders Wiegand 1998, Engelberg/Lemmitzer 2001). Anzumerken ist in diesem Zusammenhang, dass derzeit relativ wenig über die Benutzung elektronischer Wörterbücher bekannt ist und daher viele Erkenntnisse aus Rückschlüssen gewonnen werden müssen. Hier herrscht sicherlich Forschungsbedarf.

2.5 Quellen und Datengewinnung

Für die Erarbeitung von ELDIT wurden besonders Sekundärquellen als Referenzwerke bei der Ermittlung von Wortbedeutungen, Kollokationsangaben etc. herangezogen (z.B. vergleichende Analyse bestehender Standard- und Lernerwörterbücher, Sprachlernmaterialien usw.). Zudem wurden Internetressourcen verwendet, soweit dies möglich war und sinnvoll erschien (z.B. das *Repubblica*-Korpus, DWDS, COSMAS, das Wortschatzportal der Universität Leipzig, Abgleich und Kontrolle mit Google-Ergebnissen). *ellexiko* beruht vorwiegend auf Primärquellen, wobei der Rückgriff auf die hauseigenen Korpora

⁹ Entscheidungsleitend waren dabei finanziell-pragmatische Gründe.

des IDS möglich ist. Zudem wird auf Sekundärquellen zurückgegriffen, etwa bei der Ermittlung von Informationen für die Rubrik „Sprachreflexion“.

Die Datengewinnung in ELDIT erfolgt vorwiegend manuell, aber auch zunehmend computergestützt (z.B. aus ELDIT-Texten, WordManager). Die Daten für *elexiko* hingegen werden vor allem computergestützt und korpusbasiert gewonnen (besonders für die Lesartendisambiguierung, die Auswahl der semantischen Mitspieler und die typischen Verwendungen). Zudem ist eine nachträgliche, kompetenzgestützte Kontrolle und Sichtung nötig. Manuell erfolgt auch die Belegstelleninterpretation.

2.6 Typus

Während es sich bei *elexiko* um ein ausschließlich einsprachiges Wörterbuch handelt, eignet sich zur Beschreibung von ELDIT keine der geläufigen Etikettierungen, weshalb es als „crosslinguales Wörterbuch“ (vgl. z.B. Abel 2006, S. 38) bezeichnet wurde. ELDIT zeichnet sich dadurch aus, dass es eine Verbindung aus ein- und zweisprachigem Wörterbuch darstellt. Es enthält Informationen in der Zielsprache des Benutzers wie ein einsprachiges Wörterbuch (Bedeutungsangaben, Beispiele usw.), aber auch Informationen in der Ausgangssprache des Benutzers (Übersetzungen, metasprachliche Erklärungen usw.), wodurch man besonders den Bedürfnissen des Lerners im Anfängerstadium entgegenkommen möchte. In jedem Wörterbuchteil, im deutschen wie im italienischen, wird die jeweilige Sprache „von sich heraus“ beschrieben; über die Übersetzungsäquivalente ist es möglich, vom einen in den anderen Wörterbuchteil hinüberzuwechseln (vgl. dazu z.B. Abel/Weber 2005 und Abel/Ralli 2007).

2.7 Medium

ELDIT und *elexiko* sind gleichermaßen von vornherein für die elektronische Publikation konzipiert, was im Bereich der elektronischen Wörterbücher selten ist, und eignen sich nicht für eine Printausgabe. Bei *elexiko* ist eine Druckansicht einzelner Einträge möglich, ELDIT eignet sich kaum dafür.

Beide Wörterbücher weisen Ähnlichkeiten in Bezug auf die modulare Datenstruktur auf (mehr dazu siehe im Abschnitt 2.12).

Was die Hypermedia-Elemente betrifft, so sind in ELDIT alle Elemente, in *elexiko* hingegen ausgewählte Elemente, vor allem explizite Verweise (z.B. die mit „Info“ bezeichneten), verlinkt. In beiden Wörterbüchern gibt es die Möglichkeit, einzelne Elemente ein- oder auszublenden, was zur Entlastung der Mikrostruktur beiträgt (z.B. „Info“, Belege, Kommentare in *elexiko*; z.B. Angaben in den Wortfeld-Grafiken in ELDIT¹⁰). ELDIT enthält Audiodateien, durch die der Benutzer die Aussprache der Lemmata anhören kann, während in *elexiko* bislang auf Audioelemente verzichtet wurde. Des Weiteren beinhaltet ELDIT interaktive Grafiken zur Veranschaulichung semantischer Relationen (Wortfeld-Modul) und bewegliche Bilder zur Beschreibung der Verbalenz im Sinne einer semiotischen Didaktik (vgl. Abel 2002, siehe Abbildung 1).

¹⁰ in der Rubrik „Wortbeziehungen“

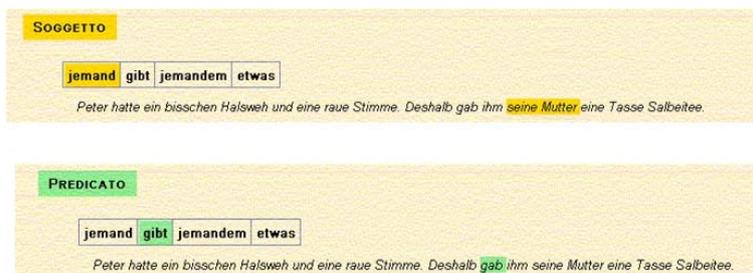


Abbildung 1: Darstellung der Verbvalenz in ELDIT; das Strukturmuster ist in einer vereinfachten linguistischen Metasprache formuliert und wird durch die Unterlegung mit Farben einprägsam dargestellt, die den einzelnen Satzgliedern zugewiesen sind und sich ändern, sobald der Benutzer mit der Maus über das Muster fährt.

In *ellexiko* spielt die grafische bzw. farbliche Darstellung der sinnverwandten Wörter sowie der Grammatik des Verbs keine Rolle. Beide Wörterbücher haben hingegen damit begonnen, (ansatzweise) veranschaulichende Abbildungen/Fotos einzubinden.

2.8 Benutzungshinweise

Um dem Benutzer zu erklären, wie er das Wörterbuch verwenden kann, werden in ELDIT verschiedene Strategien angewandt: Einmal kann er über einen Hilfebutton auf ein ausführliches Tutorial auf Deutsch, Italienisch und Englisch zugreifen, wo er über ein Inhaltsverzeichnis Erklärungen samt Screenshots aufrufen kann. Zudem gibt es ein kurzes Video, das ihn durch das Wörterbuch begleitet. Außerdem helfen Pop-ups¹¹ dabei, zu verstehen, was sich hinter einzelnen Schaltknöpfen verbirgt.

In *ellexiko* kann der Benutzer über einen entsprechenden Button Benutzungshinweise aufrufen, unterteilt in die Rubriken „Nachschlagen/Suchen“, „Angaben im Wortartikel“ und „Glossar“ (letzteres mit Erläuterungen der in *ellexiko* verwendeten Fachbegriffe).¹²

2.9 Zugriffsstrukturen

Beide Wörterbücher erlauben eine eingabebasierte Suche über einer Suchmaske; dabei ist die Suche mit Platzhaltersymbolen möglich, in ELDIT zudem eine Einzel- und Mehrwortsuche (z.B. führt die Eingabe „ganz Ohr“ zur phraseologischen Wendung „ganz Ohr sein“) sowie eine fehler- und flexionstolerante Suche (z.B. führt „Schwiggermutter“ zu „Schwiegermutter“, „Häuser“ zu „Haus“). Außerdem können in beiden Wörterbüchern eine indexbasierte Suche über eine Lemmalaufleiste (Liste) und eine textbasierte Suche (über Links) durchgeführt werden (zu den Suchmöglichkeiten vgl. Engelberg/Lemnitzer 2001).

2.10 Metasprache

In ELDIT wird versucht, im Hinblick auf die Zielgruppe linguistische Fachtermini weitgehend zu vermeiden. Werden Fachausdrücke verwendet, so werden sie – über einen direkten Link – in einem Glossar („Kurzgrammatik“) erklärt. Die metasprachlichen Be-

¹¹ Erklärungen erfolgen jeweils in der Ausgangssprache des Benutzers, nicht in der L2.

¹² In ELDIT hingegen sind die verwendeten linguistischen Termini in der so genannten „Kurzgrammatik“ erfasst.

schreibungen erfolgen möglichst in der L1 des Benutzers (d.h. im deutschsprachigen Wörterbuchteil auf Italienisch und umgekehrt). In *elexiko*, das sich auch an ein Fachpublikum wendet, werden Fachtermini verwendet und bei deren Erklärungen (über ein Glossar bzw. einen Infobutton) wird eine Kombination aus laien- und linguistenbezogenen Formulierungen angestrebt (vgl. Haß (Hg.) 2005, S. 3). Beide Wörterbücher verzichten großteils auf Abkürzungen und „kryptische“ Symbole. In ELDIT möchte man mit einem möglichst hohen Explizitheitsgrad in der Beschreibung sowie mit Anschaulichkeit den Bedürfnissen des Lernalters und Laien entgegenkommen (siehe z.B. Verbvalenzbeschreibung).

2.11 Angabeklassen – Informationsarten

Die Angabeklassen der beiden Wörterbücher sind aus der folgenden Tabelle ersichtlich:

ELDIT:	<i>elexiko</i> :
<ul style="list-style-type: none"> • Lemma • Morphologische Angaben (kurze und gesamte Flexionsparadigmen) • Schreibung • Aussprache • Bedeutungserläuterung • Übersetzungsangaben • Stilistische Markierungen • Diatopische Markierungen • Sprachenkontrastive Angaben (falsche Freunde etc.) • Typische Verbindungen • Valenzangaben (Verben – semiotisch-didaktische Angaben) • Phraseologismen • Lexikographische Beispiele • Paradigmatische Relationen – Wortfelder • Wortbildung: Wortfamilien (Derivate, Komposita) • Abbildungen 	<ul style="list-style-type: none"> • Lemma • Grammatik (Wortart, Flexionsparadigmen; Steigerungsformen etc.) • Schreibung • Worttrennung • Wortbildung • Diachrone Angaben / Etymologie • Bedeutungserläuterung (kurz vs. ausführlich) • Konzeptfamilie und Homonymie • Diatopische Markierungen • Sprachreflexion (Hinweise) • Besonderheiten des Gebrauchs • Typische Verwendungen • Belege • Paradigmatische Relationen – sinnverwandte Wörter • Semantische Umgebung und lexikalische Mitspieler • Wortklasse • Abbildungen

2.12 Benutzerschnittstelle

Eine etwas ausführlichere Betrachtung verdienen die beiden Benutzerschnittstellen, die bei allen Unterschieden dennoch eine erstaunliche Vielzahl an Gemeinsamkeiten aufweisen.

Die Benutzeroberfläche von ELDIT ist dreigeteilt (vgl. dazu Abel/Weber 2000, Abel 2003): Generell kann zwischen einem Orientierungsbereich (1), einem Explikationsbereich (2) und einem erweiterten Explikationsbereich (3) unterschieden werden (siehe Abbildung 2). Der Orientierungsbereich (1) befindet sich im oberen Teil der Benutzeroberfläche und ist zweigeteilt, und zwar in eine Ebene mit Suchmaske, erweiterter Suche und Stichwortliste, die quasi als Bindeglied zwischen Mikro- und Makrostruktur fungieren, sowie mit weiteren Informationen (z.B. Hilfe, Grammatik, Inhalt) und in eine Ebene mit einer Reihe virtueller Karteikartenreiter, die den Zugang zu einzelnen Angabeklassen erlauben (Verwendung,¹³ Redewendungen, Wortbeziehungen, Deklination/Konjugation, Wortbildung, N.B., Bild), wodurch der modulare Aufbau des Wörterbuchs besonders deutlich wird.

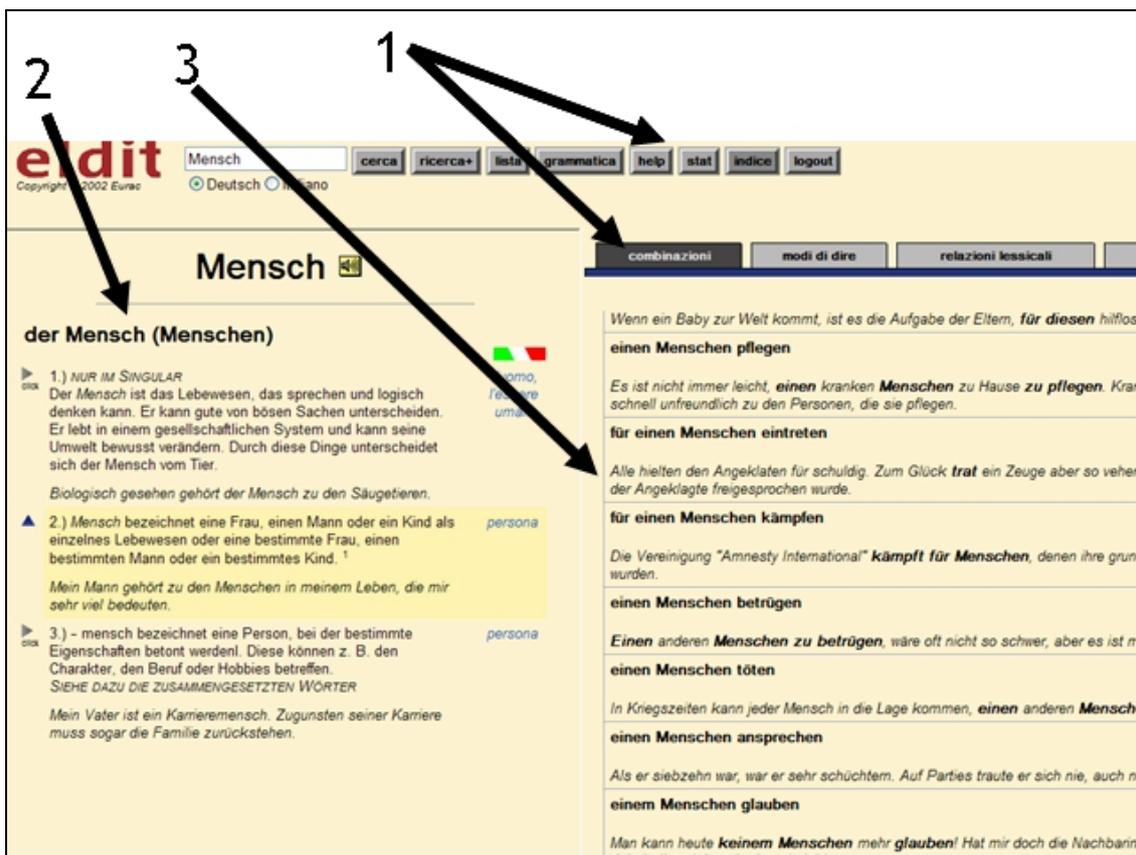


Abbildung 2: Benutzeroberfläche ELDIT

Der Explikationsbereich (2) im linken Teil der Benutzeroberfläche enthält lesartenübergreifende Angaben, z. B. Lemma, morphologische Angaben und Aussprache, sowie grundlegende lesartenbezogene Angaben in Form von Bedeutungserklärungen und Übersetzungsäquivalenten.

¹³ Bei der Wahl der Etikettierungen der Reiter wurde versucht, möglichst transparente Bezeichnungen (auf Kosten der exakten linguistischen Termini) zu wählen: In der Rubrik „Verwendung“ findet der Benutzer typische Verwendungen des Lemmas bzw. Kollokationen (siehe dazu z.B. Abel/Campi et al. 2005); in der Rubrik „Wortbeziehungen“ findet der Benutzer Grafiken mit semantisch verwandten Wörtern, wobei Anregungen dazu aus der Wortfeldtheorie stammen (siehe dazu z.B. Abel/Campogiani/Reichert 2004); in der Rubrik „Wortbildung“ findet der Benutzer Derivate und Komposita (siehe dazu z.B. ten Hacken/Abel/Knapp 2006).

Der erweiterte Explikationsbereich (3) nimmt den rechten und größten Teil des grafischen Interface ein. Er füllt sich erst dann mit Inhalten, wenn der Benutzer diese durch Anklicken der entsprechenden Reiter aufruft. Einige Angaben sind lesartenübergreifend (Redewendungen, Deklination/Konjugation, Wortbildung: Derivate, Anmerkungen) und jederzeit einsehbar, einige Angaben hingegen sind lesartenbezogen und können nur dann aktiviert werden, wenn der Benutzer vorher die entsprechende Lesart durch Anklicken auswählt (Verwendung, Wortbeziehungen, Wortbildung: Komposita, Bild).

Die Benutzeroberfläche von *elexiko* kann als (zweimal) zweigeteilt beschrieben werden: Einmal (A) gibt es einen Orientierungsbereich (1), der im oberen Teil des Bildschirms aus einer Suchmaske und verschiedenen Schaltknöpfen in mehreren Reihen (z.B. Stichwortliste, Hilfe, alphabetische Liste) und im linken Teil des Bildschirms aus einer Lemmalaufliste besteht. Zum anderen gibt es den Explikationsbereich im Zentrum/rechts am Bildschirm (2), der sowohl lesartenübergreifende Angaben (z.B. Orthografie, Wortbildung) als auch lesartenbezogene Angaben in Form von Kurzetikettierungen (z.B. „Individuum“, „Gattung“ etc. zum Lemma „Mensch“) enthält (siehe Abbildung 3).

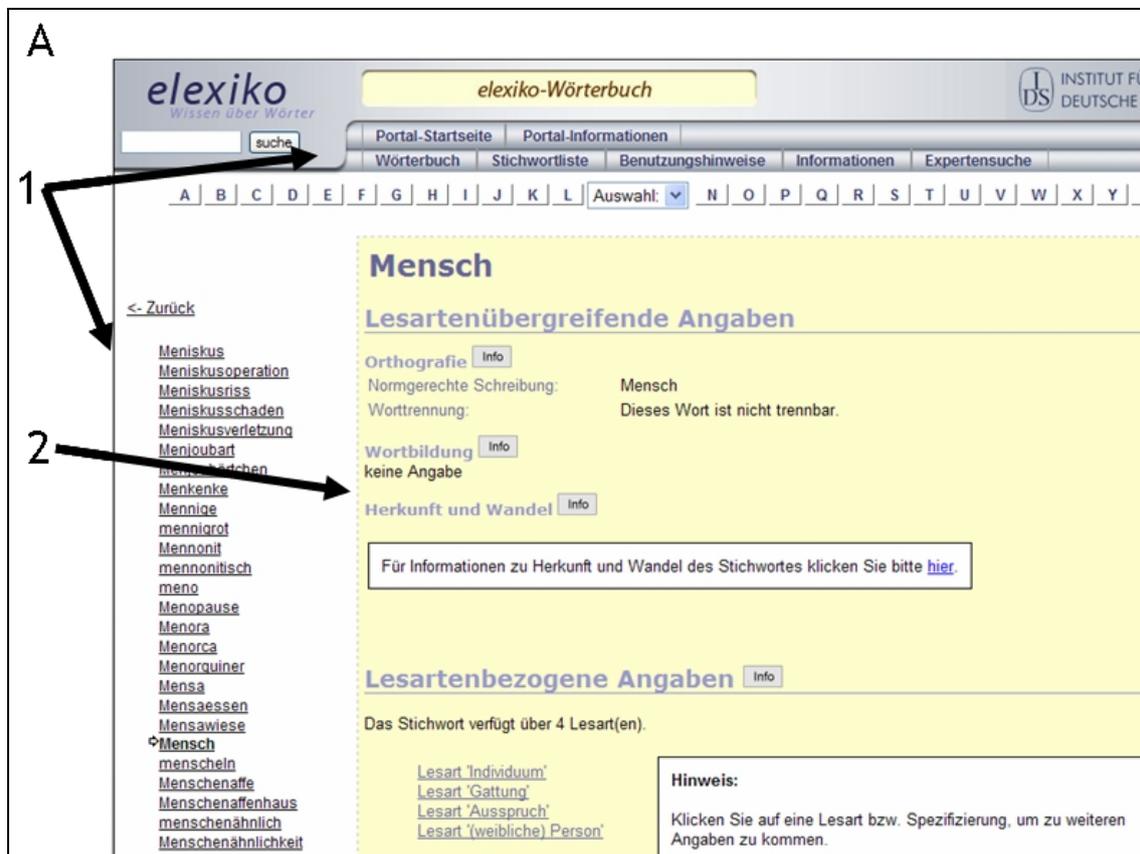


Abbildung 3: Benutzeroberfläche *elexiko* (A)

Durch die Auswahl einer Lesart gelangt der Benutzer auf eine neue Seite (B): Während sich am Orientierungsbereich (1) nichts ändert, erscheint nun anstelle des einfachen der erweiterte Explikationsbereich (2) mit einem Überblick über die Lesarten (Kurzetikettierungen) und verschiedenen Angabeklassen im zentralen Teil des Bildschirms. Genau wie in ELDIT wird auch in *elexiko* eine Karteikartenmetapher bemüht, über die der Benutzer auf einzelne Informationen gezielt zugreifen kann (Bedeutungserläuterung, Se-

mantische Umgebung und lexikalische Mitspieler, Typische Verwendungen, Sinnverwandte Wörter, Besonderheiten des Gebrauchs, Grammatik), die modular strukturiert sind (siehe Abbildung 4).



Abbildung 4: Benutzerschnittstelle *elexiko* (B)

In beiden Wörterbüchern muss der Benutzer zunächst eine Lesart durch Anklicken auswählen, bevor er zu weiteren, ausführlicheren Angaben gelangt. Die Auswahl erfolgt in *elexiko* über Kurzetikettierungen, in ELDIT über die Bedeutungserklärungen; die Funktion der Kurzetikettierungen in *elexiko* übernehmen in ELDIT z.T. die Übersetzungsäquivalente, welche die Bedeutungserläuterungen flankieren.¹⁴ Die Trennung zwischen lesartenübergreifenden und lesartenbezogenen Informationen wird in *elexiko* stärker hervorgehoben (durch explizites Angeben) als in ELDIT.

Einfacher und erweiterter Explikationsbereich befinden sich in ELDIT jeweils auf einer Bildschirmseite, in *elexiko* hingegen auf verschiedenen Seiten. Der Orientierungsbereich ist bei beiden Benutzerschnittstellen prinzipiell invariabel, mit dem Unterschied, dass sich bei *elexiko* die Struktur der Benutzeroberfläche beim Aufrufen der Hilfefunktion – vom Wörterbuch ausgehend – insofern ändert, als dann die Benutzungshinweise anstelle der Wörterbuchseite erscheinen und sich zugleich im oberen Bildschirmbereich eine Fläche mit weiterführenden Links auftut, während sich in ELDIT lediglich ein neues Fenster öffnet (bzw. kurze Pop-ups erscheinen).

¹⁴ Interessant ist in diesem Zusammenhang die Kombination aus kurzer vs. langer Bedeutungserklärung in der elektronischen Lernerlexikografie, die in DAFLES (Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde) zu finden ist: www.kuleuven.be/dafles.

3. Schlussbemerkung

Der kurze Beitrag kann lediglich einen sehr allgemeinen Überblick über die Wörterbücher ELDIT und *elexiko* gewähren. Für vertiefende Einblicke etwa in einzelne Angabeklassen muss an andere Stellen verwiesen werden.

Wie in der Synopse zu erkennen ist, weisen beide Systeme einerseits eine Reihe von Gemeinsamkeiten auf, die unter anderem in der modularen Struktur und in der Art des Zugangs zu einzelnen Angabeklassen, aber auch in den eingesetzten Technologien und der Hypermedianutzung bestehen, andererseits unterscheiden sie sich wesentlich voneinander. Aus den möglichen Konstellationen verschiedener Parameter ergeben sich nämlich jeweils spezifische Anforderungen, die eine differenzierte Ausgestaltung nicht nur dieser, sondern ein- oder mehrsprachiger Benutzerschnittstellen im Allgemeinen auf mehreren Ebenen nötig machen,

und zwar auf der Benutzerebene (Adressat: Laie vs. Experte), der Fachebene (Fachgebiet: allgemein- vs. fachsprachlich) und der Funktionsebene (Zweck/Benutzungssituation). Zu diesen Parametern hat sich ein weiterer gesellt, der bislang in der wissenschaftlichen Diskussion noch nicht ausreichend berücksichtigt worden ist und der sich auf die Darstellungsebene (Medium: gedruckt vs. elektronisch) bezieht. (Abel/Ralli 2007, S. 242)

Variablen von Benutzerschnittstellen, wie etwa Sprache, Struktur, Inhalte usw. müssen auf diese Parameter reagieren (vgl. Abel/Ralli 2007, S. 246; Abel/Voltmer 2003).

Dass ein Vergleich von Benutzerschnittstellen wie ELDIT und *elexiko* anspruchsvoller ist, als man zunächst vermuten möchte, wurde bereits eingangs deutlich: Denn die Komplexität und Vielgestaltigkeit machen die neuen Instrumente, die vielfach von einfachen Wörterbüchern oder Korpora zu Ressourcenpools angewachsen sind und in denen neue, multiple (Misch-)Formen von Werkzeugen sowie Nutzungs- und Nutzerbedingungen entstehen, schwer fass- und beschreibbar (vgl. dazu z.B. Abel 2006b, S. 49ff.).

4. Daten zur technischen Umsetzung (*Judith Knapp*)

	1998/99		2006	
	ELDIT	<i>elexiko</i>	ELDIT	<i>elexiko</i>
Data Modeling and Storage				
<i>Datamodel structure</i>	DTD with encoding on sentence level	First basic concepts of the DTD, in 2002 specific <i>elexiko</i> data modeling (Müller-Spitzer 2005, Müller-Spitzer 2006)	DTD with encoding on word level (Gamper/Knapp 2003)	Development of further DTDs for further <i>elexiko</i> -portals; granular and individual access possible
<i>Datamodel description language</i>	Document Type Definition (DTD)	Document Type Definition (DTD)	XML non schema-based using XML binary objects	Document Type Definition (DTD) [internally in database: XML-schema]

	1998/99		2006	
	ELDIT	<i>elexiko</i>	ELDIT	<i>elexiko</i>
<i>Data description language</i>	XML	XML	XML	XML
<i>Data types</i>	Manually developed ELDIT content	<i>elexiko</i> - and neologism-dictionary	ELDIT content re-used (Knapp et al. 2004) + Word-Manager (ten Hacken/Domenig 1996)	<i>elexiko</i> -portal (Müller-Spitzer 2007)
<i>Data storage</i>	XML-files on disk	On disk	Oracle XML DB (Abel/Bracco 2006)	Oracle XML DB
<i>Data Authoring</i>				
<i>Content authoring</i>	All data is inserted by hand	Data is inserted by hand	Additionally, data is added electronically (Knapp 2004)	Data is inserted partially by hand, partially electronically
<i>Editor used</i>	XML-editor	XML-editor	WEB-based authoring tool (Brusilovsky et al. 2006, Abel/Bracco 2006)	XML-editor
<i>Data administration</i>	Single Java and Perl programs	---	Oracle + Java programs	Oracle
<i>Data Access and Presentation</i>				
<i>Operating system</i>	Linux	Unix (Sun Solaris)	Linux/Windows	Unix (Sun Solaris)
<i>Web application server</i>	Java webserver	Apache Tomcat	Java webserver	Apache
<i>Logic part</i>	Java (servlet technology)	PL/SQL	Java (servlet technology)	PL/SQL
<i>Data searches</i>	Indexing with Lucene (Knapp 2004)	---	Oracle XML binary technology associated to Oracle XML index	Oracle SQL + XPATH
<i>Data Access</i>	DXML	DXML	JDOM/JDBC-OCI	DOM/JDBC-OCI
<i>Interface-generation and -interaction language</i>	JavaScript (Gamper/Knapp 2003)	JavaScript	JavaScript (Gamper/Knapp 2003)	JavaScript
<i>Interface description language</i>	HTML+CSS	HTML+CSS	DHTML+CSS	DHTML+CSS
<i>Browser</i>	Internet Explorer	Internet Explorer	Internet Explorer, Mozilla	Internet Explorer, Netscape, Mozilla, Opera

	1998/99		2006	
	ELDIT	<i>elexiko</i>	ELDIT	<i>elexiko</i>
<i>Operating system</i>	Linux	Unix (Sun Solaris)	Linux	Unix (Sun Solaris)

5. Verwendete und weiterführende Literatur

- Abel, Andrea (1999): Semantische und syntagmatische Normen des Lexikons – ihre Darstellbarkeit in einem computergestützten Lehrwerk. In: Börner, Wolfgang/Vogel, Klaus (Hg.): Lehrwerke im Fremdsprachenunterricht. Lernbezogene, interkulturelle und mediale Aspekte. (= Fremdsprachen in Lehre und Forschung 23). Bochum. S. 286-314.
- Abel, Andrea (2002): Ein neuer Ansatz der Valenzbeschreibung in einem elektronischen Lern(er)wörterbuch Deutsch – Italienisch (ELDIT). In: Dolezal, Fredric F.M./Rey, Alain et al. (Hg.): Lexicographica. International Annual for Lexicography 18/2002. Tübingen. S. 147-167.
- Abel, Andrea (2003): Alte und neue Problematiken der Lernerlexikographie in Theorie und Praxis. Univ. Diss. Innsbruck.
- Abel, Andrea (2006a): Aspekte der Verständlichkeit in der elektronischen Lernerlexikographie. In: Engberg, Jan/Ditlevsen, Marianne Grove/Kastberg, Peter/Stegu, Martin (Hg.): New directions in LSP teaching. (= Linguistic Insights. Studies in Language and Communication 55). Bern/Berlin u.a. S. 163-188.
- Abel, Andrea (2006b): Elektronische Wörterbücher: Neue Wege und Tendenzen. In: San Vicente, Felix (Hg.): Lessicografia bilingue e Traduzione: metodi, strumenti e approcci attuali. Monza, S. 35-56. Internet: <http://www.polimetrica.com/form/form2150.php> (Stand: Juni 2008).
- Abel, Andrea/Bracco, Stefano (2006): From an online dictionary to an online dictionary writing system. In: de Schryver, Gilles-Maurice (Hg.): DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems. Pretoria. S. 25-34.
- Abel, Andrea/Campi, Sara et al. (2005): ELDIT – Elektronisches Lernerwörterbuch Deutsch – Italienisch: einige Aspekte der syntagmatischen und paradigmatischen Ebene. In: Di Meola, Claudio/Hornung, Antonie/Rega, Lorenza (Hg.): Perspektiven Eins. Akten der ersten Tagung Deutsche Sprachwissenschaft in Italien (Rom, 06.–07.02.2004). Rom. S. 587-606.
- Abel, Andrea/Campogianni, Stefania (2005): Facetten der Bedeutungsbeschreibung – ein integrativer Ansatz in der elektronischen Lernerlexikographie (aufgezeigt am Beispiel von ELDIT). In: Mård-Miettinen, Karita/Niemilä, Nina (Hg.): Fachsprachen und Übersetzungstheorie. Vakki-Symposium XXV, Vörå, 12.-13.02.2005. (= Publikationen der Studiengruppe für Fachsprachenforschung der Universität Vaasa). Vaasa. S. 62-72.
- Abel, Andrea/Campogianni, Stefania/Reichert, Julia (2004): Wortfelder in einem zweisprachigen elektronischen Lernerwörterbuch: Darstellung der paradigmatischen Bedeutungsbeziehungen in der pädagogischen Lexikographie am Beispiel von ELDIT. In: Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004. Bd. II. Stuttgart. S. 437-442.
- Abel, Andrea/Gamper, Johann/Knapp, Judith/Weber, Vanessa (2003a): Describing verb valency in an electronic learner's dictionary: linguistic and technical implications. In: Lassner/Mcnaught (Hg.), S. 1202-1209.
- Abel, Andrea/Gamper, Johann/Knapp, Judith/Weber, Vanessa (2003b): New answers to old questions about lexicon acquisition and dictionary use. In: Lassner/Mcnaught (Hg.), S. 1218-1224.

- Abel, Andrea/Ralli, Natascia (2007): Bedeutungs- bzw. Begriffserklärungen in der modernen Lexikographie und Terminographie. In: Di Meola, Claudio/Gaeta, Livio et al. (Hg.): Perspektiven Zwei. Akten der zweiten Tagung Deutsche Sprachwissenschaft in Italien (Rom, 9.–11.02.2006). Rom. S. 235-249.
- Abel, Andrea/Voltmer, Leonhard (2003): Variablen multilingualer Kommunikation an der Benutzerschnittstelle. In: Seewald-Heeg, Uta (Hg.): Sprachtechnologie für die multilinguale Kommunikation – Textproduktion, Recherche, Übersetzung, Lokalisierung. Beiträge der GLDV-Frühjahrstagung 2003 (= Sprachwissenschaft, Computerlinguistik und Neue Medien 5). Sankt Augustin. S. 184-198.
- Abel, Andrea/Weber, Vanessa (2000): ELDIT – A Prototype of an Innovative Dictionary. In: Heid, Ulrich/Evert, Stefan/Lehmann, Egbert/Rohrer, Christian (Hg.): Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Bd. II. Stuttgart. S. 807-818.
- Abel, Andrea/Weber, Vanessa (2005): ELDIT – Electronic Learner's Dictionary of German and Italian: Semibilingual, bilingualised or a totally new type? In: Gottlieb, Henrik/Mogensen, Jens Erik/Zettersten, Arne (Hg.): Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography. May 2-4, 2002 at the University of Copenhagen. (= Lexicographica: Series Maior 115). Tübingen. S. 73-84.
- Bergenholtz, Henning/Tarp, Sven (2002): Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. In: Lexicographica 18, S. 253–263.
- Bergenholtz, Henning/Tarp, Sven (2003): Two opposing theories: On H. E. Wiegand's recent discovery of lexicographic functions. In: Hermes 31, S. 171-196.
- Bergenholtz, Henning/Tarp, Sven (2005a): Electronic dictionaries: old and new lexicographic solutions. In: Hermes 34, S. 7-9.
- Bergenholtz, Henning/Tarp, Sven (2005b): Wörterbuchfunktionen. In: Barz, Irmhild (Hg.): Schreiben, Verstehen, Übersetzen und Lernen. Frankfurt a.M.
- Brusilovsky, Peter/Knapp, Judith/Gamper, Johann (2006): Supporting teachers as content authors in intelligent educational systems. In: International Journal of Knowledge and Learning 2(3/4), S. 191-215.
- Engelberg, Stefan/Lemmitzer, Lothar (2001): Lexikographie und Wörterbuchbenutzung. (= Stauffenburg Einführungen 14). Tübingen.
- Gamper, Johann/Knapp, Judith (2003): A data model and its implementation for a Web-based language learning system. In: Proceedings of the 12th International World Wide Web Conference (WWW 2003). Budapest. S. 217-225.
- Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *ellexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York.
- Ignatova, Kateryna/Abel, Andrea (demn.): The use of context vectors for word sense disambiguation within the ELDIT Dictionary. In: Proceedings of the 13th EURALEX International Congress.
- Knapp, Judith (2004): A new approach to CALL content authoring. Univ. Diss. Hannover.
- Knapp, Judith/Gamper, Johann/Brusilovsky, Peter (2004): Reuse of lexicographic examples in a Web-based learners' dictionary. In: Richards, Griff (Hg.): Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. Chesapeake, VA. S. 776-783.
- Lassner, David/Mcnaught, Carmel (Hg.) (2003): Proceedings of Ed-Media 2003 World Conference on Educational Multimedia, Hypermedia & Telecommunications, June 23-28, 2003, Honolulu, Hawaii, USA. Norfolk (USA).
- 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: <http://www.ids-mannheim.de/pub/laufend/opal/privat/opal06-2.html> (Stand: Juni 2008).
- Müller-Spitzer, Carolin (2007): Das *elexiko*-Portal – Ein neuer Zugang zu lexikografischen Arbeiten am Institut für Deutsche Sprache (IDS). In: Rehm, Georg/Witt, Andreas/Lemmitzer, Lothar (Hg.): Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of the Biennial GLDV Conference 2007 (April 11-13, 2007, Eberhard Karls Universität Tübingen). Tübingen. S.179-188.
- ten Hacken, Pius/Abel, Andrea/Knapp, Judith (2006): Word formation in an electronic learners' dictionary: ELDIT. In: International Journal of Lexicography 19(3), S. 243-256.
- ten Hacken, Pius/Domenig, Marc (1996): Reusable dictionaries for NLP: The word manager approach. In: Lexicology 2, S. 232-255.
- Wiegand, Herbert Ernst (1987): Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. In: Dolezal, Fredric F. M./Rey, Alain/ Roelcke, Thorsten et al. (Hg.): Lexicographica: International Annual for Lexicography 3/1987. Tübingen. S. 178-227.
- Wiegand, Herbert Ernst (1998). Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd. Berlin/New York.
- Zöfgen, Ekkehard (1994): Lernerwörterbücher in Theorie und Praxis: ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen. Tübingen.

6. Wörterbücher – Onlinere Ressourcen

www.eurac.edu/eldit

www.owid.de/elexiko_/index.html

www.kuleuven.be/dafles

www.owid.de