Erschienen in: Haß, Ulrike (Hrsg.): Grundfragen der elektronischen Lexikographie.
elexiko - das Online-Informationssystem zum deutschen Wortschatz. -
Berlin [u.a.]: de Gruyter, 2005. S. 1-17.
(Schriften des Instituts für Deutsche Sprache 12)

Ulrike Haß

elexiko – das Projekt

1. Allgemeines 1
2. Ziele und Adressaten 2
3. Grundlagen und Methoden 7
4. Projektschritte und Module 13
4.1 Der Demonstrationswortschatz 15
4.2 Neologismenwortschatz 15
4.3 Usuelle Wortverbindungen und Mehrwortlemmata 16
5. Literaturverzeichnis 16
5.1 Forschungsliteratur 16
5.2 Wörterbücher 17
5.3 Internetressourcen 17

1. Allgemeines

Dieses Buch handelt von der Konzeption des Online-‚Wörterbuchs' *elexiko*[1] (http://www.elexiko.de), das man auch als lexikalisches Informationssystem bezeichnen kann, weil Nutzer und Nutzerinnen auf bestimmte Einzelinformationen gezielter zugreifen können, wenn sie wollen, als es bei der vom Buch her vertrauten Lektüre eines ganzen Wortartikels der Fall ist.

elexiko ist ein wissenschaftliches Projekt des Instituts für Deutsche Sprache, Mannheim (Mitglied der Leibniz-Gemeinschaft). Die Ergebnisse dieses Projekts werden seit 2004 schrittweise der interessierten Öffentlichkeit, nicht nur der sprachwissenschaftlichen Fachöffentlichkeit, zugänglich gemacht. Wissenschaftliches Ziel des Projekts ist, auf dem aktuellen Stand der lexikografischen, lexikologischen und korpuslinguistischen Forschung und Methodik ein langfristig angelegtes Nachschlageinstrument zu Fragen des deutschen Gegenwartswortschatzes zu entwickeln, es auszubauen und immer wieder zu aktualisieren. Dabei soll ein elektronisches Wörterbuch für Muttersprachler wie, mit Einschränkungen, auch für Deutschlerner entstehen, aber ebenso eine Wort-Datenbank, d. h. ein Instrument für die Wortschatzforschung.

Dieses Buch erklärt die linguistische und die lexikografische Konzeption von *elexiko* in allen wesentlichen Aspekten[2] und will dadurch auch dem besse-

---

1 Der Name *elexiko* ist zu verstehen als zusammenfassende Bildung zu „elektronisches, lexikalisch-lexikologisches korpusbasiertes Informationssystem".

2 Die Reihenfolge der einzelnen Kapitel in diesem Buches entspricht dabei der Abfolge der einzelnen Informationseinheiten in den Wortartikeln.

ren Verständnis des lexikalischen Informationsangebots vor allem für Fachleute aus Sprachwissenschaft, Lexikologie und angewandter Korpuslinguistik dienen. Unmittelbare Hinweise für die Benutzung hingegen sind im Internetangebot von *elexiko* selbst enthalten. Die in diesem Buch dargestellten konzeptionellen Entscheidungen des Projekts sind – anders als bei den meisten lexikografischen Programmschriften der deutschen Wörterbuchgeschichte (vgl. Reichmann 1989) – bereits durch eine über ein paar ‚Probeartikel' hinaus gehende Praxis geprüft und können anhand der Praxis veranschaulicht und nachvollzogen werden (siehe Abschnitt 4. Projektschritte und Module). Gleichwohl geht die Darstellung der *elexiko*-Konzeption von Arbeitspapieren früherer Projektstadien aus und soll vermitteln, „wie es gedacht und geplant war bzw. ist".

Die im Laufe der Jahre personell nicht ganz konstant gebliebene Projektgruppe besteht aus Linguisten und Linguistinnen, die die Konzeption arbeitsteilig entworfen und dann gemeinsam realisiert haben. Zur Realisierung gehören nicht unerhebliche Kompetenzen im Umgang mit diverser Software ‚jenseits' der üblichen Office-Programme, Texttechnologie, Datenorganisation und nicht zuletzt das Verfassen von Wortartikeln, an dem alle beteiligt waren. Während der Fertigstellung und Drucklegung dieses Buchs wird weiter an neuen Wortartikeln sowie am Ausbau der Recherchemöglichkeiten gearbeitet. Dies ist der Grund dafür, dass in diesem Buch nicht nur strikt über den im Sommer 2004 erreichten Stand des Ausbaus berichtet, sondern immer wieder auch die mittel- und langfristige Weiterentwicklung angesprochen wird.

2. Ziele und Adressaten

Als Datenbank kann und soll *elexiko* zweierlei zugleich sein: ein Online-Wörterbuch zum Nachschlagen und ein Informationssystem zum gezielten Recherchieren. Im ersten Fall steht das Einzelwort im Zentrum des Interesses, im zweiten Fall geht es um einzelne Eigenschaften des Wortschatzes. Beide Funktionen von *elexiko* sind in seiner Konzeption, insbesondere in der Datenstruktur, von Anfang an gleich gewichtet und berücksichtigt worden. Allerdings sind die Realisierungen beider Funktionen nicht zu jedem Zeitpunkt auf dem gleichen Stand.

Lexikografisch gesehen ist *elexiko* im Kern ein einsprachiges Wörterbuch des Deutschen in allen seinen nationalen Varianten, das um fremdsprachige Äquivalentangaben ergänzt oder mit diesen vernetzt werden könnte. Wörterbuchgegenstand ist die geschriebene deutsche Standardsprache der Gegenwart seit der 2. Hälfte des 20. Jahrhunderts, so wie sie in überregionalen und regionalen Zeitungen verwendet wird. Im Zuge der regelmäßigen Aktualisierung des *elexiko*-Korpus wird auch der Wörterbuchgegenstand an die jeweilige Ge-

genwart herangeführt. Die Konzentration auf Mediensprache (verstanden als ‚Sprache, die in den Medien verwendet wird', v. a. Zeitungen, Zeitschriften) soll erhalten bleiben. Der Umfang ist auf ca. 300.000 Stichwörter angesetzt.

Grundsätzlich ist *elexiko* so konzipiert, dass es auf viele verschiedene Nutzungsinteressen antworten und damit mehr Wörterbuchfunktionen abdecken kann, als es bei einem gedruckten Wörterbuch sinnvoll ist. In Übereinstimmung mit Wiegand (2001) gilt für *elexiko*, dass „viele lexikographische Entscheidungen sich sinnvoll und einheitlich nur mit Bezug auf Wörterbuchfunktionen begründen [lassen]" (Wiegand 2001, 217). Allerdings verbinden sich in *elexiko* Plurifunktionalität, die für die konzeptionelle und Produktionsseite gilt, mit spezifischer bzw. Monofunktionalität auf Seiten des Nutzers/der Nutzerin in einer bestimmten Nutzungssituation. Die Autoren erarbeiten ein Informationspotenzial, das unterschiedliche Funktionen des einsprachigen Wörterbuchs und unterschiedliche Nutzungsinteressen unterschiedlicher Adressatengruppen abdecken kann. Erst durch die hypertextuelle Darstellung gegenüber den Nutzern werden je nach Situation und Interesse Informationen ausgewählt. 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 des Wörterbuchs bestimmen.[3] Gleichzeitig müssen die AutorInnen teilweise aber auch schon während des Artikelschreibens bei jeder Angabeart entweder laienbezogen oder linguistenbezogen formulieren. In der bisherigen *elexiko*-Praxis werden alle narrativen Angabearten, von der semantischen Paraphrase über diverse Kommentare bis hin zu Hinweisen (zur Unterscheidung vgl. dazu C. Müller-Spitzer, Die Modellierung, in diesem Band) so geschrieben, dass Sprachinteressierte sie mühelos verstehen, wohingegen die zahlreichen linguistischen Kategorien entweder durch ein reines Zuweisungsverfahren markiert oder in linguistischer Terminologie vermittelt werden.

Im Ergebnis kann man *elexiko* grundsätzlich bei der Textproduktion, bei der Textrezeption und zum Zwecke des Wortschatzlernens nutzen, als Rechtschreibwörterbuch, als rückläufiges Wörterbuch, als Bedeutungswörterbuch, als grammatisches Wörterbuch, als Wortbildungswörterbuch, als Kollokationswörterbuch, als onomasiologisches Wörterbuch, als phraseologisches Wörterbuch und/oder als Belegwörterbuch. Damit sind zwar zunächst nur klassische Wörterbuchtypen aufgezählt, aber diese korrespondieren mit bestimmten Nutzungssituationen und damit mit Zweckbestimmungen des Wörterbuchs. Viele weitere speziellere Nutzungsanlässe werden sich bestimmen lassen, sobald die Recherchefunktion von *elexiko* voll ausgebaut ist. Grundsätzlich kann

---

[3] Da die *elexiko*-Wortartikel sehr komplex sind, online als Hypertexte angeboten werden und bewusst von Nutzern nicht nur als Ganzes, sondern auch selektiv rezipiert werden können, ist eine (grafische) Gesamtdarstellung eines Wortartikels im Rahmes dieses Buches nicht möglich.

auch die Recherche und ihre Kategorien einerseits für Laien und andererseits für Linguisten eingerichtet werden; derzeit und vorläufig orientiert sich die hypertextuelle Wortartikelansicht eher (aber nicht vollkommen strikt) an Laien, die Recherchefunktion mittels ausgewählter Parameter eher an linguistischen Fragestellungen.

Viele Urteile darüber, welches Wörterbuch für welche Nutzer oder Nutzungssituationen geeignet oder ungeeignet ist bzw. wie ein Wörterbuch und erst recht: ein elektronisches Online-Wörter-‚Buch' für welche Nutzer oder Nutzungssituationen gestaltet werden sollte, sind spekulativ und von individuellen Erfahrungen und Ansichten geprägt und faktisch immer noch ungeklärt. Zwar war es zweifellos ein Fortschritt der Wörterbuchforschung der letzten Jahrzehnte, ‚den Nutzer' und seine Bedürfnisse ins Bewusstsein der Lexikografen zu holen. Eine objektive, durch empirische Untersuchungen gestützte Wörterbuchbenutzungsforschung gibt es für elektronische, insbesondere für hypertextuelle Werke jedoch nicht, und auch die von Wiegand begründete Erforschung zur Nutzung gedruckter ein- und zweisprachiger Wörterbücher (vgl. Wiegand 1998) ist nur partiell auf ein Instrument wie *elexiko* übertragbar. Die z. B. bei Wiegand (2001, 232 f.) zusammengestellten Benutzungssituationen decken die theoretischen Möglichkeiten der Nutzung ab. Ob das von *elexiko* präsentierte Informationsangebot jedoch für alle, einige oder nur eine dieser Möglichkeiten genutzt wird, könnte im Grunde nur durch eine empirische Nutzungsstudie festgestellt werden, deren Ergebnisse – selbst wenn die Studie breit und differenziert genug angelegt wäre – sich mit der immer noch fortschreitenden Etablierung des Mediums Internet und mit den noch unfesten Rezeptionsgewohnheiten verändern würden.

Die Alternative, ökonomische Merkmale wie Absatz und Marktanteile bestimmter Wörterbuchtypen der Verlagslexikografie als Indikatoren für dasjenige heranzuziehen, was ‚die Nutzer wollen', würde zu vorhersagbaren Ergebnissen führen und Wörterbuchfunktionen auf einige wenige traditionelle Aspekte reduzieren. Innovative Elemente müssen in rein ökonomischer Perspektive von vornherein als ‚nicht nutzungsbezogen' erscheinen, weil Nutzer keinen Bedarf an Angabetypen äußern (können), die sie noch nicht kennen. Dies wäre einer wissenschaftlichen Lexikografie, die ihre technologischen, kulturellen und gesellschaftlichen Bedingungen mit reflektiert, nicht angemessen.

Die MacherInnen von *elexiko* sind überzeugt, dass die Geschichte der Lexikografie insofern weitergeht, als die elektronischen Medien auch neue Nutzungsarten ermöglichen und dass letztere angeboten werden müssen, wenn sie nach eingehender Prüfung für plausibel und ‚interessant' genug gehalten werden. Erst nach einem längeren Zeitraum kann entschieden werden, ob die Sprach- und Kulturgemeinschaft diese und jene Besonderheit von *elexiko* nicht braucht. Für die Konzeption von *elexiko* wurde daraus der Schluss gezogen,

- in der Anlage des Nachschlagewerks viele, aber nicht alle Wörterbuchfunktionen in gleicher Weise und Gewichtung anzulegen,
- NutzerInnen bei ihren Gewohnheiten abzuholen, ohne dort stehen zu bleiben,
- die Realisierung der jeweiligen Funktion des Wörterbuchs den NutzerInnen zu überlassen
- und darüber hinaus neuartige Informationsangebote zu machen, die der Lexikografie der heutigen Zeit eine kulturelle, nicht nur auf Verwertung angelegte Funktion zuweisen.

Die Entscheidung über nutzergruppen- bzw. nutzungssituationsspezifische Anpassungen des Informationsangebots hat in *elexiko* derzeit zunächst ‚nur' das Ziel einer *zweifachen* Differenzierung:

(1) allgemein verständliche Informationen für linguistische Laien und Studierende und
(2) Vertiefungsinformationen für linguistisch Vorgebildete.

Die spezifische Anpassung an diese zwei Nutzungssituationen und Nutzervoraussetzungen wird wie oben erwähnt nicht schon in der Datenbasis geschaffen, sondern erst in verschiedenen Präsentationsformen bzw. über Navigationsstrukturen. Letztere werden mit sog. XSL-Stylesheets realisiert, in denen jeweils festgeschrieben wird, welche Angabeart überhaupt und wenn ja, mit welchem Layout, in welcher Reihenfolge und mit welcher (adressatengruppengerechten) Beschreibungsterminologie dargeboten werden soll. Mittelfristig soll es gemäß der genannten zweifachen Differenzierung zwei solcher Stylesheets geben, eines für Laien und eines für Linguisten. Diese Differenzierung ist derzeit, d. h. beim Demonstrationswortschatz und bei den ‚Neologismen' noch nicht umgesetzt. Es ist erst eine einzige Form der Präsentation mittels eines Stylesheets realisiert, das in seiner Beschreibungsterminologie an sprachinteressierten Nicht-LinguistInnen ausgerichtet ist, zugleich aber auch solche Angabearten enthält und nicht unterdrückt, die eher für LinguistInnen interessant sind. Unterstützend sind zu schwer verständlichen linguistischen Termini Erläuterungstexte (mit „Info" betitelte Knöpfe) eingebaut.

elexiko kann im Ganzen gesehen viele der Fragen beantworten, die bei der **Textrezeption** entstehen: semantische, pragmatische, grammatische und andere. Es kann aber auch und vielleicht vor allem dazu genutzt werden, um das Wissen über Wörter (so der frühere Projektname) in seinen diversen Dimensionen zu vertiefen, kann also sehr gut zum **Wortschatzlernen** eingesetzt werden. *elexiko* unterstützt die Reflexion über Wortgebrauch und Wortschatz und

enthält sicher auch Antworten auf bisher ungestellte Nutzerfragen – wie erwähnt, ein sehr bewusster Versuch.

Einige Angabearten, z.B. die zu den paradigmatischen Relationen, zu syntagmatischen Anschlüssen, zum Frame, zur Valenz und bedingt solche zur Wortbildung sind mit einer gewissen Einschränkung für die Nutzung im Rahmen der **Textproduktion** geeignet.

Für die Nutzung durch **Deutschlerner** bzw. **Nicht-Muttersprachler** ist *elexiko* prinzipiell nur eingeschränkt geeignet. Erstens würde ein Lernerwörterbuch für Anfänger ein anderes Korpus zugrunde legen müssen als die im *elexiko*-Korpus dominanten Zeitungstexte. Die Sprache, die Anfänger lernen, ist durch Alltagssituationen, Mündlichkeit und eine didaktisch notwendige Komplexitätsreduktion der Beispieltexte und -sätze geprägt. Deshalb legt ein Wörterbuch für den DaF-Gebrauch, wie das Langenscheidt Großwörterbuch Deutsch als Fremdsprache (1993) ein Korpus zugrunde, in dem neben Zeitungen und Literatur gesprochene Sprache und DaF-Lehrbücher einen größeren Stellenwert besitzen (ebenda S. VIII; vgl. Zöfgen 1994; Wiegand 1998; Wiegand 2002); und natürlich werden Korpusbelege hier nicht unverändert zitiert. Erst in einem späteren Lern-Stadium kann man Fremdsprachenlerner an unveränderte Originaltexte wie Zeitungsartikel heranführen. *elexiko* ist hingegen ein stark korpusbasiertes und korpusdokumentierendes Wörterbuch.

Zweitens hätte eine konsequente Ausrichtung von *elexiko* an Bedürfnissen von Fremdsprachenlernern ein kontrolliertes Beschreibungsvokabular erzwungen, bei dem nur solche Wörter verwendet werden, die auch selbst in der Stichwortstrecke vorkommen. Dies ist zwar bei *elexiko* mit hoher Wahrscheinlichkeit der Fall, schon allein, weil die Stichwortmenge überdurchschnittlich groß ist, doch ist beispielsweise die Art der Verwendung von Ausdrücken wie *Person, Gegenstand* oder *Handlung* innerhalb der bedeutungserläuternden Texte von *elexiko* in gewisser Weise spezifisch und stimmt nicht überein mit dem Gebrauch derselben Ausdrücke in den Texten des *elexiko*-Korpus. Anders ausgedrückt: die Textsorte ‚Bedeutungserläuterung eines Worts im Wörterbuch' gehört nicht zum Korpus und nicht zu der von *elexiko* dokumentierten und beschriebenen Spielart von Sprache. Dies dürfte im Übrigen für die allermeisten Gebrauchswörterbücher ebenso gelten und nur bei gezielt für Lernzwecke verfassten Wörterbüchern eine Ausnahme finden.

Drittens hätte eine konsequente Ausrichtung von *elexiko* an Bedürfnissen von Fremdsprachenlernern bedeutet, Art und Menge der Lesarten polysemer Lexeme auf solche zu beschränken, die alltags- und allgemeinsprachlich dominant sind, und alle diejenigen auszuschließen, die an speziellere Diskurse gebunden sind.[4] Durch die Korpusbasiertheit dokumentiert *elexiko* auch man-

[4] So in VALBU (2004).

che fachsprachlichen Lesarten, insofern sie in bestimmten gesellschaftlich aktuellen Diskursen eine Rolle spielen; andererseits fehlen in *elexiko* Lesarten (und Redewendungen, z. B. *Gute Reise!*), die typischerweise in mündlicher Kommunikation, nicht oder nur zitatweise in schriftlicher Kommunikation vorkommen. Das für Lesarten Gesagte gilt in ähnlicher Weise auch für Valenzen von Verben und Pluralformen – die Konzeption jeder einzelnen Angabeart müsste auf die speziellen Bedürfnisse von Fremdsprachenlernern abgestimmt sein.

Viertens macht ein Lernerwörterbuch eine spezifische Gewichtung der Angabearten, ein besonderes Informationsprofil erforderlich. Viel größeres Gewicht als bei *elexiko* momentan der Fall müssten Angaben zur Aussprache haben. Hingegen stieße eine kritisch dokumentierende und teilweise relativierende Auseinandersetzung mit soziopragmatisch begründeten Gebrauchsnormen, auf die *elexiko* großen Wert legt, bei Fremdsprachenlernern auf gewisse Widerstände, geht es bei Anfängern doch zunächst um das Erlernen des Normengefüges.

Wäre *elexiko* ein Lernerwörterbuch, stünde zuviel ‚Überflüssiges' und darum Verwirrendes darin. Für fortgeschrittene Deutschlerner, die den Zugang zu aktuellen Medien des deutschsprachigen Raums suchen, dürfte *elexiko* hingegen sehr hilfreich sein, da der Mediensprachgebrauch sehr explizit erläutert wird.

3. Grundlagen und Methoden

Das wohl wichtigste methodische Prinzip in *elexiko* ist die **Korpusbasiertheit**. Es bedeutet, dass in *elexiko* keine einzige Angabe enthalten ist, die allein aus der individuellen Sprecherkompetenz der LexikografInnen abgeleitet wurde. Die Angaben in *elexiko* werden strikt aus dem 1,3 Milliarden Textwörter umfassenden und regelmäßig aktualisierten *elexiko*-Korpus gewonnen (vgl. P. Storjohann, in diesem Band). Wörterbuchgegenstand von *elexiko* ist wie erwähnt das Deutsche vom Ende der 1940-er Jahre bis in die Gegenwart in allen seinen nationalen Varianten. Aus Gründen der technologischen Entwicklung und der daraus resultierenden Textverfügbarkeit deckt das *elexiko*-Korpus (siehe P. Storjohann, Das *elexiko*-Korpus, in diesem Band) die Jahre quantitativ nicht gleichmäßig, sondern ‚aufsteigend' ab. Damit simuliert das Korpus gewissermaßen die diachrone Gewichtsverteilung im anzunehmenden allgemeinen Sprachbewusstsein: Gegenwart und jüngste Vergangenheit der 1990-er Jahre sind darin stärker präsent als die weiter zurück liegenden Jahre.

Das Prinzip der Korpusbasiertheit gewährleistet methodische Konsequenz und Transparenz. *elexiko* ist ein dokumentierendes, kein präskriptives Nach-

schlagewerk. Dennoch werden Nutzer auch über Normfragen informiert, und zwar so, dass sie sich eine eigene Meinung bilden können, d. h explizit und nicht-wertend (vgl. A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band). Nur eine einzige Angabeart in *elexiko* ist implizit präskriptiv und weicht systematisch von Korpusbefunden ab: die orthografische Gestalt des Lemmazeichens, die den Regeln der neuen Rechtschreibung ab 1997 folgt. Da der Großteil der im *elexiko*-Korpus enthaltenen Texte von vor 1997 stammen und also in ‚alter' Rechtschreibung vorliegen, hätte eine Entscheidung über die Gestalt des Lemmazeichens nach Häufigkeit im Korpus eine unter sprachkulturellen Gesichtspunkten nicht verantwortbare Stichwortliste ergeben. Es werden also auch Stichwörter in neuer Rechtschreibung angesetzt, die u.U. so noch gar nicht im Korpus belegt sind. Dennoch haben NutzerInnen die Möglichkeit, sich jeweils genau darüber zu informieren, welche orthografische Variante im Korpus belegt ist oder nicht und ob sie der aktuellen Norm entspricht oder nicht (vgl. A. Klosa, Orthografie und morphologische Varianten, in diesem Band).

Das wichtigste Instrument im Umgang mit diesem riesigen Korpus ist die COSMAS-Kookkurrenzanalyse (siehe: http://www.ids-mannheim.de/cosmas2/), die zu jedem Lexem eine Liste signifikanter Kontextpartner und Syntagmen liefert (vgl. U. Schnörch, Der Autoren-Arbeitsplatz, in diesem Band). Primär Kontextpartner und Syntagmen, nur sekundär einzelne Belege werden im Hinblick auf verschiedene Angabearten lexikografisch interpretiert. Nachträglich wird kompetenzgestützt kontrolliert: Gibt es noch eine weitere Lesart? noch ein Synonym? noch eine weitere semantische Rolle? Wenn ja, müssen auch diese sich, so die redaktionelle Regel, im Korpus nachweisen lassen, um aufgenommen zu werden.

Ist die Umsetzung des Prinzips der Korpusbasiertheit bei Angaben wie der der orthografischen oder morphologischen Varianten, der belegten Lesarten eines Lexems oder dessen Rektion leicht vorstellbar – es werden nur die tatsächlich belegten Varianten, Lesarten und Rektionsbeziehungen angegeben[5] –, so bedeutet das Prinzip bei einigen Angabearten eine zumindest partielle Neukonzipierung der der Angabe zugrunde liegenden Kategorien. Z. B. müssen Synonymie, Antonymie und andere Begriffsrelationen anders und abweichend von der strukturell geprägten linguistischen Tradition bestimmt werden, wenn man sich absichtsvoll darauf beschränkt, solche Relationen aus dem Korpus

5 Dass für automatisch generierte Angaben, wie sie in Zukunft auch bei *elexiko* integriert werden sollen, auch noch anderes, z. B. externe Wortlisten, als Korpusbefunde herangezogen werden muss, weil automatisch nie heißen kann ‚auf der Basis von Belegstelleninterpretation', steht auf einem anderen, nämlich technischen Blatt. Sobald ein Lexikograf oder eine Lexikografin sich eines bestimmten Stichworts annimmt, werden die bis dahin automatisch generierten Angaben anhand von Belegstelleninterpretation überprüft.

und damit aus der syntagmatischen Ebene des Textes und des Diskurses zu (re)konstruieren (vgl. P. Storjohann, Paradigmatische Relationen, in diesem Band).

Das Prinzip der Korpusbasiertheit ist bei den einzelnen Modulen (beschriebenen Wortschatzbereichen, s. u.) leicht abweichend realisiert. Dies gilt insbesondere für das Modul der Neologismen; die Neuartigkeit der hier behandelten Wörter und Lesarten verlangt die Einbeziehung noch aktuellerer Quellen, als das elektronische Korpus sie bieten kann, selbst wenn dies halbjährlich ‚upgedatet' wird. Selbstverständlich werden solche Quellen angegeben, so dass im Grunde nur das Korpus erweitert wird.

Ein von Modul zu Modul leicht abweichender Umgang mit dem Prinzip der Korpusbasiertheit kann auch mit deren spezifischen Informationsprofilen zusammenhängen. Ein Modul wie der Demonstrationswortschatz, das weniger Stichwörter tiefer beschreibt, wird die Ergebnisse der automatischen Korpusanalysen, vor allem die Listen der Kookkurrenzpartner, kritischer sichten und linguistisch aufwändiger interpretieren als ein Modul, das mehr Stichwortartikel unter Einsatz überwiegend (teil-)automatischer Verfahren mit Angaben füllen will. Die einzelnen Beiträge in diesem Band werden auf der Basis der Erfahrungen mit dem Modul ‚Demonstrationswortschatz' jeweils genau angeben, wie das Prinzip der Korpusbasiertheit bei jeder lexikografischen Angabeart angewendet wurde, wie mit den von COSMAS II gelieferten Ergebnissen verfahren wurde, wie sie bewertet und ggf. ausgewählt wurden. Folgende Arten von Ergebnissen stellt das Korpusrecherchesystem COSMAS II des IDS linguistischen NutzerInnen bei Bedarf zur Verfügung:[6]

- sog. Expansionslisten, das sind Listen der Rechts- und Linkserweiterungen des Suchworts mit Trefferzahlen (wichtig v. a. für Wortbildungsfragen)
- sog. KWIC (Key Word in Context)-Zeilen in chronologisch oder anders sortierten Listen (wichtig z. B. für Valenzen, Rektionen)
- Belege in wählbarem Kontextumfang (v. a. wichtig für Semantik und Pragmatik)
- die Kookkurrenzanalyse; sie liefert nach Signifikanz geordnete Listen mit den Kookkurrenzpartnern des jeweiligen Suchworts oder einer gesuchten Wortgruppe und zeigt auch die (typische) syntagmatische

---

6 Die Ergebnisse sind je nach linguistischem Interesse modifizierbar; auf solche Feinheiten kann hier nicht eingegangen werden. Sie werden in den Online-Hilfe-Texten von COSMAS II erläutert. Auf nachfolgender Internetseite kann eine Client-Software heruntergeladen werden, mit der sich das Programm und damit der Zugriff auf die IDS-Korpora starten lässt. Innerhalb des Programms wird die Suchsyntax erläutert: http://www.ids-mannheim.de/cosmas2/

Einbettung der Kollokationen an (wichtig v. a. für Vorstufen der semantischen Disambiguierung, der Bestimmung von Diskursspezifik, Argumentstruktur/Frames, Sinnverwandtschaften, Phraseologisch-Idiomatischem)

- Trefferlisten und Belege zu beliebigen Wortpaaren und -gruppen, die das kookkurrente Verhalten gewünschter Lexeme in einem festgelegten Kontext anzeigen (wichtig v. a. für Kontrollen und Hypothesenüberprüfung).

Die Struktur der *elexiko*-Wortartikel ist im Hinblick auf das oben erläuterte Ziel eines plurifunktionalen Informationspotenzials ungewöhnlich ausdifferenziert. Dahinter steht der Versuch, der metalexikografischen und lexikologischen Forschung der letzten 30 Jahre gerecht zu werden. Die Verwendung der XML-Technologie sorgt hierbei zudem für Einheitlichkeit der Bearbeitung und für eine konsistente Datenbasis (vgl. C. Müller-Spitzer, Modellierung lexikografischer Daten, in diesem Band).

Das Informationsprofil von *elexiko* wird in den Beiträgen dieses Bandes eingehend beschrieben und begründet. Es deckt von der Semantik und Pragmatik über Wortbildung und Grammatik bis zur Orthografie alles ab, was von einem einsprachigen gegenwartsbezogenen Wörterbuch erwartet werden kann, das nicht nur Fragen von Laien, sondern auch von Linguisten beantworten will.

Zu den allgemeinen lexikografischen Grundsätzen von *elexiko* gehört, den unbegrenzten ‚Druckraum' mit größtmöglichem Vorteil für die Rezipienten zu nutzen und gleichzeitig eine punktgenaue, datenbankangemessene und damit standardisierte Codierung der in Angaben, Belegen, Kommentaren und Hinweisen enthaltenen Informationen zu gewährleisten. Als Vorteil für Rezipienten wird angesehen, wenn sie zu dem jeweils interessierenden lexikologischen Aspekt das Maximum an differenzierter und verständlich-anschaulicher Information ohne die üblichen Textverdichtungen durch Abkürzungen oder Symbole erhalten und ohne, dass sämtliche anderen Informationselemente zwangsläufig mit gelesen werden müssen, wie es in gedruckten Wörterbüchern der Fall ist.

Die methodische Lösung des doppelten Anspruchs an ‚Lesbarkeit' und an ‚Recherchierbarkeit' sieht wie folgt aus: Die Artikelstruktur enthält an allen irgendwie sinnvollen Stellen die Möglichkeit, zu jedem standardisiert eingetragenen Angabe-Element einen lexikografischen Kommentar, einen Literatur- oder Projekthinweis, Belege, Abbildungen, Faksimiles oder Tondokumente zu geben. D. h. *elexiko* setzt die metalexikografische Forderung in die Praxis um,

nach der Angaben von Kommentaren unterschieden werden sollen:[7] Angaben enthalten die Ergebnisse der Korpusanalyse durch lexikografische Experten oder auch automatisch gewonnene Korpusanalyseergebnisse; Kommentare beziehen sich immer auf eine Angabe und dienen der Reflexion, Relativierung oder Ergänzung der Angabe aus lexikografischer Sicht und unter Einbeziehung sonstiger Quellen. Ferner werden mit dieser Lösung die multimedialen Möglichkeiten ausgeschöpft.

Druckt man sämtliche Informationen zu einem Stichwort aus, was zwar möglich ist, aber eigentlich dem Angebot der selektiven Nutzung widerspricht und zu falschen Urteilen über *elexiko* führen kann, stellt man fest, dass die Wortartikel im Durchschnitt deutlich länger sind als die in gedruckten Wörterbüchern, selbst mehrbändigen (vom Grimmschen Wörterbuch einmal abgesehen). Es dauert daher auch länger, einen *elexiko*-Wortartikel zu verfassen, nämlich zwischen einer halben Stunde und zwei Tagen. Da die Arbeitszeit pro Wortartikel begrenzt bleiben sollte, wurden einige wenige Informationsarten vorläufig von der Bearbeitung ausgeschlossen, obwohl sie in der Artikelstruktur bereits angelegt sind und jederzeit nachgetragen werden könnten. Ausgeschlossen sind derzeit (d. h. im Modul ‚Demonstrationswortschatz' und z. T. auch im 2005 neu beginnenden Modul ‚Lexikon zum öffentlichen Sprachgebrauch')

- Angaben zur Diachronie, d. h. zur Vorgeschichte von Form und Bedeutung des Stichworts. Die *elexiko*-Konzeption enthält hierfür drei Stufen. Erstens eine Mikrodiachronie, die das Stichwort in der Diskursgeschichte der der Gegenwart unmittelbar vorangehenden Jahrzehnte behandelt, sofern es hier signifikante Entwicklungen gibt. Dies ist möglich durch eine diachrone Interpretation der Ergebnisse der Korpusanalyse. Zweitens eine ebenso diachron angelegte Analyse historischer Korpustexte. Das historische Korpus des IDS ist derzeit noch relativ klein, sodass eine Anwendung der korpusanalytischen Methoden auf ein historisches Vergleichskorpus zu einem späteren Zeitpunkt sinnvoller erscheint. Drittens sollen Angaben zur Diachronie sich auch auf die älteren Jahrhunderte bis ins Althochdeutsche oder Indoeuropäische erstrecken. Hierzu existieren allerdings zahlreiche historische bzw. etymologische Wörterbücher und Einzeldarstellungen, deren Angaben in *elexiko* allenfalls referiert und zitiert, keinesfalls jedoch korpusbasiert neu erstellt werden könnten. Hinweise auf sprachhistorische Forschungsliteratur werden bei Bedarf auch jetzt bereits in die synchronen Angaben integriert.

---

[7] Vgl. hierzu Reichmann (1986) und Wiegand (1989).

- Angaben zur Wortbildungsproduktivität, d. h. zu den typischen und musterhaften Wortbildungen, die ein lexikografisch beschriebenes Stichwort, oft ein Simplex, unter dem Einfluss aktueller gesellschaftlicher Diskurse und Themenkarrieren ‚anregt', z. B. Bildungen mit *mobil* (vgl. A. Klosa, Wortbildung, in diesem Band).

In *elexiko* wurden spezifische Artikelstrukturen zu **drei Stichworttypen** (Lemmatypen) entworfen; von ihnen sind vorläufig nur zwei realisiert, nämlich Einwortlemmata und Mehrwortlemmata. Der dritte Typ bezieht sich auf Wortelementlemmata, zu denen einmal die bekannten Affixe und Konfixe gehören, zum zweiten aber auch lexikalische Einheiten, die in besonders produktiver Weise reihenbildend sind wie *Gen-, Riesen-, bio-/Bio-* usw. und die in Zusammensetzungen u. U. eine andere Bedeutung und Pragmatik zeigen als in ihrer Verwendung als Einzellexem. Da die indigenen Wortbildungselemente teilweise bereits im Grammatischen Informationssystem des IDS behandelt sind, und sekundär auch aus Kapazitätsgründen wird in *elexiko* auf die Behandlung von Wortelementen vorläufig verzichtet.

Zum lexikografischen Programm von *elexiko* gehört wesentlich auch die medienspezifische Konzeption der sog. Mediostruktur (vgl. Wiegand 1989), d. h. eine gegenüber gedruckten Wörterbüchern deutlich ausgebaute und differenzierte Verweisstruktur in Form von Links. Hierdurch soll die Vernetzung des Wortschatzes auf seinen verschiedenen Ebenen bzw. in seinen Dimensionen visuell deutlich gemacht werden. Nicht nur Assoziationen, die die Wortbildung und die Phraseologie zwischen Lexemen stiften, sollen als Links realisiert werden, sondern auch und vor allem semantisch und pragmatisch begründete Relationen zwischen Lesarten. Naturgemäß hängt der Ausbau der Vernetzungskomponente in einem elektronischen Wörterbuch wie *elexiko* von der Menge der differenziert beschriebenen Stichwörter ab. Daher ist es möglich, dass NutzerInnen derzeit den hohen Stellenwert der Vernetzung noch nicht in der Weise erkennen können, wie sie konzeptionell angelegt ist und wie mit dem fortschreitenden Ausbau des Informationsangebots sichtbar werden wird.

Das ‚**Technische**': *elexiko* wird in einer Datenbank (derzeit Oracle 9i) realisiert, die XML-fähig ist. Die Artikelstruktur ist mittels XML, einer formalen Sprache zur inhaltlichen Modellierung von Textstrukturen, definiert (vgl. C. Müller-Spitzer, Modellierung lexikografischer Daten, in diesem Band). Lexikografische AutorInnen schreiben Wortartikel mit Hilfe eines XML-Editors, der am Ende des Schreibprozesses überprüft, ob die Artikelstruktur ‚valide' ist, d. h. der Strukturvorgabe entspricht und von der Datenbank verarbeitet werden kann. Viele notwendige Vereinheitlichungen, die sonst von den redaktionellen Richtlinien eines Wörterbuchs garantiert werden sollen, sind durch diese Struktur schon gesichert. Anschließend wird der Wortartikel in die Datenbank

gespeichert. Man kann ihn dort jederzeit wieder herausholen, überarbeiten und erneut einspeichern. Ein Redaktionssystem für Konsistenzprüfungen, Aktualisierungen und Koordination mehrerer Arbeitsgruppen wird mit wachsender inhaltlicher Füllung notwendig. Zum Lesen in einem Internet-Browser wurde ein bzw. wurden einige sog. Stylesheets in XSLT erarbeitet. Die Recherche in der Datenbank geschieht ebenfalls über einen Internet-Browser, der originale Datenbank-Abfrage-Befehle in eine Fensteransicht umsetzt, in der NutzerInnen die sie interessierenden Parameter bequem auswählen können. Hierbei kommt u. a. Java-Skript zum Einsatz. Selbstredend werden bei der Projektorganisation auch die üblichen Office-Programme genutzt.

4. Projektschritte und Module

Die Überlegungen zu einer lexikalischen Datenbank reichen im Institut für Deutsche Sprache, Mannheim, schon eine Reihe von Jahren zurück und lagen gewissermaßen allgemein schon lange in der Luft. Doch erst mit der technologischen Entwicklung Mitte der neunziger Jahre, mit den Möglichkeiten des World Wide Web zeigten sich Wege zur Realisierung eines langfristigen und umfangreichen Online-Wörterbuchs auf. Die Hypertext-Grammatik GRAMMIS bzw. das Grammatische Informationssystem des IDS ging mit gutem Beispiel darin voran. Erste Skizzen zu einem lexikalischen Informationssystem entstanden Ende 1997. Der Aufbau einer Arbeitsgruppe, die Exploration der computertechnischen Möglichkeiten und die Konkretisierung eines den personellen wie technischen Möglichkeiten angepassten Konzepts mündeten in die linguistische Entwicklung und texttechnologische Realisierung einer Wortartikelstruktur in Form einer ungewöhnlich komplexen XML-DTD. Sie enthielt und enthält an die 400 Elemente je Lesart eines Stichworts, wenn man die Enden sämtlicher möglicher Zweige zählt.

In der darauf folgenden Phase wurde diese DTD getestet, evaluiert und insgesamt dreimal überarbeitet. Ihre Funktion beschränkt sich ja nicht darauf, für gut gegliederte, lesbare Wortartikel zu sorgen, sondern besteht auch und wesentlich darin, alle interessanten lexikologischen Informationen über Datenbankrecherche zugänglich zu machen (Näheres dazu siehe C. Müller-Spitzer, Die Modellierung lexikografischer Daten, in diesem Band). In diese Konzeptionsphase fallen auch sämtliche Bemühungen um eine mittelfristig stabile Lösung der gesamten Softwarearchitektur eines solches Vorhabens, die auch heute nicht und vermutlich nie ganz wunschgemäß gegeben ist. Anfang 2003 war die Konzeptionsphase sowohl linguistisch-lexikografisch als auch com-

putertechnisch[8] so weit abgeschlossen, dass mit der Realisierung erster Wortschatzbereiche (Module, siehe unten), d. h. mit dem Verfassen von Wortartikeln und ihrer Speicherung in der Datenbank begonnen werden konnte.

In die Realisierungsphase ist auch die Erarbeitung der knapp 300.000 Wörter umfassenden Stichwortliste aus dem IDS-Korpus einbezogen, die nicht nur die üblichen Lemmatisierungsprobleme zu lösen, sondern auch die Anpassung an die seit 1997 übergangsweise, ab 2005 alleinig geltende Orthografie zu leisten hatte (vgl. U. Schnörch, Die *elexiko*-Stichwortliste und A. Klosa, Orthografie und morphologische Varianten, in diesem Band). Lexikografisches Ziel dieses Lexembestandes war, den geschrieben verwendeten deutschen Standardwortschatz aus einem definierten Korpus, d. h. empirisch zu ermitteln und bis zu einer hohen Qualitätsstufe, d. h. manuell zu lemmatisieren. Quantitativ sollte er mehr als die meisten Gebrauchswörterbücher abdecken und eine Vergleichsgröße für letztere darstellen.

Seit Januar 2004 sind diese 300.000 gebräuchlichen Einwortausdrücke mit orthografischen und morphologischen Angaben öffentlich zugänglich. Als weitere Lemmatypen werden zunächst Mehrworteinheiten und später eventuell auch produktive Wortelemente hinzu kommen. Diese 300.000 Stichwörter sind auch rückläufig sortierbar, so dass *elexiko* auch ein neues und umfangreiches rückläufiges Wörterbuch darstellt.

elexiko wird auch in Zukunft modular weiterentwickelt und ausgebaut, d. h. nicht entlang dem Alphabet, sondern vertikal nach Teilwortschätzen bzw. horizontal nach bestimmten Angabearten. Vertikal bedeutet hier, dass zu jedem Lexem eines in Angriff genommenen Wortschatzausschnitts das Spektrum sämtlicher Informationsarten, die in der Artikelstruktur vorgesehen sind oder aber ein begründeter Ausschnitt dieses Spektrums abgearbeitet wird. Alle für das Lexem relevanten Felder sind ausgefüllt und liefern so das Repertoire der Angaben, aus denen in einem davon getrennt gehaltenen Arbeitsschritt je Nutzungssituation und Adressatentyp die passenden ausgewählt und in geeignetem Layout präsentiert werden. Das erste Beispiel für einen vertikal ausgearbeiteten Teilwortschatz ist der 240 Lexeme umfassende Demonstrationswortschatz.

Im Unterschied zu solchen ‚tief' ausgearbeiteten Wortartikeln bedeutet „horizontal", dass einzelne Angabearten unter Einsatz computertechnischer Verfahren quasi über Nacht zu sämtlichen Stichwörtern eingetragen werden. Schlägt man ein solches Stichwort auf, findet man nur wenige, z. B. orthografische Informationen. Der große Nutzen der horizontalen Füllung liegt darin,

---

[8] Mit *computertechnisch* ist aus der Wissenschaftlerperspektive alles dasjenige gemeint, was nicht unter die rein sprachwissenschaftlichen Aufgaben gehört, auch wenn dieser Ausdruck aus informatischer Sicht undifferenziert ist und so viele unterschiedliche Kompetenzen vereint, dass kaum ein Einzelner, sondern nur ein Team sie aufweisen kann.

dass der gesamte Stichwortbestand nach den verschiedenen Attribuierungen einer solchen Angabeart durchsucht werden kann, z. B. danach, welche Wörter durch den Unterschied zwischen alter und reformierter Orthografie betroffen sind, welche Wörter über eine oder mehrere variante Schreibungen verfügen oder welche Wörter über eine morphologische Variante verfügen. Es gehört zur mittelfristigen Weiterentwicklung von *elexiko*, mittels computer- und korpuslinguistischer Verfahren Möglichkeiten der horizontalen Füllung zu finden, zu testen und zu realisieren. Im Grundsatz strebt *elexiko* hierbei eine sehr hohe Korrektheitsquote an. Der Einsatz digitaler Verfahren sollte u. E. nicht zu einem Verlust an lexikografischer Qualität führen, etwa dadurch, dass zu hochgradig polysemen bzw. zu homonymen Lexemen per Zufallsauswahl Belege ermittelt werden, die die semantische Heterogenität in aller Regel unterschlagen. Bei monosemen, niedrig frequenten und/oder zusammengesetzten Lexemen können Zufallsbelege hingegen sinnvoll sein.

4.1 Der Demonstrationswortschatz

Seit Juni 2004 stehen die 240 Artikel des sog. Demonstrationswortschatzes online zur Verfügung; er ist in die Gesamtstichwortliste integriert und dort über einen speziellen Rechercheknopf in der Suche-Funktion (vgl. http://www.elexiko.de) als Liste zu erhalten. Dieser Wortschatz wurde definiert und bestimmt, um die Möglichkeiten der lexikografischen Konzeption von *elexiko* in größtmöglicher Breite zu demonstrieren und auch, um für die weitere Planung eine Grundlage zu erhalten.

Der Demonstrationswortschatz besteht aus zwei Teilmengen: einmal aus den signifikanten Kontextpartnern des willkürlich nach gesellschaftlicher Relevanz ausgewählten Zentral-Lexems *Mobilität*, und zum anderen aus systematisch ergänzten Lexemen. Letztere sollten garantieren, dass alle Wortarten, Wortbildungstypen, Alphabetstrecken und dass hoch- und niedrigfrequente Wörter in diesem Wortschatzausschnitt in realistischer Verteilung vorkommen.

4.2 Neologismenwortschatz

Neologismen, d. h. neue Wörter und neue Bedeutungen ‚alter' Wörter werden am IDS kontinuierlich erforscht und in Zehnjahresschritten beschrieben. Die Neologismen der neunziger Jahre erschienen 2004 zunächst in Form eines gedruckten Buchs (Herberg/Kinne/Steffens 2004), wurden kurz darauf aber in erweiterter Form in *elexiko* integriert. Erweitert bedeutet hier, dass die elektronische Version mehr Belege enthält und darüber hinaus in vielfältigerer Weise

nutzbar ist. Die Publikation der Neologismen des ersten Jahrzehnts des 21. Jahrhunderts in *elexiko* kann schon vor Ablauf dieser ‚Nullerjahre' beginnen.

Das Modul unterscheidet sich in konzeptioneller Hinsicht leicht von den übrigen, insbesondere von den *elexiko*-Kernmodulen ‚Demonstrationswortschatz' und (zukünftig) ‚Lexikon zum öffentlichen Sprachgebrauch', insofern es ein über das *elexiko*-Korpus hinausgehendes, noch mehr auf Aktualität bedachtes Korpus zugrunde legt, in das grundsätzlich auch nicht-elektronisch verfügbare Belege einbezogen sind. Ferner macht der Neologismenwortschatz die Berücksichtigung spezifischer Kategorien wünschbar, wenn nicht nötig.[9]

4.3 Usuelle Wortverbindungen und Mehrwortlemmata

Dieses Modul wird strikt korpusbasiert mehr oder weniger feste Syntagmen (Kollokationen, Phraseme, Funktionsverbfügungen, Slogans usw.) identifizieren und als eigenen Stichwort- und Artikeltyp im Rahmen von *elexiko* und mit den Einwortausdrücken vernetzt lexikografisch ‚tief' beschreiben. Ziel ist darüber hinaus die Erfassung möglichst vieler fester Verbindungen und die Extraktion automatischer ‚horizontal-flacher' Angaben aus einem Korpus. Näheres unter http://www.ids-mannheim.de/lexik/UsuelleWortverbindungen/.

5. Literaturverzeichnis

5.1 Forschungsliteratur

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

Reichmann, Oskar (1989): Geschichte lexikographischer Programme in Deutschland. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (1989): Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilbd. S. 230-246. Berlin/New York. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1).

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

---

9 Aktuelle Informationen zu diesem Modul unter http://www.ids-mannheim.de/lexik/Neologie/

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

Wiegand, Herbert Ernst (2001): Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung. In: Lexicographica 17, S. 217-248.

Wiegand, Herbert Ernst (Hg.) (2002): Perspektiven der pädagogischen Lexikographie des Deutschen I: Untersuchungen anhand des „de Gruyter Wörterbuchs Deutsch als Fremdsprache". Tübingen. (Lexicographica / Series Maior; 110).

Zöfgen, Ekkehard (1994): Lernerwörterbücher in Theorie und Praxis: Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen. Tübingen. (Lexicographica / Series Maior; 59).

5.2 Wörterbücher

FWB = Frühneuhochdeutsches Wörterbuch. Hrsg. von Reichmann, Oskar/Goebel, Ulrich/Anderson, Robert A. Berlin/New York 1986 ff.

Herberg, Dieter/Kinne, Michael/Steffens, Doris (2004): Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi. Berlin/New York.

Langenscheidt Großwörterbuch Deutsch als Fremdsprache (1993). Hrsg. von Götz, Dieter/Haensch, Günther/Wellmann, Hans. 1. Auflage. Berlin/München.

VALBU (2004) = Schumacher, Helmut/Kubzcak, Jaqueline/Schmidt, Renate/de Ruiter, Vera: VALBU – Valenzwörterbuch deutscher Verben. Mannheim.

5.3 Internetressourcen

COSMAS II: http://www.ids-mannheim.de/cosmas2/ (letzter Zugang Oktober 2004).

Projekt GRAMMIS: http://hypermedia.ids-mannheim.de/grammis/ (letzter Zugang Oktober 2004).

Projekt Usuelle Wortverbindungen: http://www.ids-mannheim.de/lexik/UsuelleWortverbindungen/ (letzter Zugang Oktober 2004).

Projekt Neologismen: http://www.ids-mannheim.de/lexik/Neologie/ (letzter Zugang Oktober 2004).