

PETRA STORJOHANN

# Mit digitalem Textmaterial die innere Ordnung des Wortschatzes entdecken

## 1. Vorüberlegungen

Die linguistische Arbeit mit digitalen Textsammlungen (so genannten Korpora) hat in den letzten Jahren stark zugenommen. Zum einen deshalb, weil elektronische Daten immer besser zugänglich sind, Analysewerkzeuge den Zugang verbessern, zum anderen, weil das Potenzial umfangreicher elektronischer Datensammlungen zunehmend anerkannt wurde. Nach einigen Anfangsschwierigkeiten liegen nun auch im deutschsprachigen Raum eine Vielzahl an Untersuchungen vor, die zeigen, dass der Einsatz von Korpora in vielen Bereichen der theoretischen und angewandten Sprachwissenschaft Einzug gehalten hat (u. a. in der Übersetzungswissenschaft, Diskursanalyse, Lexikografie sowie im EFL- und DAF-Bereich). Dieser Beitrag soll über die Arbeit mit Korpora und Korpusmethoden informieren und exemplarisch zeigen, zu welchen Erkenntnissen man durch den Einsatz von Korpora für Wortschatzuntersuchungen gelangen kann. Er versucht Anregungen zu geben, wie Korpora für den Deutschunterricht eingesetzt werden können und zeigt auch, wie Korpora gewinnbringend für Wortschatzbeschreibungen in der Lexikografie eingesetzt werden. Zugleich wird ein kritischer Blick auf existierende Nachschlagewerke geworfen, die z. T. mit anderen Arbeitsgrundlagen und nicht korpusgeleiteten Methoden erarbeitet wurden. An anderer Stelle in diesem Heft wird eine Liste mit diversen Korpora, Wörterbüchern und Online-Ressourcen zur Verfügung gestellt.

## 2. Mit Korpora auf Entdeckungsreise

„Die Korpuslinguistik wertet das Korpus systematisch und nicht nur exemplarisch aus.“

TEUBERT 1999, 302

Die Arbeit mit digitalen Textkorpora ermöglicht Linguist(inn)en nicht nur schneller und durch den Einsatz spezieller Software unkomplizierter auf sprachliche Daten zuzugreifen, sondern sie stellt auch einen neuen methodischen Ansatz dar. Der größte Vorteil dieses Ansatzes liegt darin, dass Erkenntnisse auf empirischer Basis gewonnen werden, d. h., dass linguistisches Wissen über direkte Beobachtungen authentischen Sprachgebrauchs, wie er im Korpus abgebildet ist, erlangt wird. Deshalb geht die Korpuslinguistik vor allem der Frage nach, was im usuellen Sprachgebrauch wahrscheinlich ist und nicht, was theoretisch möglich ist. Ein weiterer Vorteil von Korpora liegt in der Möglichkeit, holistisch und ohne Erwartungen an Sprachdaten herangehen zu können. Im Gegensatz zu introspektiven Analysen, die ausschließlich auf eigener (mutter)sprachlicher Kompetenz fußen, sind Korpora systematisch zugänglich, und sie können Aussagen darüber treffen, was typisch und zentral in einer Sprache vorkommt. Korpora ermöglichen uns einen Blick hinter Sprachstrukturen, der vor 20 Jahren nicht möglich war. Sie können Erkenntnisse über Sprache bestätigen oder vertiefen, aber auch viel Neues und Unerwartetes ans Licht bringen. Beobach-

tungen zeigen, dass intuitive Erwartungen häufig von tatsächlichen Korpusfunden abweichen (vgl. Tognini-Bonelli 2001). Die Auswertung von Korpusdaten kann daher eine qualitative Veränderung unseres Verständnisses von Sprache und deren Strukturen herbeiführen. Die Arbeit mit Korpora macht deshalb eine Lücke besonders deutlich, nämlich die Diskrepanz zwischen dem Verständnis von und der Theorie über Sprache (einschließlich ihrer Strukturen) und der Vermittlung von Sprache in der Praxis. Dies gilt sowohl für den Deutschunterricht an Schulen als auch für das Fach Deutsch als Fremdsprache an Universitäten.

## 2.1 Methoden der Korpusanalyse

„Strictly speaking, a corpus by itself can do nothing at all, being nothing other than a store of used language.“

HUNSTON 2002, 3

Korpora allein sind nur eine Sammlung von Sprachdaten aus dem tatsächlichen Sprachgebrauch und mit wachsendem Umfang stellt sich die Frage, wie diese Datenmengen bewältigt werden können. Neben der Beschäftigung mit Korpusaufbau und Korpusrepräsentativität standen daher auch immer Fragen der Korpusrecherche und -analyse im Mittelpunkt der Korpuslinguistik. Heute können Korpusinteressierte auf Analysetools zurückgreifen, die mittels diverser Erschließungsmethoden Daten gezielt recherchier- und analysierbar machen. Mit Korpuswerkzeugen genießt man einen holistischen Zugang zur Sprache, d.h. Masendaten werden systematisch und ganzheitlich erfasst und mithilfe statistischer Auswertungen strukturiert. Dadurch können sprachliche Strukturen repräsentativer ermittelt und einzelne sprachliche Zufälligkeiten ausgeschlossen werden. Einfache Suchabfragen gewähren u. a. Einblicke in die Frequenz eines Wortes, geben Auskünfte über die chronologische Belegsituation eines Wortes oder dessen Erstbelegung. Recherche- und Analysetools, die mit mathematisch-statistischen Methoden arbeiten, helfen, semantisch Auffälliges innerhalb großer Textmengen aufzuspüren und regelhafte Zusammenhänge zu erkennen, indem sie die Rechercheergebnisse syste-

matisiert und strukturiert zur eigenen Interpretation zur Verfügung stellen. Solche automatisierten Strukturierungen enthalten u. a. Informationen zum syntagmatischen Gebrauch eines Wortes (z. B. zu typischen Verwendungsmustern, Idiomen oder Phrasenologismen) sowie zu paradigmatischen Verwendungen (z. B. Synonyme, Antonyme etc.).

In Abhängigkeit vom Zweck der Untersuchung können Korpustools auf verschiedene Weise eingesetzt werden. Sind Lehrende mit Korpora und deren Tools vertraut, können sie selbstständig überprüfen, ob Textkorpora auch zur Unterrichtsvorbereitung oder sogar im Deutsch- bzw. Medien-Unterricht selbst einsetzbar sind. Der Lernbereich Reflexion über Sprache, wie er in Richtlinien und Lehrplänen gefordert wird, soll z. B. den Zusammenhang zwischen Form (Wörter, grammatische Mittel und Strukturen) und Funktion (Adressaten- und Situationsangemessenheit) behandeln, er soll v. a. im Bereich der Rechtschreibung und der Grammatik gültige Sprachnormen vermitteln, aber zugleich als im Prinzip veränderbare Setzungen verständlich machen und er soll Einsichten über Zweck und Probleme des Varietätengebrauchs, z. B. der Fachsprachen, vermitteln. Korpora bieten mit zahlreichen Textbelegen Lehrenden wie Lernenden die Möglichkeit, selbst zu erkennen, was Usus, was Vorschrift und was Abweichung ist, was häufig und was selten ist sowie was stilistische Markiertheit heißen könnte. Sie lassen teilweise auch Sprachwandel erkennen, z. B. die Tendenz zum Wechsel der Kasusreaktion bei vielen Präpositionen, das langsame oder stürmische Aufkommen neuer Wörter, seien sie entlehnt oder aus deutschen Elementen gebildet. Ob Korpora hierbei Informationen aus Wörterbüchern ergänzen oder vielmehr ersetzen (auch: kritisch kommentieren) können, bleibt einer kreativen Unterrichtspraxis überlassen. Aber auch der Teilbereich Wortfeld und Arbeit an „Begriffen“ (Bedeutungen) ließe sich mithilfe von Korpora womöglich lebensnaher gestalten. Das lesende Erarbeiten von Sachinformationen aus Texten könnte auch dadurch gestaltet werden, dass den Schülerinnen und Schülern eine Reihe von Satzbelegen zu einem schillernden Wort (Modewort, Fachwort, usw.) aus

einzelnen Frau der Fall, kann sie eine gesonderte	<b>Abstimmung</b>	von Frauen und Männern beantragen.
da sich alle anderen Abgeordneten vor der	<b>Abstimmung</b>	unter Protest zurückzogen.
Die bei der	<b>Abstimmung</b>	am Sonntagabend unterlegene Gruppe warf
Im hiesigen Staatswesen findet nun die	<b>Abstimmung</b>	am Kiosk statt.
SPD und Grüne beantragten eine namentliche	<b>Abstimmung</b>	über die Entlassung des Finanzministers
Für die heutige	<b>Abstimmung</b>	im Bundestag wird zwar erwartet, daß
von SPD und Grünen stimmen in namentlicher	<b>Abstimmung</b>	für die Abberufung von Kämmerer
des Halleiner Stadtverkehrs sollte eine bessere	<b>Abstimmung</b>	auf die Zubringerlinien von KWD und
worden, die seinerzeit willkürlich und ohne	<b>Abstimmung</b>	mit denen getroffen wurde, die für den
Brisch gepflegten engen Kontakte, die bis in die	<b>Abstimmung</b>	bei Ankäufen gegangen sein sollen,
Das eingeschossige Gebäude wurde in	<b>Abstimmung</b>	zwischen den Architekten,
einer geringen Beteiligung ging die landesweite	<b>Abstimmung</b>	zur Abschaffung des Regierungsprozesses
Presseberichte, wonach es unter den Musikern eine	<b>Abstimmung</b>	gegen den 67jährigen gegeben habe

Beispiel 1:  
Ausschnitt von  
Konkordanzen  
für das Suchwort  
*Abstimmung*

einem Korpus vorgelegt wird und sie eine Bestimmung von Bedeutung, Gebrauch und „Konnotationen“ vornehmen. In einem einzelnen Text kommt ein interessierendes Wort zu selten vor, um das gleiche leisten zu können, und Wörterbuchinformationen abstrahieren und werten den Sprachgebrauch oft schon zu sehr.

Recherchewerkzeuge mit unterschiedlichen Funktionen werden in vielen wissenschaftlichen Institutionen und Firmen entwickelt. Dennoch verfügen die meisten über Funktionalitäten wie eine statistische Auswertung der Daten und die Erfassung kontextueller Partnerwörter. Die grundsätzlichen Verfahren (vor allem zur Erfassung von Wortschatzstrukturen, z. B. in Wortfeldern oder semantischen Netzen) sollen kurz im Folgenden erläutert werden. Dazu werden exemplarisch die Korpora geschriebener Sprache des Instituts für deutsche Sprache Mannheim<sup>1</sup> und das Werkzeug *COSMAS* mit seiner integrativen Software *Statistische Kollokationsanalyse und Clustering*<sup>2</sup> zur Illustration herangezogen.

### 2.1.1 Konkordanzen – alle Wortumgebungen auf einen Blick

Die meisten KorpusnutzerInnen untersuchen Wörter oder lexikalische Beziehun-

gen mithilfe von Konkordanzen im KWIC<sup>3</sup>-Format. Das Suchwort wird zeilenweise so dargestellt, dass es mittig positioniert ist, umgeben von seinem unmittelbaren Kontext. Diese Darstellung zeigt vor allem die syntagmatische Einbettung des Suchwortes und ermöglicht es, schnell sprachliche Regularitäten in der Wortverwendung zu erfassen, variante und invariante Strukturen aufzuspüren sowie unterschiedliche Verwendungsmuster zu erkennen. Die automatisch gewonnenen KWIC-Zeilen lassen sich unterschiedlich sortieren (z. B. nach Textquellen, chronologisch, alphabetisch nach Vorgänger- oder Nachfolgewort) und können wie folgt aussehen (**Bsp. 1**).

Dadurch, dass das Suchwort in seiner unmittelbaren kontextuellen Umgebung zu sehen ist, lassen sich in Abhängigkeit des Suchwortes und seiner Wortart unterschiedliche Wortschatzstrukturen beobachten. In Beispiel 1 kann z. B. gezeigt werden, welche präpositionalen Anschlüsse in Verbindung mit *Abstimmung* möglich sind. Aber auch andere syntagmatische Relationen können anhand von Konkordanzen erkannt werden, z. B.

1) typische Subjekt (Agens)- oder Objektbesetzungen (Patens, Experiens etc.) (bei Verben),

(1) Siehe Tabelle mit Korpora und Nachschlagewerken in diesem Heft.

(2) Die *Statistische Kollokationsanalyse und Clustering* wurde von Cyril Belica (1995-2002) am IDS Mannheim entwickelt (<http://corpora.ids-mannheim.de/cosmas>).

(3) KWIC=Key Word in Context.

SPD und Grüne	<b>üben</b>	heftige <b>Kritik</b> an Zimmermanns Plänen zur
Die Bundesgrünen	<b>üben</b>	scharfe <b>Kritik</b> und kündigen eigenes
Wir	<b>üben</b>	ja auch eine grundlegende <b>Kritik</b> an diesen
Die „Panther“	<b>üben</b>	auch inhaltlich herbe <b>Kritik</b> an Arkenstette und
Vorfälle zum Anlaß, grundsätzliche	<b>Kritik</b>	am Gesundheitssystem zu <b>üben</b> .
die jüdischen Menschen nicht berechtigt, offene	<b>Kritik</b>	am Staate Israel zu <b>üben</b> und gegen seine
nach vorne los, weil sie plötzlich ernsthafte	<b>Kritik</b>	an der populären Kultur <b>üben</b> musste.
bei der IG Metall gefordert Delegierte	<b>üben</b>	deutliche <b>Kritik</b> am Führungsstil des
Ich muß allen Ernstes harsche	<b>Kritik</b>	an der taz <b>üben</b> .
Aids-Aktivistin	<b>üben</b>	harte <b>Kritik</b> an den Studio-Managern.
ruhig, und die Gewerkschaftsverbände	<b>üben</b>	nur sanfte <b>Kritik</b> .
weil sie geeignet sind, versteckte	<b>Kritik</b>	am Regime zu <b>üben</b> .
Journalisten nicht den Mut haben, fundamentale	<b>Kritik</b>	an der Regierung zu <b>üben</b> .
sich gegenseitig anzuerkennen und konstruktive	<b>Kritik</b>	zu <b>üben</b> .
kann nicht die Rede sein, da Eltern erhebliche	<b>Kritik</b>	am Kindergarten <b>üben</b> .
Japan einfach nicht möglich, starke, explizite	<b>Kritik</b>	an diesem System zu <b>üben</b> oder auch
argumentierten, konnte der Ingenieur fundierte	<b>Kritik</b>	an der geplanten Anlage <b>üben</b>
Allerdings nutzt Hampel die Gelegenheit, zarte	<b>Kritik</b>	zu <b>üben</b> : Im Vergleich mit anderen deutschen
Es hilft nichts, lautstarke	<b>Kritik</b>	zu <b>üben</b> “, sagte Trittin.
SPÖ und Grüne	<b>üben</b>	vernichtende <b>Kritik</b> an Grassers Budget
Geharnischte	<b>Kritik</b>	<b>üben</b> S-18-Gegner am Plan von Landeshauptmann

- 2) attributive Hinzufügungen (bei Substantiven und Adjektiven),
- 3) verbale Bezugswörter (bei Substantiven und Adjektiven),
- 4) valenzabhängige Anschlüsse (bei Verben und Verbalsubstantiven),
- 5) Bestandteile von Funktionsverbgefügen oder Redewendungen.

Besonders vorteilhaft können Konkordanzanzen z. B. auch zur Untersuchung von variablen syntagmatischen Beziehungen eingesetzt werden. Beispiel 2 illustriert die automatische Erfassung unterschiedlicher attributiver Erweiterungen innerhalb des Funktionsverbgefüges *Kritik üben* (Bsp. 2).

Dieses Beispiel zeigt, wie man mit Schülern und Schülerinnen mögliche Attribuierungen (z. B. *heftig*, *scharf*, *grundlegend*, *grundsätzlich*, *ernsthaft* etc.) zu *Kritik* im Verbgefüge *Kritik üben* ermitteln könnte. Auch wenn Konkordanzanzen schnell die Verwendung eines Wortes und seine Be-

ziehungen zu unmittelbaren kontextuellen Mitspielern erkennen lassen, ist ihr Nutzen eingeschränkt, da bei hochfrequenten Suchwörtern (mit mehr als 500 Belegen) nicht alle Konkordanzanzen „manuell“ ausgewertet werden können. Mithilfe von Konkordanzanzen lassen sich keine Aussagen über Häufigkeiten und Signifikanz treffen und sie vermögen es nicht, Information zu kategorisieren. Für solche Ansprüche sind zusätzliche Funktionalitäten von Korpustools heranzuziehen, die die Analyse so genannter Kollokationen<sup>4</sup> durchführen können.

### 2.1.2 Kollokationen – oder aufschlussreiche Wortnachbarschaften

Eine lexikalisch-semantische Strukturierung der im Korpus abgebildeten Kontexte eines Wortes lässt sich systematisch über Kollokationen (auch Kookkurrenzen genannt) erfassen. Dabei handelt es sich um die statistische Ermittlung lexikalischer Mitspieler,

Beispiel 2:  
Ausschnitt der Konkordanzanzen für *Kritik üben*

(4) Der Begriff *Kollokation* bezieht sich in diesem Beitrag nicht auf syntaktisch binäre Relationen wie *blondes Haar* oder *eingefeischter Junggeselle*, sondern lehnt sich an den Begriff *collocation* der Korpuslinguistik an. Er wird als loses Miteinandervorkommen von Wörtern innerhalb eines Kontextes verstanden.



Anzahl	LLR <sup>5</sup>	Kookkurrenzen	Struktur	Typ
330	2351	<b>Reiter</b>	bedeut.konst. Rel.	thematischer Mitspieler
60	1035	<b>Kleintiere</b>	paradigm. Rel.	Kohyponym
128	1030	<b>Rücken</b>	paradigm. Rel.	Partonym
36	617	<b>Trojanisches</b>	syntagm. Rel.	MWV/Metapher
68	557	<b>Stall bestes</b>	bedeut.konst. Rel.	thematischer Mitspieler
35	460	<b>falsche</b> gesetzt	syntagm. Rel.	MWV: <i>aufs falsche Pferd setzen</i>
30	431	<b>Ponys</b>	paradigm. Rel.	Kohyponym
45	356	<b>reiten</b>	syntagm. Rel.	verbales Bezugswort
18	266	<b>galoppierenden</b>	syntagm. Rel.	Attribut
33	264	<b>Kutsche</b>	bedeut.konst. Rel.	thematischer Mitspieler
18	236	<b>aufzäumen</b>	syntagm. Rel.	MWV: <i>das Pferd am/vom/beim Schwanz/ von hinten aufzäumen</i>
1927	213	<b>Koppel</b>	bedeut.konst. Rel.	thematischer Mitspieler
2122	171	<b>scheu</b>	syntagm. Rel. syntagm. Rel.	MWV: <i>die Pferde [...] scheu machen</i> Attribut
2315	153	<b>Schwanz</b>	paradigm. Rel.	Partonym
2351	150	<b>Hufe</b>	bedeut.konst. Rel.	thematischer Mitspieler
2412	131	<b>beschlagen</b> wurden	syntagm. Rel.	verbales Bezugswort
2486	114	<b>satteln</b>	syntagm. Rel.	verbales Bezugswort
2635	105	<b>Stallungen</b>	bedeut.konst. Rel.	thematischer Mitspieler
2639	106	<b>edle</b>	syntagm. Rel.	Attribut
2772	102	<b>Fohlen</b>	paradigm. Rel.	Hyponym
2781	102	<b>traben</b>	syntagm. Rel.	verbales Bezugswort
2858	101	<b>Sattel</b>	bedeut.konst. Rel.	thematischer Mitspieler
3703	74	<b>Hufeisen</b>	bedeut.konst. Rel.	thematischer Mitspieler
3796	71	<b>wiehern</b>	syntagm. Rel.	verbales Bezugswort
3926	64	<b>Stufenbarren</b>	paradigm. Rel.	Kohyponym
3928	64	<b>Schwebebalken</b>	paradigm. Rel.	Kohyponym
4262	53	<b>Lipizzaner</b>	paradigm. Rel.	Hyponym
4494	47	<b>rassige</b>	syntagm. Rel.	Attribut

Beispiel 3:  
exemplarischer  
Kollokations-  
ausschnitt für das  
Suchwort *Pferd*

die zusammen mit dem Suchwort in einem bestimmten wählbaren Wortabstand in Zusammenhang stehen. Die automatische Kollokationsanalyse ordnet die ermittelten miteinander vorkommenden Partnerwörter nach lexikalischer Kohäsion (bestimmt nach Frequenz und lexikalischer Nähe) und deckt dadurch Regelmäßigkeiten in der Verwendung eines Suchwortes auf. Neben binären lexika-

lischen Relationen werden außerdem mehrgliedrige usuelle Wortverbindungen sowie ganze Phraseologismen erfasst. Die größten Vorteile der Kollokationsanalyse sind die automatisch gewonnene Strukturierung der im Korpus abgebildeten Kontexte eines Suchwortes, die statistische Erfassung von typischen und zentralen Mustern sowie die synoptische Präsentation der Ergebnisse.

(5) LLR = Log Likelihood Ratio ist das mathematisch-statistische Maß, mit dem die lexikalische Kohäsion erfasst wird.

Anzahl	LLR	Kookkurrenzen	Struktur	Typ
65	8383	<b>Kreditkarten</b>	syntagm. Rel.	semantische Rolle: Effiziertes Obj.
62	4177	<b>Bevölkerung</b>	syntagm. Rel.	semantische Rolle: Agens
558	1583	<b>Entscheidung</b>	syntagm. Rel.	semantische Rolle: Effiziertes Obj.
92	1074	<b>zähneknirschend</b>	syntagm. Rel.	Attribut
88	1010	<b>stillschweigend</b>	syntagm. Rel.	Attribut
235	732	<b>Vorschlag</b>	syntagm. Rel.	semantische Rolle: Effiziertes Obj.
155	480	<b>Gesellschaft</b>	syntagm. Rel.	semantische Rolle: Agens
52	460	<b>widerwillig</b>	syntagm. Rel.	Attribut
125	310	<b>Mehrheit</b>	syntagm. Rel.	semantische Rolle: Agens
8	111	<b>Gesprächspartner</b>	syntagm. Rel.	semantische Rolle: Patiens
17	77	<b>Partner</b>	syntagm. Rel.	semantische Rolle: Patiens
3	45	<b>respektieren</b>	paradigm. Rel.	Kohyponym
4	27	<b>integrieren</b>	paradigm. Rel.	Kohyponym

Beispiel 4:

exemplarischer Kollokationsausschnitt für das Suchwort *akzeptieren*

Anzahl	LLR	Kookkurrenzen	Struktur	Typ
347	3807	<b>hohe</b>	syntagm. Rel. syntagm. Rel.	Attribut Funktvgb. <i>hohe Akzeptanz genießen</i>
218	3206	<b>breite</b> stößt	syntagm. Rel. syntagm. Rel.	Attribut Funktvgb: <i>auf breite Akzeptanz stoßen</i>
12	776	<b>Toleranz</b>	paradigm. Rel.	Kohyponym
116	715	<b>erhöhen</b>	syntagm. Rel.	verbales Bezugswort
20	86	<b>Anerkennung</b>	paradigm. Rel.	Synonym
3	85	<b>Nichtakzeptanz</b>	paradigm. Rel.	Antonym
20	67	<b>verbessern</b>	syntagm. Rel.	verbales Bezugswort
11	67	<b>Glaubwürdigkeit</b>	paradigm. Rel.	Kohyponym
16	60	<b>Ablehnung</b>	paradigm. Rel.	Antonym

Beispiel 5:

exemplarischer Kollokationsausschnitt für das Suchwort *Akzeptanz*

Die Untersuchung kontextueller Mitspieler bietet besonders fruchtbaren Boden für die Ermittlung von unterschiedlichen Wortschatzbeziehungen in sprachlichen Kontexten. Kollokatoren erlauben uns u. a. Einblicke in die thematisch-diskursiven, referentiellen, paradigmatischen und syntagmatischen Verbindungen, die ein Wort innerhalb des Wortschatzes verorten. Daher kann über Kollokationen das semantische Profil (inklusive semantischer Besonderheiten) eines Wortes erschlossen werden. Die-

se Strukturen sind sehr heterogen, schließen variable oder feste Muster unterschiedlicher Art ein und können je nach Suchwort und Wortart vor allem die Folgenden sein:

- paradigmatische Beziehungen: Sinnverwandte Wörter (Synonyme, Antonyme, Hyperonyme etc.),
- syntagmatische Zusammenhänge: unmittelbares Bezugswort, Attribuierungen, Bestandteile von Funktionsverbgefügen, Bestandteile von Mehrwortverbindungen (MWV), semantische Rollen (z. B.



Suchwortes vornehmen oder auch konnotative Verwendungsmerkmale (z. B. Sprechereinstellung, Situationsbedingungen etc.) aufspüren wollen. Auch für Lexikografinnen, die diskursive Wortbeschreibungen und kontextuelle Veränderungen im Wortgebrauch nachvollziehen wollen, ist die Volltextanalyse unabdingbar. Einige Nachschlagewerke (z. B. *Brisante Wörter, Zeitgeschichtliches Wörterbuch*)<sup>7</sup> bieten diskursgeschichtliche Beschreibungen lexikalischer Einheiten an. Im Vordergrund stehen dabei Erläuterungen textkohärenter Besonderheiten, d. h. dass die Verwendung eines Wortes innerhalb eines thematischen Kontextes zu einem bestimmten Zeitpunkt unter Berücksichtigung kultureller Faktoren untersucht wird. Diese stellen historisch-soziokulturelle Dokumentationen gemeinschaftlichen Sprachgebrauchs dar und beschreiben Wortschatz zeit- und sprecher (gruppen)abhängig. Das folgende Beispiel<sup>8</sup> soll die Ergebnisse einer diskursiven Analyse verdeutlichen (siehe dazu auch U. Haß in diesem Heft, insbesondere Abschnitt 4) (Bsp. 7).

Mit den einzelnen Methoden der Korpuserschließung, von Konkordanzen über Kollokatoren bis hin zur Analyse umfangreicherer Textzusammenhänge lassen sich im Korpus unterschiedliche Wortschatzstrukturen ermitteln. Die hier geschilderten Fälle beziehen sich dabei ausschließlich auf lexikalische Strukturen, die semantische Beziehungen zwischen Wörtern erkennen lassen. Mithilfe korpusgestützter Ansätze können auch differenzierte Beobachtungen zu syntaktischen Gebrauchsbesonderheiten gemacht werden. Es lässt sich z. B. ermitteln, ob ein Verb häufig aktivisch oder passivisch gebraucht wird, ob ein bestimmtes Adjektiv attributiv, prädikativ oder adverbial verwendet wird, ob ein Substantiv häufiger im Plural als im Singular belegt ist oder welche präpositionalen Anschlüsse für ein Suchwort möglich sind (vgl. Beispiel 1). Auf die Vorteile computergesteuerter Analysen von syntaktischen Relationen wird hier aber nicht näher eingegangen.

Das, was im *ellexiko*-Korpus als **Heimat** gekennzeichnet wird, wird vorwiegend positiv bewertet. Besonders deutlich wird das in Kontexten, in denen **Heimat** zusammen mit Ausdrücken wie **Geborgenheit, Identität, das Zuhause, Traditionen, Kindheit** oder auch das **Gute und Schöne** thematisiert wird. Wenn Sprecher z. B. mittels Negation etwas als Verlust von Heimat charakterisieren, werden gleichzeitig oft die Ausdrücke **Sehnsucht, Liebe und Verbundenheit** gebraucht.

**Heimat** wird im *ellexiko*-Korpus häufig in politischen Kontexten verwendet, besonders in Kontexten, in denen Kriegssituationen geschildert werden, in denen Flüchtlinge gezwungen sind, ihre Heimat zu verlassen. **Heimat** wird aber auch in gesellschaftspolitischen Kontexten verwendet, in denen u. a. die Situation von Asylbewerbern, die von Abschiebung betroffen sind, thematisiert wird.

Ein als **Heimat** gekennzeichnete Ort wird im *ellexiko*-Korpus häufig zugleich als Ort der Traditionen, heimischer Kultur und des Brauchtums dargestellt. Damit einhergehend werden mit **Heimat** viele Erinnerungen, Liebe und Sehnsucht assoziiert.

### 3. Korpora in der Lexikografie

„Corpora have so revolutionised the writing of dictionaries [ ...].“

HUNSTON 2002, 96

Beispiel 7:  
Beschreibung  
des Stichworts  
*Heimat* in typischen  
Diskursen

Die Richtlinien und Lehrpläne lassen erkennen, dass zur Sprachreflexion im Deutschunterricht auch die reflexiv-kritische Auseinandersetzung mit Sprachnormen gehört. Deshalb kann die Frage, wie Wörterbücher entstehen und auf welchen Grundlagen die in ihnen kodifizierten Normen formuliert wurden, nicht ausgeschlossen bleiben. Mit elektronischen Textkorpora zu arbeiten, heißt auch, die Bedingungen der Wörterbucharbeit zu erfahren und den Prozess der Normenkodifizierung mittels Wörterbüchern und Grammatiken durchschauen zu lernen. Während in der englischen Lexikografie Korpora bereits seit ca. 20 Jahren die Auswertung und Beschreibung sprachlicher Da-

(7) Vgl. dazu U. Haß in diesem Heft.

(8) Das Beispiel *Heimat* und seine diskursive Beschreibung wurde dem Internetwörterbuch *ellexiko* entnommen. (Siehe dazu [www.ellexiko.de](http://www.ellexiko.de).)

ten verändert haben, werden sie in der deutschen Lexikografie erst seit einigen Jahren genutzt.<sup>9</sup> Was Teubert bereits 1998 anmerkte, gilt heute immer noch: Deutschland verfügt bis dato nicht über eine korpusbasierte Standardlexikografie. Das Korpus als Datengrundlage wird in der deutschen Lexikografie sehr unterschiedlich eingesetzt.<sup>10</sup> Einige Projekte erstellen Nachschlagewerke auf Korpusbasis, deren Inhalte ausschließlich automatisch mittels informationstechnologischer Verarbeitung extrahiert und nicht weiter linguistisch ausgewertet werden (z. B. WORTSCHATZ-LEXIKON). Dann gibt es LexikografInnen, die elektronische Textsammlungen z. T. in ihre Arbeit mit einbeziehen, aus denen schnell und komfortabel Belegquellen extrahiert werden (z. B. DUDEN-Wörterbücher). Darüber hinaus liegen traditionelle Printwörterbücher vor, die nachträglich digitalisiert und mit Korpusbelegen angereichert wurden (z. B. das DIGITALE WÖRTERBUCH DER DEUTSCHEN SPRACHE DES 20. JAHRHUNDERTS), und es gibt Wörterbücher, die ihre lexikografischen Angaben ausschließlich durch die Auswertung von Korpora ermitteln, dabei auf korpusverarbeitende Software zurückgreifen und die extrahierten Daten einer linguistischen Auswertung und Interpretation unterziehen (z. B. ELEXIKO). Welchen Zweck und welches Ziel auch jedes Wörterbuchprojekt verfolgt, es wird deutlich, dass Korpora bei der lexikografischen Arbeit eine wichtige Rolle einnehmen.

### 3.1 Erkenntnisse aus Korpora im Wörterbuch?

„Besonders deutschen standardsprachlichen Wörterbüchern ist anzusehen, daß sie viel zu sehr auf tradierten Wortbeschreibungen, die sich über ganze Wörterbuchgenerationen fortschreiben, als auf eigener Beleganalyse beruhen.“

TEUBERT 1998, 154

Ein ausgewogenes Korpus als Arbeitsgrundlage zu nutzen, um ohne Vorannahmen an sprachliche Daten heranzugehen und die Korpusdaten lexikografisch auszuwer-

ten und zu interpretieren, ist Ziel weniger Wörterbücher oder Grammatiken. Obwohl mit einem möglichst repräsentativen Korpus Daten systematisch und holistisch ausgewertet und interpretiert werden können und Lexikograf(innen) nicht auf ihr persönliches Sprachgefühl zurückgreifen müssen, um sämtliche Verwendungen eines Wortes und seine Beziehungen zu anderen Wörtern zu analysieren, werden Korpora derzeit kaum in den Prozess der Erkenntnisgewinnung einbezogen, um neue Ansätze und Beschreibungsmöglichkeiten zu finden. Die deskriptive korpusgeleitete Beschreibung der Sprachwirklichkeit, also das, was tatsächlich im Sprachgebrauch konkret realisiert wird, kann dabei von Wörterbuchangaben, die auf einer anderen Basis gewonnen wurden, stark abweichen. Diese Abweichungen können sich auf unterschiedliche Wörterbuchinformationen beziehen. Einige Abweichungen, die sich auf die Beschreibung von paradigmatischen und syntagmatischen Wortschatzbeziehungen, d. h. von Wortfeldern und Wortverbindungen im Satz beziehen, sollen nachfolgend exemplarisch vorgeführt werden.

#### 3.1.1 Paradigmatische Wortschatzstrukturen in Wörterbüchern

Im DUDEN 8 (Synonymwörterbuch) wird unter dem Eintrag *Anerkennung* (2b) das Synonym *Duldung* gegeben. In verschiedenen Korpora des IDS lässt sich die synonymische Verwendung aber nicht nachweisen. Zwar kommen beide Suchwörter häufig miteinander in gemeinsamen Texten vor, sie sind aber in den entsprechenden Kontexten nicht substituierbar. Statt semantischer Identität oder Nähe, wie sie von DUDEN 8 vorgeschlagen wird, werden die Stichwörter häufig in Kontexten miteinander kontrastiert. Der folgende Korpusbeleg soll das demonstrieren:

GAL-Fraktionschefin Antje Möller weist auf die „Stichwortgeber“ für die Rechten hin. Sie zitiert CDU-Politiker, die den Neonazis durch „leichtfertigen Umgang

(9) Vgl. z. B. Scholze-Stubenrecht (2001) zu Veränderungen der lexikografischen Arbeit in der Dudenredaktion und Sinclair (1987) zum Korpuseinsatz für das COBUILD-Wörterbuch.

(10) Näheres zu einzelnen Wörterbüchern und Korpora siehe Tabelle „Elektronische Textsammlungen und Korpora zur deutschen Sprache“ in diesem Heft.

mit Worten“ Vorschub geleistet haben. „Ich wünsche mir einen Wahlkampf, in dem die Ausländer kein Thema sind, und in dem es um **Anerkennung** statt um **Duldung** geht“, verlangt sie und fordert, dass Zivilcourage nicht zur Worthülse verkommen dürfe.

die tageszeitung, 07.09.2000, S. 21,  
Ressort Hamburg Aktuell;  
Mehr oder weniger gegen rechts

Auch die Verben *akzeptieren* – *dulden*, *hinnehmen* sind im DUDEN 8 als Synonyme verzeichnet. Im Korpus lassen sich tatsächlich eine Anzahl synonymischer Verwendungen nachweisen, diese sind allerdings seltener belegt als Verwendungen, in denen die Suchwörter deutlich voneinander abgegrenzt werden, wie folgende Belege zeigen:

Wer in der einen Szene Opfer ist, kann in der nächsten schon Täter sein. So der bosnische Serbe, als Flüchtling in Belgrad mehr **geduldet** als **akzeptiert**.

Züricher Tagesanzeiger, 18.06.1999,  
S. 20, Bilder der Gewalt

Reformen und Veränderungen im Bildungswesen und in der Gesellschaft erfolgen heute sehr schnell. Das ist eine Tatsache, die man akzeptieren muss. **Akzeptieren** heisst für mich aber nicht, es einfach resignierend **hinzunehmen**.

St. Galler Tagblatt, 25.03.2000,  
Mitsprache mit Mitverantwortung

Ähnlich verhält es sich mit dem Substantiv *Kauf*, für das im DUDEN 8 in der 3. Einzelbedeutung die Synonyme *Bestechung*, *Korruption*, *Schmierer* angegeben sind. Bei genauer Analyse vieler Korpusbelege wird aber deutlich, dass es sich bei den Referenzstellen um völlig unterschiedliche Bezugswörter handelt und dass eher eine kausale Beziehung zwischen den Partnerwörtern vorliegt.

Bislang wurden 14 Personen aus dem olympischen Zirkel bekannt, die vom Bewerbungskomitee der Olympischen

Winterspiele in Salt Lake City Zahlungen oder Leistungen im Wert von knapp 800 000 Dollar in Anspruch genommen haben. Die **Bestechung** der IOC-Mitglieder war laut einem Bericht der „New York Times“ sogar von langer Hand geplant. Die der Zeitung vorliegenden Dokumente sollen beweisen, daß sich die Offiziellen der Olympia-Stadt 2002 schon 1991 darauf verständigt hatten, Gelder gezielt für den **Kauf** von Stimmen einzusetzen.

Berliner Zeitung, 01.02.1999,  
IOC weitet Untersuchungen aus, S. 35.

Diese Beziehung kann aufgrund der Belegauswertung nicht als Synonymie definiert werden, sondern lässt sich wie folgt beschreiben:<sup>11</sup>

Ähnlich wie bei dem Verb **kaufen** in der Lesart ‚bestechen‘, wird durch **Kauf** auf eine Handlung Bezug genommen, bei der eine Person etwas erwirbt, nachdem der Tatbestand der Korruption oder der Bestechung vorlag. Somit unterscheidet sich **Kauf** von **Bestechung** und **Korruption** dadurch, dass in gemeinsamen Kontexten ein anderes Objekt verwendet wird und erst nachdem z. B. die Bestechung eines Politikers stattfand, damit der Kauf von Wählerstimmen erfolgen konnte. In den Texten des *lexiko*-Korpus ist **Kauf** als Folge von **Bestechung** oder **Korruption** thematisiert. Darüber hinaus konnte anhand der Texte des *lexiko*-Korpus beobachtet werden, dass **Kauf** häufig, wenn auch nicht ausschließlich, ein unbelebtes Objekt verlangt, während **Bestechung** und **Korruption** belebte Objekte fordern.

Ähnliche Beschreibungslücken gibt es auch in Nachschlagewerken, die automatisch gewonnene Daten ohne lexikografische Auswertung zur Verfügung stellen. So verfügt das WORTSCHATZ-LEXIKON zwar über automatisch generierte Synonyme und Antonyme, die aber mangels linguistischer Auswertung nur bedingt nutzbar sind. Zum Stichwort *Anerkennung* wird das Gegenwort *Nichtanerkennung* gebucht. Die Extrahierung solcher Partnerwörter basiert auf einem Verfahren, das statistisch signifikante Wörter innerhalb eines bestimmten Wortabstandes (häufig max. 5 Wörtern) ermittelt

(11) Dieses Beispiel wurde dem korpusbasierten Internetwörterbuch *lexiko* entnommen.



oder eine Suche nach vorgegebenen Negationspräfixen durchführt. Die lexikografische Praxis zeigt aber, dass viele Synonyme und auch Antonyme häufig in einem wesentlich größeren Abstand voneinander stehen und viele sinnverwandte Wörter deshalb mit diesem Verfahren nicht erfasst werden. Das wäre eine Erklärung dafür, dass Mitspieler wie *Ablehnung*, *Missbilligung* und *Skepsis* nicht gebucht sind, von denen zwei in den IDS-Korpora häufiger in gemeinsamen Kontexten belegt sind als *Nichtanerkennung*.<sup>12</sup>

### 3.1.2 Syntagmatische Strukturen

Auch in Bezug auf syntagmatische Wortschatzstrukturen, d. h. Wortverbindungen im Satz, lassen sich im Korpus Abweichungen zu in Wörterbuch angeführten Informationen nachweisen. Das Verb *kicken* wird im Duden GWDS wie folgt paraphrasiert:

**kijcken** <sw. V.; hat> [engl. to kick, Kick] (ugs.): **1.** *Fußball spielen*: die Kinder gehen k.; er hat schon seit frühester Jugend gekickt; sie kickt für den HSV. **2.** *(einen Ball) mit dem Fuß stoßen, schießen*: er hat den Ball ins Tor gekickt; Ü ... höhnt auch Richy und kickt eine Konservendose statt eines Fußballs so, dass sie haarscharf an Adams Nase vorbeizischt (Ossowski, Flatter 93).

Ein Blick in die Konkordanzen verrät, dass für eine Handlung, die mit *kicken* bezeichnet wird, auch andere Handlungsträger neben *Fuß*, wie es in der Bedeutungserläuterung festgehalten ist, vorkommen können. Darüber hinaus lassen sich weitere Verwendungen des Suchwortes nachweisen (z. B. *das kickt extrem*), die auf Einzelbedeutungen hinweisen, die nicht im GWDS gebucht sind (**Bsp. 8**).

Auch bei der Ermittlung von variablen syntagmatischen Zusammenhängen, die in einem Nachschlagewerk als Varianten gebucht werden sollen, kann die korpusgestützte Ermittlung von Mustern und deren Frequenz Vorteile geltend machen. Die Mehrwortverbindung *dumm wie Bohnenstroh* ist im Duden 11 (Redewendungen) mit der Variante *dumm wie die Nacht sein* eingetragen. Im

Korpus können eine Vielzahl an Varianten aufgespürt werden, wie folgende Konkordanzen belegen (**Bsp. 9**).

Die Ermittlung der Häufigkeit eines einzelnen Musters ergibt, dass z. B. die lexikalische Einheit *dumm wie Brot* deutlich frequenter ist (14 Treffer) als die im Wörterbuch eingetragene Variante *dumm wie die Nacht sein* (1 Treffer). Sprachinteressierte, die sich für den Variantenreichtum und die sprachliche Kreativität bestimmter Syntagmen interessieren, können besonders mithilfe von Konkordanzen Muster ermitteln, die häufig (auch aus Platzgründen) nicht Eingang in ein Wörterbuch gefunden haben.

## 4. Ausblick

„Now corpora, large and small, are seen by many teachers as useful tools, and are being put into use more and more every day. Access has become easy on standard small computers, user-friendly software is available for most normal tasks, websites are accumulating fast, and corpora are almost part of the pedagogical landscape.“

SINCLAIR 2004, 2

Das oben aufgeführte Zitat von Sinclair mag wohl für den englischsprachigen Raum zutreffen, die Arbeit mit Korpora im deutschsprachigen Raum ist noch weit davon entfernt, in die Sprachwissenschaft, geschweige denn in den Alltag der pädagogischen Sprachvermittlung des Deutsch- und Medienunterrichts oder auch des sonstigen Fachunterrichts (dort, wo es um sprachliche Anforderungen geht) integriert zu sein. Computergestützte Softwareprogramme werden zwar z. T. im Sprachunterricht erfolgreich eingesetzt, diese basieren jedoch auf Übungen, die mithilfe existierender Grammatiken und Nachschlagewerke konzipiert wurden und neueste korpusgestützte Spracherkenntnisse ignorieren. Dabei kann besonders der gezielte Einsatz von Korpora dazu führen, neue und vor allem selbstständig erarbeitete Erkenntnisse über den Wort-

(12) Abstand maximal 1 Satz, gemeinsame Kontexte *Anerkennung* – *Nichtanerkennung*: 31 Treffer, *Anerkennung* – *Ablehnung*: 217 Treffer, *Anerkennung* – *Missbilligung*: 7 Treffer, *Anerkennung* – *Skepsis*: 51 Treffer.

Weiterbildungsträger wollen Private Schulen aus dem Markt	<b>kicken</b>	
mehr den Boden berühren und nicht mit Armen oder Händen	<b>gekickt</b>	oder gefangen werden.
sich die Gelegenheit bietet, den Ball mit der Hacke zu	<b>kicken,</b>	nimmt er sie wahr.
Am häufigsten wurde mit Hüften und Gesäß	<b>gekickt,</b>	aber auch mit Händen, Kopf und Knien.
Außer mit den Händen darf der Ball mit allen Körperteilen	<b>gekickt</b>	und aufgefangen
Das	<b>kickt</b>	extrem.
die Grünen wohl kaum absichtlich selbst aus dem Gremium	<b>kicken</b>	wollten, bleibt der Verdacht, daß
nicht weiß, was Thomas gleich mit ihr machen wird, das	<b>„kickt“</b>	sie.
Wer meint, das	<b>kickt</b>	nicht, ist selber Schuld.
um einen Partyrausch mit Sonne, Sand und Meer. Das	<b>kickt</b>	extrem.
Dabei ist die nächste Party immer die	<b>kickendste,</b>	wildeste und energischste.

Beispiel 8:

verkürzter Ausschnitt der Konkordanzen zum Suchwort *kicken*

scheiden lassen, einer Frau, die selbst stolz bekennt, daß sie	<b>„dumm wie</b>	ein Ziegelstein“ ist und
Toilette steht: „Skins: Hart wie Kruppstahl, zäh wie Leder,	<b>dumm wie</b>	Stulle“.
und forderten eine saubere Heimat(t)scheibe.	<b>Dumm wie</b>	Brot, schrieb FDP- Mann
Zuletzt glaubte die arme Di, sie sei	<b>„dumm wie</b>	„ne Erbse“, erkrankte an
des Vorspiels war mir so bang: Da mußten Edelleute paradien,	<b>dumm wie</b>	Lipizzaner, da hampelt das
Moderator Steinbrecher schleimend, nichtfragend,	<b>dumm wie</b>	ein Pfund Mehl unterm Fön.
bereut Diana, einmal zu einem Kind gesagt zu haben, sie sei „so	<b>dumm wie</b>	ein Stück Holz“, weil das
hätte den Krieg gewonnen oder Princess Diana wäre nicht ganz so	<b>dumm wie</b>	Bohnenstroh – wie stünde es
Weil er allerdings zwar auch schon 13, aber	<b>dumm wie</b>	ein Storchküken war,
In einer Brandrede zäh wie Leder und	<b>dumm wie</b>	Windhund („Das Wurzelwerk
irgendwann hat der Franzose einen Knoten in den Beinen, steht	<b>dumm wie</b>	Stangenweißbrot vor dem
spätestens in Januar bei uns ausziehen werde, weil wir alle	<b>dumm wie</b>	Grießbrei seien.
Nehmen wir einen ganz alltäglichen Satz: „Der Bernt, der ist	<b>dumm wie</b>	Schifferscheiße“.
Gastkommentar	<b>Dumm wie</b>	Brötchen Warum der wahre
Ein Computer, der nur auf dem Tisch steht, ist zwangsläufig	<b>dumm wie</b>	Rosenkohl.
Oder wenn er über die Elsässer herzieht, die einfach nur	<b>dumm wie</b>	Stroh seien.
der alte Plutzer, von der Wissenschaft nicht länger behelligt,	<b>dumm wie</b>	ein Kürbis, wie man in
Antlitz gestreift, da bist du schon dumm, daß Gott erbarmt,	<b>dumm wie</b>	ein Esel, daß man dich
h alles sein, was er weiß, und wer nur das weiß, der ist so	<b>dumm wie</b>	die Nacht, und wäre ihm
mit der populären Radlerdroge Erythropoietin (Epo), sagt er.	<b>Dumm wie</b>	Knäcke Brot
Sessel in der Académie Française nachfolgte, Hugo sei doch so	<b>dumm wie</b>	der Himaaya
Dann heißt es „Schön wie ein Schaf“, nicht	<b>„Dumm wie</b>	ein Schaf“.
„Double- D’s“ (deutsche Erklärung dieser Abkürzung: sie sind	<b>dumm wie</b>	Dackel.

Beispiel 9:

verkürzter Ausschnitt der Konkordanzen zu *dumm wie*

schatz und seine Strukturen zu gewinnen sowie Lehrende und Lernende zur Reflexion über Sprache „so wie sie ist, statt so, wie sie im Lexikon steht“ anzuregen.

Korpora sind häufig kostenlos im Internet zugänglich und die meisten Recherche-werkzeuge bedürfen keiner langwierigen

Einarbeitung. Deshalb besteht, wie Sinclair (2004, 188) anmerkt, “[...] no need to wait for the new textbooks and reference books. [...] the results can be illuminating and very helpful.” Auch wenn ein Korpus nicht 100-prozentig verlässlich, umfassend genug oder gar in jeder Hinsicht repräsen-

tativ sein kann, so stellen sich trotzdem nur wenige die Frage, ob die für den Unterricht konzipierten Lehrbücher, Grammatiken und Wörterbücher diesen Kriterien entsprechen. Jede Informationsquelle sollte daher genau untersucht und ausgewertet werden. Auch bei einem Korpus sollte man sich über Inhalte und Zusammensetzung informieren, um extrahierte Ergebnisse nachvollziehbar zu machen. Das ist im Gegensatz zu vielen existierenden Unterrichtsmaterialien möglich (vgl. Sinclair 2004). Für Lehrende und Lernende gilt daher folgende Aufforderung:

„[...] for the present any teacher or student can readily enter the world of the corpus and make the language useful in learning“.

SINCLAIR 2004, 297 ■

## 5. Literatur

### 5.1 Forschungsliteratur

- Belica, Cyril (1995): Statistische Kollokationsanalyse und Clustering. Korpusanalysemodul. Mannheim: Institut für Deutsche Sprache.
- Haß-Zumkehr, Ulrike (2002): Das Wort in der Korpuslinguistik – Chancen und Probleme empirischer Lexikologie. In: Vilmos Ágel (Hg.): Das Wort: seine strukturelle und kulturelle Dimension. Festschrift für Oskar Reichmann zum 65. Geburtstag. Tübingen: Niemeyer, S. 45–70.
- Hunston, Susan (2002): Corpora in Applied Linguistics. Cambridge: Cambridge University Press.
- Scholze-Stubenrecht, Werner (2001): Das Internet und die korpusgestützte praktische Lexikographie. In: Jarmo Korhonen (Hg.): Von der mono- zur bilingualen Lexikographie für das Deutsche. (Finnische Beiträge zur Germanistik, Bd. 6). Frankfurt a. M.: Peter Lang, S. 43–64.
- Sinclair, John (Hg.) (1987): Looking Up: An Account of the COBUILD Project in Lexical Computing. London: HarperCollins.
- Sinclair, John (2004): Introduction: In: John Sinclair (Hg.): How to use Corpora in Language Teaching. Amsterdam/Philadelphia: John Benjamins, S. 1–10.
- Sinclair, John (2004): New evidence, new priorities, new attitudes. In: John Sinclair (Hg.): How to use Corpora in Language Teaching. Amsterdam/Philadelphia: John Benjamins, S. 272–299.
- Teubert, Wolfgang (1998): Korpus und Neologie. In: Wolfgang Teubert (Hg.): Neologie und Korpus. Tübingen: Niemeyer, S. 129–170.
- Teubert, Wolfgang (1999): Korpuslinguistik und Lexikographie. In: Deutsche Sprache 27/1999, S. 292–313.
- Tognini-Bonelli, Elena (2001): Corpus Linguistics at Work. Amsterdam/Philadelphia: John Benjamins.

### 5.2 Nachschlagewerke

DWDS: Digitales Wörterbuch der deutschen Sprache des 20. Jahrhunderts: <http://www.dwds.de/>.

DUDEN 8 (2004): Das Synonymwörterbuch. 3., völlig neu erarbeitete Auflage Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

DUDEN 11 (2002): Redewendungen. Wörterbuch der deutschen Idiomatik. 2., neu erarbeitet und aktualisierte Auflage, Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

DUDEN GWDS (2000): Das große Wörterbuch der deutschen Sprache. 10 Bände auf CD-Rom. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

elexiko – Das lexikalisch-lexikologische korpusbasierte Informationssystem des IDS: [www.elexiko.de](http://www.elexiko.de).

inclair, John (Hg.) (1987): COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY, London: Collins.

Stötzel, Georg; Eitz, Thorsten (Hg.) (2002): ZEITGESCHICHTLICHES WÖRTERBUCH DER DEUTSCHEN GEGENWARTSSPRACHE. Hildesheim: OLMS.

Stößel, Gerhard/Haß, Ulrike/Harras, Gisela (1989): BRISANTE WÖRTER VON AGITATION BIS ZEITGEIST: Ein Lexikon zum öffentlichen Sprachgebrauch. Berlin/New York: de Gruyter.

Quasthoff, Uwe (Hg.) (2004): DORNSEIFF – Der deutsche Wortschatz nach Sachgruppen. 8., völlig neu bearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage. Mit einer lexikographisch-historischen Einführung und einer ausgewählten Bibliographie zur Lexikographie und Onomasiologie von Herbert Ernst Wiegand. Berlin/New York: de Gruyter.

WORTSCHATZ-LEXIKON: <http://wortschatz.informatik.uni-leipzig.de/>.

### 5.3 Internetressourcen

COSMAS: Corpus Search, Management and Analysis System <http://www.ids-mannheim.de/cosmas2/>.

Projekt Methoden der Korpusanalyse und -erschließung <http://www.ids-mannheim.de/kl/projekte/korpora/>.

Statistische Kollokationsanalyse und Clustering: <http://corpora.ids-mannheim.de/cosmas>, zu Parametereinstellungen siehe auch <http://www.ids-mannheim.de/kl/misc/tutorial.html>.