

Der Einsatz verschiedener Korpusmethoden und -verfahren zur Qualitäts- und Konsistenzsicherung am Beispiel der Ermittlung und Dokumentation von Synonymen und Antonymen

- | | | | |
|-----|--|-----|--|
| 1. | Einleitung | 3. | Methoden konsistenter sinnrelationaler Vernetzungsstrukturen |
| 2. | Der Einsatz verschiedener korpuslinguistischer Verfahren | 4. | Schlussbemerkungen |
| 2.1 | Methodologische Vorüberlegungen | 5. | Literatur |
| 2.2 | Chancen und Probleme der korpusgesteuerten Ermittlung | 5.1 | Wörterbücher |
| 2.3 | Chancen und Probleme der korpusbasierten Ermittlung | 5.2 | Literatur |
| | | 5.3 | Elektronische Quellen |

Abstract

Dieser Beitrag beleuchtet die Vor- und Nachteile korpusgestützter lexikografischer Methoden zur Ermittlung und Dokumentation sinnrelationaler Ausdrücke eines Stichwortes. Konkrete Beispiele aus der Praxis des *lexiko*-Wörterbuchs dienen der Veranschaulichung von Chancen als auch von konkreten Problemen, die die eingesetzten Methoden mit sich bringen. Für die Gewinnung potentieller Synonyme und Antonyme nutzt *lexiko* zwei unterschiedliche Verfahren, die mit verschiedenen Prämissen an ein Korpus als Datengrundlage herantreten (cf. Tognini-Bonelli 2001). Das korpusgesteuerte/korpusgeleitete Verfahren der Kollokationsanalyse und die zugrunde liegende Ermittlung von Ausdrücken mit verwandten Kollokationsprofilen (related profiles) (cf. Belica 2011) dienen der empirischen und statistischen Absicherung von sprachlichen Phänomenen. Sie erweisen sich aber als lückenhaft in Bezug auf einige Kontexte, in denen semantisch-konzeptuelle Beziehungen der Ähnlichkeit oder des Gegensatzes realisiert, aber nicht mit Korpus-Tools erfasst werden. Mit der Anwendung der in *lexiko* komplementär genutzten korpusbasierten Vorgehensweise können diese Lücken teilweise gefüllt werden. Das Zusammenspiel beider Korpusansätze hat sich in der lexikografischen Praxis prinzipiell als vorteilhaft erwiesen, bringt jedoch auch Erkenntnisse zum Vorschein, die bisher weder linguistisch erfasst noch lexikografisch dokumentiert wurden und löst nicht, wie teilweise angenommen, das Problem inhaltlicher Inkonsistenzen (cf. Paradis/Willners 2007). Diese Aspekte werden anhand von konkreten Korpusbeispielen und Wörterbucheinträgen illustriert.

Als Online-Wörterbuch profitiert *lexiko* von seinen schnellen Navigationsmöglichkeiten über Verlinkungen. Diese werden auch für sinnrelationale Partnerwörter wie Synonyme und Antonyme angelegt, um diverse Vernetzungsstrukturen nachvollziehbar zu machen. Die Arbeit mit einem Korpus kann bis zu einem gewissen Grad die Konsistenz der bidirektionalen Vernetzungen gewährleisten, sie aber nicht vollständig absichern. In diesem Beitrag wird auch die Frage beantwortet, inwieweit die erwähnten Korpusmethoden dazu beitragen, das gegenseitige Dokumentieren zwischen Synonym- oder Antonympaaren sicherzustellen. Anhand des für diese Zwecke entwickelten Tools *vernetziko*, einem Vernetzungsmanager, wird gezeigt, warum die Unterstützung zusätzlicher Software für eine konsistente Verlinkung zwischen paradigmatisch miteinander verbundenen Stichwörtern unerlässlich ist (Storjohann/Meyer 2012).

1. Einleitung

Sinn- und sachverwandte Ausdrücke wie Synonyme und Antonyme¹ können mit umfangreichen Korpora und den entsprechenden Analysewerkzeugen sehr systematisch und zielgerichtet extrahiert, ggf. linguistisch untersucht und daraufhin lexikografisch beschrieben werden. Dennoch beschreibt kaum ein korpusgestützt erarbeitetes Wörterbuch, das derartige lexikalische Beziehungen zwischen zwei Ausdrücken dokumentiert, seine zugrunde liegenden linguistischen Kriterien, die gewählten korpusmethodologischen Verfahren sowie die dafür eingesetzten Werkzeuge zur Ermittlung von Synonymen bzw. Antonymen. Nur über eine genauere Analyse der Wörterbucheinträge selbst lässt sich feststellen, ob und wenn ja, welche Prinzipien für die Aufnahme von sinnrelationalen Ausdrücken vorliegen. Die Erwartungen an korpusfundierte Angaben schließen neben einer objektiveren Beschreibung des tatsächlichen Sprachgebrauchs vor allem auch ein konsistentes Dokumentieren von Beziehungen für die beteiligten Partnerwörter ein (cf. Paradis/Willners 2007). Die lexikografische Arbeit mit Korpora lässt nicht nur vermuten, dass relevantes Material empirisch, systematisch und auch schneller erfasst werden kann, sondern auch, dass die zugrunde liegende Bidirektionalität einer synonymen oder antonymen Beziehung sich in der konsistenten gegenseitig rückverweisenden Dokumentation zwischen den einzelnen Stichwörtern niederschlagen sollte, vorausgesetzt, sie gehören jeweils zum Stichwortbestand des Nachschlagewerkes.

Ziel einer Untersuchung von Paradis/Willners (2006, 2007) war es, im ersten korpusgestützten Nachschlagewerk des Englischen, dem *Collins COBUILD Advanced Learner's English Dictionary* (2003), herauszufinden, ob das Vorhandensein oder das Fehlen eines Antonyms bestimmten linguistischen Kriterien unterliegt und wie konsequent Rückverweise in den entsprechenden Einträgen angelegt wurden. Die Annahmen der Autorinnen waren, dass die Arbeit mit Korpusdaten solche Auswahlprinzipien nachvollziehbar macht, weil die Analyse des Sprachmaterials stärker an statistische Auswertungen und Parameter wie die Signifikanz gekoppelt ist. Das Fazit dieser systematischen Studie ergab, dass nur 37% aller Antonyme konsequent in beide Richtungen verzeichnet sind und dass die Prinzipien zur Aufnahme von semantischen Gegensatzpaaren nicht transparent sind.

We fail to see a clear pattern in the choice of antonymy and the symmetry of presentation [...].
(Paradis/Willners 2006, 216)

Die Mängel wurden besonders bei solchen Paaren deutlich, die den so genannten kanonischen, also äußerst prototypischen Paaren (cf. Jones et al. 2012) wie beispielsweise *billig* → *teuer* (*cheap* → *expensive*), *schwer/schwierig* → *leicht* (*difficult* → *easy*), *alt* → *neu* (*old* → *new*), *Hass* → *Liebe* (*hate* → *love*), *traurig* → *glücklich* (*sad* → *happy*), *ungerade* → *gerade* (*odd* → *even*) zugeschrieben werden, bei denen die Beziehung jeweils nur in einer Richtung angegeben wurde.

¹ Für Synonymie wird nachfolgend ein allgemeinsprachliches Verständnis der Bedeutungsäquivalenz zugrunde gelegt. Unter Antonymie werden nicht nur semantisch oder syntaktisch graduierbare Adjektive (cf. Cruse 1986) verstanden, sondern sämtliche Arten des konträren oder kontradiktorischen Beziehungsgegensatzes subsumiert.

In our opinion as non-native speakers of English, all these pairs deserve to be reversed, e.g. *big* should be listed as an antonym of *small* and *cheap* of *expensive*. Again, it is not clear to us why these pairings should be helpful for the learner in the above directions only. (Paradis/Willners 2006, 217)

Paradis/Willners (2006, 2007) interpretieren ihren Befund als methodologisches Defizit hinsichtlich der genutzten Verfahren und sie vermuten, dass diese Einträge primär durch korpusbasierte Arbeit zustande gekommen sind. Im tatsächlichen Sprachgebrauch tauchen Antonyme, insbesondere antonyme Adjektive, regelhaft in unmittelbarer kontextueller Umgebung auf (cf. Justeson/Katz 1991, Jones 2002). Lexikografisch sollte man sich daher stärker einer Herangehensweise verpflichten, die die Ermittlung von Kookkurrenzen bzw. statistisch signifikanten sprachlichen Konstruktionen mit einem bestimmten Suchwort zum lexikografischen Prinzip erklärt. Paradis/Willners (2006) schlugen vor, aufgrund des kontextuellen Verhaltens von Oppositionspaaren bei der Ermittlung solcher Beziehungen stärker korpusgesteuert/korpusgeleitet statt korpusbasiert vorzugehen. Sie nehmen an, dass somit stringente Kriterien und eine systematische Vorgehensweise bei der Analyse binärer Beziehungen herangezogen werden können.

Nachfolgend soll anhand des deutschen Nachschlagewerkes *lexiko*, einem Online-Wörterbuch zur deutschen Gegenwartssprache, gezeigt werden, welche Ergebnisse zu sinnverwandten Ausdrücken aus Korpusdaten erzielt werden, wenn sowohl korpusbasierte als auch korpusgesteuerte Methoden zum Einsatz kommen. Der Beitrag beleuchtet die Vor- und Nachteile beider lexikografischer Ansätze anhand konkreter Beispiele aus der Wörterbuchpraxis und weist auf Chancen und konkrete Probleme hin, die die eingesetzten Herangehensweisen mit sich bringen. Besonderes Augenmerk liegt auf der Beantwortung der Fragestellung, ob der Einsatz korpusgesteuerter Verfahren die von Paradis/Willners (2006) aufgeführten Probleme der Inkonsistenz bidirektionaler Beziehungen löst und die Mängel bezüglich der symmetrischen Aufnahme gegenseitiger Partnerwörter beseitigt. Es soll gezeigt werden, in wie weit damit ein Instrument zur Qualitätssicherung und zur konsistenten Dokumentation von semantisch konträren bzw. äquivalenten Wortpaaren vorliegt.

2. Der Einsatz verschiedener korpuslinguistischer Verfahren

2.1 Methodologische Vorüberlegungen

Der korpusgesteuerte und der korpusbasierte Ansatz werden im engen korpuslinguistischen Sinn als zwei unterschiedliche, und nicht selten auch gegensätzliche, Herangehensweisen an Korpusdaten verstanden (cf. Tognini-Bonelli 2001). Die Gegensätzlichkeit beruht auf der unterschiedlichen Zielsetzung der jeweiligen Ansätze. Bei der korpusbasierten Methode werden Korpusdaten als Beweismaterial für bestehende Annahmen genutzt. Das Korpus fungiert also als Prüf-, Quantifizierungs- und Validierungsinstrumentarium. Bestehende linguistische Kategorien oder Theorien werden nicht hinterfragt und wie Tognini-Bonelli (2001, 81) hervorhebt „the commitment to the data as a whole is not ultimately very strict or systematic“. Bei der korpusgesteuerten Methode tritt man dagegen ohne Vorannahmen an Korpusdaten heran und wertet diese empirisch mithilfe statistischer Verfahren aus (z.B. mit

einer Kookkurrenzanalyse), um regelhafte Strukturen zu ermitteln, auf deren Grundlage man zu Beschreibungen oder linguistischen Generalisierungen gelangt.

Während oftmals zwischen beiden Verfahren ein grundlegender Widerspruch gesehen wird (cf. Sinclair 2002), wurde in der lexikografischen Praxis des Projektes *ellexiko* die Erfahrung gemacht, dass sich die beiden Methoden für verschiedene Untersuchungszwecke unterschiedlich eignen. Die explorative automatische Kontextanalyse sichert für frequente Stichwörter objektivere Auswertungen zu kontextuellen Verwendungen und unmittelbaren Partnerwörtern. Für die Ermittlung sinnrelationaler Partner eines zu beschreibenden Ausdrucks hat sich aber gezeigt, dass beim Einsatz beider Korpusverfahren die bestmöglichen Ergebnisse erzielt werden. Für die Gewinnung von Synonymen und auch für die Beschreibung sinnverwandter Ausdrücke verbaler Stichwörter ist eine strikte Trennung der durchaus komplementären Ansätze unvorteilhaft.

2.2 Chancen und Probleme der korpusgesteuerten Ermittlung

Bei der korpusgesteuerten Methode geht man davon aus, dass eine bestehende Synonymie- oder auch Antonymierelation, so sie regelhaft im Sprachgebrauch vorkommt, als sprachliches Konstrukt emergent aus dem Korpus hervortritt und dass entsprechende Analysewerkzeuge diese auf der Basis statistischer Verfahren aufdecken, ohne dass der Lexikograf/die Lexikografin gezielt nach einer bestimmten Beziehung zwischen zwei Ausdrücken sucht. Auch *ellexiko* nutzt die Ermittlung von Kollokationen/Kookkurrenzen² und anderen statistisch signifikanten Konstruktionen, um bei der semantischen Analyse eines Ausdrucks u. a. auf bedeutungsähnliche bzw. -gegensätzliche Ausdrücke zu stoßen. Dies geschieht mit Hilfe des Korpusanalysemoduls „Statistische Kollokationsanalyse und Clustering“ (Belica 1995). Die Ergebnisse der korpusgesteuerten Untersuchung zeigen zum Teil große Diskrepanzen zu Synonym- bzw. Antonymangaben in Wörterbüchern, die nicht auf Korpusbasis erarbeitet wurden (cf. Storzjohann 2006).

Besonders zielführend ist das korpusgesteuerte Verfahren bei der Suche nach adjektivischen Gegensatzwörtern. Es besteht bei Gegensatzpaaren im Allgemeinen die Tendenz kontextuell eng in bestimmten sprachlichen Konstruktionen miteinander aufzutauchen (Justeson/Katz 1991). Korpusorientierte Studien zeigen, dass die sprachlichen Konstruktionen von Antonymierelationen bestimmte semantische, text- bzw. diskursbezogene Funktionen ausüben (cf. Jones 2002). Die mit Abstand häufigste sprachliche Realisierung stellt dabei die koordinierende Antonymie dar, beispielsweise *kleine und große Kinder, schöne oder hässliche Gebäude, sowohl öffentliche als auch private Investoren, entweder feste oder freie Mitarbeiter* etc.³ Auch Konstruktionen mit negierender Funktion wie etwa, *nicht mit legalen, sondern illegalen Mitteln* oder *nicht dauerhaft, sondern temporär einrichten* und mit vergleichender Funktion wie *eher dick als dünn* oder *mehr tot als lebendig* zeigen, dass sprachlich die beteiligten Ausdrücke in enger semantischer Umgebung miteinander auftauchen. Das gleiche gilt für Muster, die sich in semantischer Hinsicht auf einen Unterschied

² Die Termini „Kollokation“ und „Kookkurrenz“ werden in diesem Beitrag synonymisch verwendet und lehnen sich an das korpuslinguistische Verständnis im sinclairischen Sinne an (cf. Sinclair 1991).

³ Cf. auch Jones' (2002) Untersuchung zu ‚coordinated antonymy‘ und ‚ancillary antonymy‘.

beziehen, z.B. zwischen *arm und reich*, *ältere Frauen im Gegensatz zu jüngeren Frauen* oder Syntagmen, die einen Übergang markieren, wie *der Optimismus schlägt bald in Pessimismus um*. Aufgrund des miteinander Kookkurrierens lassen sich derartige Strukturen gezielt aus dem Korpus extrahieren, ohne Vorannahmen darüber zu haben, welche potentiellen Gegensatzwörter kontextuell mit einem Ausdruck *X* sprachlich realisiert werden. Darüber hinaus gewährleistet ein enges kotextuelles Miteinandervorkommen und die statische Auswertung die Aufdeckung des Partnerwortes *Y* für das Suchwort *X* und umgekehrt.

Besonders effektiv ist eine Suche bei monosemen bis gering polysemen Ausdrücken wie etwa *privat*. Eine korpusgeleitete Suche zweier solcher Muster ergab im *lexiko*-Korpus zahlreiche, z.T. auch unerwartete Ergebnisse (siehe Abbildungen 1 und 2). An das Korpusrecherche und -analysewerkzeug COSMAS II wurden dafür z.B. folgende Suchanfragen gestellt⁴:

1. &privat / Abstand (2w Max) sondern
2. &privat / Abstand (2w Max) ‘UND‘.

Die beiden Suchanfragen (Abbildung 1 und 2) zeigen Überlappungen, aber auch Unterschiede in ihren Ergebnissen. Die Wahl der Suchanfragen hat maßgeblichen Einfluss auf ermittelte Ergebnislisten und damit auch auf die mögliche lexikografische Auswahl. Es gilt also, sich der Möglichkeiten typischer Antonymkonstruktionen bewusst zu werden und diese gezielt für die datengeleitete Ermittlung von Gegensatzpaaren einzusetzen.

LLR	Kookkurrenzen	syntagmatische Muster
399	beruflich	61% nicht nur beruflich [...] sondern auch privat
277	öffentliche	62% nicht nur öffentliche [...] sondern auch private
115	staatliche	63% nicht keine staatliche [...] sondern auchleine private
86	dienstlich	90% nicht nur dienstlich [...] sondern auch privat
45	gewerblichen	83% allerdings keine gewerblichen [...] sondern nur private
33	geschäftliche	66% geschäftliche Termine haben sondern
22	gesellschaftliches	100% kein privates sondern ein gesellschaftliches
20	öffentlich-rechtliche	33% sondern ... öffentlich-rechtliche
17	landeseigene	50% private ... landeseigene
16	politisch	100% privat ... politisch
12	offizieller	75% offizieller ... sondern
11	berufliches	100% sondern ... berufliches
9	städtischer	50% privater ... städtischer

Abbildung 1: Gegensatzwörter zu *privat* in Konstruktion nicht nur X, sondern auch Y

⁴ Die Korpusrecherche beschränkt sich auf die Textsuche, in denen das Lemma *privat* mit all seinen Wortformen zusammen mit den Ausdrücken *sondern* und *und* innerhalb eines Wortabstands von maximal zwei Wörtern auftritt. Darüber hinaus gibt es zahlreiche weitere typische Muster, die Antonympaare syntagmatisch konstruieren (cf. Jones 2002). Für generelle Beispiel-Suchanfragen in COSMAS II siehe auch unter <http://www.ids-mannheim.de/cosmas2/web-app/hilfe/seite/suchanfrage/beispiele.html>.

LLR	Kookkurrenzen	syntagmatische Muster
11724	öffentlichen	58% öffentlichen [...] und privaten ...
5475	beruflichen	58% aus beruflichen [...] und privaten Gründen
2740	öffentlich-rechtlichen	59% öffentlich-rechtlichen [...] und privaten ...
1963	gewerblichen	67% privaten [und] gewerblichen
1814	staatlichen	63% staatlichen [...] und privaten ...
1510	betriebliche	48% die betriebliche [...] und private ...
1459	institutionelle Anleger	53% für private und institutionelle Anleger
1063	gesetzlichen	65% zwischen der gesetzlichen [...] und privaten Kassen/ Krankenkassen
951	geschäftliche	61% für private und geschäftliche
903	dienstliche	49% dienstliche [...] und private ...
513	kommunale	50% kommunale [...] und private ...
300	genossenschaftlichen	80% der die privaten [und] genossenschaftlichen Banken ...
316	Sicherheitsdiensten	71% von Polizei und privaten Sicherheitsdiensten
238	institutioneller	61% privater und institutioneller Kunden ...
167	freigemeinnützigen	50% freigemeinnützigen [...] und privaten ...

Abbildung 2: Gegensatzwörter zu *privat* in Konstruktion X und Y

Die großen Vorteile dieser Herangehensweise liegen in folgenden Aspekten. Das Datenmaterial wird explorativ ohne Vorannahmen untersucht, und es werden mithilfe bestimmter Suchanfragen von typischen Gegensatzkonstruktionen investigativ und systematisch alle entsprechenden sprachlichen Realisierungen herausgefiltert. Die linguistischen Regularitäten werden von Analysewerkzeugen aufgedeckt, die Ergebnisse werden nach Signifikanz geordnet und so einer möglichen nachfolgenden lexikografischen Auswertung vorklassifiziert zur Verfügung gestellt. Die Erfahrung, die im Umgang mit solchen Abfragen in *lexiko* gesammelt wurden, haben gezeigt, dass auch unerwartete Funde zutage treten, die außerhalb der eigenen linguistischen Kompetenz und Erwartbarkeit liegen und daher mit anderen Methoden möglicherweise nicht erfasst worden wären.

Deutlich schlechtere Ergebnisse erzielt die Suche nach Kollokatoren zur Extrahierung potentieller Kandidaten bei hochpolysemen Ausdrücken wie *frei*, bei denen einige wenige Lesarten stark im Korpus dominieren zu Ungunsten zahlreicher wenig auffälliger Lesarten. Ebenso schwierig ist diese Herangehensweise auch für Verben, die generell dazu tendieren, valenzartige Subjekt- oder Objektstrukturen in unmittelbarer Umgebung einzugehen, die mit Substantiven oder auch mit Adjektiven kollokieren und weniger systemhafte Konstruktionen mit Verben des gleichen Paradigmas aufweisen. Und deutlich eingeschränkter ist auch die Suche nach Synonymen, die teilweise in enger semantischer Nachbarschaft, aber tendenziell in größerer kontextueller Umgebung miteinander vorkommen.⁵ So zeigt etwa eine Kollokationsanalyse des Verbs *akzeptieren* zahlreiche erwartbare Substantive wie *Entschuldigung*, *Friedensplan*, *Verhandlungspartner* oder *Spielregeln* und auch adverbial

⁵ Zu einigen typischen Synonymkonstruktionen in unmittelbarem Kontext siehe Storjohann (2010).

gebrauchte Adjektive wie *klaglos*, *stillschweigend*, *widerspruchslos*, *widerwillig*, *vorbehaltlos* etc. aber kaum Verben, die eine Bedeutungsäquivalenz oder einen semantischen Gegensatz bezeichnen (siehe Abbildung 3).

Ergebnisse von 'LEM(akzeptieren)' in Korpus: elexiko - virt. Korpus des Projektes elexiko					
Ansicht Andocken Extras Hilfe					
#	LLR	kumul.	Häufig	Kookkurrenzen	syntagmatische Muster
9	744	3966	196	klaglos	45% klaglos [...] akzeptiert
10	738	4192	226	stillschweigend	52% stillschweigend [...] akzeptiert
11	509	4472	280	allseits	38% allseits [...] akzeptierten
12	481	4897	425	Entschuldigung	38% die Entschuldigung [...] akzeptiert
13	432	5039	142	Murren	50% ohne Murren [...] akzeptiert
14	417	5147	108	widerspruchslos	53% widerspruchslos akzeptiert
15	380	5302	155	widerwillig	44% nur widerwillig [...] akzeptiert
16	318	5388	86	Schlichterspruch	53% den Schlichterspruch [zu] akzeptieren
17	299	5504	116	vorbehaltlos	48% Rechtsordnung vorbehaltlos [zu] akzeptieren
18	296	5584	80	Dr	28% nicht zu akzeptieren [so ...] Dr
19	293	5670	86	Anderssein	66% das Anderssein [...] zu akzeptieren und
20	261	5800	130	Friedensplan	33% den Friedensplan [...] zu akzeptieren
21	240	6001	201	respektieren	60% zu akzeptieren [und zu] respektieren
22	226	6128	127	Verhandlungspartner	50% als Verhandlungspartner [...] akzeptiert ... werden
23	215	6155	27	Toulas	100% und Ian Toulas griechische Großfamilie akzeptieren
24	214	6156	1	akzept humane	100% akzeptierende ... humane ... akzept
			6159	3	akzept - für akzeptierende
25	197	6214	55	Strafbefehle	56% die Strafbefehle [nicht] akzeptiert
26	195	6286	72	anstandslos	54% anstandslos [...] akzeptiert
27	187	6474	188	Spielregeln	51% die Spielregeln [...] zu akzeptieren
28	170	6533	59	Mehrheitsentscheidungen	67% Mehrheitsentscheidungen [zu] akzeptieren
29	155	6582	49	bzw	28% bzw [...] nicht] akzeptiert werden
30	141	6625	43	Mehrheitsentscheidung	41% eine die Mehrheitsentscheidung [zu] akzeptieren
31	140	6655	30	EC-Karten	66% Euro EC-Karten [werden] akzeptiert
32	135	6786	131	respektiert	57% akzeptiert [und] respektiert werden
33	131	6834	48	Schiedsspruch	47% den Schiedsspruch [...] zu akzeptieren
34	128	6854	20	Mehrheitsentscheide	60% die ... Mehrheitsentscheide [zu nicht] akzeptieren
35	127	8200	1346	Urteil	33% das Urteil [...] nicht zu] akzeptieren
36	127	8246	46	kritiklos	56% kritiklos [...] akzeptiert werden

Abbildung 3: Auszug aus der Kollokationsanalyse des Verbs *akzeptieren*

In solchen Fällen nutzt *elexiko* ein korpusgesteuertes Verfahren, bei dem Kollokationsprofile zwischen Ausdrücken miteinander kontrastiert werden, um semantisch ähnliche Ausdrücke zu einem Suchwort zu ermitteln (cf. Belica 2011). In *elexiko* wird dafür die Kookkurrenzdatenbank CCDB⁶ (Belica 2001ff.) mit der entsprechenden Funktionalität „related collocation profiles“ eingesetzt. Auf diese Weise werden u.a. für verbale Stichwörter Verben ermittelt, die über ein ähnliches Kollokationsprofil verfügen und potentielle Kandidaten für eine Synonymie- oder Gegensatzrelation sind. Diese sind empirisch und statistisch abgesichert, erhalten aber zunächst in *elexiko* nur den Status eines Kandidaten, den es zu prüfen gilt. Für das Suchwort *akzeptieren* sind es u.a.: *respektieren*, *zustimmen*, *anerkennen*, *hinnehmen*, *abfinden*, *ablehnen*, *beugen*, *abrücken*, *annehmen*, *ignorieren*, *widersetzen*, *tolerieren*, *zulassen*, *opponieren*, *aufzwingen*, *unterstützen*, *zurückweisen* (siehe Abbildung 4).

⁶ Kookkurrenzdatenbank CCDB siehe <http://corpora.ids-mannheim.de/ccdb/>.

The screenshot shows the CCDB interface for searching synonyms. The search term is 'akzeptieren'. The results are displayed in a table with columns for frequency, percentage, and related terms. The interface also includes a search bar, filters, and a list of related terms with their respective frequencies and percentages.

Frequency	Percentage	Related Term
4	100%	nicht [...] länger akzeptieren
5	80%	wollte ... partout nicht akzeptieren
555	54%	wollte [...] nicht [...] akzeptieren
202	68%	nicht [...] länger [...] akzeptieren
26	88%	partout [...] nicht akzeptieren
21468	45%	nicht [...] akzeptieren
5	60%	Kreditkarten Euroschecks werden akzeptieren
119	62%	Kreditkarten [...] werden akzeptieren
283	62%	könne [...] nicht akzeptieren
15	60%	Euroschecks [...] werden akzeptieren
7446	64%	akzeptiert [...] werden akzeptieren
137	49%	als [...] Gesprächspartner akzeptieren
80	53%	als [...] Verhandlungspartner akzeptieren
35	80%	als vollwertige ... akzeptieren
9657	40%	als akzeptieren
148	65%	von allen [...] akzeptieren
13	46%	wird [...] akzeptieren
21	52%	wird [...] nicht über akzeptieren
5498	46%	wird akzeptieren
9	77%	nach allgemein akzeptieren
11	72%	ist inzwischen allgemein akzeptieren
8	62%	auf allgemein [...] akzeptieren

Abbildung 4: Ausschnitt aus der CCDB, verbale Ausdrücke mit ähnlichem Kollokationsprofil wie *akzeptieren*

Die angebotene Liste ist nicht das Ergebnis lexikografischer Interpretation, sondern ausschließlich das Ergebnis softwaregesteuerter Analysen. Daher ist ihr endgültiger Status zunächst ungeklärt. Je nach Ziel des Nachschlagewerkes wird unterschiedlich mit solchen Daten umgegangen. In *lexiko* werden die Ergebnisse nicht einfach automatisiert übernommen. Hier wird der Übergang zum korpusbasierten Vorgehen unerlässlich, um verlässliche Aussagen darüber treffen zu können, ob es sich um relevante Synonyme/Antonyme handelt. Nur über anschließende, gezielte datenbasierte Prüfanfragen kann u.a. geklärt werden, ob die automatisch vorgeschlagenen Partnerwörter hinreichend oft kontextuell belegt sind oder periphere Erscheinungen sind (siehe dazu Abschnitt 2.3).

Dagegen werden z.B. im Wortschatzportal der Universität Leipzig⁷ rein statistisch ermittelte, mögliche Relationspartner in alphabetisch geordneten Listen angeboten. In Abbildung 5 sind die Relationen des Ausdrucks *frei*, die automatisiert gewonnen werden, illustriert.

Die hier angegebenen Korpusdaten enthüllen eine beeindruckende Vielfalt an möglichen Synonymen. Eine Analyse bezüglich der Symmetrie der angegebenen Synonyme zeigt ebenso, dass unter jedem aufgelisteten Synonym von *achtbar* bis *zwanglos* auch das Partnerwort *frei* als Rückverweis vorzufinden ist. Aber über eine quantitative korpusgesteuerte Analyse gehen diese Ergebnisse nicht hinaus. Die Synonyme erscheinen unkommentiert und sind kontextlos, und ohne semantische Zuordnung, wie sie ein stark polysemes Adjektiv wie *frei*

⁷ Wortschatzportal Universität Leipzig siehe: <http://wortschatz.uni-leipzig.de/>.

Relationen zu anderen Wörtern:

Synonyme: achtbar, allein, alleinstehend, astrein, autark, autonom, befreit, bieder, brav, disponibel, durchgängig, ehrbar, ehrenhaft, emanzipiert, entlassen, erlöst, formlos, fortschrittlich, frankiert, freigemacht, gebührenfrei, gelöst, gratis, großzügig, hemdsärmelig, improvisiert, kostenlos, lässig, ledig, leer, leger, modern, nachlässig, natürlich, neuartig, offen, offen, progressiv, rein, salopp, selbständig, selbstverantwortlich, sinngemäß, souverän, supermodern, umsonst, unabhängig, unbebaut, unbefangen, unbehindert, unbelastet, unbeschränkt, unbeschwert, unbesetzt, uneingeschränkt, unentgeltlich, ungebunden, ungeniert, ungenutzt, ungezwungen, unkontrolliert, unkonventionell, unverdächtig, unverheiratet, unvorbereitet, unzeremoniell, vakant, verfügbar, zwanglos

vergleiche: emanzipiert, frank, gratis, vakant

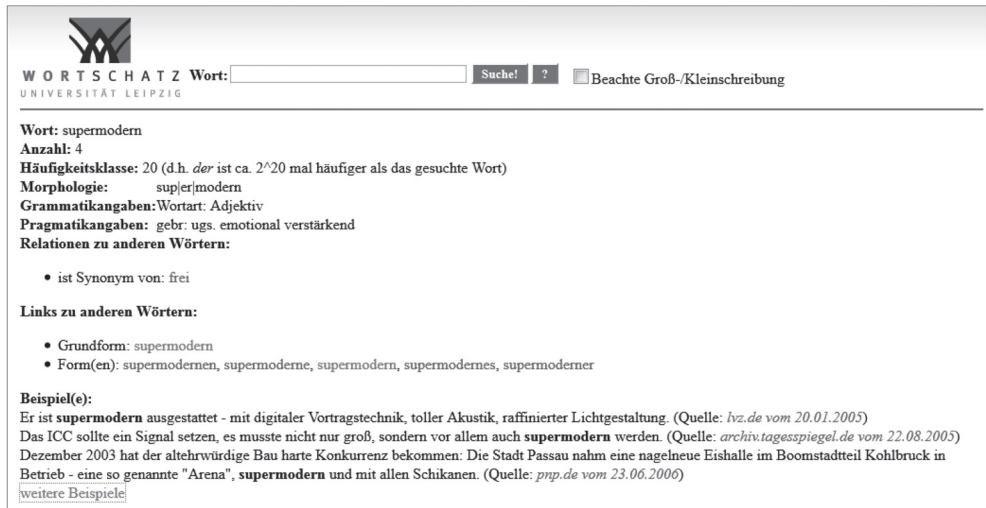
ist Synonym von: abgespannt, ausgestoßen, ausgezogen, auswendig, autark, autonom, bloß, ehelos, eigenständig, emanzipiert, entblößt, entfesselt, enthemmt, enthüllt, entkleidet, entlastet, entlaubt, entspannt, familiär, gebührenfrei, geflohen, geschminkt, gratis, hemdsärmelig, hemmungslos, herrenlos, hüllenlos, improvisiert, informell, kahl, kostenfrei, kostenlos, kostenlos, lässig, ledig, leer, leger, licht, locker, los, lose, nackt, natürlich, nonchalant, offen, offen, offen, offenbar, offenherzig, salopp, selbständig, single, Stegreif, unabhängig, unbeaufsichtigt, unbefangen, unbehindert, unbekleidet, unbelastet, unbeobachtet, unbeschränkt, unbesetzt, unbestraft, unentgeltlich, ungebunden, ungehemmt, ungeniert, ungezwungen, unkontrolliert, unumwunden, unverheiratet, unverkrampft, unvorbereitet, unzeremoniell, vakant, verfügbar, vertraulich, vertraut, zügellos, zwanglos

wird referenziert von: gratis, ledig, offen, schrankenlos, unabhängig, ungebunden, vakant

Abbildung 5: Auszug aus dem Eintrag *frei* des Wortschatzportals

erfordern würde, ist eine derartige Auflistung für jeden Nichtmuttersprachler „positively dangerous“ (Partington 1998, 47). Bei genauerer Betrachtung der dazugehörigen Stichworteinträge werden Details deutlich, die die Verlässlichkeit der angebotenen Informationen in Frage stellen. Ein Beispiel: Folgt man dem Eintrag zu *supermodern* ist dort das Synonym *frei* als einziges Synonym korrekt symmetrisch verzeichnet. Weiter im Eintrag gibt es Korpusbelege, die die Bedeutung des monosemen Ausdrucks *supermodern*, mit dem *frei* ja bedeutungsähnlich sein soll, veranschaulichen. In keinem der Belege ist aber *frei* als Synonym denkbar (cf. Abbildung 6).

Ausschließlich automatisiert extrahierte Daten maschineller Prozesse auf der Basis korpusgesteuerter Verfahren bergen das Risiko, fehlerhaft oder unzuverlässig zu sein. Dies steht in klarem Widerspruch zu den Ansprüchen der NutzerInnen an ein Nachschlagewerk. Laut einer Umfrage sehen NutzerInnen in der Zuverlässigkeit der Wörterbuchangaben eines der wichtigsten Kriterien (cf. Müller-Spitzer u.a. 2011). Eine korpusgesteuerte Ermittlung von sinnverwandten Ausdrücken nutzt auch *lexiko*, aber die Ergebnisse unterliegen der lexikografischen Analyse und Interpretation, einer Zuordnung zu entsprechenden Lesarten sowie einer kontextuellen Überprüfung eines jeden einzelnen Partnerwortes und der Bereitstellung eines entsprechenden Beleges, der jede dokumentierte Beziehung illustriert (detaillierter siehe Storjohann 2005). Die redaktionelle Arbeit in *lexiko* hat in der Vergangenheit gezeigt, dass, in Abhängigkeit des zu beschreibenden Stichwortes, die korpusgesteuerte Methode allein nicht der Qualitätssicherung genügen kann. Sie ist kein Garant dafür, alle relevanten Synonyme bzw. Antonyme im Korpus zu erfassen, um diese lexikografisch zu dokumentieren. Darüber hinaus stellt ein korpusgeleitetes Prozedere auch keine Gewährleistung dafür



W **O** **R** **T** **S** **C** **H** **A** **Z** **W** **o** **r** **t**: **Suche!** Beachte Groß-/Kleinschreibung
 UNIVERSITÄT LEIPZIG

Wort: supermodern
Anzahl: 4
Häufigkeitsklasse: 20 (d.h. *der* ist ca. 2²⁰ mal häufiger als das gesuchte Wort)
Morphologie: sup|er|modern
Grammatikangaben: Wortart: Adjektiv
Pragmatikangaben: gebr: ugs. emotional verstärkend
Relationen zu anderen Wörtern:

- ist Synonym von: frei

Links zu anderen Wörtern:

- Grundform: supermodern
- Form(en): supermodernen, supermoderne, supermodern, supermodernes, supermoderner

Beispiel(e):
 Er ist **supermodern** ausgestattet - mit digitaler Vortragstechnik, toller Akustik, raffinierter Lichtgestaltung. (Quelle: *lvz.de* vom 20.01.2005)
 Das ICC sollte ein Signal setzen, es musste nicht nur groß, sondern vor allem auch **supermodern** werden. (Quelle: *archiv.tagesspiegel.de* vom 22.08.2005)
 Dezember 2003 hat der altherwürdige Bau harte Konkurrenz bekommen: Die Stadt Passau nahm eine nagelneue Eishalle im Boomstadteil Kohlbruck in Betrieb - eine so genannte "Arena", **supermodern** und mit allen Schikanen. (Quelle: *pnj.de* vom 23.06.2006)
[weitere Beispiele](#)

Abbildung 6: Eintrag *supermodern* mit Synonymangabe und Belegen

dar, Ausdrücke, die eine Antonymie- bzw. Synonymiebeziehung regelhaft eingehen, auch konsequent symmetrisch in einem Nachschlagewerk zu verzeichnen. Können diese beiden Mängel denn durch den Einsatz eines korpusbasierten Verfahrens im Sinne einer komplementären Methode behoben werden?


2.3 Chancen und Probleme der korpusbasierten Ermittlung


In einigen Fällen ist der ausschließliche Einsatz der korpusgeleiteten Methode nicht hinreichend nutzbar, um das semantische Beziehungsgeflecht, das ein Ausdruck zu seinen Bedeutungsäquivalenten und Gegensätzen eingeht, eruieren zu können. Bei der korpusbasierten Methode nutzen Lexikografinnen und Lexikografen das Korpus als quantitative Datenquelle und Prüfinstrumentarium, um lexikografische Annahmen zu einer bestehenden Sinnverwandtschaft zu bestätigen oder zu verwerfen, um nach Belegen für bestehende Hypothesen zu suchen oder um ein Phänomen quantitativ zu evaluieren. In solchen Fällen dienen introspektive Kompetenz und Angaben aus Sekundärquellen zunächst der Hypothesenbildung, die anhand der Korpusdaten gesucht, quantifiziert, interpretiert und verifiziert werden. In *ellexiko* werden die Ergebnisse der datengeleiteten Kookkurrenzanalyse, die Resultate der Ermittlung von Ausdrücken mit ähnlichem Kollokationsprofil (siehe Abschnitt 2.2, Abbildung 4) sowie die Angaben weiterer Sekundärquellen gezielt kontextuell im Korpus überprüft. Das heißt, dass *ellexiko* das korpusbasierte Verfahren als zusätzliche, nachgestellte und komplementäre Methode der Gewinnung von sinnverwandten Ausdrücken nutzt.


Korpusbasierte Untersuchungen stellen in manchen Fällen nur „Bonusmaterial“ (cf. Tognini-Bonelli 2001, 66) dar, in anderen Fällen sind sie die einzige Möglichkeit, umfangreiche und essenzielle Angaben bezüglich sinnverwandter Ausdrücke machen zu können. Dies gilt vor allem für stark polyseme und hochfrequente Ausdrücke, für die die korpusgesteuerte Methode Ergebnisse hervorbringt, die nur einer oder wenigen Lesarten zugeord-



Beziehung(en) der Bedeutungsgleichheit/-äquivalenz


i **Synonym(e):**

anerkennen 



annehmen 


billigen 

dulden  


 *dulden* wird nur eingeschränkt synonymisch zu *akzeptieren* verwendet. Häufig setzt die Handlung, die mit *dulden* bezeichnet wird, einen geringeren Grad an Zustimmung des Handlungsträgers voraus, als es bei einer Handlung, die mit *akzeptieren* bezeichnet wird, der Fall ist. (Vgl. dazu den folgenden Beleg.)


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**. (Zürcher Tagesanzeiger, 18.06.1999, S. 20, Bilder der Gewalt.)

hinnehmen  

 *hinnehmen* wird nur manchmal synonymisch zu *akzeptieren* verwendet. Häufig setzt die Handlung, die mit *hinnehmen* bezeichnet wird, einen geringeren Grad an Zustimmung des Handlungsträgers voraus, als es bei einer Handlung, die durch *akzeptieren* bezeichnet wird, der Fall ist. (Vgl. dazu den folgenden Beleg.)


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.)


zulassen 


zustimmen 


Beziehung(en) des Bedeutungsgegensatzes

i **komplementäre(r) Partner:**

ablehnen 

kritisieren 

verurteilen 

verweigern 


weigern 

Abbildung 7: Ausschnitt der sinnverwandten Wörter zu *akzeptieren* in *lexiko*

net werden können, weil bestimmte Verwendungsweisen im Korpus sprachlich dominieren. Dann ist eine gezielte Korpusprüfung unerlässlich, um auch für weniger signifikante Verwendungen Informationen zu Relationswörtern angeben zu können. Gerade Fremdsprachenlernende haben ein Interesse, im Wörterbuch Synonymangaben auch bei weniger geläufigen Gebräuchen zu finden. So wird beispielsweise ein potentielles Synonym *Y* eines Suchwortes *X*, im Abstand von ca. 2 Sätzen im Korpus miteinander in beliebiger Reihenfolge gesucht. Die in Abbildung 4 dargestellten potentiellen synonymen Kandidaten zum Suchwort *akzeptieren* werden hinsichtlich ihrer Relevanz lexikografisch „manuell“ überprüft. Eine korpusbasierte Abfrage ist in *ellexiko* auch immer gleichzeitig die Suche nach geeigneten Belegen zur authentischen Illustrierung von sprachlichen Beziehungen. Bei der Überprüfung der gemeinsamen Textstellen wird z.B. deutlich, dass Ausdrücke wie *dulden* oder *hinnehmen* Verwendungsrestriktionen aufweisen. Alle gewonnenen Erkenntnisse werden daraufhin in der lexikografischen Beschreibung der Relationswörter u. a. in Form von Belegen und Kommentaren integriert (siehe Abbildung 7).

Aber neben dem Prüfen automatisch extrahierter Kandidaten und den zusätzlichen Erkenntnissen zu synonymen Einschränkungen wird die korpusbasierte Methode genutzt, um mögliche, vorwiegend weniger signifikante Beziehungen auch quantitativ zu erfassen und auszuwerten. Das hochpolyseme Stichwort *frei* (siehe Wörterbuchartikel unter: <http://www.owid.de/artikel/272789>), mit seinen 15 unterschiedlichen kontextuellen Verwendungen kann nicht allein auf der Basis eines korpusgesteuerten Verfahrens disambiguiert werden, und entsprechend bietet auch die statistische Auswertung unmittelbarer Kontextwörter nicht hinreichend Informationen zum sinnrelationalen Geflecht sämtlicher Lesarten und ihrer semantischer Nuancierungen. So konnten z.B. die Synonyme *nackt*, *unbegrenzt*, *ungeregelt* nicht durch automatische Analysen generiert werden. Sie sind z.B. auch nicht im entsprechenden Eintrag des Wortschatzportals aufgelistet. Die korpusbasierte Recherche nach diesen Ausdrücken zeigt aber, dass diese Ausdrücke mehrfach in bedeutungsähnlichen Kontexten belegt sind, als relevant eingestuft werden können und damit lexikografisch kontextuell eingeordnet, dokumentiert und belegt werden können. In *ellexiko* sind sie daher für das Stichwort als Synonyme mit folgenden Belegen gebucht:

nackt – eingeordnet in der Lesart: ‚nicht umgeben oder unbedeckt‘, Spezifizierung: ‚nackt‘

1. Wenn in den Sportumkleidekabinen die High-School-Jungs nicht nur *freien* Oberkörper, sondern auch *nackten* Hintern zeigen, kreischen die Mädels im Zuschauersaal laut auf, als wär man bei den „Chippendales“. (die tageszeitung, 12.11.1998, S. 29)

unbegrenzt – eingeordnet in der Lesart: ‚unbehindert‘

2. Für Gray sind die neuen Erlösungsfantasien von erschreckender Simplizität: Der *unbegrenzte* Weltmarkt als Heimstatt der aufgeklärten Vernunft, der Homo oeconomicus als Völlender der Moderne. Irrig ist dies laut Gray vor allem deshalb, weil der *freie* Markt alles andere als eine historische Notwendigkeit sei. (Frankfurter Rundschau, 29.10.1999, S. 7)

ungeregelt – eingeordnet in der Lesart: ‚unbehindert‘:

3. Der Beinahe-Zusammenbruch des Dualen Systems zeigt, wie berechtigt das Mißtrauen in die Kräfte des *freien* Marktes bei Umweltfragen ist. In einer absolut *ungeregelten* Marktwirtschaft würden die Schwachen untergehen, zu ihrem Schutz haben wir die soziale Marktwirtschaft. (Frankfurter Allgemeine Zeitung, 00.00.1993)

Die Synonym- bzw. Antonymangaben sind also nicht zentral für ein Stichwort erfasst, sondern sie werden kontextabhängig dokumentiert. Ohne korpusbasierte Suchanfragen hätten diese Synonyme nicht in *ellexiko* erfasst werden können.

In der lexikografischen Praxis von *ellexiko* hat sich der kombinierte Einsatz unterschiedlicher Strategien (multi-method approach) und die konsequente Dateninterpretation als sehr gewinnbringend erwiesen. Die lexikografisch angelegte Disambiguierung und die Zuordnung synonymmer/antonymer Ausdrücke in die entsprechenden Lesarten stellen einen erheblichen Mehrwert gegenüber rein automatisierten Listen dar. Die computergesteuerte korpusgeleitete Gewinnung von Synonymen oder Antonymen bringt statistisch signifikante Regularitäten systematisch zum Vorschein. Die korpusgesteuerte Ermittlung von Ausdrücken mit ähnlichem Kollokationsprofil stellt einen zusätzlichen Gewinn für solche Suchwörter dar, für die die statistische Kookkurrenzanalyse wenig bis gar keine paradigmatischen Partnerwörter extrahieren kann. Aber nur die lexikografische Überprüfung der Daten ermöglicht Einblicke in Verwendungsbeschränkungen, gewährleistet die entsprechende kontextuelle Einordnung und bietet somit insgesamt eine deutlich höhere Authentizität und Zuverlässigkeit der Wörterbuchangaben. Die dokumentierten Wörterbuchinformationen sind empirisch fundiert, inhaltlich bestätigt, quantitativ evaluiert und lexikografisch kontextuell eingeordnet. Aber gewährleistet diese Strategie nun Wörterbucheinträge mit einer stärkeren Konsistenz von bilateralen Beziehungen und mit konsequent symmetrischen Verweisen?

3. Methoden konsistenter sinnrelationaler Vernetzungsstrukturen

Im Jahr 2010 wurde in *ellexiko* eine Auswertung vorgenommen, bei der der damalige Stichwortbestand von ca. 1400 Stichwörtern mit all ihren Lesarten und ihren sinnverwandten Wörtern, die mit den oben erwähnten Methoden gewonnen wurden, dahingehend überprüft wurde, ob alle ein- und ausgehenden Relationen in beiden Richtungen korrekt eingetragen sind.⁸ Das Ergebnis war, dass 2558 bidirektionale Vernetzungen (1279 Wortpaare) korrekt auf Lesartenebene angegeben wurden. 1700 Vernetzungen waren dagegen nicht symmetrisch miteinander vernetzt. Die aufgedeckte Lückenhaftigkeit ist auf mehrere Ursachen zurückzuführen. Zum einen passierte es, dass ein Partnerwort schlichtweg während des lexikografischen Prozesses vergessen wurde. Außerdem kam es vor, dass auf eine falsche Lesart, also auf einen falschen Kontext bei polysemen Ausdrücken verlinkt wurde. Die Mehrzahl der Fehler aber unterliefen, weil zahlreiche Stichwörter *A*, *B*, *C* etc. zum Zeitpunkt der Erarbeitung eines Stichwortes *X* nicht redaktionell ausgearbeitet sind, weil *ellexiko* kein abgeschlossenes Projekt ist. Zum Zeitpunkt der Bearbeitung von *X* kann daher nicht auf *A*, *B* oder *C* verlinkt werden, weil diese Stichwörter noch nicht über ein ausgearbeitetes Bedeutungsspektrum verfügen. Wurde eines der bereits als Beziehungspartner dokumentierten Stichwörter *A*, *B* oder *C* später ausgearbeitet, wurde vermutlich vergessen, die für einen korrekten Verweis zusätzlich benötigten Informationen, wie Angabe der entsprechenden Lesart, im Stichwort *X* nachzutragen.

⁸ Derzeit sind ca. 1700 Stichwörter in *ellexiko* vollständig redaktionell ausgearbeitet.

Das Ergebnis der Auswertung zeigt deutlich, dass nicht ein bestimmtes Prozedere oder die Kombination verschiedener Korpusverfahren das Instrument der Konsistenzsicherung sind. Selbst wenn sämtliche relevante Beziehungen auf die eine oder andere Art ermittelt werden, ist das kein Garant dafür, dass diese auch konsequent symmetrisch dokumentiert werden, auch dann nicht, wenn das Wörterbuch inhaltlich abgeschlossen ist und alle Stichwörter bearbeitet sind. Hält man sich noch einmal die Vielzahl an sinnrelationalen Beziehungen für ein hochpolysesmes und hochfrequentes Stichwort wie *frei* vor Augen (siehe Abbildung 8)⁹, wird deutlich, dass diese Geflechte nicht intuitiv systematisch abrufbar sind.

Die Komplexität des Beziehungsnetzes für den Ausdruck *frei* illustriert die folgende Abbildung, welche Beziehungen zusammengehöriger Kontexte aufgliedert.

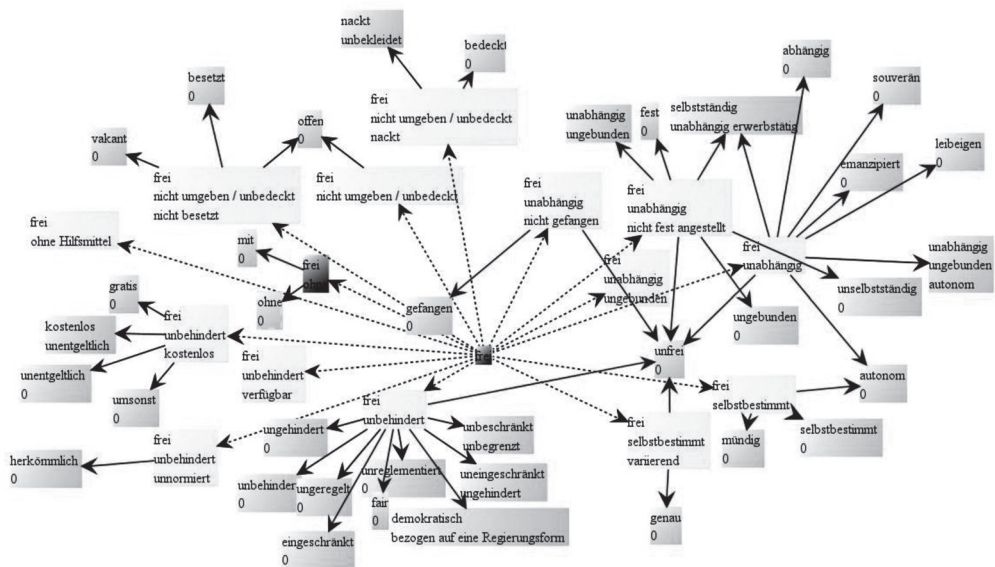


Abbildung 8: Ausgehende lesartengebundene synonyme Vernetzungen des Ausdrucks *frei*

Die größten Bemühungen auf redaktioneller Seite scheitern beim Versuch, alle Vernetzungen introspektiv korrekt abzurufen und daraufhin symmetrisch anzulegen. Müller-Spitzer (2010, 150f.) weist darauf hin, dass es sich schlichtweg um menschliches Versagen handelt. Alle sinnverwandten Ausdrücke sind, sofern sie selbst Stichwort des Wörterbuches sind, als lesartenbezogener Link angelegt, um eine schnelle Navigation zu diesen Stichwörtern in

⁹ Die Abbildung hat Dr. Peter Meyer, wissenschaftlicher Mitarbeiter des IDS, erstellt. Die Rechtecke enthalten den Relationspartner (oben) und darunter die entsprechende Lesart und, falls vorhanden, die relevante semantische Spezifizierung, für die eine Beziehung besteht. Bei der mit Schrägstrich markierten Bezeichnung einer Lesart handelt es sich um eine Doppelbenennung. Die in der Grafik mit „0“ markierten Partnerwörter verfügen noch nicht über ein ausgearbeitetes Lesartenspektrum. Die Beziehung wird daher zunächst auf Stichwortebene und nicht auf Kontextebene (Lesartenebene) angesetzt.

ihren jeweils relevanten sinnrelationalen Verwendungen zu gewährleisten. Die Verlinkungen sind also nicht auf der Ebene der Stichwörter angelegt, sondern auf Lesartenebene, um spezifische Beziehungen rein kontextuell zu erfassen. Damit gelangt man von einem Stichwort in einer bestimmten Verwendung zu einem seiner Synonyme oder Gegensatzwörter in den entsprechenden Kontexten und erfasst, in welchen konkreten Verwendungssituationen beispielsweise eine bedeutungsähnliche oder oppositionelle Beziehung vorliegt. Anders als bei rein lemmabasierten Zuordnungen (z.B. bei Wörterbucheinträgen im Leipziger Wortschatzportal), liegt bei *ellexiko* die Herausforderung darin, dass bei der Erarbeitung eines Wörterbuchartikels Informationen zu den entsprechenden Verlinkungen sehr detailliert und lesartengebunden in der Artikelstruktur hinterlegt werden müssen.

Diese Art der Vernetzung ist nicht ohne Computerunterstützung zu leisten. Daher sollte für die Überprüfung von Wörterbuchvernetzungen eine Management-Software zum Einsatz kommen, die Beziehungen zu verschiedenen Stichwortebenen und zwischen verschiedenen Beziehungsarten prüfen kann (cf. Müller-Spitzer 2010). Für ein elektronisches Wörterbuch müssen dafür bestimmte technische Voraussetzungen geschaffen werden.

Vernetzungen werden in den *ellexiko*-Artikeln technisch durch XML-Elemente im verweisenden Artikel repräsentiert; das Verweisziel wird darin durch eindeutige Kennungen (Zahlen bzw. Wörter oder Wortfolgen) spezifiziert, die dem Artikel als ganzen sowie auch z.B. seinen einzelnen Lesarten über XML-Attribute zugeordnet sind (cf. auch Müller-Spitzer 2010). Um die Bidirektionalität und Konsistenz etwa einer Antonymie-Vernetzung zu prüfen, müssen die zugehörigen XML-Elemente in den beiden Artikeln aufgesucht und die Vernetzungsinformationen miteinander verglichen werden. (Storjohann/Meyer 2012, 130)

Zur Sicherstellung von sinnrelationalen Beziehungen wurden in *ellexiko* alle dafür notwendigen technischen Gegebenheiten der Redaktionsumgebung hergestellt, wie z.B. eine Schnittstelle zum XML-Editor, Zugriff auf die Datenbank, in der die Artikelinstanzen hinterlegt sind, etc. (siehe Meyer 2010, Klosa 2011b: 14–16, Müller-Spitzer 2010, Müller-Spitzer/Schneider 2009). Im Rahmen des Projektes BZV*ellexiko*¹⁰ konnte dann ein solches Tool, der Vernetzungsmanager *vernetziko*, entwickelt werden, der seit 2010 bei der täglichen Artikelarbeit eingesetzt wird. Abbildung 9 demonstriert am Beispiel des Artikels *frei* die Benutzeroberfläche des Vernetzungsmanagers. Im oberen Bereich befinden sich diejenigen Stichwörter, die eine Beziehung auf den Artikel *frei* dokumentiert haben. Zusätzlich werden alle Informationen angegeben, die an diese Beziehung gebunden sind, wie die relevante Lesart oder eine relevante semantische Spezifizierung. Im unteren Bereich werden alle ausgehenden Beziehungen bzw. Beziehungsarten von *frei* für die entsprechenden Lesarten aufgeführt, die zum Zeitpunkt der Erarbeitung dokumentiert wurden. Der Vernetzungsmanager erteilt Auskunft über den Status der Beziehung (siehe Spalte „Vernetzungsstatus“), also darüber, ob eine Zuweisung der entsprechenden Lesart eines Stichwortes nicht erfolgt oder etwa, ob bei einem synonymen Ausdruck der Verweis in die Gegenrichtung fehlt. Informationen zum jeweiligen Vernetzungsstatus werden nach Auswählen einer Beziehung unterhalb der Tabelle zusätzlich kurz erläutert.

¹⁰ Zum Projekt BZV*ellexiko* (Benutzeradaptive Zugänge und Vernetzungen in *ellexiko*) siehe <http://www.ids-mannheim.de/lexik/BZVlexiko/>. Der Vernetzungsmanager wurde in Zusammenarbeit mit Dr. Peter Meyer entwickelt.

frei - vernetziko: Vernetzungsmanager und Redaktionshelfer für elexiko

paradigmatische Vernetzungen | andere Vernetzungen | Datenbanksuche | Visualisierung

Eingehende Vernetzungen geprüfte Vernetzungen ausblenden ↓ Tabellen maximieren

geprüft?	Vernetzungstyp	Quell-Lemma	Quell-Lesart	Vernetzungsstatus	Zieladresse
<input type="checkbox"/>	inkompatibilität	autonom *	'ohne äußeren Einfluss' / Spez. 'politisch'	↔	frei 'unabhängig'
<input type="checkbox"/>	inkompatibilität	autonom *	'eigenständig'	↔	frei 'selbstbestimmt'
<input type="checkbox"/>	inkompatibilität	geheim *	'vertraulich'	⊕↔	frei 'selbstbestimmt'
<input type="checkbox"/>	synonymie	kostenlos	'unentgeltlich'	↔	frei 'unbehindert' / Spez. 'kostenlos'
<input type="checkbox"/>	synonymie	leer *	'nicht belebt'	↔	frei 'nicht umgeben oder unbedeckt' / Spez. 'nicht
<input type="checkbox"/>	inkompatibilität	selbstständig	'autonom'	⊕↔	frei 'unabhängig'
<input type="checkbox"/>	inkompatibilität	selbstständig	'ohne äußere Ermischung'	⊕↔	frei 'selbstbestimmt'
<input type="checkbox"/>	synonymie →	unabhängig	'ungebunden'	↔	frei 'unabhängig'

Vernetzungsstatus: Zieladresse bekannt. Eine genau passende, vollständige Vernetzung in Gegenrichtung gefunden.

LA-Spektrum | Browser | Vernetzung korrigieren | RefIDs/LZA | <relpartnerA> | <relpartner> | im Editor zu Zieladresse

Quelle im Editor aus Quelle einfügen

Ausgehende Vernetzungen geprüfte Vernetzungen ausblenden Vernetzungen auf unbearbeitete Artikel ausblenden

geprüft?	Vernetzungstyp	Zieladresse	Vernetzungsstatus	Ziel-Lemma	Ziel-Lesart
<input type="checkbox"/>	komplementarität	frei 'unabhängig'	↔	abhängig	[nicht red. bearb.]
<input checked="" type="checkbox"/>	inkompatibilität	frei 'unabhängig'	↔	autonom *	'ohne äußeren Einfluss' / Spez. 'politisch'
<input checked="" type="checkbox"/>	inkompatibilität	frei 'selbstbestimmt'	↔	autonom *	'eigenständig'
<input type="checkbox"/>	komplementarität	frei 'nicht umgeben oder unbedeckt' / Spez. 'nicht'	↔	bedeckt	[nicht red. bearb.]
<input type="checkbox"/>	inkompatibilität	frei 'unbehindert'	⊕↔	demokratisch *	'bezogen auf Regierungsform'
<input type="checkbox"/>	komplementarität	frei 'unbehindert'	↔	eingeschränkt	[nicht red. bearb.]
<input type="checkbox"/>	inkompatibilität	frei 'unabhängig'	↔	emanzipiert	[nicht red. bearb.]
<input type="checkbox"/>	inkompatibilität	frei 'unbehindert'	↔	fair	[nicht red. bearb.]
<input type="checkbox"/>	komplementarität	frei 'unabhängig' / Spez. 'nicht fest angestellt'	↔	fest	[nicht red. bearb.]

Vernetzungsstatus: Zieladresse bekannt. Eine genau passende, vollständige Vernetzung in Gegenrichtung gefunden.

LA-Spektrum | Browser | auschecken und anzeigen | im Editor zu relpartnerA

Ziel

alle Bausteine | passende Bausteine | Notizen | aktualisiere alle Anzeigen | Einstellungen | prüfe Artikel | Ordner einchecken | Hilfe

LA-Spektrum anzeigen für: XML-Quelltext in Zwischenablage

Abbildung 9: Vernetzungsmanager für *ellexiko*-Wörterbuchartikel *frei* bei Korrekturarbeiten

Ellexiko hat seitdem zahlreiche Fehler beseitigen, relevante Angaben, die zur Vernetzungsinformation gehören, nachtragen oder Partnerwörter ergänzen können. Gegenwärtig ist das Tool integrativer Bestandteil der redaktionellen Artikelarbeit, um bereits während des lexikografischen Prozesses alle nötigen Schritte zur Konsistenzherstellung durchführen zu können.

4. Schlussbemerkungen

In *ellexiko* unterliegt die Auswahl der sinnverwandten Wörter primär Kriterien, die an den sprachlichen Usus gebunden sind und im Korpus eine ermittelbare Signifikanz aufweisen. Ziel von *ellexiko* ist es, neben einem breiten Spektrum an Angaben vor allem authentische und zuverlässige Informationen für Nachschlagende kontextgebunden bereitzustellen. Dafür erweist sich das Zusammenspiel aus korpusgesteuertem und korpusbasiertem Ansatz als besonders geeignet, um optimale Erkenntnisse über verschiedene paradigmatische Konstruktionen zu erhalten, wie sie auch im tatsächlichen Sprachgebrauch realisiert werden. Bei der täglichen lexikografischen Arbeit mit einem Korpus und dem gewählten methodologischen Vorgehen werden die Diskrepanzen zur eigenen Intuition, zu stärker traditionell ori-

entierten Synonymiken oder zu ausschließlich automatisch extrahierten Relationsangaben sichtbar. Intuitiv Erwartbares wird manchmal bestätigt, manchmal widerlegt, und Unerwartetes wird zum Teil als reguläres Phänomen evident. Beides wird mit Korpusdaten systematisierbar und gezielt überprüfbar gemacht. Diese „unexpectedness of the findings“ (Tognini-Bonelli 2001, 48) stellt einen großen Gewinn bei der Arbeit mit einem umfangreichen Korpus in *lexiko* dar. Aber die Korpusbasis, die entsprechenden Analyse- und Retrieval-Tools und die gewählten Ansätze garantieren nicht, wie Paradis/Willners (2006) vermuten, „a symmetry of presentation“. Für ein Printwörterbuch oder für ein elektronisches Nachlagewerk gilt, dass konsistente Verweis- oder Vernetzungsstrukturen nur mit zusätzlicher Softwareunterstützung gewährleistet werden können, denn nur diese kann Fehler oder inhaltliche Lücken ermitteln und korrigieren, für die man aus metalexikografischer Sicht nach vermeintlich nachvollziehbaren und transparenten Prinzipien sucht. Die Gewinnung von Synonymen oder Antonymen kann aus linguistischer Sicht durchaus sehr strikten linguistischen oder methodologischen Kriterien unterliegen, die sich jedoch nicht immer in der lexikografischen Präsentation nachvollziehen lassen.

Der lexikografische Aufwand zur Gewinnung, Analyse und Dokumentation von sinnrelationalen Ausdrücken ist trotz leistungsstarker Computerwerkzeuge in *lexiko* hoch. Der Einsatz von Recherchewerkzeugen und die Entwicklung maßgeschneiderter Tools haben die lexikografische Praxis in *lexiko* positiv verändert und die Wörterbuchbeschreibungen objektiver und inhaltlich konsistenter gemacht. Die Weiterentwicklung solcher oder ähnlicher Software wird auch in Zukunft die Arbeit der Lexikografinnen und Lexikografen verändern. Besonders in der englischen Lexikografie werden bereits Werkzeuge genutzt, die Daten inhaltlich nach diversen Merkmalen stärker ‚stromlinienförmig‘ bündeln.¹¹ Die computationellen Prozesse der so genannten „Streamline-Tickboxlexikografie“ vermögen es, sprachliche Daten hinsichtlich unterschiedlicher linguistischer Kriterien automatisiert zu extrahieren, inhaltlich zu bündeln und stark vorselektiert anzubieten. Aufgrund ihrer fortlaufenden „Lernfähigkeit“ vermögen sie es außerdem, die kompetenzgesteuerte Selektion der gewonnenen Daten seitens der LexikografInnen nachzuvollziehen und abzuspeichern, um bei weiteren Anfragen die Suchergebnisse weiter nach erlernten Wünschen zu optimieren. Die Daten können dann automatisiert, ohne weiteres Eingreifen des Lexikografen/der Lexikografin in das entsprechende Dokument an die ausgewiesenen Stellen überführt werden. Diese Methode wurde z.B. im Projekt Dante - A lexical database for English erprobt. Neue Prozesse mit automatischer Datenauswahl werden die Wörterbucharbeit, und damit wohl auch die Methoden zur Ermittlung von sinnverwandten Ausdrücken, künftig weiter verändern. Neue lexikografische Modelle werden bereits entwickelt (cf. Rundell/Kilgariff 2011).

In this model, we envisage a change from the current situation, where the corpus software (some version of the word sketches) presents data to the lexicographer in (as we have seen) intelligently pre-digested form, to a new paradigm where the software selects what it believes to be relevant data and actually populates the appropriate fields in the dictionary database. In this way of working, the lexicographer's task changes from selecting and copying data from the software, to validating – in the dictionary writing system – the choices made by the computer. Having deleted or adjusted

¹¹ Einen derartigen computergesteuerten Prozess bietet z.B. das Tool *Sketch Engine* (<http://www.sketchengine.co.uk/>) an, ein Korpusanalysesystem, das die Verwendungsmöglichkeiten eines Ausdrucks nach grammatischen und kollokationalen Kriterien zusammenfasst.

anything unwanted, the lexicographer then tidies up and completes the entry. (Rundell/Kilgarriff 2011, 278)

Die gegenwärtige und wohl auch künftige Wörterbucharbeit ist ohne große elektronische Korpora, ohne Computertechnologien und ohne komplexe Automatisierungsprozesse undenkbar. Sie sollte aber die lexikografische Dateninterpretation nicht ausschließen müssen. Ob es künftig überhaupt noch LexikografInnen geben wird (cf. Grefenstette 1998, Rundell 2002, Rundell/Kilgarriff 2011), hängt zum einen davon ab, wie zuverlässig und authentisch NutzerInnen ihre Wörterbücher haben möchten, aber zum anderen auch davon, wie gut es LexikografInnen gelingen wird, ihre veränderte Rolle bei der Entstehung von Wörterbüchern vorteilhaft wahrzunehmen.

5. Literatur

5.1 Wörterbücher

Collins COBUILD 2003 = Sinclair, John (ed.): English Advanced Learner's Dictionary. 4. Auflage. Glasgow: Harper Collins Publishers, 2003.

elexiko = *elexiko*. <http://www.owid.de/wb/elexiko/start.html>

Wortschatzportal = Wortschatzlexikon. Wortschatzportal der Universität Leipzig. <<http://wortschatz.uni-leipzig.de/>>

5.2 Literatur

Belica 1995 = Belica, Cyril: Statistische Kollokationsanalyse und Clustering. Korpusanalysemodul. Mannheim: Institut für Deutsche Sprache, 1995.

Belica 2001ff. = Belica, Cyril: Kookkurrenzdatenbank CCDB – V3.2. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. Mannheim: Institut für Deutsche Sprache, 2001ff. <<http://corpora.ids-mannheim.de/ccdb/>>

Belica 2011 = Belica, Cyril: Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In: Abel, Andrea/Zanin, Renata (edd.): Korpora in Lehre und Forschung. Bozen-Bolzano: Freie Universität, 2011, 155–178.

Cruse 1986 = Cruse, Alan: Lexical Semantics. Cambridge: Cambridge University Press, 1986.

Grefenstette 1998 = Grefenstette, Gregory: The Future of Linguistics and Lexicographers: Will there be Lexicographers in the year 3000? In: Fontenelle, Thierry et al. (edd.): Proceedings of the 8th EURALEX International Congress on Lexicography in Liege, Belgium. Liege: University of Liege, English and Dutch Departments, 1998, 25–41.

Jones 2002 = Jones, Steven: Antonymy: a corpus-based perspective. London/New York: Routledge, 2002.

Jones et al. 2012 = Jones, Steven et al.: Antonyms in English: Construals, Constructions and Canonicity. Studies in English Language. Cambridge: Cambridge University Press, 2012.

Justeson/Katz 1991 = Justeson, John S./Katz, Slava M.: Redefining Antonymy: The Textual Structure of a Semantic Relation. In: Literary and Linguistic Computing 7, 1991, 176–184.

Meyer 2011 = Meyer, Peter: *vernetziko*: A Cross-Reference Management Tool for the Lexicographer's Workbench. In: Conference Proceedings of eLEX2011 (Electronic Lexicography in the 21st

- Century: New Applications for New Users), 10–12 November 2011, Bled, Slovenia. Bled, eLEX, 2011. <<http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-25.pdf>>
- Müller-Spitzer 2010 = Müller-Spitzer, Carolin: The Consistency of Sense-Related Items in Dictionaries: Current Status, Proposals for Modelling and Applications in Lexicographic Practice. In: Storjohann, Petra (ed.): *Lexical-Semantic Relations. Theoretical and Practical Perspectives*. Amsterdam: John Benjamins, 2010, 145–162 (*Linguisticæ Investigationes Supplementa* 28).
- Müller-Spitzer/Koplenig/Töpel 2011 = Müller-Spitzer, Carolin/Koplenig, Alexander/Töpel, Antje: What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project. In: Conference Proceedings of eLEX2011 (Electronic Lexicography in the 21st Century: New Applications for New Users), Bled, Slovenia. Bled, eLEX, 2011. <www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-27.pdf>
- Paradis/Willners 2006 = Paradis, Carita/Willners, Caroline: What a Corpus Based Dictionary Tells us about Antonymy. In: Corino, Elisa/Marello, Carla/Onesti, Christina (edd.): *Proceedings of the 12th EURALEX International Congress 2006*, Turin. Band 1. Turin: Edizioni dell’Orso Alessandria, 2006, 213–219.
- Paradis/Willners 2007 = Paradis, Carita/Willners, Caroline: Antonyms in Dictionary Entries: Methodological Aspects. In: *Linguistica* 61/3, 2007, 261–277.
- Rundell 2002 = Rundell, Michael: Good Old-Fashioned Lexicography: Human Judgment and the Limits of Automation. In: Corréard, Marie-Hélène (ed.): *Lexicography and Natural Language Processing. A Festschrift in Honour of B. T. S. Atkins*. Grenoble: EURALEX, 2002, 138–155.
- Rundell/Kilgarriff 2011 = Rundell, Michael/Kilgarriff, Adam: Automating the Creation of Dictionaries: Where will it all End? In: Meunier, Fanny et al. (edd.): *A Taste for Corpora*. In Honour of Sylviane Granger. Amsterdam/Philadelphia: John Benjamins, 2011, 257–281.
- Sinclair 1991 = Sinclair, John: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
- Sinclair 2002 = Sinclair, John: Intuition and annotation – the discussion continues. In: Aijmer, Karin/Altenberg, Bengt (edd.): *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*. Göteborg 2002, 39–59.
- Storjohann 2005 = Storjohann, Petra: Paradigmatische Relationen. In: Haß, Ulrike (ed.): *Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz*. Berlin/New York: de Gruyter, 2005, 249–264.
- Storjohann 2006 = Storjohann, Petra: New Lexicographic Approaches to the Description of Sense Relations. In: Corino, Elisa/Marello, Carla/Onesti, Christina (edd.): *Proceedings of the 12th EURALEX International Congress 2006*, Turin. Band 2. Turin: Edizioni dell’Orso Alessandria, 2006, 1201–1212.
- Storjohann 2010 = Storjohann, Petra: Synonymy in Corpus Texts – Conceptualisation and Construction. In: Storjohann, Petra (Hg.): *Lexical-Semantic Relations. Theoretical and Practical Perspectives*. Amsterdam: John Benjamins, 2010, 69–94 (*Linguisticæ Investigationes Supplementa* 28).
- Storjohann/Meyer 2012 = Storjohann, Petra/Meyer, Peter: Konsistent vernetzte Wörterbucheinträge. In: *Zeitschrift für Angewandte Linguistik* 56, 2012, 121–134.
- Tognini-Bonelli 2001 = Tognini-Bonelli, Elena: *Corpus Linguistics at Work*. New York/Amsterdam: John Benjamins, 2001.

5.3 Elektronische Quellen

- BZ*Velexiko* = BZ*Velexiko*: Benutzeradaptive Zugänge und Vernetzungen in *elexiko*. <<http://www1.ids-mannheim.de/lexik/BZVelexiko/>>
- CCDB = Kookkurrenzdatenbank CCDB. <<http://corpora.ids-mannheim.de/ccdb/>>
- COSMAS II = Corpus Search, Management and Analysis System (COSMAS). <<http://www.ids-mannheim.de/cosmas2/uebersicht.html>>
- Dante = Dante. A Lexical Database for English. <<http://www.webdante.com/index.html>>
- Sketch Engine = Sketch Engine. <<http://www.sketchengine.co.uk/?page=Website/SketchEngine>>

