

Christian Kreuz/David Römer

EDV-gestützte Arbeit mit Korpora. Eine praxisorientierte Handreichung¹

Stand: 15.03.2013

¹ Noah Bubenhofer danken wir für viele Hinweise und zahlreiche Tipps. Außerdem verweisen wir auf seine ausführliche und übersichtliche Online-Einführung in die Korpuslinguistik: <http://www.bubenhofer.com/korpuslinguistik/kurs/>, zuletzt abgerufen am 15.03.2013.

Inhalt

1	Einleitung.....	1
2	Ingwer: Eine Datenbank für Diskurs- und Korpusanalyse	2
3	Open Corpus Workbench.....	8
3.1	CQP	8
3.2	Suchanfragen und Suchsyntax.....	9
3.2.1	Standardabfrage (Standard query).....	9
3.2.2	Die eingeschränkte Abfrage (Restricted query).....	18
3.2.3	Ergebnisdarstellung und Weiterverarbeitung.....	18
3.3	Weitere Funktionen	23
3.3.1	Word lookup.....	23
3.3.2	Frequency lists.....	24
3.3.3	Keywords und Teilkorpora.....	24
3.3.4	Suchverlauf und gespeicherte Suchanfragen.....	25
4	Fazit.....	26
5	Literatur.....	26
6	Korpora	28
7	Die Autoren.....	28

1 Einleitung

In diversen empirischen Studien (nicht nur) diskurslinguistischer Provenienz² wurde bewiesen, dass Presstexte lohnende Untersuchungsquellen sind, um öffentliche Diskurse zu erforschen.³ Mit zunehmender digitaler Verfügbarkeit von Zeitschriften und Zeitungen⁴ steigen nicht nur ihr Nutzen und die Möglichkeiten für linguistische Analysen. Gleichzeitig steigt das Bedürfnis nach einfach bedienbaren technischen Hilfsmitteln, um die Daten zu

² Vgl. u.a. die Arbeiten der „Düsseldorfer Schule“: Böke (1996a, 1996b, 1997); Stötzel/Wengeler (1995); Wengeler (2003). Einen Überblick über diesen Forschungsansatz verschaffen Jung (2000) und Wengeler (2005). Empirische Untersuchungen an Presstextkorpora neueren Datums sind u.a.: Wengeler/Ziem (2010); Wengeler/Scholz (2012); Ziem/Scholz/Römer (2013); Kuck/Römer (2012), Römer (2013). Den aktuellen Forschungsstand zur linguistischen Diskursanalyse geben die Bände von Warnke (2007) und Warnke/Spitzmüller (2008) sowie das Einführungsbuch von Spitzmüller/Warnke (2011) wieder.

³ Ähnliches konstatiert Thomas Niehr (1999: 205): „Zur Erforschung des öffentlichen Diskurses bzw. des öffentlichen Sprachgebrauchs ist die Analyse von großen Zeitungskorpora in hervorragender Weise geeignet“.

⁴ So sind etwa der SPIEGEL und die ZEIT in Online-Datenbanken frei zugänglich. Auch andere „leitende“ Printmedien, wie SÜDDEUTSCHE ZEITUNG (SZ), FRANKFURTER ALLGEMEINE ZEITUNG (FAZ) oder das Populärmedium BILD, lassen sich zumindest ab den 1990er Jahren über Universitätsbibliotheken und auf CD-ROM digital recherchieren.

„beherrschen“, den Überblick über das Sprachmaterial zu behalten und um die Textmengen computergestützt möglichst exhaustiv auswerten zu können.

In dieser kleinen Handreichung stellen wir eine Datenbank- und Analysesoftware vor, die im Rahmen des DFG-Projektes „Sprachliche Konstruktionen sozial- und wirtschaftspolitischer „Krisen“ in der BRD von 1973 bis heute“⁵ in Zusammenarbeit mit semtracks⁶ und DiskursNetz⁷ für diskurs- und korpuslinguistisches Arbeiten entwickelt wurde. Anschließend stellen wir eine korpuslinguistische Standardsoftware vor, die IMS Open Corpus Workbench⁸ und zeigen, wie sich in linguistisch annotierten Korpora große Textsammlungen – im Hinblick auf die grafische Schnittstelle (CQPweb) zur Open Corpus Workbench – durchforsten lassen.

2 Ingwer: Eine Datenbank für Diskurs- und Korpusanalyse

Die Verwaltung und Analyse von Massendaten stellt eine besondere Herausforderung dar. Datenbanken bieten verschiedene arbeitspraktische Vorteile: Einerseits lassen sie sich als „modernen Zettelkasten“ nutzen, d.h. zur Verwaltung von Texten und Belegen sowie für Recherchen verschiedener Art. Auf der anderen Seite sind heutige Datenbanken webfähig, sodass mit ihnen kollaborativ gearbeitet werden kann. Einmal in einer Datenbank gespeicherten Daten lassen sich außerdem, je nach vorhandenen Schnittstellen, zur Weiterverarbeitung in andere Systeme transferieren.

Das sind einige gute Gründe, weshalb wir uns im Rahmen des eingangs erwähnten Forschungsprojektes dazu entschieden haben, zur Verwaltung und computergestützten qualitativen wie quantitativen Analyse des im Projekt entstandenen Korpus‘ zu sozialen, wirtschaftlichen und politischen „Krisen“ (KriKo)⁹ eine Datenbank (Ingwer) zu entwickeln.

Die Idee von Ingwer ist aus dem Bedürfnis gewachsen, große Textmengen handhabbar zu machen und gleichzeitig mit verschiedenen Methoden untersuchen zu können. Dabei hielten wir es für angemessen, so wenige individuelle Strategien und spezifische vorbereitende Beschlüsse in der Handhabung mit den Daten wie möglich zu zulassen¹⁰, uns aber gleichzeitig an gewissen Standards im Umgang mit ihnen zu orientieren (z.B. TEI¹¹). Mit Ingwer ist es uns gelungen, trotz unterschiedlicher Forschungsinteressen und -methoden der daran Mitwirkenden, eine konsistente Datenbasis zu schaffen, die den heterogenen Voraussetzungen gerecht wird, d.h. die divergierenden wissenschaftlichen Einstellungen und methodischen

⁵ Für nähere Informationen verweisen wir auf den Internetauftritt des Projektes: <http://www.uni-trier.de/index.php?id=45453>, zuletzt abgerufen am 15.03.2013.

⁶ <http://www.semtracks.com/web/>, zuletzt abgerufen am 15.03.2013.

⁷ <http://www.diskursanalyse.net/wiki.php?wiki=DFG-MeMeDa::DiskursNetz>, zuletzt abgerufen am 15.03.2013.

⁸ <http://cwb.sourceforge.net/index.php>, zuletzt abgerufen am 15.03.2013.

⁹ Das KriKo umfasst mehr als 10.000 Presstexte aus BILD, FAZ, SZ, SPIEGEL und ZEIT über einen Zeitraum von 1973 bis 2009.

¹⁰ So hat beispielsweise die Aufbereitung der Daten je nach Fragestellung und Methode anders zu erfolgen oder wird von bestimmten Tools auf eine bestimmte Weise verlangt, wodurch zahlreiche Vorentscheidungen über spätere Analysemöglichkeiten getroffen werden, was immer auch zur Folge hat, dass andere Verfahrensweisen ausgeschlossen werden. Nicht zuletzt wird durch solche Vorentscheidungen auch das Innovations- und Kreativitätspotential der Forscher gehemmt.

¹¹ <http://www.tei-c.org/index.xml>, zuletzt abgerufen am 15.03.2013.

Zugänge gleichwertig behandelt, ohne eines als Primat anzusehen. Man kann sich auf Augenhöhe begegnen, weil etwa die Software sowohl einem hermeneutischen Zugang zum Datenmaterial als auch einem quantitativ-berechnendem Zugang dient. Dabei können beide Zugänge nebeneinander herlaufen oder aber ineinander integriert werden. Damit eignet sich Ingwer unserer Meinung nach besonders für Projektarbeit und für das bunte Feld der linguistischen Diskursanalyse.



Abb. 1: „Start“

Ingwer ist serverbasiert und besteht im Wesentlichen aus einer MySQL Datenbank mit graphischer Benutzerschnittstelle (vgl. Abb. 1-4), die sich aus drei Teilkomponenten zusammensetzt:

1. Eine der zentralen Funktionen von Ingwer ist, dass mit Hilfe der Datenbank digital verfügbare Zeitungsartikel – und alle darin enthaltenen Annotationen – im Volltext verwaltet und durchsucht werden können. Dadurch besteht die Möglichkeit, Teilmengen oder Teilkorpora aus dem gesamten Datenbestand bilden zu können, was eine wichtige Voraussetzung für Analysen des betreffenden Sprachgebrauchs in Bezug auf bestimmte Dimensionen wie Diachronie, Akteure oder Textsorten ist.
2. Eine weitere Komponente von Ingwer ist das Modul zur manuellen Annotation, d.h. zur Anreicherung der Texte oder bestimmter Textstellen (beliebiger Länge) mit linguistischen Interpretationen der sprachlichen Phänomene. Mit dieser Funktion können Belege zu frei erstellbaren Analysekatoren, die sich nicht oder nur begrenzt auf der Ebene der sprachlichen Oberfläche algorithmisch berechnen lassen, etwa (konzeptuelle) Metaphern oder Argumentationstopoi, lesend erkannt, „per Hand“ ausgezeichnet und auf diese Weise gesammelt werden.
3. Das dritte Kernstück von Ingwer besteht aus verschiedenen Importfunktionen und Exportroutinen, um für spezialisierte Analysen beliebige Mengen von Artikeln und Belegen aus der Datenbank in andere Systeme zu überführen.

Es wurde bereits erwähnt, dass ein häufiges Problem im Umgang mit elektronischen Korpora die Frage ist, wie sich große Textmengen nach bestimmten Einflussgrößen effizient und komfortabel (auch für informatische Laien) verwalten lassen. Ein Blick auf die „Such-Seite“¹² von Ingwer zeigt, dass in der Datenbank verschiedene Felder angelegt wurden, die den Metadaten¹³, d.h. bestimmten Metainformationen über die Texte, entsprechen.

Suchen

+ **Zeitung**
 Spiegel
 SZ

+ **Datum** 2003-01-01 zwischen 2003-12-31 TT.MM.JJJJ

+ **Titel**

+ **Untertitel**

+ **Autor**
 Gerhard Mauerer
 Gerhard Schröder

+ **Ressort**
 Aktuelles
 Anleihen, Devisen und Genussscheine

+ **Presstextsorte**
 Anzeige
 Bericht

+ **Geopolitischer Bezug**
 Abu Dhabi
 Afrika

+ **Seite**

+ **Serie**
 Der Münchhausen-Test
 Der Volkswirt

+ **word_count**

+ **lemma_count**

+ **Textkörper**

+ **Kategorie**

Ergebnisse:

id	Zeitung	Datum	Textbeginn
14197	Spiegel	2003-04-19	"Dann lasst uns streiten!" Bundeskanzle...
14241	Spiegel	2003-07-07	Viel Kraft in diesem Land Bundeskanzler...
			Alle auswählen / Auswahl entfernen

Anzahl Datensätze: 2 Markierte Datensätze: 0 Anzahl Wörter: 4861 Anzahl Lemma: 4313

Abb. 2: „Suchen“

Welche Metadaten erhoben werden, ist natürlich abhängig vom jeweiligen Erkenntnisinteresse und damit grundsätzlich projektspezifisch.¹⁴ Das KriKo weist folgende Metadatenstruktur auf: Jeder Zeitschriftenartikel/Zeitungsartikel enthält einen Metadatenkopf mit Informationen zu Zeitung, Datum, Titel, Untertitel, Autor, Ressort, Presstextsorte, geopolitischer Bezug, Seite und Serie. Diese Informationen wurden den Texten mit Hilfe der Auszeichnungssprache XML hinzugefügt. Mit Ingwer lassen sich die XML-Dokumente komfortabel verwalten, indem die Metadaten als Suchkriterium herangezogen werden können.

¹² Die folgende Beschreibung der einzelnen Komponenten folgt teilweise der von semtracks verfassten Erklärung unter dem Menü-Reiter „Hilfe/Impressum“.

¹³ Vgl. zu Metadaten Scherer (2006: 9).

¹⁴ Bibliographische Grundinformationen wie Zeitung und Datum sollten u.E. mindestens berücksichtigt werden.

Auf dem Screenshot oben (Abb. 2) haben wir die Suche nach den Parametern Zeitung (SPIEGEL), Datum (zwischen dem 01.01.2003 und dem 31.12.2003) und Autor (Gerhard Schröder) eingeschränkt. Die Resultate einer Suche werden ebenfalls auf der „Such-Seite“ in einer Ergebnisliste unterhalb der Datenbankfelder angezeigt. Im obigen Beispiel wurden zwei Artikel gefunden. (Ein Klick auf eine Ergebniszeile öffnet den entsprechenden Artikel.)

Die Ergebnisse einer Suche können als Teilkorpus gespeichert und später wieder geladen werden, um damit weiterzuarbeiten. Über die Suche nach Metadaten hinaus ist es zusätzlich möglich, in den Daten nach lexikalischen Einheiten zu recherchieren, auch mit Hilfe sog. Boolescher Operatoren, etwa nach dem Lexem *Krise* sowie nach allen im Korpus enthaltenen Annotationen (dazu später mehr).

Unter dem Menü-Reiter „Lesen/Bearbeiten“ lassen sich die Artikel eines Suchresultates lesen und bearbeiten. Weiterhin können an dieser Stelle Texte und Metadaten verändert werden. Das ist beispielsweise nützlich, für eine manuelle Nachbearbeitung von Daten, die mit Hilfe von Texterkennungsprogrammen digitalisiert wurden oder, um nachträglich Metainformationen zu ergänzen bzw. falsche Eingaben zu korrigieren.

Wesentlich interessanter für (diskurs-)linguistisches Arbeiten ist die Analysekomponente „Annotation“. Hier können den Daten nach einem frei bestimmbareren Kategoriensystem manuelle Annotationen hinzugefügt werden. Unter einer manuellen Annotation verstehen wir zunächst ganz allgemein das Einflechten von linguistischen Interpretationen zu einem Textsegment, wie es wohl Philologen aus der täglichen Praxis kennen dürften: Textstellen markieren und Markiertes kommentieren (denkbar ist ein Zwischenschritt, indem Markiertes erst klassifiziert wird und dann kommentiert oder erst kommentiert und anschließend werden die Anmerkungen klassifiziert). Dieses Verfahren ergänzend, werden in Ingwer manuelle Annotationen direkt in der Datenbank am digitalen Text vorgenommen und als sogenannter XML-Tag hinterlegt. So ist es möglich, die Annotationen in der Datenbank zu verwalten (über die „Such-Seite“ wiederzufinden) und darüber hinaus in Beziehung zu allen anderen Metainformationen zu setzen (Beispiel: „Suche manuelle Annotation X in Relation zu Zeitung Y in Relation zu Presstextsorte Z“). Das heißt, sie stehen ebenfalls, wie die Metadaten, als Such-Parameter zur Verfügung.

Start Suche Lesen/Bearbeiten Annotation Verwaltung Hilfe/Impressum

Annotation

<< < 2 / 2 > >> mit Part of Speech (POS)

id: 1848
 Datum: 1973-12-17
 Untertitel: Ölkrise und Konjunkturabschwung treiben die Arbeitslosenzahl
 -- nachlassende Kauf- und Investitionslust, gleichzeitig Produktionsbesch
 Ressort: Titel
 Geopolitischer Bezug: BRD
 Serie:
 lemma_count: 3512
 nur annotierten Text anzeigen

Arbeitslose: So knüppeldick war's noch nie

Ölkrise und Konjunkturabschwung treiben die Arbeitslosenzahlen hoch
 Investitionslust, gleichzeitig Produktionsbeschränkung wegen Öl Mangels

Die Experten waren einer Meinung. "Schwieriger denn je zuvor" und
 westdeutschen Arbeitsmarkt 1974.

Ihr "Alternativen-Tableau" enthält Voraussagen, die -- unter verschiedenen

Abwärtsgang der Konjunktur und Ölkrise bescherten den Wirtschafts

entlasten, in der Bundesrepublik sei die Zahl der Arbeitslosen gering wie in kaum einem anderen Land, so droht jetzt der westdeut

Dort grassiert seit Jahren schon die Stagnation: Arbeitslosen-Quoten zwischen zwei und fünf Prozent bei Preissteigerungsraten von

Auch der Kanzler möchte seine alte Arbeitsplatz-Garantie nicht mehr in reiner Form aufrechterhalten: Vor dem Bundestag ließ

Arbeitsplatz, sondern nur noch auf "einen"

Helmut Schmidt verbreitete im Kreis seiner Beamten offen seine Befürchtung, im kommenden Jahr könnte der Bund mehr als je

Wirtschaft Arbeitslosigkeit zu bekämpfen

Bearbeiter:

- Frameanalyse
- + Metapheranalyse
- Toposanalyse
- Topoi
- düstere Zukunft/Prognose
- stützende Elemente
- Autoritäten
- Ländervergleich

+ Kategorie hinzufügen

Abb. 3: „Annotation“

Um Textstellen zu annotieren, müssen zuallererst Kategorien festgelegt werden, die sich in der rechten Spalte (s. Abb. 3) frei erstellen lassen.¹⁵ Mit diesen Kategorien können anschließend im Text beliebige Passagen ausgezeichnet werden. Um eine Kategorie anzulegen, muss „Kategorie hinzufügen“ ausgewählt werden. Es erscheint ein Fenster, in dem sich die Bezeichnung der Kategorie festlegen lässt (z.B. „düstere Zukunft/Prognose“), es kann eine Bemerkung dazu hinterlegt werden, die Position in der Hierarchie der Kategorien muss bestimmt werden (im Beispiel oben ist die Kategorie „düstere Zukunft/Prognose“ Kindelement von „Topoi“. „Topoi“ ist wiederum Kindelement von „Toposanalyse“) und die Farbe kann festgelegt werden (grün). Um eine bestimmte Textpassage zu annotieren, muss sie mit dem Mauszeiger markiert werden, anschließend klickt man auf die Kategorie im Kategorienbaum, mit der die Passage im Text markiert werden soll. Mit den Häkchen neben den Kategorien in der Kategorienliste rechts kann man steuern, welche Kategorien im Text

¹⁵ Die Definition von Kategorien sollte induktiv während des Lesens geschehen, denn aus unserer Sicht ist es problematisch mit im Vorfeld festgelegten Kategorien zu arbeiten, wenn man Neues entdecken und nicht Altes prüfen möchte. Selbstverständlich kann es auch das Ziel einer Analyse sein, Hypothesen zu verifizieren, zu falsifizieren oder zu modifizieren (vgl. u.a. Busse/Teubert (1994) und Hermanns (2007)).

angezeigt werden sollen. Die Textpassagen, die mit dieser Kategorie ausgezeichnet sind, werden im Text mit der zuvor festgelegten Farbe hervorgehoben. Zudem zeigt ein Balken am Seitenrand die Stelle der Annotation im Text an. Mit einem Klick auf den Balken kann zur im Text gewählten Annotation eine Bemerkung hinzugefügt werden.

In Abb. 3 wurde ein Topos der düsteren Zukunft/Prognose annotiert (grün gekennzeichnete Textstelle). Innerhalb dieses Argumentationstopos wurden zusätzlich auf einer Hierarchieebene darunter als die Argumentation stützende Elemente ein Autoritätstopos und ein Ländervergleichs-Topos annotiert (hell- und dunkelgelbe Markierung im Text). Wie die Beispiel-Annotation zeigt, können sich Annotationen im Text überschneiden und beliebige Textmengen umfassen, d.h. ein oder mehrere Wörter, ein Satz, ein Textabschnitt oder ein ganzer Text kann farblich markiert und einer Kategorie zugewiesen werden. Grundsätzlich kann eine Kategorie Element von beliebig vielen anderen Elementen sein. Hierarchisierungen und Mehrfachzuordnungen sind in freier Anordnung (auch im Nachhinein) möglich.

Schließlich lässt sich im Feld „Bearbeiter“ rechts oben ein Name eintragen, der automatisch allen hinzugefügten Annotationen zugewiesen bzw. in das Eigenschaftfenster der Annotation übertragen wird. Dadurch wird kollaboratives Arbeiten organisierbar.

Auf der Seite „Verwaltung“ lassen sich einzelne Texte, ganze Korpora oder kleinere Teilkorpora importieren und exportieren.

Abb. 4: „Verwaltung“

Mit Ingwer ist es möglich, Texte in unterschiedlichen Formaten, z.B. ein strukturiertes Textformat¹⁶ oder ein XML-Format, in die Datenbank zu importieren. Mit dem Import der Texte in Ingwer werden die Wörter mithilfe des TreeTaggers¹⁷ automatisch lemmatisiert und mit Wortart-Informationen versehen (Part-of-speech-tagging¹⁸). Das ist ein wichtiger Arbeitsprozess, um jene quantitativen Analysen durchführen zu können, die wir im zweiten Teil dieser Handreichung exemplarisch vorstellen. Eine Statusmeldung gibt Aufschluss darüber, ob der Import erfolgreich war bzw. welche Datensätze aufgrund welcher Fehler nicht erfolgreich eingelesen werden konnten.

(Teil-)Korpora lassen sich mitsamt Metadaten und allen Annotationen in verschiedene Formate (XML, Corpus Workbench, Lexico 3, CSV) exportieren, um mit ihnen (in anderen Systemen oder mit anderer Software) weiterzuarbeiten. Für den vorliegenden Beitrag haben wir den „Workflow“ durchlaufen von der Erstellung eines Korpus in Ingwer hin zur quantifizierenden Untersuchung dieser Korpusdaten in der Open Corpus Workbench. Das heißt, es wurden alle Texte, die das KriKo bereithält, über die Exportfunktion in Ingwer automatisch in das korpuslinguistische Analysetool überführt. Denkbar wäre auch eine Weiterverarbeitung und Auswertung der Daten in Excel. Hierfür steht das CSV-Format zur Verfügung. Im nun folgenden Abschnitt möchten wir aber an einigen Beispielen einen Eindruck davon vermitteln, wie sich in den Daten quantitativ recherchieren lässt bzw. welche Analysemöglichkeiten die Open Corpus Workbench, insbesondere das Web-Interface CQP-Web, im Einzelnen bietet.¹⁹

3 Open Corpus Workbench

3.1 CQP

Als grafisches Interface zur Abfrage der in der Workbench hinterlegten Daten wird CQPweb verwendet. Neben diesem steht auch der Zugriff über eine Befehlszeile zur Verfügung. Die Befehlszeile bietet zurzeit zwar einen größeren Funktionsumfang, die Bedienung gestaltet sich aber schwieriger.²⁰ Im Sinne einer komfortablen, d.h. menügesteuerten Recherche wird

¹⁶ Hierbei handelt es sich um eine projektspezifische Aufbereitung der Texte. Die nach einem bestimmten Muster als txt-Dateien aufbereiteten Daten werden von der Datenbank in ihrer Struktur erkannt und können deshalb automatisch eingelesen werden.

¹⁷ <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>, zuletzt abgerufen am 15.03.2013.

¹⁸ In Ingwer werden Wortarten nach dem Stuttgart/Tübinger Tagsets (STTS) zugeordnet. Dementsprechend sind die jeweiligen Abkürzungen der im Internet verfügbaren Tabelle (<http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html>, zuletzt abgerufen am 15.03.2013) zu entnehmen.

¹⁹ Für weitere Informationen über den Export zur Open Corpus Workbench/CQPweb und zu Analysemöglichkeiten verweisen wir an dieser Stelle nochmals auf den Menü-Reiter „Hilfe/Impressum“ in Ingwer sowie auf das entsprechende Kapitel in der bereits genannten Online-Einführung von Noah Bubenhofer. (Eine Demoversion von Ingwer ist unter folgender Adresse zu finden: <http://www.semtracks.com/web/index.php?id=1&id2=3&level=1>, zuletzt abgerufen am 15.03.2013.)

²⁰ Das Interface CQPweb hat neben dem eingeschränkten Funktionsumfang außerdem noch einige Fehler, die z.B. beim Import der Korpora mit Metadaten und Annotationen entstehen. Aus diesem Grund haben wir uns dazu entschieden in Absprache mit Noah Bubenhofer neben dem zugrunde gelegten KriKo zur Darstellung weiterer Suchfunktionalitäten, die auf dem aktuellen Stand mit diesem Korpus nicht möglich waren, am Text-Berg-Korpus (<http://www.textberg.ch/>, zuletzt abgerufen am 15.03.2013) zu zeigen. Die jeweiligen Abweichungen von KriKo werden mit der Anmerkung „Text-Berg-Korpus“ angegeben und sind in den Screenshots auch an der Farbe zu erkennen: braun=KriKo und blau=Text-Berg-Korpus. Es ist generell zu empfehlen, die Installation von Ingwer, der CWB und der Korpora von Server-Erfahrenen durchführen zu lassen und sich als programmier-

hier deshalb die Suche über das menügesteuerte Interface CQPweb vorgestellt. Die Abfragesprache nennt sich CQP (Corpus Query Processor) und ist teilweise mit dem Interface kompatibel. (Auch natürlichsprachige Suchterme können verwendet werden.) Das Suchfenster gliedert sich ähnlich dem vieler WWW-Suchmaschinen und Webseiten in einen linken Navigationsbereich und einen Inhalts- bzw. Arbeitsbereich.

3.2 Suchanfragen und Suchsyntax

Es gibt zwei verschiedene Abfragemasken, die über die linke Navigation zu erreichen sind: Die Standardabfrage (Standard query) und die eingeschränkte Abfrage (Restricted query), die im Vergleich zur Standardabfrage die Möglichkeit, Metadaten (text-type-restrictions wie Zeitung, Datum oder Textsorte) des abzufragenden Korpus einzubinden, erlaubt.

3.2.1 Standardabfrage (Standard query)

Die Standardabfrage erlaubt die Spezifizierung einer einfachen Abfrage (simple query) durch Berücksichtigung der Groß- und Kleinschreibung des Suchterms und der darauf bezogenen Ergebnismenge (case-sensitive oder ignore-sensitive) – dies entspricht der üblichen schnellen Suche bei Suchmaschinen und arbeitet mit natürlichsprachigen Suchtermen. Außerdem ist über das Dropdown-Menü Restriction eine Einschränkung auf in der CWB gebildete Teilmengen eines Korpus (Teilkorpora) zu machen. Als natürlichsprachigen Suchterm geben wir z.B. *kanzler* mit der Einstellung simple query (ignore case) ein und erhalten alle Dokumente, in denen *kanzler* groß oder klein geschrieben wird. Die gleiche Suche mit der Einstellung simple query (case-sensitive) ergibt nur Dokumente, in denen *kanzler* klein geschrieben wird. Kommen wir aber zur Suche mit der CQP-Syntax²¹, bevor auf die Darstellung und Weiterverarbeitung der Ergebnislisten eingegangen wird. Die zu suchenden Wörter (Lexeme oder syntaktische Wörter), man könnte diese auch allgemeiner als Zeichenketten zusammenfassen, werden in eckige Klammern und Anführungszeichen, sogenannte Metazeichen, geschrieben.²² Sucht man nach dem Lexem *Kanzler*, sollte der einfache Suchterm folgendermaßen aussehen:

```
[lemma="kanzler"]
```

Der Suchterm, bzw. der reguläre Ausdruck setzt sich aus verschiedenen Metazeichen (z.B. die eckigen Klammern), einem Attribut (lemma) und dem in Anführungszeichen eingefassten Attributwert (*kanzler*) zusammen. Ein Regulärer Ausdruck ist damit eine bestimmte Zeichenkette, ein Muster, nach dem wir suchen²³, welches einfacher oder komplexerer Art sein kann.

unerfahrener Linguist auf die Analyse zu konzentrieren. Dies war auch eine Motivation diese Handreichung zu schreiben und die potentielle Handhabung solcher Systeme in der Praxis für (diskurs- und korpus)linguistische Zwecke darzustellen.

²¹ Vgl. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPTutorial/cqp-tutorial.pdf> oder http://cwb.sourceforge.net/files/CQP_Tutorial.pdf (zuletzt abgerufen am 15.03.2013). Die Dokumentationen sind leider nur mit der Abfrage über die Befehlszeile vollständig kompatibel, weshalb hier bloß die wichtigsten Befehle verwendet werden, die über CQPweb funktionieren.

²² Die CQP-Syntax arbeitet Zeichensensitiv, demnach sollte auf Groß- und Kleinschreibung geachtet werden.

²³ Vgl. zu regulären Ausdrücken: The IEEE and The Open Group (2004): The Open Group Base Specifications Issue 6, IEEE Std 1003.1

Dementsprechend wird für die Suche nach einem syntaktischen Wort das Attribut *lemma* ersetzt durch *word*²⁴. Der entsprechende einfache reguläre Ausdruck ist dann:

```
[word="Kanzler"]
```

Um die Suche weiter zu unterstützen, bietet CQP eine Reihe von Zeichen, die entweder als Platzhalter (Wildcards) oder zur Bestimmung der Vorkommenshäufigkeit von Zeichen und Zeichengruppen (Quantifizierung) fungieren. Ist man auf der Suche nach Dokumenten, in denen der ehemalige bayrische Kultusminister Hans Maier erwähnt wird, man aber nicht weiß, ob dieser mit *ei* oder *ai* geschrieben wird, ist die Setzung eines Platzhalters (Wildcard) für einen Buchstaben von Nutzen:

```
[lemma="M.ier"]
```

Wir suchen hier nach allen flektierten Formen des Attributwertes *M.ier*, wobei der Punkt als Platzhalter dient und ein beliebiges Zeichen ersetzt. Weiter einschränken kann man die Suche, indem man noch den Vornamen als syntaktische Form angibt:

```
[word="Hans"][lemma="M.ier"]
```

Wir sprechen dann schon von einem komplexen regulären Ausdruck. Die logische Verknüpfung zwischen diesen beiden Suchausdrücken entspricht voreingestellt dem Booleschen UND. Daneben ist das Boolesche ODER, welches mit dem senkrechten Strich angegeben wird, zu setzen, wenn eine Auswahl zwischen zwei oder mehr Ausdrücken angegeben werden soll:

```
[word="Hans"]|[lemma="M.ier"]
```

Hier werden alle Texte gefunden, in denen das syntaktische Wort *Hans* oder das Lexem *Herr* jeweils gefolgt vom Lexem *M.ier* vorkommt.

```
([word="Hans"]|[lemma="Herr"])[lemma="M.ier"]
```

Das Verwenden der runden Klammern entspricht der logischen Zuordnung einer Menge von einfachen regulären Ausdrücken zueinander und setzt diese in Relation zu anderen, die außerhalb der Klammer stehen, und ist mit der Verwendungsweise u.a. in der Mathematik zu vergleichen. Neben dem Gebrauch von Wildcards als Platzhalter für Zeichen in einem Attributwert ist auch das Weglassen von Attribut und Attributwert zwischen den eckigen Klammern als nicht weiter bestimmter Suchausdruck mit einem bestimmten oder weiteren unbestimmten Ausdrücken kombinierbar:

```
[ ][lemma="M.ier"]
```

```
[ ][ ][lemma="M.ier"]
```

```
[ ][ ][ ]
```

(http://pubs.opengroup.org/onlinepubs/009695399/basedefs/xbd_chap09.html#tag_09_03_05, zuletzt abgerufen am 15.03.2013).

²⁴ Um zur Abfragemaske zurückzukehren, muss rechts oben New query ausgewählt und mit Go bestätigt werden.

Im letzten Fall sprechen wir dann schon von Trigrammen als Kombination dreier verschiedener Zeichenketten. Zweckmäßig könnte auch ein Trigramm mit einem bestimmten Startausdruck sein:

```
[lemma="Herr"][][]
```

Diese Suche nach Trigrammen oder allgemeiner n-Grammen scheint auf den ersten Blick eine wahllos zusammengestellte begrenzte Menge an Zeichenketten zu ergeben (s. Beispiel unten), kann aber über weitere Befehle spezifiziert werden und wird bei der Darstellung und Weiterverarbeitung der Ergebnislisten noch interessant werden.

	WIRTSCHAFT 20 Prozent	Lohnzuschlag für 1974 ?
WIRTSCHAFT	20 Prozent Lohnzuschlag	für 1974 ? Bonns Wirtsch
WIRTSCHAFT 20	Prozent Lohnzuschlag für	1974 ? Bonns Wirtschaf
WIRTSCHAFT 20 Prozent	Lohnzuschlag für 1974	? Bonns Wirtschaftslenk
TSCHAFT 20 Prozent Lohnzuschlag	für 1974 ?	Bonns Wirtschaftslenker
CHAFT 20 Prozent Lohnzuschlag für	1974 ? Bonns	Wirtschaftslenker hoffen
T 20 Prozent Lohnzuschlag für 1974	? Bonns Wirtschaftslenker	hoffen auf ein rasches Ab
20 Prozent Lohnzuschlag für 1974 ?	Bonns Wirtschaftslenker hoffen	auf ein rasches Abflauen
ozent Lohnzuschlag für 1974 ? Bonns	Wirtschaftslenker hoffen auf	ein rasches Abflauen der
g für 1974 ? Bonns Wirtschaftslenker	hoffen auf ein	rasches Abflauen der Kc
974 ? Bonns Wirtschaftslenker hoffen	auf ein rasches	Abflauen der Konjunktur
? Bonns Wirtschaftslenker hoffen auf	ein rasches Abflauen	der Konjunktur . Nur da

Abb. 5: Suchanfrage „[][][]“

Die Suche kann je nach vorheriger Aufbereitung des Korpus weiter spezifiziert werden. Wie bereits erwähnt, ist für den vorliegenden Aufsatz das Korpus zuvor in Ingwer aufbereitet worden und die einzelnen Dokumente mit verschiedenen inhaltlich und kontextuell relevanten Metainformationen und Annotationen versehen worden. Eine der Vorteile der Aufbereitung in Ingwer ist z.B., dass über das Part-of-speech-tagging eine Konkretisierung des Suchausdrucks auf Wortarten durch das Attribut `pos` ermöglicht wird. Wir suchen nach Adjektiven, die Hans Maier näher beschreiben. Der komplexe reguläre Ausdruck setzt sich aus der Anweisung, ein Adjektiv `[pos="ADJA"]`, welchem das Lexem *Maier* `[lemma="Maier"]` folgt, zu suchen, zusammen:

```
[pos="ADJA"][lemma="Maier"]
```

Wird nach Lexem *Maier* gesucht, dem kein Adjektiv vorangeht, kann das Attribut `pos` mit dem Ausrufezeichen negiert werden:

```
[pos!="ADJA"][lemma="Maier"]
```

Wissen wir nicht, ob sich noch irgendein Wort zwischen dem Adjektiv und *Maier* befindet, setzen wir eckige Klammern und geben an, dass das Vorkommen nur fakultativ ist. Dies machen wir mit dem Fragezeichen ?:

[pos="ADJA"]?[lemma="Maier"]

Für die Angabe (der Häufigkeit) des Vorkommens oder Nicht-Vorkommens, der Quantifizierung eines einfachen oder komplexen Ausdrucks oder einer Wildcard, können nachgestellt das Fragezeichen (? der voranstehende Ausdruck kommt einmal oder nicht vor), der Asterisk (* der voranstehende Ausdruck kommt beliebig oft oder nicht vor) und das Pluszeichen (+ der voranstehende Ausdruck kommt einmal oder beliebig oft vor) als Quantoren verwendet werden. Die Suche nach [pos="ADJA"]?[lemma="Kanzler"] ergibt alle Belege, in denen das Lexem *Kanzler* mit oder ohne einem vorangestellten Adjektiv vorkommt:

msch-Telegramm . Der boshafte Claqueur , der den	Kanzler	im Genossenkreis " Hindent
h in der Presse sinkt das Brandt-Renommee . " Des	Kanzlers	Sterne leuchten blasser " , s
e und fragt bekommen : " Wäre Brandt wirklich der	richtige Kanzler	für eine richtige Krise ? " De
" sei " entrückt " und " abgeschlafft " ; " der	Kanzler	badet gern lau -- so in einer
ich Wehner " beim Gute-Nacht-Sagen " von seinem	sprachlosen Kanzler	mit der Bitte , " es noch einn
als der Samstagabend-Entertainer Rudi Carrell den	Kanzler	als komische Figur präsentie
ragen auf den politischen Olymp , muß der nunmehr	60jährige Kanzler	fürchten , unversehens vom
von der SPD zur FDP " . In der Krise macht der	Kanzler	Ferien . So nimmt es kaum

Abb. 6: Suchanfrage „[pos="ADJA"]?[lemma="Kanzler"]“

Die Suche nach [pos="ADJA"]+[lemma="Kanzler"] zeigt:

er Bürgermeister zu Zeiten des Mauerbaus , als	Bonner Kanzler	ohne Mehrheit - , sich in Bonn
führungsmacht geschwächt , aber noch immer	attraktiven Kanzler	als " Vehikel " (Altsozialist H
ugt , um zu meinen , er könne der Republik ein	besserer Kanzler	sein . Aber zugleich weiß er ,
bstimmung des Machtwechsels provozierte der	neue Kanzler	die Unions-Opposition mit de
h die wilden Streiks ausbreiteten , fürchtete der	aufgebrachte Kanzler	, die detaillierten TV-Berichte
kann keiner genau sagen . Rudolf Augstein Der	führungsschwache Kanzler	Seinen 60. Geburtstag erlebt
aßhalteappelle der Regierung verwandelten die	stärksten Verbündeten Kanzler	Brandts aus dem Wahljahr 19
och verschärft werden . (Siehe auch Seite 4 .)	Ernster Kanzler	Bar . In der Bundesrepublik d
nländern bezeichnet hat , meinte Kreisky . Der	österreichische Kanzler	verhehlte jedoch nicht seine S
a rasch hinter die FDP zurückfallen würde . Ein	starker Kanzler	wurde ohnedies nur nach der
ihnen wohl gut , dem der Entscheidungs-Scheu	verdächtigen Kanzler	die Chance eines donnernden
Objekt der Begierde des Marktführers , als der	damalige Kanzler	Schröder sie der Deutschen E

Abb. 7: Suchanfrage „[pos=“ADJA“]+[lemma=“Kanzler“]“

Die Quantoren können auch in Verbindung mit einer voranstehenden Wildcard im Wort oder am Ende bzw. Anfang als Trunkierung verwendet werden, wodurch die Wildcard auf ein oder mehrere aufeinanderfolgende beliebige Zeichen funktional erweitert wird:

[lemma=“Maie.*“]

m rund 200 Prozent . Diese Kostenlawine , so überlegt Herr	Maier	, trifft ihn sicherlich nicht nur als privaten Verbra
osmetika . Dabei wird es jedoch nicht bleiben , überlegt Herr	Maier	. Statt Öl werden jetzt andere Güter (Energietr
e werden noch deutlicher: Metall-NRW-Chef Horst-Werner	Maier-Hunke	findet die Lohnforderung schon jetzt „schwer na
beitgeberverbandes in Nordrhein-Westfalen, Horst-Werner	Maier-Hunke,	sagte, mit einem Angebot sei zwar noch im Lau
ngfristig Arbeitsplätze und Wettbewerbsfähigkeit zu sichern.	Maier-Hunke	kündigte an, der Gewerkschaft erneut einen Kr
ie Kreditinstitute das Geld lieber woanders anlegen. Thomas	Maier,	Volkswirt der Deutschen Bank, hält diesen We
ommen Beileidsbekundungen . Kein Geringerer als Gerhard	Maier	, der frühere Landesbischof der evangelischen I

Abb. 8: Suchanfrage „[lemma=“Maie.*“]“

Interessant erscheint bei dieser Möglichkeit, Platzhalter zu erweitern, die Suche nach Ismen, denen evtl. noch ein Adjektiv vorangeht:

[pos=“ADJA“]?[lemma=“.+ismus“]

in dem Heizölpreis von 25 Mark je Tonne .	SOZIALISMUS	Kleiner Wurf Der Nationalöko
mit der Aufhebung der Widersprüche des	Kapitalismus	" auch alle sozialen Antagonism
Widersprüche des Kapitalismus " auch alle	sozialen Antagonismen	und sozialen Zwänge beseitigt :
die neueren Theoretiker der Analyse des	Kapitalismus	" so selten ein ökonomisches R
luste gebracht , als unter Bedingungen des	Marktmechanismus	überhaupt je entstehen . " Die
zu den unschamhaften Begriffen , die mit "	Sozialismus	" diffamiert worden sind . Jetzt
zu entlarven , was er ist : ein Tarnwort für	Sozialismus	. " (CSU-MdB Friedrich Zimm
inzustellen haben . " Die machen zwar in	Optimismus	" , umriß ein Bonner Staatssek

Abb. 9: Suchanfrage „[pos=“ADJA“]?[lemma=“+.ismus“]“

Ebenso können damit auch Komposita zu einem Attributwert gefunden werden:

[lemma=“Krise.+“]

In der Ergebnisliste finden wir Dokumente mit Komposita *Krisenplan*, *Krisenmanagement*, *Krisensituation*, *Krisenplan*, etc. Weitere bestimmte Quantoren werden in geschweiften Klammern notiert. Dadurch lässt sich die Anzahl der in einer Zeichenkette nacheinander wiederholt auftretenden Zeichengruppen definieren. Bei der Eingabe {n}, {min,}, {min,max} oder {0,max} wird nach Zeichenketten gesucht, in denen der voranstehende Ausdruck n-mal, mindestens n-mal, mindestens n-mal und maximal n-mal oder n- bis max-mal vorkommt. Für n, min und max müssen numerische Werte angegeben werden. Suchanfragen könnten damit so aussehen:

[pos=“ADJA“]{2} entspricht {n}

[pos=“ADJA“]{2,} entspricht {min,}

[pos=“ADJA“]{2,3} entspricht {min,max}

[pos=“ADJA“]{0,3} entspricht {0,max}

Diese Suche ergibt Adjektivanhäufungen bestimmter oder unbestimmter Menge. Sie könnte deshalb sinnvoll kombiniert werden mit einem weiteren Attribut pos und dem passenden Attributwert, mit einem bestimmten Lexem oder syntaktischen Wort oder leeren eckigen Klammern als Platzhalter wie in diesem Beispiel:

[pos=“ADJA“]{2,}[]

Die Suche nach einem bestimmten Lexem oder syntaktischen Wörtern kann zusätzlich mit dem Attribut pos, welches sich direkt auf das Lexem oder das Wort bezieht, eingeschränkt werden:

[lemma=“Maier“ & pos=“NE“]

Der Ausdruck besteht aus den schon bekannten Elementen für das Lemma *Maier*, der Anweisung nach Wortarten (Eigennamen) zu suchen und der Verknüpfung &. Alle Elemente

werden in eckige Klammern gefasst. Das Gegenteil erreicht man, wenn man das Attribut pos mit dem Ausrufezeichen negiert. Hierfür ein geeigneteres Beispiel:

```
[lemma=".+ist.+“ & pos!="ADJA“]
```

Oder:

```
[lemma=".+ist.+“ & pos!="NN“]
```

Auf der Basis dieser Ausdrücke sind mit Setzen von runden Klammern, wie zuvor schon gezeigt, auch komplexere Suchanfragen und Ausdrücke realisierbar. Neben n-Grammen sind auch grammatische Strukturen, beispielsweise Phrasen, in einer Suche inhaltlich-semantisch relativ undeterminiert, obschon man sich bewusst sein sollte, dass die Suche nach bestimmten Strukturen und Konstruktionen einem strikten induktivem Vorgehen nicht entspricht. Als Beispiel nehmen wir die Suche nach Präpositionalphrasen. Die passende Suchanfrage würde lauten: Eine Präposition möglicherweise gefolgt von einem Artikel, gefolgt von keinem oder beliebig vielen Adjektiven. Jedem Adjektiv ist möglicherweise ein Adverb vorangestellt. Darauf folgt ein Substantiv. Die runden Klammern dienen dabei der Zuordnung des Fragezeichens und des Asterisken zu bestimmten Ausdrücken:

```
[pos="APPR“]([pos="ART“]?([pos="ADV“]?[pos="ADJA“])*[pos="NN“]
```

nns Wirtschaftslenker hoffen	<u>auf ein rasches Abflauen</u>	der Konjunktur .
konjunktur . Nur dann nämlich ,	<u>bei den ersten Anzeichen</u>	einer Arbeitslosigkeit
Gewerkschaften bereit , sich	<u>mit vergleichsweise bescheidenen Lohnabschlüssen</u>	zufriedenzugeben
n . Die Standardforderungen	<u>in den Betrieben</u>	lauten 15 bis 20 Prozent
bis 20 Prozent Lohnzuschlag	<u>für 1974. LÖHNE</u>	Reine Nerventöterei
st und Metallindustrie stehen	<u>vor der härtesten Lohnrunde</u>	der letzten Jahre .
Wirtschafts-Elite der Nation	<u>über Preise</u>	und Produktion ,
ion versammelt hatten , einig	<u>über das konjunkturpolitische Gebot</u>	der Stunde : nicht
bot der Stunde : nichts tun . "	<u>In unerhörter Freundlichkeit</u>	" (ein Konferenz
eine Welle spontaner Streiks	<u>über die Republik</u>	geschwappt . so .

Abb. 10: Suche nach Präpositionalphrasen

Ebenso bleibt auch die Suche nach grammatischen Oberflächenstrukturen mit Hilfe von regulären Ausdrücken, die durch die Verwendung des Attributes pos zusammengestellt werden, inhaltlich-semantisch unspezifisch. Syntaktische Strukturen wie z.B. eine Genetivkonstruktion in der Form $[pos="NN"][pos="ART"][pos="NN"]$ ²⁵ können jedoch Aufschluss über semantische Signifikanzen geben, wie die unten stehende Ergebnisliste erahnen lässt:

²⁵ Natürlich ergibt diese Suchanfrage auch Zeichenketten, die Genetivkonstruktionen nicht entsprechen. Die Anfrage bleibt in dieser Form aber offener.

lenker hoffen auf ein rasches	<u>Abflauen der Konjunktur</u>	. Nur dann nämlich , bei den
kann nämlich , bei den ersten	<u>Anzeichen einer Arbeitslosigkeit</u>	, seien die Gewerkschaften t
vielen Betrieben fordern die	<u>Arbeitnehmer einen Zuschlag</u>	von 20 Prozent . Knapp sie
eben Stunden debattierte die	<u>Wirtschafts-Elite der Nation</u>	über Preise und Produktion
ber das konjunkturpolitische	<u>Gebot der Stunde</u>	: nichts tun . " In unerhörter I
. " Und Hans-Günther Sohl "	<u>Präsident des Bundesverbandes</u>	der Deutschen Industrie . urt
te Stabilitätsprogramm in der	<u>Geschichte der Bundesrepublik</u>	verkündet . ein dreiviertel Ja
noch immer darauf , daß die	<u>Konjunktur einen Knacks</u>	bekommt und die Preissteige
mehr so stark wie im gleichen	<u>Monat des Vorjahres</u>	. Und eine " skeptischere Be
res . Und eine " skeptischere	<u>Beurteilung der Geschäftsentwicklung</u>	durch die Unternehmer " . sc
en Stabilitätsprogramms . die	<u>Dämpfung der Investitionsneigung</u>	" zumindest in der Industrie e

Abb. 11: Suchanfrage „[,pos=“NN“][pos=“ART“][pos=“NN“]“

Als Variante folgt ein Beispiel für „Bestimmungssätze“²⁶, in denen einer Referenz eine Prädikation zugeordnet wird:

[pos=“ART“][lemma=“krise.*“][pos=“VAFIN“]

[pos=“ART“][lemma=“krise.*“][pos=“VAFIN“][

[pos=“ART“][lemma=“krise.*“][pos=“VAFIN“][+]

Oder:

[pos=“ART“][lemma=“krise.*“][pos=“VAFIN“][pos=“ADJD“]

Dies funktioniert auch bei der Zuordnung von Eigennamen und deren Prädikation:

[pos=“NE“][pos=“VAFIN“][pos=“ADJD“]

²⁶ Unter Bestimmungssätzen verstehen wir die Verbindung eines Wortes in Subjektposition mit einem Kopula-verb und einem Nomen oder Adjektiv. Semantisch-pragmatisch beschreibt dies die Beziehung von Referenz und Prädikation.

n .	Bonn war ahnungslos	. D
s ,	Frank habe unmittelbar	na
nd	Ägypten waren offenbar	vo
rt ,	Saudi-Arabien sei moralisch	ver
In	Frankreich wird fröhlich	we
."	Willy ist echt	sa
Der	Willy ist echt	sa
in	Rußland sind höher	. D

Abb. 12: Suchanfrage „[pos=“NE“][pos=“VAFIN“][pos=“ADJD“]“

Zuletzt ist noch auf die Suche nach und in Annotationen, die zuvor z.B. in Ingwer eingegeben wurden und nun in den XML-Strukturen verfügbar sind, einzugehen. Wir gehen davon aus, dass wir beispielsweise nach Topoi (Argumentationsmuster)²⁷ suchen. Den entsprechenden Textteilen haben wir zuvor in Ingwer die Kategorie „Topos“ zugeteilt. Nun sind wir auf der Suche nach diesen Topoi respektive nach musterhaftem Wortschatz oder Mustern innerhalb der Textteile, die wir mit der Kategorie „Topos“ annotiert haben. Suchanweisungen dieser Art sehen so aus:

```
<annotation_categoryname></annotation_categoryname>
```

Die Suchanfrage besteht aus Anweisungen, die in Pfeilen notiert werden. Es gibt einen öffnenden Teil (<annotation_categoryname>) und einen schließenden Teil (</annotation_categoryname>). Innerhalb dieser Teile kann ein regulärer Ausdruck eingesetzt werden:

```
<annotation_categoryname>[]+</annotation_categoryname>
```

Diese Suchanweisung entspricht einer Suche nach allen Annotationen, die in den XML-Strukturen hinterlegt sind, und mit mindestens einer zusammenhängenden Zeichengruppe ([]+) verknüpft sind.

Suchen wir nach einer bestimmten Annotation, z.B. „Topos“, weisen wir diese als Wert dem öffnenden Teil zu:

```
<annotation_categoryname =“Topos“>[]+</annotation_categoryname>
```

Gefunden werden damit alle Zeichengruppen, die mit der Kategorie „Topos“ annotiert sind. Möchte man jedoch ein Wort innerhalb des Topos suchen entfallen Pfeile und Klammern:

```
“Krise“ within annotation
```

```
“Krise“ :: match.annotation_categoryname=“Topos“
```

²⁷ Vgl. zum hier zugrunde gelegten Topos-Begriff u.a. Wengeler 2003.

3.2.2 Die eingeschränkte Abfrage (Restricted query)

Die Suchmaske der eingeschränkten Abfrage ist aufgebaut wie die der Standardabfrage, erlaubt aber im unteren Bereich Spezifizierungen (text-type-restrictions). Die Ergebnismenge kann z.B. auf Dokumente eingegrenzt werden, die bestimmten Metadaten zugeordnet sind.²⁸ Damit ist je nach Aufbereitung des Korpus eine Einschränkung auf Jahr, Textsorte, Sprache, etc. denkbar.

Text+Berg-Korpus Release 145: powered by CQPweb

Restricted Query

Query mode: [Simple query language syntax](#)

Number of hits per page:

Select the text-type restrictions for your query:

g_id	g_level	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> g_1	<input type="checkbox"/> geo	<input type="checkbox"/> 0
<input type="checkbox"/> g_10		<input type="checkbox"/> 10489842
<input type="checkbox"/> g_101		<input type="checkbox"/> 10517096
<input type="checkbox"/> g_1014		<input type="checkbox"/> 17361331
<input type="checkbox"/> g_103		<input type="checkbox"/> 18299284

Abb. 13: Die eingeschränkte Abfrage (Text-Berg-Korpus)

Daneben bleiben außer der Einschränkung auf Teilkorpora, die nicht einstellbar ist, alle Suchfunktionalitäten der Standardsuche erhalten.

3.2.3 Ergebnisdarstellung und Weiterverarbeitung

Zur Darstellung der Ergebnisse und deren Weiterverarbeitung ziehen wir die oben aufgeführten Beispiele hinzu. Wir geben beispielsweise den Ausdruck für ein Trigramm ([[][]]) ein und erhalten als Ergebnis eine Auflistung aller Trigramme, d.h. von jeweils drei zusammen vorkommenden Zeichengruppen. Die nach der Suchanfrage erscheinende Ergebnisliste gliedert sich in einen Kopfbereich. Dieser zeigt oben die Suchanfrage und die Anzahl der Ergebnisse in Relation zu den Texten, in denen die gesuchten Ausdrücke vorkommen. Darunter folgt der Menübereich, der Befehle zur Weiterverarbeitung beinhaltet, und der Inhaltsbereich mit der Anzeige der Ergebnisliste. Diese Liste kann über die Menüleiste wählbar entweder in der Reihenfolge, in der die Dokumente im Korpus (Show in corpus order) vorkommen, oder in willkürlicher Reihenfolge (Show in random order) sortiert

²⁸ Die in den XML-Strukturen der Dokumente hinterlegten Metadaten sind im Menü Corpus info unter View corpus metadata aufgelistet.

werden. Außerdem kann die Ansicht je nach Voreinstellung²⁹ von der Zeilenansicht (Line View) in die Keywords-in-kontext-Ansicht (KWIC View) gewechselt werden. Über eine Pfeil- und eine numerische Navigation kann man sich je nach Ergebnismenge durch die ganze Liste bewegen. Über die Spalte Filename ist der jeweilige Link zu den durchsuchten Dokumenten und deren Metadaten verfügbar. Für die Weiterverarbeitung stellt die Menüleiste ein Dropdown-Menü bereit, welches die Funktionen New Query, Thin, Frequency Breakdown, Distribution, Sort, Collocations, Download, Categorise und Save current Set of hits beinhaltet. Als Grundfunktionen könnte man die Eingabe, eine neue Suchanfrage zu stellen (New Query), das Herunterladen der Suchergebnisse (Download) und das Speichern der aktuellen Ergebnisse in der CWB (Save current Set of hits) bezeichnen. Kommen wir zu den Funktionen zur Weiterverarbeitung.

Thin und Sort – große Ergebnislisten reduzieren und sortieren

Ist die Liste der Suchergebnisse für die Weiterverarbeitung zu groß³⁰, lässt sie sich mit der Funktion Thin reduzieren.

Abb. 14: Ergebnislisten reduzieren

Die Anzahl oder der Anteil der Suchergebnisse, auf die bzw. auf den reduziert wird, kann rein numerisch (9 entspricht neun willkürlich ausgewählten Treffern) oder numerisch in Prozent (9 % entspricht neun Prozent willkürlich ausgewählter Treffer der ganzen Ergebnisliste) angegeben werden. Die Einstellung random (selection is reproducible/not reproducible) gibt an, ob die willkürliche Auswahl wiederholbar sein soll oder nicht. Die Auswahl ist wie angeführt willkürlich. Liegen zu viele Ergebnisse vor, sollte man über eine Reformulierung der Suchanfrage nachdenken, bevor man die Ergebnisliste kürzt und damit evtl. wichtige Treffer ausschließt.

Mit der Funktion Sort kann eine Ergebnisliste sortiert werden. Die Sortierung kann nach der Position eines Kollokators rechts, links oder innerhalb der Keywords in Kontext, nach Annotationen (z.B. ADJA) oder nach einem frei wählbaren Startausdruck vorgenommen werden. Die Einschränkung auf Annotationen und einem frei wählbaren Startausdruck kann aus der Sortierweise auch einzeln ausgeschlossen werden. So ist es beispielsweise möglich, sich eine Sortierung anzeigen zu lassen, in der wir nach dem Artikel *die* sortieren, der sich in den Keywords in Kontext befindet.

²⁹ Voreinstellungen dieser Art sind im User controls unter dem Menüpunkt User settings zu machen.

³⁰ Dies ist von der Art der Suchanfrage, der Größe des Korpus und den Servereinstellungen abhängig. Eine Fehlermeldung wird ausgegeben.

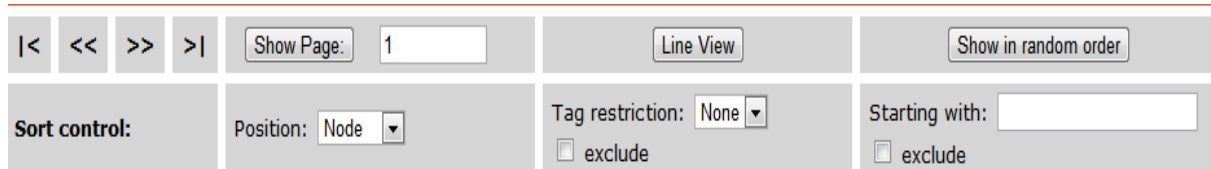


Abb. 15: Ergebnislisten sortieren (Text-Berg-Korpus)

Frequency Breakdown

Mit der Funktion Frequency Breakdown lässt sich die Frequenz der Suchergebnisse innerhalb des Korpus anzeigen. Man erhält z.B. bei der Suche nach Trigrammen einen guten Überblick, welche Trigramme im Korpus hoch- oder niedrigfrequent verwendet werden. Spezifiziert man das Trigramm ([pos="ART"][lemma="krise.*"][pos="VAFIN"], sodass man die oben schon angeführten Zeichenketten in Form von „Bestimmungssätzen“ erhält, zeigt sich diese Verteilung bezogen auf auf das Korpus:

Search result	No. of occurrences	Percent
Die Krise ist	30	23.62%
der Krise ist	24	18.9%
der Krise sind	12	9.45%
die Krise sei	10	7.87%
der Krise war	6	4.72%
der Krise waren	5	3.94%

Abb. 16: Frequenzwerte ermitteln

Die Links unter Search result führen zu den entsprechenden Treffern, die dieses Trigramm enthalten, was eine Sichtung des Kontextes, der Kollokatoren und eine weitere Verarbeitung ermöglicht.

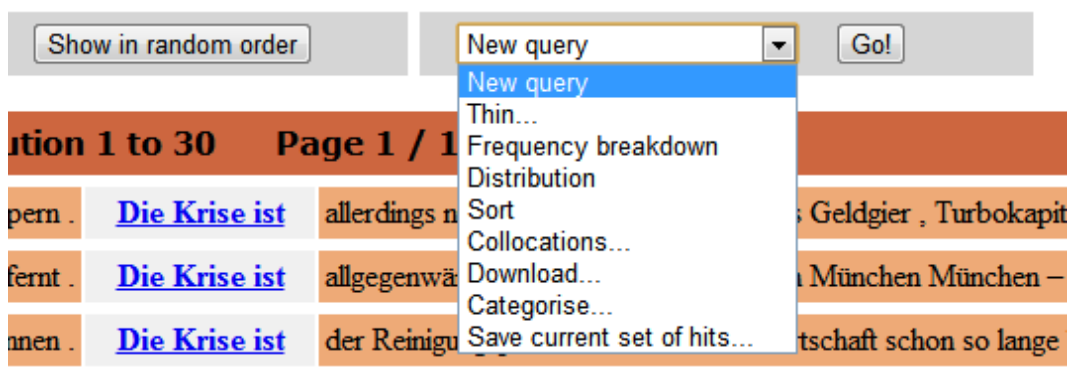


Abb. 17: Auflistung bestimmt-frequenter Zeichenketten

Distribution der Ergebnisse

Die Distribution (Verteilung) der Ergebnisse einer Suchanfrage (hier ein Trigramm) über das durchsuchte Korpus kann mit der Funktion Distribution angezeigt werden. Die Verteilung kann durch die Einstellung unter Categories und Category for crosstabs an den Metadaten (je nach Aufbereitung des Korpus: Sprache, Textsorte, usw.)³¹ ausgerichtet werden.

³¹ Die Einstellung General Information umfasst alle im Korpus als Metadaten hinterlegten Informationen.

Categories:	<input type="text" value="g_id"/>	Show as:	<input type="text" value="Distribution table"/>
Category for crosstabs:	<input type="text" value="g_id"/>		
Based on classification: g_id			
Category [!]	Words in category	Hits in category	Dispersion (no. files with 1+ hits)
g_1	12777	12777	3 out of 3
g_10	290	290	1 out of 1
g_101	4866	4866	1 out of 1
g_1014	864	864	1 out of 1

Abb. 18: Distribution der Ergebnisse (Categories) (Text-Berg-Korpus)

Categories:	<input type="text" value="g_id"/>	Show as:	<input type="text" value="Distribution table"/>
Category for crosstabs:	<input type="text" value="No crosstabs"/>		
Based on classification: g_id			
Category [!]	Words in category	Hits in category	Dispersion (no. files with 1+ hits)
g_1	12777	12777	3 out of 3
g_10	290	290	1 out of 1
g_101	4866	4866	1 out of 1

Abb. 19: Distribution der Ergebnisse (Category for crosstabs) (Text-Berg-Korpus)

Zusätzliche Einstellungen unter Category for crosstabs erlauben sogar eine Verschränkung zwei verschiedener Kategorien (z.B. Sprache und Datum) als Kreuztabelle. Als grafisches Ausgabeformat ist unter Show as eine Verteilungstabelle oder ein Balkendiagramm, welches aber nur eine Kategorie und keine Verschränkung darstellen kann, auswählbar.

Categories:	<input type="text" value="text_year"/>	Show as:	<input type="text" value="Show distribution"/>	
Category for crosstabs:	<input type="text" value="No crosstabs"/>	<input type="text" value="Bar chart"/>	<input type="text" value="New Query"/> <input type="button" value="Go!"/>	
Based on classification: g_id				
Category [!]	Words in category	Hits in category	Dispersion (no. files with 1+ hits)	Frequency [!] per million words in category
g_1	12777	12777	3 out of 3	1000000
g_10	290	290	1 out of 1	1000000
g_101	4866	4866	1 out of 1	1000000
g_1014	864	864	1 out of 1	1000000

Abb. 20: Tabellarische Darstellung der Distribution der Ergebnisse (Text-Berg-Korpus)

Aus der diachronen Perspektive ist die Ausrichtung an der Kategorie Datum interessant. Möchte man dagegen einen intersprachlichen Vergleich bzw. Verlauf skizzieren, ist eine Orientierung an der Kategorie Sprache sinnvoll.

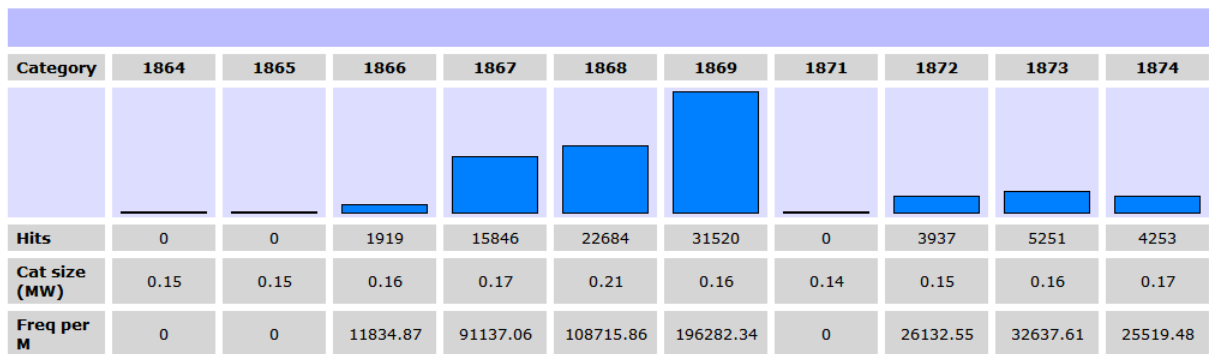


Abb. 21: Distribution der Ergebnisse als Balkendiagramm nach Datum (Text-Berg-Korpus)

Collocations/Kollokationen

Die Kollokationen als „Mitspieler“³² einer zu suchenden Zeichenkette sind über die Funktion Collocations herauszufinden. Die Auswahl dieser Funktion führt durch zwei Menüs, von denen das erste Voreinstellungen wie die Einbeziehung oder den Ausschluss von Wortarteninformationen und der Spezifikation auf Lexeme oder syntaktische Wörter sowie die maximale Spannweite der anzuzeigenden Kollokationen um die gesuchte Zeichenkette (hier *kanzler*) ermöglicht. Das sich anschließende Fenster zeigt neben der unten stehen Ausgabe der Kollokatoren die Informationen Anzahl im gesamten Korpus (Total no. in whole corpus), erwartete Frequenz (Expected collocate frequency), beobachtete Frequenz (Observed collocate frequency), Anzahl der Texte, in denen die Kollokationen mit der gesuchten Zeichenkette gefunden wurden, (In no. of texts) und dem Wert der zugrunde gelegten statistischen Berechnungsweise (hier: Log-likelihood value). Darüber befindet sich das zweite Menü, welches eine Modifikation der Berechnung der Kollokatoren zulässt.

Collocation controls						
Collocation based on:	Word form	Statistic:	Log-likelihood			
Collocation window from:	5 to the Left	Collocation window to:	5 to the Left			
Freq(node, collocate) at least:	5	Freq(collocate) at least:	5			
Filter results by:	specific collocate:	and/or tag:	no restriction	Submit changed parameters	Go!	
There are 6,135 different words in your collocation database for "[word="kanzler"%c]". (Your query "kanzler" returned 2,653 matches in 1051 different texts) [1.809 seconds - retrieved from cache]						
No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood value
1	Der	225,655	926.734	1,958	752	913.701
2	Kohl	2,369	9.729	142	108	504.99
3	dem	30,906	126.927	402	275	382.084
4	Schroder	2,337	9.598	118	89	380.932
5	Schmidt	3,495	14.354	112	89	268.039
6	Helmut	1,920	7.885	82	66	238.923

Abb. 22: Kollokationen

³² Vgl. zu Kollokationen und weiteren z. T. korpuslinguistischen Termini neben der schon angeführten Online-Einführung von Bubenhofer u.a. Lemnitzer/Zinsmeister (2006) und Scherer (2006).

Die Berechnung kann weiter modifiziert werden:

- Unter Collocation based on und Statistic ist eine Spezifizierung der Ausgabe nach Wortarten, Lexemen oder syntaktischen Wörtern sowie der zugrunde gelegten statistischen Berechnungsweise möglich.
- Zusätzlich kann die Mindestfrequenz, in der Kollokationen (Freq(node, collocate) at least) oder Kollokatoren (Freq(collocate) at least) auftreten, eingestellt werden.³³
- Die Einstellungen unter Collocation window from und Collocation window to beschränkt die abzusuchende Spannweite um die gesuchte Zeichenkette.
- Darunter steht eine Reihe an Filtern (Filter results by) (nach bestimmten Kollokatoren oder Wortarten) zur Verfügung.

Categorise

Unter Categorise kann eine Suche mit selbst erstellten Kategorien versehen und gespeichert werden. Diese Kategorien können als Standardwerte (Default category) für weitere Suchen verwendet werden. Die gespeicherten kategorisierten Suchanfragen sind später über das Menü User Control unter dem Menüpunkt Categorised queries verfügbar.

3.3 Weitere Funktionen

Neben den beiden Suchen stehen im gleichen Menü noch die Menüpunkte Word lookup, Frequency lists und Keywords zur Verfügung.

3.3.1 Word lookup

Die Funktion Word lookup erlaubt einen Einblick in das Korpus mit Bezug darauf, wie viele und welche token es zum Suchterm gibt. Dieser Einblick kann mit der Einstellung Show only words starting with/ending with/containing/matching exactly the pattern you specified eingeschränkt oder erweitert werden. Außerdem können natürlichsprachige Suchausdrücke und alle Typen von Wildcards verwendet werden, die, wie oben angeführt, auch in der Suche verwendet werden können.

Abb. 23: Word lookup - Einblick in das Korpus

³³ Diese Einstellungsmöglichkeit entspricht der Sichtweise, dass Relevanz meist mit einer hohen Verwendungsfrequenz korreliert. Dies ist unter semantisch-diskursanalytischer Sichtweise mehr als fraglich.

3.3.2 Frequency lists

Unter Frequency lists kann man sich die Frequenzlisten aller Wörter des Korpus oder der angelegten Teilkorpora³⁴ ansehen. Hierzu steht eine Auswahl von Filtern bereit, die eine Verwendung von Wildcards zum Teil ausschließen. Die Auswahl an Einstellungen, die man hier machen kann, ähnelt dem Menü der Funktion Frequency Breakdown in der Weiterverarbeitung von Ergebnislisten.

kriko: powered by CQPweb

Frequency lists

You can view the frequency lists of the whole corpus and frequency lists for subcorpora you have created. [Click here to create/view subcorpus frequency lists.](#)

View frequency list for ... Whole of krikoweb ▾

View a list based on ... Word forms ▾

Frequency list option settings

Filter the list by *pattern* - show only words/tags ... starting with ▾

Filter the list by *frequency* - show only words/tags ... with frequency between and

Number of items shown per page: 50 ▾

List order: most frequent at top ▾

Abb. 24: Frequenzlisten

3.3.3 Keywords und Teilkorpora

Unter Keywords können quantitativ ermittelte und automatisch erstellte Schlüsselwortlisten durch Vergleich von Frequenzlisten der angelegten (Teil-)korpora abgefragt werden. Über den Link „Click here to create/view frequency lists“ können Teilkorpora eingesehen und erstellt sowie weitere Frequenzlisten angelegt werden.

³⁴ Neben dem oben stehenden Link können über das Menü User controls unter dem Menüpunkt Create/edit subcorpora Teilkorpora angelegt werden, auf die sich die Frequenzlisten und Schlüsselwortlisten beschränken lassen (View a list based on).

kriko: powered by CQPweb

Keywords and key tags

Keyword lists are compiled by comparing frequency lists you have created for different subcorpora. [Click here to create/view frequency lists.](#)

Select frequency list 1: Select frequency list 2:

Compare:

Options for keyword analysis:

Min freq (list 1): Min freq (list 2):

Comparison statistic: Significance threshold:

Options for comparing frequency lists (by filtering):

Display words that only occur in:

Abb. 25: Schlüsselwortlisten

Diese nach bestimmten Kriterien definierten Teilkorpora dienen neben der Zusammenstellung von Schlüsselwortlisten und Frequenzlisten u.a. der Standardsuche als Vergleich von Dokumenten aus Korpus und Teilkorpus oder aus Teilkorpus und Teilkorpus. Auf diesem Weg erhält man über absolute Werte hinaus auch relative Werte. Zwei interessante Möglichkeiten wären z.B. Teilkorpora zwei verschiedener Textsorten, oder verschiedener Zeiträume anzulegen, wobei ein Korpus als Referenzkorpus dienen könnte.

Create and edit subcorpora

Define new subcorpus via:

Existing subcorpora

Name of subcorpus	No. of texts	No. of words	Frequency list	Actions		Delete
bericht	4,764	0	Compile	[copy]	[add]	[x]
Kommentare	1,936	0	Compile	[copy]	[add]	[x]

Abb. 26: Teilkorpora erstellen und abrufen

3.3.4 Suchverlauf und gespeicherte Suchanfragen

Als letzten Punkt möchten wir noch die Speicher- und Protokollfunktion zeigen. Die in der CWB gemachten Suchanfragen sind über den Menüpunkt Query history im Menü User controls zu erreichen. An dieser Stelle sind alle gemachten Suchanfragen mit Angabe der Trefferanzahl und des Abfragedatums einzusehen. Zudem ist die Suchanfrage dadurch wiederholbar. Sind Suchanfragen über das Menü der Ergebnisliste mit der Funktion Save current set of hits unter einem Namen gespeichert worden, können diese über den Menüpunkt Saved queries wieder aufgerufen werden.

4 Fazit

Zweck dieser Handreichung ist es, eine Datenbank- und Analysesoftware vorzustellen, um zu zeigen, wie mit größeren Datenmengen im Rahmen eines diskurs- und korpuslinguistischen Forschungsprojektes umgegangen werden kann. Darüber hinaus zeigten wir verschiedene Möglichkeiten, die mit Ingwer verwalteten Daten in der Open Corpus Workbench quantitativ zu recherchieren. Dabei war es nicht unser Anliegen eine Korpusanalyse zu präsentieren, sondern Hilfestellungen zu geben in der Handhabung dieser Programme. Gerade was die Suche nach sprachlichen Mustern betrifft, sind die Möglichkeiten zum einen begrenzt durch die Art der Aufbereitung der Korpora und die darin enthaltenen Annotationen. Deshalb sollte man sich bereits vor der Korpuserstellung Gedanken über die Verwaltung und Recherche machen. So ist es beispielsweise eher ein Notbehelf über lineare Wortfolgen nach Phrasen zu suchen, statt mit syntaktisch getaggten Korpora (Chunking und Parsing) zu arbeiten. Hier wäre eine Erweiterung der Funktionalitäten der Open Corpus Workbench wünschenswert. Zum anderen sind die Möglichkeiten begrenzt durch die Fähigkeiten und den Willen der Forscher, sich in die technischen Gegebenheiten einzuarbeiten. Denn erst wenn man beispielsweise ein gewisses Verständnis für die Abfragesprache entwickelt hat, ist man auch in der Lage Ideen zu entwickeln, wonach sich suchen lässt.

Zum Schluss sollte nicht unerwähnt bleiben, dass quantitative Recherchen zwar schnell von der Hand gehen, der Weg dort hin, das heißt: von der Erstellung eines Korpus, über die Aufbereitung der Daten und Annotation, bis hin zur Analyse seine Zeit und Mühen beansprucht. Außerdem sollte klar sein, dass der Forschungsprozess zumindest für diskurslinguistisches Arbeiten um hermeneutische Verfahren zu ergänzen ist.

5 Literatur

- Böke, Karin (1996a): Politische Leitvokabeln in der Adenauer-Ära. Zu Theorie und Methodik. In: Böke, Karin/Liedtke, Frank/Wengeler, Martin (Hrsg.): *Politische Leitvokabeln in der Adenauer-Ära*. Berlin, New York: de Gruyter, 19–47.
- Böke, Karin (1996b): Überlegungen zu einer Metaphernanalyse im Dienste einer „parzellierten“ Sprachgeschichtsschreibung. In: Böke, Karin/Jung, Matthias/Wengeler, Martin (Hrsg.): *Öffentlicher Sprachgebrauch. Praktische, theoretische und historische Perspektiven. Georg Stötzel zum 60. Geburtstag gewidmet*. Opladen: Westdeutscher Verlag, 431–452.
- Böke, Karin (1997): Die „Invasion“ aus den „Armenhäusern Europas“. Metaphern im Einwanderungsdiskurs. In: Jung, Matthias/Wengeler, Martin/Böke, Karin (Hrsg.): *Die Sprache des Migrationsdiskurses. Das Reden über „Ausländer“ in Medien, Politik und Alltag*. Opladen: Westdeutscher Verlag, 163–192.
- Jung, Matthias (2000): Diskurshistorische Analyse als linguistischer Ansatz. In: *Sprache und Literatur* 86, S. 20–38.
- Kuck, Kristin/Römer, David 2012: Argumentationsmuster und Metaphern im Mediendiskurs zur „Finanzkrise“. In: Lämmle, Kathrin/Peltzer, Anja/Wagenknecht, Andreas (Hrsg.):

- Krise, Cash und Kommunikation – Die Finanzkrise in den Medien*. Konstanz: UVK, S. 71-95.
- Lemnitzer, Lothar/Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung* (2., durchges. und aktual. Aufl.). Tübingen: Narr.
- Niehr, Thomas (1999): Halbautomatische Erforschung des öffentlichen Sprachgebrauchs oder Vom Nutzen computerlesbarer Textkorpora. In: *ZGL* 27 (1999), S. 205-214.
- Römer, David (2013): „Politikversagen!“ – Relationale Toposanalyse: Überlegungen zu einem Verfahren linguistischen Interpretierens und dessen sprachkritischer Anwendbarkeit am Beispiel eines Diskursausschnitts zu „Krisen“. In: *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 8 (2012), Heft 3, S. 193-216
- Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter.
- Scholz, Ronny/Wengeler, Martin (2012): „Steuern runter macht Deutschland munter“ und „Kriegen die Pleitebanker auch noch einen Bonus?“. Zwei Wirtschaftskrisen in BILD. In: *Osnabrücker Beiträge zur Sprachtheorie* (Bd. 81), S. 155-176.
- Spitzmüller, Jürgen/Warnke, Ingo. H. (2011): *Diskurslinguistik: Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin: De Gruyter.
- Stötzel, Georg/Wengeler, Martin (1995) (Hrsg.): *Kontroverse Begriffe. Geschichte des öffentlichen Sprachgebrauchs in der Bundesrepublik Deutschland*. Berlin/New York: de Gruyter.
- Warnke, Ingo. H. (Hrsg.) (2007): *Diskurslinguistik nach Foucault: Theorie und Gegenstände*. Berlin: De Gruyter.
- Warnke, Ingo. H./Spitzmüller, Jürgen (Hrsg.) (2008): *Methoden der Diskurslinguistik: Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*. Berlin: De Gruyter.
- Wengeler, Martin (2003): *Topos und Diskurs. Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960-1985)*. Tübingen: Niemeyer.
- Wengeler, Martin (Hrsg.) (2005): *Sprachgeschichte als Zeitgeschichte*. Hildesheim: Olms.
- Ziem, Alexander/Scholz, Ronny/Römer, David (2013). Korpusgestützte Zugänge zum öffentlichen Sprachgebrauch: Spezifisches Vokabular, semantische Konstruktionen, syntaktische Muster in Diskursen über „Krisen“. In: Felder, Ekkehard (Hrsg.): *Faktizitätsherstellung in Diskursen: Die Macht des Deklarativen*. Berlin: De Gruyter, S. 219-321.

5.1 Internetquellen

The IEEE and The Open Group (2004): The Open Group Base Specifications Issue 6, IEEE Std 1003.1

(http://pubs.opengroup.org/onlinepubs/009695399/basedefs/xbd_chap09.html#tag_09_03_05, zuletzt abgerufen am 15.03.2013).

Bubenhofer, Noah (2006-2011): Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge. Elektronische Ressource:

<http://www.bubenhofer.com/korpuslinguistik/>, zuletzt abgerufen am 15.03.2013.

6 Korpora

Kriko: DFG-Projekt „Sprachliche Konstruktionen sozial- und wirtschaftspolitischer „Krisen“ in der BRD von 1973 bis heute“.

Text+Berg-Korpus: Noah Bubenhofer/Martin Volk/Adrian Althaus/Magdalena Jitca/David Klaper/Rico Sennrich (Hgg.) (2011): Text+Berg-Korpus (Release EdA Version 01). Digitale Edition des Echo des Alpes 1872-1924. Institut für Computerlinguistik, Universität Zürich.

Text+Berg-Korpus: Noah Bubenhofer/Martin Volk/Adrian Althaus/Magdalena Jitca/Maya Bangerter/Rico Sennrich (Hgg.) (2011): Text+Berg-Korpus (Release 145). Digitale Edition des Jahrbuch des SAC 1864-1923 und Die Alpen 1925-2009. Institut für Computerlinguistik, Universität Zürich.

7 Die Autoren

David Römer

Universität Vechta
 Institut für Geistes- und Kulturwissenschaften
 Germanistische Sprachwissenschaft
 Driverstraße 22-26
 49377 Vechta
david.roemer@uni-vechta.de

Christian Daniel Kreuz

Universität Trier
 Fachbereich II Germanistik
 Germanistische Linguistik
 Universitätsring 15
 54286 Trier
kreuz@uni-trier.de