

## Linguistik im Internet

Eva Gredel\* u. Ruth M. Mell\*

# Digitale Ressourcen und ihr Potential für korpuslinguistisch informierte Diskursanalysen

DOI 10.1515/zgl-2015-0018

## 1 Digitale Ressourcen und Diskurslinguistik

Im Programm einer linguistischen Diskursanalyse, wie es erstmals Busse 1987 in Anlehnung an den Foucault'schen Diskursbegriff skizzierte (vgl. Busse 1987), kam in den 1990er Jahren die Frage nach dem empirischen Zugang zu Diskursen und somit nach methodischen und methodologischen Aspekten auf: Unter dem Titel „Ist Diskurs ein sprachwissenschaftliches Objekt?“ diskutierten Busse und Teubert Anfang der 90er Jahre die Möglichkeiten, Diskurs- und Korpusanalyse zu verbinden (vgl. Busse/Teubert 1994). Dieser programmatische Artikel kann als Ausgangspunkt für die Etablierung eines neuen Forschungsfeldes unter dem Begriff *Diskurslinguistik* verstanden werden, das stark durch die Engführung von Diskurs- und Korpuslinguistik geprägt ist (vgl. Bubenhofer 2009, Busse/Teubert 2013), die mit korpuslinguistischen Methoden den empirischen Zugang zu Diskursen ermöglicht. Spitzmüller und Warnke legten 2011 die erste Einführung in die Diskurslinguistik vor und definierten die Diskurslinguistik als Teildisziplin der Linguistik mit „Interesse an textübergreifenden, also transtextuellen Sprachstrukturen“ (Spitzmüller/Warnke 2011: 22). Zentral für korpuslinguistisch informierte Diskursanalysen ist, welche Ressourcen Zugang zu großen digital verfügbaren Textsammlungen, also Korpora, bieten, um diskursiv etablierte und diskurspezifische sprachliche Muster zu rekonstruieren. Für solche Diskursanalysen können dabei die unterschiedlichsten Korpusypen verwendet werden, wobei hier die Kriterien „Persistenz“ und

---

\***Kontaktperson: Eva Gredel:** Universität Mannheim, Philosophische Fakultät, Seminar für deutsche Philologie, Lehrstuhl Germanistische Linguistik, D-68131 Mannheim, E-Mail: eva.gredel@phil.uni-mannheim.de

\***Kontaktperson: Ruth M. Mell:** Universität Mannheim, Philosophische Fakultät, Seminar für deutsche Philologie, Lehrstuhl Germanistische Linguistik/Institut für Deutsche Sprache, D-68161 Mannheim, E-Mail: r.mell@mail.uni-mannheim.de

„Verfügbarkeit“ (Lemnitzer/Zinsmeister 2010: 105/107) mitthematisiert werden.<sup>1</sup> Die folgenden Ausführungen geben Einblick in die diskurslinguistische Brauchbarkeit digitaler Ressourcen, die interessierten Nutzerinnen und Nutzern vorbehaltlich technischer und rechtlicher Schutzmaßnahmen zur Verfügung stehen.

## 2 Primär diskurslinguistische, digitale Ressourcen

Die **Diskurswörterbücher<sup>2</sup> von Heidrun Kämper zum Schulddiskurs 1945–55** (<http://www.owid.de/wb/disk45/einleitung.html>) und zum **Demokratiediskurs 1967/68** (<http://www.owid.de/wb/disk68/start.html>) sind über OWID – das Online-Wortschatz-Informationssystem Deutsch ([www.owid.de](http://www.owid.de)) – digital verfügbar. Insofern die Wörterbücher selbst Resultat diskurslexikographischer Verfahren sind, können sie in vielfältiger Weise diskurslinguistischen Erkenntnisinteressen zuarbeiten (vgl. Mell i. Dr.). So geben Diskurswörterbücher Aufschluss über Diskurswortschätze, Diskursthemen und Diskursbeteiligte. Sie bieten zudem die Möglichkeit, die Bedeutungsentwicklung von Schlüsselwörtern sowie deren Gebrauch im Sinne der Diskursanalyse anhand themenbezogener Diskurswortschätze zu rekonstruieren und in die eigene Forschung mit einzubeziehen. So stellt ein Diskurswörterbuch „die semantische Prägung lexikalischer Einheiten unter den Bedingungen des Diskurses dar“ (Kämper 2006: 338) und ermöglicht damit eine lexikologische wie auch lexikographische Perspektivierung des Diskurses. In den Wörterbüchern zum Schulddiskurs 1945 und zum Demokratiediskurs 1968 sind diejenigen Lexeme versammelt, die als Schlüsselwörter gelten können, „in deren Bedeutung sich dieser Diskurs verdichtet“ (<http://www.owid.de/wb/disk45/einleitung.html>; 16.4.2015). Damit stellt z. B. das Schulddiskurs-Wörterbuch denjenigen Wortschatzbereich dar, „der hinsichtlich Frequenz oder Funktion den Schulddiskurs der frühen Nachkriegszeit repräsentiert“ (ebd.). Bei dem im Wörterbuch zum Demokratiediskurs verzeichneten Wortschatz handelt es sich um diejenigen lexikalischen Einheiten, die – anders als beim Schulddiskurswörterbuch – hinsichtlich der Funktion den kritischen Demokratiediskurs der späten 1960er Jahre konstituieren und verdichten (<http://www.owid.de/wb/disk68/start.html>; 16.4.2015).

---

1 Dieser Beitrag orientiert sich an der Korpusmorphologie von Lemnitzer/Zinsmeister (2010: 102ff.).

2 Die Diskurswörterbücher sind auch als Printversionen verfügbar, vgl. Kämper 2007, 2012.

**Das Heidelberger Korpus (HeideKo, [http://www.gs.uni-heidelberg.de/sprache02/hd\\_korpus.html](http://www.gs.uni-heidelberg.de/sprache02/hd_korpus.html))** wurde am Lehrstuhl für Germanistische Linguistik der Universität Heidelberg als eine gemeinsame Infrastruktur für Spezialkorpora konzipiert. Technische Grundlage ist eine hierarchisch strukturierte Access-Datenbank, die dazu dient, gesellschaftliche Konflikte im Spiegel der Sprache zu analysieren. Die Verfügbarkeit ist aus urheberrechtlichen Gründen jedoch auf den Offline-Zugriff beschränkt. In HeideKo verfügbar sind mehrere Medientypen, darunter Text, Transkript, Audiofile, Videofile, Bild, die es ermöglichen, multimediale Diskursrealitäten zu rekonstruieren.

### 3 Digitale Ressourcen mit Potential für diskursanalytische Auswertungen

Obwohl das **Wortschatzlexikon** ([www.wortschatz.uni-leipzig.de](http://www.wortschatz.uni-leipzig.de)) über die Elemente gedruckter Lexika und Wörterbücher hinaus durch korpus-basierte Elemente (z. B. Angaben zu absoluten Häufigkeiten) erweitert wird und sich somit für erste diskurslexikographische Anfragen qualifiziert, bleibt die diskursanalytische Brauchbarkeit begrenzt. Die Auswahl der mitgelieferten Beispielsätze zum Suchlexem erscheint willkürlich, und es werden nur minimale Kontexte geliefert. Der über die Benutzeroberfläche zur Verfügung gestellte Kookkurrenzgraph lässt nur optische, aber keine statistisch fundierten Deutungen des visualisierten Kookkurrenzprofils zu (vgl. Gredel 2014: 107f.). Problematisch erscheint hinsichtlich der Gütekriterien empirischen Arbeitens auch die mangelhafte Reproduzierbarkeit aufgrund der fehlenden Dokumentation zur Persistenz.<sup>3</sup>

Das Projekt „**Digitales Wörterbuch der deutschen Sprache**“ (DWDS: <http://www.dwds.de/>) der Berlin-Brandenburgischen Akademie der Wissenschaften hat zum Ziel, ein digitales lexikalisches System zu erstellen, das in Form einer jedem Benutzer zugänglichen Datenbank Auskunft über den deutschen Wortschatz in Vergangenheit und Gegenwart gibt. Das diskursanalytische Potential ergibt sich vor allem über das Panel „Wortprofil“: Das hier implementierte Kookkurrenz-Tool integriert die Klassifikation syntaktischer Relationen in die Analyse signifikanter Kookkurrenzpartner. Es werden dabei jeweils separat diejenigen Kookkurrenzpart-

---

<sup>3</sup> Lemnitzer und Zinsmeister unterscheiden hinsichtlich der Persistenz zwischen statischen Korpora und dynamischen Korpora. In diesem Zusammenhang verweisen sie auf die Notwendigkeit, die verwendeten Versionen (Releases) dieser Korpora ausführlich zu dokumentieren (Lemnitzer/Zinsmeister 2010: 105).

ner aufgeführt, die in einer vordefinierten syntaktischen Relation stehen. Das Wortprofil bietet somit zu zentralen Suchlexemen diskursiv etablierte und syntaktisch klassifizierte Muster, die frequent in den DWDS-Korpora vorkommen.

Bei **COSMAS II** (Corpus Search, Management and Analysis System: <http://www.ids-mannheim.de/cosmas2/>) handelt es sich um eine am Institut für Deutsche Sprache (IDS) konzipierte Volltextdatenbank für das linguistisch motivierte Recherchieren in den Korpora des IDS. Die über die Benutzeroberfläche COSMAS zur Verfügung stehende Textsammlung ist das sogenannte Deutsche Referenzkorpus (DeReKo): „Not being available for download, the main way to access DeReKo is via COSMAS, the Corpus Search, Management and Analysis System“ (Kupietz und Keibel 2009: 55). Zentral für diese Textsammlung ist, dass DeReKo nicht als ausgewogenes Korpus, sondern als eine Ur-Stichprobe zu verstehen ist. Um das diskursanalytische Potential von **COSMAS** auszuschöpfen, bieten sich zwei Eigenschaften dieser digitalen Sprachressource an: die Kookkurrenzanalyse sowie die Themenansicht. Die implementierte **Kookkurrenzanalyse**<sup>4</sup> bietet die Möglichkeit, diskursiv etablierte Sprachgebrauchsmuster korpuslinguistisch zu rekonstruieren. Eine zweite Möglichkeit in COSMAS diskursanalytischen Fragestellungen nachzugehen, stellt die sogenannte Themenansicht dar, über welche Texte nach in DeReKo hinterlegten Themenordnungen betrachtet werden können. Die Themenansicht erlaubt somit die Rekonstruktion „transdiskursive[r] Driften“ (vgl. Gredel 2014: 133ff.) zentraler Suchlexeme.

## 4 Nicht-linguistische, digitale Ressourcen mit Potential für diskursanalytische Auswertungen

Die Datenbanken, die im Datenbanken-Informationssystem der Universität Regensburg (DBIS: [http://rzblx10.uni-regensburg.de/dbinfo/fachliste.php?bib\\_id=ubman&lett=l&colors=&ocolors=](http://rzblx10.uni-regensburg.de/dbinfo/fachliste.php?bib_id=ubman&lett=l&colors=&ocolors=)) zur Verfügung stehen, ermöglichen es, Spezialkorpora zu erstellen. Insgesamt vier digitale Ressourcen sollen hier überblicksartig hinsichtlich ihres diskursanalytischen Potentials untersucht werden: Über die Datenbank **NEXIS**

---

<sup>4</sup> In COSMAS ist die Kookkurrenzanalyse auf der Grundlage einer bestimmten Suchanfrage automatisch möglich, d. h. der Kontext des zentralen Suchlexems wird ausgezählt und mithilfe eines Signifikanztests darauf hin evaluiert, welche der auftretenden Kookkurrenzpartner überzufällig häufig – also signifikant im mathematischen Sinne – mit dem zentralen Suchlexem auftreten. Als Signifikanztest ist der Test Log Likelihood Ratio implementiert.

(bisher Lexis Nexis: <http://www.lexisnexis.com/de/business>) sind mehr als 35.000 Quellen und über fünf Milliarden Volltext-Dokumente aus mehr als 150 Presseerzeugnissen zugänglich. Überregionale Tagesszeitungen wie die Frankfurter Allgemeine Zeitung (**F.A.Z.-BiblioNet**: <http://www.faz-archiv.de/biblio>) und die Süddeutsche Zeitung (**SZ LibraryNet**: <http://librarynet.szarchiv.de>) sind mit jeweils eigenen Datenbanken in DBIS vertreten.

Da Werbekommunikate gesamtgesellschaftlich relevante Themen aufgreifen, kann Werbung als gesellschaftlicher Teildiskurs verstanden werden. Eine diskursanalytisch brauchbare Ressource ist in diesem Zusammenhang die über DBIS verfügbare Datenbank **Produktwerbung im Plakat 1890 bis 1918** ([http://www.fr.re-di-bw.de/session/DBReklame\\_2005-199bcdfb.html](http://www.fr.re-di-bw.de/session/DBReklame_2005-199bcdfb.html)). Die Sammlung historischer Plakate ermöglicht es, multikodale Recherchen aus einer diachronen Perspektive durchzuführen und somit multikodale Text- und Diskursrealitäten zu berücksichtigen.

## 5 Fazit und Desiderata

In diesem Beitrag wurden digitale Ressourcen auf ihr Potential für diskursanalytische Auswertungen hin untersucht. Die digitalen Ressourcen bieten vielfältige Möglichkeiten des empirischen Zugangs zu Diskursen. Auffällig war jedoch, dass der *Web-as-a-Corpus*-Gedanke noch keinen Eingang in die Erstellung entsprechender Ressourcen gefunden hat. Dies wäre jedoch für die Analyse digitaler Diskurse, z. B. hypertextuell geführter Diskurse in der Wikipedia oder diskursiv ausgehandelter sowie medienpezifisch realisierter Positionierungspraktiken und Geltungsansprüche im Internet, notwendig. Ein weiteres Desiderat im Bereich diskursanalytischer Untersuchungen bleiben digitale Ressourcen, die nicht nur diskurslinguistische, sondern auch diskurssemiotische Erkenntnisinteressen bedienen, d. h. multimediale, multimodale und multikodale Diskursrealitäten berücksichtigen, wie es z. B. schon in HeideKo angelegt ist.

## Literatur

- Bubenhof, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin/Boston: de Gruyter.
- Busse, Dietrich (1987): Historische Semantik. Stuttgart: Klett-Cotta.
- Busse, Dietrich/Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hg.): Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik. Opladen: Westdeutscher Verlag, 10–28.

- Busse, Dietrich/Teubert, Wolfgang (Hg.) (2013): *Linguistische Diskursanalyse: neue Perspektiven*. Wiesbaden: Springer VS.
- Gredel, Eva (2014): *Diskursdynamiken. Metaphorische Muster zum Diskursobjekt Virus*. Berlin/Boston: de Gruyter.
- Kupietz, Marc/Keibel, Holger (2009): The Mannheim German Reference Corpus (DEREKO) as a Basis for Empirical Linguistic Research. In: *Working Papers in Corpus-based Linguistics and Language Education* (3), 53–59.
- Kämper, Heidrun (2006): Diskurs und Diskurslexikographie. Zur Konzeption eines Wörterbuchs des Nachkriegsdiskurses. In: *Deutsche Sprache* 4/06, 334–353.
- Kämper, Heidrun (2007): *Opfer – Täter – Nichttäter. Ein Wörterbuch zum Schulddiskurs 1945–1955*. Berlin/New York: de Gruyter.
- Kämper, Heidrun (2013): *Wörterbuch zum Demokratiediskurs 1967/68*. Unter Mitwirkung von Elisabeth Link. Berlin: Akademie Verlag.
- Lemnitzer, Lothar/Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- Mell, Ruth (i. Dr.): *Vernunft, Mündigkeit, Agitation*. Eine diskurslinguistische Untersuchung zur Generierung und Strukturierung von Wissen über das Konzept ‚AUFKLÄRUNG 1968‘. Bremen: Hempfen Verlag (= Reihe Sprache – Politik – Gesellschaft, 16).
- Spitzmüller, Jürgen/Warneke, Ingo (2011): *Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin/Boston: de Gruyter.