

Zeitschrift für Diskursforschung  
1. Beiheft 2015

**Zeitschrift für  
Diskursforschung**  
Journal for  
Discourse Studies

**Diskurs – Interpretation –  
Hermeneutik**

1. Beiheft

Reiner Keller | Werner Schneider | Willy Viehöver  
(Hrsg.)

**BELTZ** JUVENTA

# Inhalt

|   |     |
|---|-----|
| Einleitung  | 2   |
| <i>Daniel Wrana</i><br>Zur Lokation von Sinn. Das Subjekt als Bedingung und Gegenstand<br>von Diskursanalyse und qualitativer Forschung                       | 14  |
| <i>Rainer Diaz-Bone</i><br>Die Sozio-Epistemologie als methodologische Position<br>Foucaultscher Diskursanalysen  | 43  |
| <i>Dietrich Busse</i><br>Diskursanalyse und Hermeneutik. Ein prekäres Verhältnis  | 62  |
| <i>Werner Holly</i><br>Diskurse verstehen? Optionen linguistischer Diskurshermeneutik   | 86  |
| <i>Ekkehard Felder/Anna Mattfeldt</i><br>Linguistik als hermeneutische Wissenschaft. Das schwierige Verhältnis<br>von Text und Bild im Diskurs                | 107 |
| <i>Noah Bubenhofer/Joachim Scharloth/David Eugster</i><br>Rhizome digital: Datengeleitete Methoden für alte und neue<br>Fragestellungen in der Diskursanalyse | 144 |
| <i>Reiner Keller</i><br>Weber und Foucault. Interpretation, Hermeneutik und<br>Wissenssoziologische Diskursanalyse  | 173 |
| <i>Willy Viehöver</i><br>Narration und Interpretation. Überlegungen zum<br>hermeneutischen Strukturalismus Paul Ricoeurs                                      | 211 |
| Die Autorinnen und Autoren  | 261 |

Noah Bubenhofer/Joachim Scharloth/David Eugster

## Rhizome digital

Datengeleitete Methoden für alte und neue Fragestellungen in der Diskursanalyse<sup>1</sup>

**Zusammenfassung:** Mit dem „Data-driven Turn“ ergeben sich in der digitalen Welt neue Chancen für die linguistische Diskursanalyse: Einerseits ist es oft ein Leichtes, große Textkorpora zusammenzustellen, um Analysen auf einer breiten empirischen Basis zu ermöglichen. Andererseits bedingen diese großen Datenmengen neue Untersuchungsmethoden; die Anwendung von Erkenntnissen der Korpus- und Computerlinguistik ermöglichen datengeleitete Zugänge, mit denen auch neue Forschungsfragen angegangen werden können. Wir demonstrieren dies anhand datengeleiteter Analysen zu zeitgeschichtlichen Umbrüchen seit dem Zweiten Weltkrieg in SPIEGEL und ZEIT.

**Schlagwörter:** Linguistische Diskursanalyse, Korpuslinguistik, datengeleitete Methoden, zeitgeschichtliche Umbrüche, Frameanalyse, Kollokationsanalyse

**Summary:** The „data driven turn“ as a result of the digital world enables new approaches to linguistic discourse analysis: On the one hand, it is often much easier to build big text corpora for analyses on a broader empirical basis. On the other hand, using big data demands new methods to access it. The use of state-of-the-art techniques of corpus- and computational linguistics enables data-driven approaches, which also allows pursuing new research questions. An analysis of cultural upheavals in Germany after WWII in the newspapers SPIEGEL and ZEIT illustrates these possibilities.

**Keywords:** linguistic discourse analysis, corpus linguistics, data-driven methods, historical/cultural upheavals, frame analysis, collocation analysis

---

1 Der folgende Beitrag baut auf unserem bereits publizierten Text „Das Wuchern der Rhizome“ (Scharloth et al. 2013) auf. Einige methodisch-theoretische Überlegungen sind daraus in erweiterter, teilweise komprimierter Form, entnommen, die Datengrundlage für die Analysen wurde jedoch stark erweitert.

Die linguistische Diskursanalyse ist zweifellos ein Erfolg: Seit Dietrich Busses Dissertation „Historische Semantik. Analyse eines Programms“ (Busse 1987) hat sich die Diskursanalyse als Gegenentwurf zur historiographisch interessierten historischen Semantik zu einer eigenständigen Teildisziplin der Linguistik entwickelt. Dazu gehört ein großes Methodenrepertoire und eine Reihe von unterschiedlichen Ansätzen (vgl. Spitzmüller/Warnke 2011), die darüber hinaus an andere sozial- und kulturwissenschaftliche Disziplinen anschließbar sind (vgl. Scharloth et al. 2013).

Andererseits leidet diskursanalytische Forschung unter einem Korsett immer komplexerer Analysekatoren und Methoden: Die Umsetzung verschlingt immer mehr Ressourcen, der heuristische Mehrwert der Analysekatoren und Methoden ist aber nicht belegt (vgl. ebd.). Die Methodik zwingt zudem diskursanalytische Arbeiten oft, ihre Datenbasis „aus forschungspraktischen Gründen“ (Busse/Teubert 1994, S. 14) einzuschränken, da ansonsten durch die hermeneutische Interpretation oder Dekonstruktion manuell codierter Texte die Masse nicht zu bewältigen wäre. Erschwerend kommt hinzu, dass die komplexen Kategorienkataloge und methodischen Ansätze zwar wissenschaftliche Objektivität suggerieren, eine Validierung ihrer Messinstrumente und eine Beurteilung ihrer Reliabilität jedoch nicht erfolgt.

Obwohl die linguistische Diskursanalyse schon früh korpuslinguistische Methoden nutzte,<sup>2</sup> wird der Stand der Kunst der Korpus- und Computerlinguistik nicht berücksichtigt. So werden multivariate Verfahren oder maschinelles Lernen zur Textanalyse ebenso selten verwendet wie Methoden der automatischen Annotation von linguistischen Kategorien oder von der Computerlinguistik erarbeitete Ressourcen wie semantische Taxonomien.

Wir plädieren deswegen dafür, für die linguistische Diskursanalyse den Anschluss an die sprachtechnologische Entwicklung zu finden, um Gesellschaftsanalysen auf sprachlicher Ebene nicht der Privatwirtschaft oder anderen Disziplinen allein zu überlassen. Denn so wäre zu befürchten, dass auch das kritische Potenzial der Diskursanalyse verloren ginge (Scharloth et al. 2013).

## 1. Linguistische Diskursanalyse und der Data-driven Turn

Das Potenzial der digitalen Welt wird in der Linguistik noch wenig genutzt. Dabei sind es zwei Aspekte, die methodisch besonders interessant sind: Naheliegend ist die bessere Verfügbarkeit von großen Datenmengen. Es ist ver-

---

2 Vgl. Busse/Teubert (1994, S. 14–18) und neuerlich die konzise Einführung von Teubert/Cermáková (2007).

gleichsweise einfach, große Korpora aufzubauen, um quantitative Analysen zu machen. Der zweite Aspekt ist jedoch zunächst mit dem Problem verbunden, komplexe linguistische Kategorien für eine quantitative Analyse fruchtbar zu machen – meist ist dies sehr schwierig und unbefriedigend. Die Chance der großen Datenmenge liegt aber darin, mit neuen Fragen und neuen Methoden an die Daten heranzugehen: Denn erfolgreicher und gewinnbringender sind induktive Analysemethoden, die Strukturen in den Daten entdecken. Dies soll im folgenden Beitrag demonstriert werden.

### 1.1 Datengeleitete Korpuslinguistik

Korpuslinguistische Ansätze sind zwar inzwischen insgesamt sehr verbreitet, unterscheiden sich jedoch stark voneinander: Ein wichtiges Unterscheidungskriterium ist, ob die Ansätze eher einem corpus-based- oder einem corpus-driven-Paradigma folgen. Bei ersterem dienen digitale Korpora der Überprüfung von Forschungshypothesen, die mit bewährten interpretativen linguistischen Analysekatogorien in den Daten überprüft werden.

Für Ansätze, die einem corpus-driven-Paradigma folgen, sind Korpora jedoch weit mehr als „Belegsammlungen oder Zettelkästen in elektronischer Form“, sondern ermöglichen eine eigene „korpuslinguistische Perspektive“ (Perkuhn/Belica 2006, S. 2): nämlich aus den Daten heraus neue Hypothesen oder sogar Analysekatogorien zu bilden. Dies geschieht dadurch, dass über unterschiedliche Verfahren der statistischen Analyse für bestimmte Teilkorpora typische (oder gerade untypische) Korrelationen bestimmter linguistischer Einheiten systematisch berechnet und anschließend interpretiert werden. Ein einfacher Zugang zu solchen datengeleiteten Analysen ist das Berechnen von typischen Lemmata in einem Teilkorpus im Vergleich zu einem Referenzkorpus oder aber die Berechnung von typischen Mehrworteinheiten (Bubenhofer/Scharloth 2013a, 2013b).

Der corpus-driven- oder datengeleitete Zugang zu digitalen Korpora zeigt, dass Korpuslinguistik mehr als eine Methode, sondern eher ein Denkstil sein kann: Neue korpuslinguistische Methoden dienen (nicht nur) dem Beantworten von alten Fragen mit neuen Mitteln, sondern ermöglichen gänzlich neue Zugänge zu Sprache und den Kategorien ihrer Beschreibung.<sup>3</sup>

---

3 Zwar verzichtet das datengeleitete Paradigma auf das Formulieren von Hypothesen und auf eine Festlegung auf bestimmte Analysekatogorien, es ist jedoch offensichtlich, dass auch beim datengeleiteten Verfahren vorgängiges Wissen in den Forschungsprozess einfließt, und zwar: (1) durch die Wahl der Korpora, (2) hinsichtlich der Gestaltung der Algorithmen zur Musterberechnung, (3) bei der Festlegung dessen, was als

## 1.2 Der Data-driven Turn und seine Folgen für die linguistische Diskursanalyse

Der Wandel in der Korpuslinguistik hin zur Arbeit mit sehr großen Datenmengen („big data“) und zu strukturentdeckenden Methoden ist nicht einzigartig. Auch in anderen Disziplinen der Kultur- und Sozialwissenschaft stehen inzwischen Datenmengen zur Verfügung, die nicht mehr von einem Forscher oder einer Forscherin überblickt werden können und es werden datengeleitete Methoden in Kombination mit statistischen Verfahren eingesetzt, um neue Fragestellungen und Analysemöglichkeiten zu generieren. Beispiele sind die „Data Driven History“ in den Geschichtswissenschaften, Netzwerkanalysen in der Soziologie, um die lebensstilspezifische Gliederung sozialer Gemeinschaften aufzudecken oder Ansätze in der Kunstwissenschaft, um auf der Basis von Metadaten Sujets oder Kunststile zu berechnen. Diese Entwicklungen haben das Potenzial, die Kultur- und Sozialwissenschaften nachhaltig zu verändern, weshalb wir von einem Data-driven Turn sprechen wollen.

Die Entwicklung steht allerdings am Anfang. Die Forschergruppe *sem-tracks*, der die Autoren dieses Beitrags angehören, arbeitet seit einigen Jahren daran, Methoden der von uns mitentwickelten datengeleiteten Korpuspragmatik auch für die linguistische Diskursanalyse fruchtbar zu machen.<sup>4</sup> Dabei ist allerdings Folgendes zu betonen: Es geht bei den vorgeschlagenen Methoden nicht darum, die Analysekatoren der traditionellen Diskursanalyse korpuslinguistisch zu operationalisieren oder gar Lektüren zu simulieren. Stattdessen folgt die datengeleitete Forschung ihrer eigenen Logik und generiert Analysekatoren aus den Daten selbst.<sup>5</sup>

---

linguistische Untersuchungseinheit (token) gelten soll, und (4) bei der Festlegung dessen, welche Einheitentypen eigentlich als potenzieller Bestandteil eines Musters aufgefasst werden sollen. Schließlich ist (5) auch das Kategorisieren der Daten im Anschluss an die Musterberechnung ein interpretativer Prozess, der zwar durch statistische Verfahren teilweise objektiviert werden kann; dennoch ist die Menge der Daten meist so umfangreich, dass eine weitere Reduzierung und Gewichtung im Sinne des Forschungsinteresses vorgenommen werden muss.

4 Vgl. Bubenhofer (2009), Scharloth/Bubenhofer (2011), Bubenhofer/Scharloth (2013b), Scharloth et al. (2013).

5 Wir gehen dennoch nicht so weit wie Chris Anderson, der Herausgeber des *Wired Magazine*, der schreibt: „the opportunity is great: The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.“ (Anderson 2008) Wir halten es für wichtig, dass die

Vor dem Hintergrund des Data-driven Turns und der Verfügbarkeit großer Datenmengen in der digitalen Welt stellen wir zwei Dogmen der linguistischen Diskursanalyse in Frage (Scharloth et al. 2013):

(1) Die Thematizität bzw. den Gegenstandsbezug von Diskursen: Eine forschungspraktische Beschränkung auf thematische, gegenstands- oder wissenskomplexspezifische Bestimmung von Diskursen erscheint uns nicht geboten. Diskurse sollten vielmehr anhand strukturentdeckender Verfahren aus den Daten selbst berechnet werden (Bubenhofner 2009, S. 108).

(2) Die Forderung nach Lektüre bzw. der qualitativen Auswertung der Daten:<sup>6</sup> Die Datenmengen, mit denen nach unseren Vorstellungen im Rahmen datengeleiteter Analysen gearbeitet werden sollte, sind viel zu umfangreich, als dass sie noch durch Lektüre erschlossen, geschweige denn aufwendig kodiert werden könnten. Eine qualitative Deutung jedes einzelnen Textemplars scheidet daher von vorneherein als Option aus. Die an die Adresse korpuslinguistischer Verfahren formulierte Kritik, mit der die Notwendigkeit qualitativen Arbeitens begründet wird, lautet, dass die Korpuslinguistik Daten quasi atomistisch, d. h. losgelöst von ihrem Kontext, untersuche,<sup>7</sup> und nur durch die Lektüre ein kontextsensitives Deuten des Zeichengebrauchs möglich sei. Dieses Bild von Korpuslinguistik halten wir jedoch für überholt, denn die Analyse von Kookkurrenzen, Kollokationen und anderen kontextsensitiven Methoden gehört zum Standardrepertoire der Korpuslinguistik.

Im Folgenden soll gezeigt werden, in welche Richtung eine datengeleitete linguistische Diskursanalyse gehen könnte, und insbesondere, wie diskursive Umbrüche datengeleitet identifizierbar und analysierbar sind.<sup>8</sup>

---

Ergebnisse datengeleiteter Analysen valide und nicht lediglich statistisch signifikant sind.

6 Busse/Teubert (1994, S. 18) bezeichnen Diskursanalyse als „ein fortschreitend die Korpusbildung korrigierendes Lesen“. Spitzmüller/Warnke (2011, S. 39) fordern: „Die quantitative Organisation des Datenmaterials durch statistische Auswertungsmöglichkeiten der Korpuslinguistik sollte qualitative Analysen nicht ersetzen, sondern ergänzen“.

7 Vgl. etwa Spitzmüller/Warnke (2011, S. 38).

8 Beispielsweise im Widerspruch zur Kritik an unseren korpuslinguistischen Ansätzen in Spitzmüller/Warnke (2011, S. 39 f.).

## 2. Datengeleitete Frameanalyse am Beispiel des ZEIT/SPIEGEL-Korpus: Identifizierung und Analyse von Umbruchzeiten

### 2.1 Zum Frame-Begriff

Das Ziel der folgenden Untersuchung liegt darin, die Distribution und Vernetzung verschiedener Deutungsrahmen in der ZEIT und im SPIEGEL in den Jahren 1950 bis 2009 zu untersuchen. Dafür verwenden wir den Begriff des Frames nach Goffman (1974),<sup>9</sup> der mit seiner „Frame Analysis“ die „organization of experience“ (ebd., S. 13) theoretisch fassen möchte. Mit Ziem (2008, S. 14) geht es also darum, die Wissensstrukturen, „die es Menschen ermöglichen, ihre Erfahrungsdaten zu interpretieren“, freizulegen. Die wichtigste Funktion von Frames als sozial geteilte und kulturspezifische Wissensstrukturen (Rettie 2004) liegt darin, dem Bedeutungslosen Sinn einzuschreiben: „a primary framework is one that is seen as rendering what would otherwise be a meaningless aspect of the scene into something that is meaningful“ (Goffman 1974, S. 21). Mit dem Begriff des Framing „können all diejenigen Prozesse bezeichnet werden, bei denen Deutungsmuster in der Informationsverarbeitung aktiviert werden“ (Dahinden 2006, S. 28). Aus Sicht des öffentlichkeitstheoretischen Framingansatzes ist Framing der Prozess „by which a communication source, such as a news organization, defines and constructs a political issue or public controversy“ (Nelson et al. 1997, S. 567). In diesem wissenschaftlichen Kontext sind Frames

„kollektive Deutungsmuster, in denen bestimmte Problemdefinitionen, Kausalzusammenhänge, Ansprüche, Begründungen und Wertorientierungen in einen mehr oder weniger konsistenten Zusammenhang gebracht werden, um Sachverhalte zu erklären, Kritik zu fundieren und Forderungen zu legitimieren“ (Neidhardt/Rucht 1993, S. 108).

Mit den beiden Publikationsorganen SPIEGEL und ZEIT als Datenbasis zielt der Aufsatz darauf, Veränderungen in den Realitätskonstruktionen durch Medien zu untersuchen. Dabei steht aber nicht ein bestimmter Diskurs oder ein bestimmtes Thema und dessen Framing im Vordergrund, sondern es geht einem datengeleiteten Ansatz folgend darum, die Konjunkturen von

---

9 Wir halten eine für korpusgeleitete Zwecke geeignete Operationalisierung framesemantischer Konzepte, wie sie etwa von Konecny (1993) und Ziem (2008) elaboriert wurden, für gegenwärtig nicht zu leisten.



Frames nachzuzeichnen und den sich wandelnden Verbindungen der Frames untereinander auf die Spur zu kommen.

## 2.2 Das Untersuchungskorpus

Die folgenden Untersuchungen wurden anhand der Print-Archive der Wochenzeitung DIE ZEIT und des Nachrichtenmagazins DER SPIEGEL vorgenommen. Das vereinigte Korpus umfasste die kompletten Jahrgänge 1946 (ZEIT) bzw. 1947 (SPIEGEL) bis 2010, wie sie auf [zeit.de](http://zeit.de) und [spiegel.de](http://spiegel.de) zu finden sind. Das SPIEGEL-Korpus weist insgesamt 237.620.381 laufende Wortformen auf. Die Anzahl der laufenden Wortformen je Jahrgang variiert dabei zwischen 1.214.202 (1947) und 5.391.881 (1999). Im ZEIT-Korpus finden sich 271.439.149 laufende Wortformen und ihre Anzahl variiert zwischen 1.051.351 (1946) und 6.520.382 (2005) laufenden Wortformen.

Die Qualität der Korpora ist unterschiedlich, da jeweils die früheren Jahrgänge mittels OCR<sup>10</sup> digitalisiert wurden und dadurch typische Erkennungsfehler entstanden sind. Zudem sind Metadaten wie Autor oder Ressort nicht systematisch homogen verfügbar, daher sind z. B. Differenzierungen nach Ressort ohne aufwändige Nachbearbeitung des Korpus nicht möglich.

Das Korpus wurde mit Hilfe des TreeTaggers (Schmid 1994) tokenisiert, unter Verwendung der Standardbibliothek fürs Deutsche<sup>11</sup> mit Wortarten-Informationen annotiert und lemmatisiert. Beim verwendeten Tagset handelt es sich um das Stuttgart-Tübingen-Tagset (STTS) (Schiller et al. 1995).

Die Beschränkung des Untersuchungskorpus auf sämtliche Texte einer Wochenzeitung begrenzt natürlich auch die Aussagekraft der Analysen. Streng genommen untersuchen wir lediglich Wirklichkeitskonstruktionen, wie sie in den Texten der ZEIT und des SPIEGELS vorgenommen wurden. Eine Ausweitung auf weitere Korpora wäre natürlich wünschenswert. Die folgenden Untersuchungen sollen aber vor allem exemplarischer Natur sein und das Potenzial der angewendeten Verfahren illustrieren.

---

10 OCR, „Optical Character Recognition“: Damit wird die maschinelle Umsetzung von Text, der durch ein Pixelbild repräsentiert ist, in elektronisch bearbeitbaren Text. Dabei entstehen je nach Vorlage Erkennungsfehler.

11 „German parameter file (UTF-8)“: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/german-par-linux-3.2-utf8.bin.gz> [Abruf: 14. Mai 2014]

## 2.3 Identifikation der Frames

Wir gehen davon aus, dass Frames in den Artikeln über bestimmte auftretende Lemmata identifizierbar sind. Eine semantische Taxonomie, nämlich der „Deutsche Wortschatz nach Sachgruppen“ (Dornseiff 2004), dient dazu, die vorkommenden Lemmata einem oder mehreren Frames zuzuordnen.<sup>12</sup> Maßgeblich für die Zuteilung einer Sachgruppe (eines Frames) zu einem Text waren die folgenden Kriterien:

1. Die relative Frequenz der vorkommenden Vertreter einer Sachgruppe im Verhältnis zur Wortzahl eines Textes.
2. Die absolute Frequenz der vorkommenden Vertreter einer Sachgruppe im Text.
3. Die Abdeckung der Lemmata einer Sachgruppe (d.h. wie viele der eine Sachgruppe konstituierenden Lemmata kommen im Text vor).

In einem Text können also, auch aufgrund derselben Textstelle, mehrere Frames gleichzeitig zugewiesen werden. Zur Illustration sei zudem gezeigt, welche Vertreter in der Taxonomie zur Sachgruppe „Tausch, Handel“ gehören (Dornseiff 2004):

- austauschen, auswechseln, einhandeln, eintauschen, feilschen, handeln, kaufen, schachern, spekulieren, tauschen, umsetzen, umtauschen, verhandeln, vermitteln, vertauschen
- Absatz, Geschäft, Spekulation, Tauschhandel, Umsatz, Umschlag, Vertrieb, Warenaustausch, Warenverkehr
- Ausfuhr, Außenhandel, Bezug, Binnenhandel, Einfuhr, Einzelhandel, Export, Freihandel, Großhandel, Handel, Import, Versand, Welthandel, Zwischenhandel
- Ausgleich, Austausch, Entgelt, Ersatz, Gegenleistung, Lohn, Tausch, Umtausch, Wechsel
- Börse, Markt, Marktplatz, Umschlaghafen, Umschlagplatz, Warenterminbörse

---

<sup>12</sup> Problematisch an der Arbeit mit einer Taxonomie für eine Analyse von Texten, die wie im vorliegenden Fall über einen längeren Zeitraum verteilt sind, ist ihre Statik. Der „Dornseiff“ enthält jedoch neben dem zentralen Wortschatz eine solche Breite an teils wenig gebräuchlichen Lemmata, dass wir ihn für hinreichend robust halten, mit seiner Hilfe die Zeitgeschichte zu untersuchen.

## 2.4 Zeitreihenanalyse

Ein erster Ansatz, um Veränderungen in der Frameverteilung in den Daten festzustellen, ist die Berechnung von Zeitreihen: Für alle Frames berechnen wir jahresweise die Verteilung über das Korpus. Da die Frames jedoch sehr unterschiedliche Vorkommenshäufigkeiten aufweisen, werden die Frequenzen über eine Min-Max-Normalisierung vergleichbar gemacht. Damit ist nicht mehr sichtbar, welche Frames besonders häufig sind; dafür können ähnliche Entwicklungen (genereller Anstieg oder Abnahme über den Zeitraum etc.) gefunden werden.

Die Menge der Verlaufskurven aller Frames muss in einem zweiten Schritt aber gruppiert werden, um die Frames mit ähnlichen Verkaufskurven zu identifizieren. Dazu wurden die (normalisierten) Frequenzen jedes Frames als Vektor aufgefasst und mit einer Näherungsverteilung, also einer Ideallinie (z. B. stetige Zunahme über das Korpus), verglichen, indem die euklidische Distanz zwischen dem Näherungsverteilungsvektor und den Framevektoren berechnet wurde. Um kurzfristige Veränderungen etwas zu glätten, verwendeten wir für die framespezifischen Zeitreihen einen gleitenden Durchschnitt (Periode 5), damit die längerfristigen Trends besser sichtbar werden.

Die folgenden Abbildungen zeigen jene Frames, die im Untersuchungszeitraum näherungsweise stetig ab- (Abbildung 1) bzw. zunehmen (Abbildung 2), gemessen an einer stetigen Verteilung, die im Jahr 1950 den Wert 1 und im Jahr 2009 den Wert 0 (bzw. umgekehrt) hat.

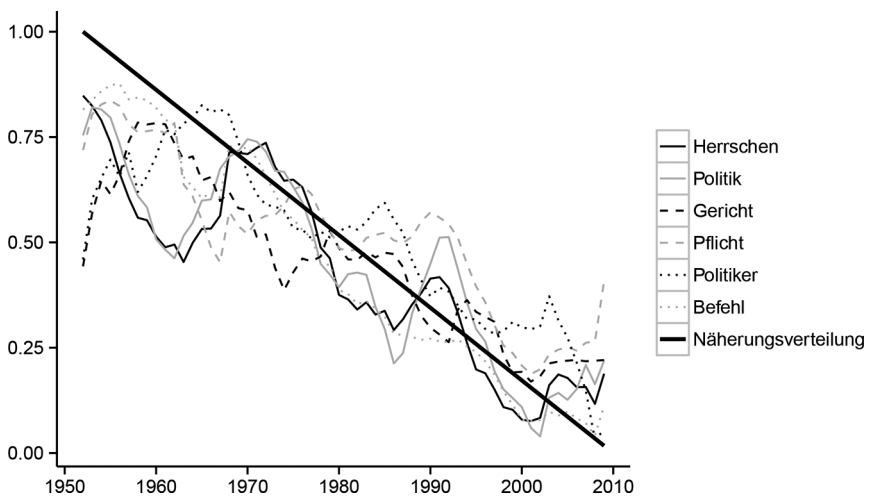


Abbildung 1: Frames, deren jahresweise relative Frequenz im Korpus annähernd stetig abnimmt.

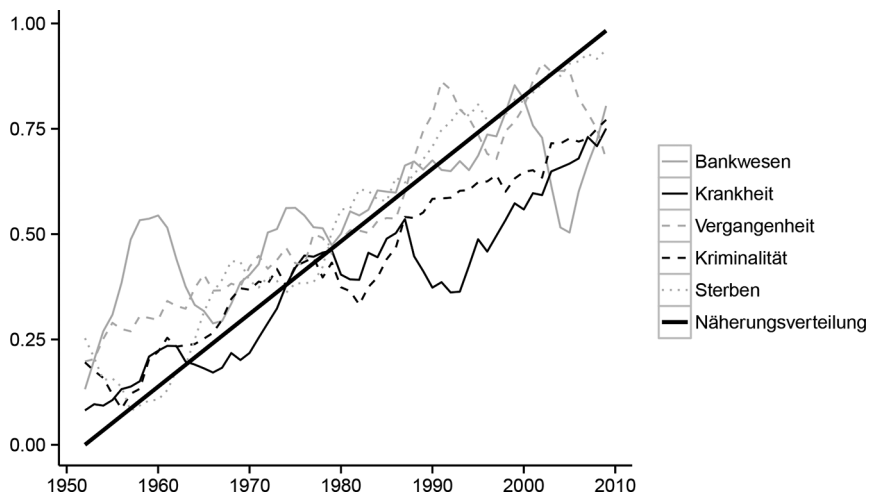


Abbildung 2: Frames, deren jahresweise relative Frequenz im Korpus annähernd stetig zunimmt.

Frames wie „Befehl“, „Politik“, „Politiker“ und „Pflicht“ verweisen auf einen eher traditionellen Wertekanon; solche Frames nehmen über den gesamten Untersuchungszeitraum hinweg ab. Auf der anderen Seite steht das Frame „Bankwesen“, das an Bedeutung zugelegt hat und für die gestiegene Bedeutung der Geldwirtschaft steht, aber auch „Krankheit“ und „Sterben“, was als verstärkte Orientierung von SPIEGEL und ZEIT hin zu religiösen Themen oder zu Themen des Alterns gedeutet werden kann. Weiter auffällig ist „Kriminalität“, was auf eine Zunahme entsprechender Berichterstattung schließen lässt.

Zudem interessierte uns, welche Frames erst in den letzten Jahren eine besondere Konjunktur erlebten. Hierfür bot sich die exponentielle Verteilung als Näherungsverteilung an (Abbildung 3).

Auch hier deutet die starke Zunahme des Frames „Gefahr“, ähnlich wie die Zunahme von „Kriminalität“, auf einen wichtiger werdenden Bedrohungsdiskurs hin. Hinter dem Frame „Fortbewegung“ könnten Mobilitätsthemen stehen, „Freizeit- und Extremsport“ deuten Veränderungen hin zu einer Freizeitgesellschaft an. Auffällig ist auch die Zunahme des Frames „Datenschutz“ in jüngster Zeit (aber vor der Snowden-Geheimdienstaffäre).

Die gezeigten Zeitreihenanalysen von Frames geben einen ersten Eindruck über gesellschaftliche Veränderungen, liefern aber noch keine befriedigende Antwort auf die Frage, wann Umbrüche stattgefunden haben.

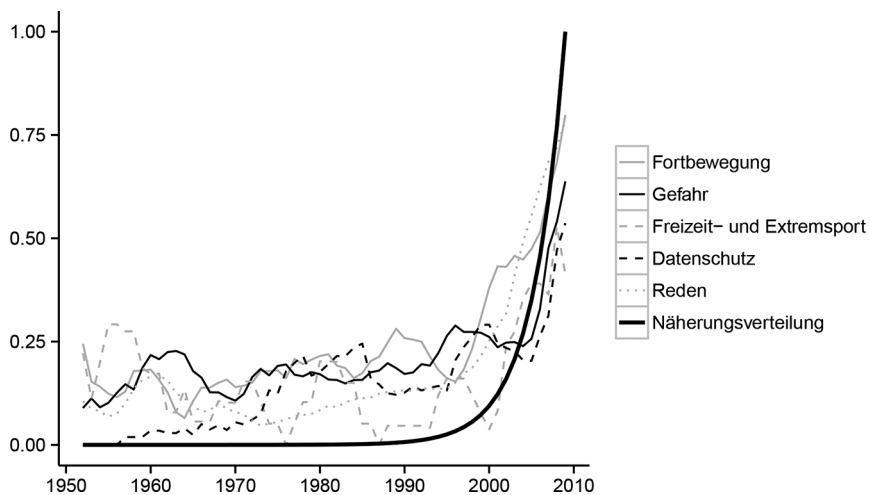


Abbildung 3: Frames, deren jahresweise relative Frequenz im Korpus annahernd exponentiell verteilt ist.

## 2.5 Sprachliche Umbruche als Indikatoren zeitgeschichtlicher Umbruche

Heidrun Kamper definiert sprachliche Umbruche als „Beginn von sprachlichen Veranderungen“, die durch „plotzliche gesellschaftliche oder politische Veranderungen“ induziert werden und Ursache fur einen langerfristigen sprachlichen Wandel sein konnen (Kamper 2007, S. 431). Sprache wird damit eng mit der Geschichte von Gesellschaft, Kultur und Politik verknupft – eine Analyse sprachlicher Umbruche tragt also zu einem vertieften Verstandnis der Qualitat historischer Veranderungen bei.

Kamper (2007, S. 432) geht bei ihren Analysen von Umbruchen aus, die von der Zeitgeschichtsschreibung postuliert werden und sucht nach korrespondierenden sprachlichen Phanomenen. Im Gegensatz dazu berechnen wir Umbruche datengeleitet. Als erste Annaherung dient eine Analyse von Frames, die Zeitbezug reflektieren. Nach Kamper ist es fur Umbruchzeiten typisch, dass die vom Umbruch betroffenen Gesellschaften ihr Selbstbild im Medium zeitreflexiver Diskurse verhandeln (vgl. Kamper, 2007, S. 429). Abbildung 4 zeigt die Entwicklung von Frames mit Zeitbezug im Korpus.

Anhand der Verteilung der Zeitbezug-Frames lieen sich durchaus Umbruchzeiten eingrenzen: etwa die spaten 1950er und fruhen 1960er Jahre, die 1970er Jahre, Anfang der 1990er Jahre und der Anfang der 2000er Jahre. Interessant ist, dass in den genannten Zeitraumen jeweils andere Frames dominieren. Wahrend der Frame „Nahe Vergangenheit“ insbesondere in den 1950er Jahren haufig in den Texten des Korpus auftrat und der Frame „Ver-

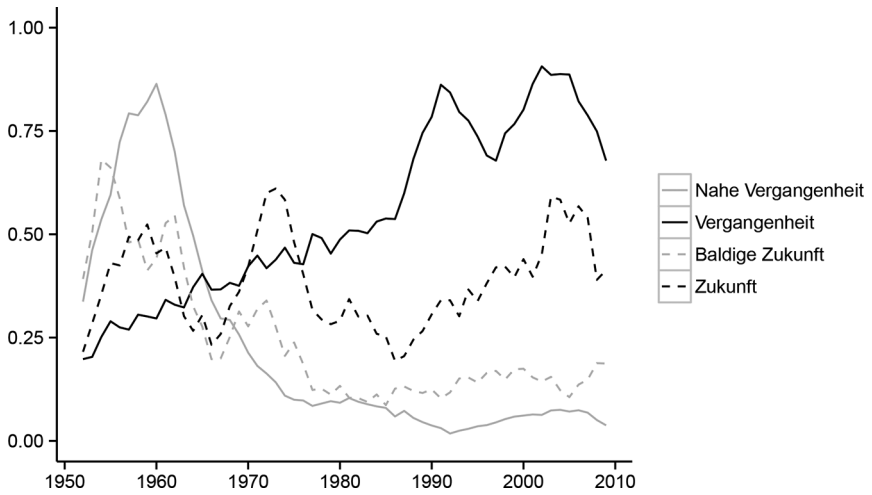


Abbildung 4: Jahresweise Verteilung von Zeitbezug-Frames im SPIEGEL/ ZEIT-Korpus

gangenheit“ Anfang der 1990er Jahre sein Maximum hat, hat der Frame „Zukunft“ in den 1970er Jahren besondere Konjunktur.

Zwar mag die Verteilung von Zeitbezug-Frames ein Indikator für das Vorliegen einer Umbruchzeit sein, für eine präzisere Bestimmung von Umbruchzeiten in der Geschichte der BRD sind wir von der Annahme ausgegangen, dass starke gesellschaftliche Veränderungen mit starken Veränderungen in der Verteilung aller Frames des Framehaushaltes korrelieren.

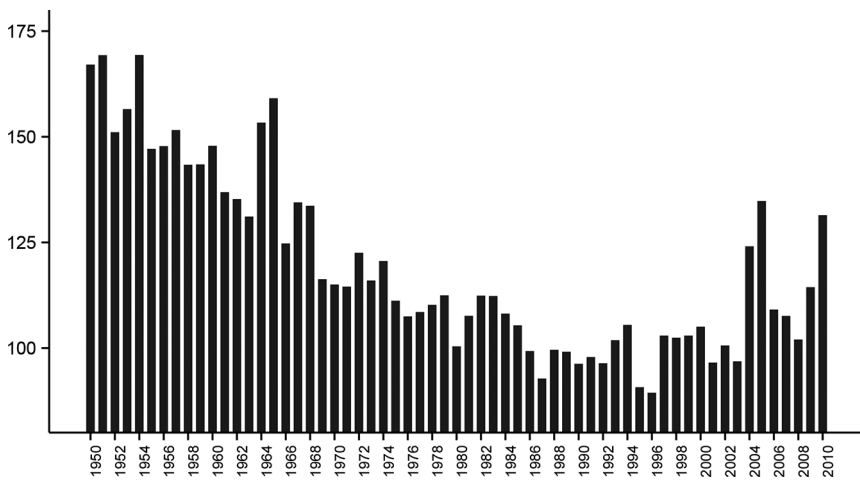


Abbildung 5: Jährliche Summen der Differenzbeträge aller Frames im Vergleich zum Vorjahr im Korpus 1950–2010, normalisierte relative Frequenzen. Zur besseren Visualisierung der Ausschläge wurde die Skalierung der y-Achse angepasst.

Während Zeiten der Stabilität davon gekennzeichnet sind, dass nur geringfügige Verschiebungen im Framehaushalt zu beobachten sind, zeichnen sich Umbrüche dadurch aus, dass ein Teil der Frames stark abnimmt, während ein anderer Teil stark zunimmt. Deshalb halten wir die Summe der Veränderungen (der normalisierten Frequenzen) im Framehaushalt für ein geeignetes Maß für die Identifizierung von Umbruchzeiten.

Wie Abbildung 5 zeigt, sind vor allem in den Jahren 1953/1954, 1963/1964 und 1965, 2003/2004 und 2005 sowie 2009/2010 große Veränderungen in der semantischen Matrix sichtbar. Daneben gibt es kleinere Ausschläge (1967/1968, 1972, 1982/1983 und 1997), wobei in der allgemeinen Tendenz der Grad der Veränderung über die Jahre kontinuierlich abnimmt, der Framehaushalt sich also verfestigt, was sich erst ab 2004 wieder verändert.

Natürlich kann darüber spekuliert werden, ob die identifizierten Umbruchjahre mit zeitgeschichtlichen Ereignissen und Entwicklungen einhergehen, allerdings muss zunächst untersucht werden, welche Frames sich in den Umbruchzeiten besonders stark verändert haben. Exemplarisch sei dies anhand der Umbrüche in den 1950er Jahren und in der ersten Dekade des 21. Jahrhunderts gezeigt.

Abbildung 6 nennt einige Frames, die ihre Maxima in den 1950er Jahren haben: Frames wie „Gute Qualität“, „Ware“, „Buchhaltung, Bilanzierung“, „Tausch, Handel“ stehen für ökonomischen Erfolg der Nachkriegsjahre und „Recht, Gerechtigkeit“, „Ehre, Ruhm“ und „Selbstlos“ für bürgerliche Tugenden der Adenauer-Ära.

Ein anderes Bild zeigt sich für den Beginn des 21. Jahrhunderts (Abbildung 7). Der Frame „Familie, Verwandtschaftsbezeichnungen“ (ggf. in

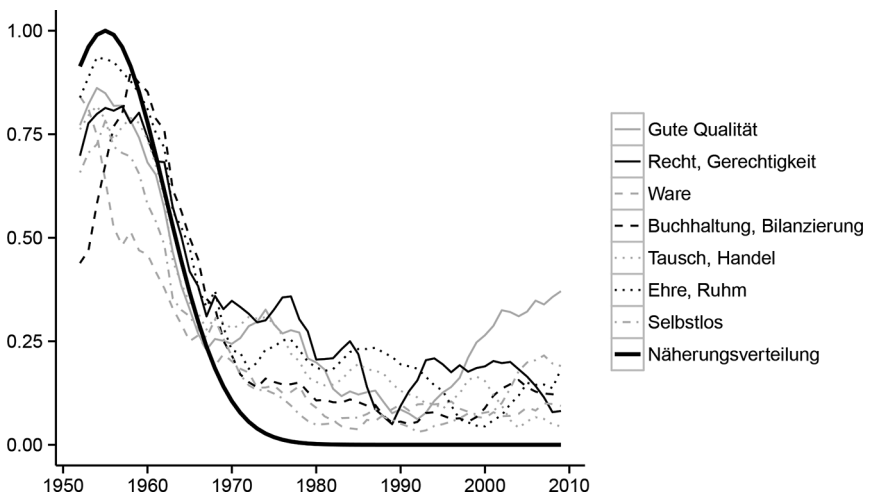


Abbildung 6: Frames im Korpus, die in den spaten 1950er-Jahren hufiger auftreten, als in den Vor- und Folgejahren.

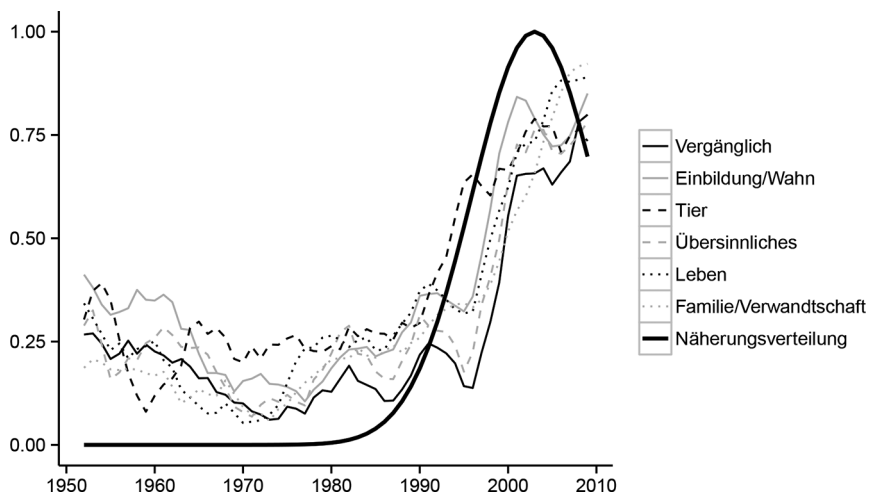


Abbildung 7: Frames im Korpus, die zu Beginn des 21. Jahrhunderts häufiger auftreten als in den Jahren davor.

Kombination mit „Leben“) könnte auf Debatten rund um die Vereinbarkeit von Familie und Beruf, Geschlechtergerechtigkeit aber auch neue Formen des familiären Lebens hinweisen. Der Frame „Übersinnliches“ umfasst neben Lemmata der fiktionalen Welt auch „Traum“ und „Vision“, die wahrscheinlich prägend für die Jahrtausendwende sind. Mit „Einbildung, Wahn“ und den dazu gehörenden Lemmata „Fehler“, „Entgleisung“, „Irrtum“, „Verfehlung“ etc. könnte eine bedeutender gewordene Skandalisierungskultur widerspiegelt werden. Mit Vertretern wie „Geschöpf“ und „Kreatur“ könnte der Frame „Tier“ in Kombination mit dem Frame „Leben“ auf Debatten in den Bereichen Biologie, Ernährung etc. hinweisen.

Wichtiger als die Entwicklung der Auftretenshäufigkeit einzelner Frames oder ihrer Kovariation aber ist die Frage, welche Frames *in einzelnen Texten* überzufällig häufig gemeinsam auftreten und ob hier im Laufe der Jahre, insbesondere während der Umbruchzeiten, Veränderungen zu beobachten sind.

### 3. Vom Wuchern der Rhizome: Das diskursive Geflecht als Kollokationsgraph

#### 3.1 Diskurs als Wissensgeflecht

Diskurse werden oft als „Geflecht“, als „Textgeflecht“ oder „Textensemble“, das einzelne Texte verbindet, beschrieben (Hermanns 1995, S. 86 f.). Diese Texte stehen in einem kommunikativen Bezug zueinander, da sie sich zitieren



ren, kopieren oder ähnliches thematisieren und gleiche oder ähnliche Äußerungen<sup>13</sup> enthalten. Der Diskurs besteht demnach aus einer Menge von über eine Vielzahl von Texten verstreuten Äußerungen, die in der traditionellen linguistischen Diskursanalyse als ein bestimmter thematisch gefasster Diskurs verstanden werden, weil sich die Äußerungen ähnlich sind. Diese Sichtweise präsupponiert Kohärenz zwischen diesen Texten über die in ihnen gefundenen Aussagen. Allerdings gehören die einzelnen Äußerungen nicht nur einem Diskurs an, sondern weisen eine Reihe von weiteren intertextuellen Verknüpfungen auf:

„Aufgrund des Evokationspotentials der Sprachzeichen, die in der sprachlichen Formulierung, die die enoncé trägt, enthalten sind, ist jede einzelne Aussage (und eigentlich jedes einzelne in einem Text eingeführte epistemische Element) eingebettet in einen Kontext weiterer Wissensrahmen und Wissensrahmen-Geflechte.“ (Busse 2006, S. 27)

Der Komplexität dieser Vernetzung, dem „Geflecht von thematisch zusammengehörigen Aussagen“ (Böke et al. 2000, S. 12), werden diskurslinguistische Arbeiten normalerweise aus forschungspragmatischen Gründen nicht gerecht. Ausgehend von einem thematisch definierten Diskurs fokussiert sich die Analyse auf einzelne Bezüge und ignoriert alle anderen. Im Analyseprozess wird aus dem Geflecht ein diskursiver Strang, das Geflecht wird unsichtbar und der Diskurs erscheint als abgeschlossene Einheit. Nimmt man die Metapher des Geflechts jedoch ernst, können Diskurse nicht als abgeschlossene Form betrachtet werden, sondern blitzen vielmehr unter gewissen Perspektiven in einem dynamischen Geflecht, das in alle Richtungen „wuchert“ (Bublitz et al. 1999), auf.

Doch wie können Diskurse als Geflecht dargestellt werden? Gilles Deleuze und Felix Guattari verwenden den Begriff des Rhizoms. Der Begriff entstammt der Botanik und beschreibt dort ein Wurzelgeflecht (z. B. Ingwer), das unterirdisch wuchert und an jeder Stelle ausschlagen kann. Deleuze und Guattari sehen den Rhizombegriff als Gegen-Denkfigur<sup>14</sup> zu den strukturalistischen Baum-Diagrammen, also zu gerichteten, hierarchischen Graphen. Das Rhizom bezeichnet für sie geflechtartige Beziehungen zwischen diversen kulturellen Praktiken, Machtstrukturen und Äußerungsarten (Deleuze/Guattari 1992, S. 12), beschreibt aber auch „kollektive Äußerungsgefüge“ (De-

---

13 Zur Differenz von Aussage und Äußerung vgl. Angermüller (2007).

14 Den Begriff der Metapher lehnen Deleuze und Guattari ab: „Wir machen absolut keinen metaphorischen Gebrauch von diesen Begriffen... Wir meinen das so, wie wir es sagen: buchstäblich“ (Deleuze/Parnet 1980, S. 25).

leuze/Guattari 1996, S. 16). Die Denkfigur des Rhizoms ermöglicht eine de-hierarchisierte und dezentrierte Repräsentation kollektiven Wissens. Jeder Punkt, d. h. jede Äußerung, jede Praktik, ist potenziell über „Linien“ mit jedem anderen verknüpft, von jedem Punkt können Linien weg verweisen, um jeden Punkt können sich neue Punkte gruppieren. Das Rhizom hat keinen Ursprung, wohl aber Bündelungszentren und fixe Achsen, aus denen es wächst (Scharloth et al. 2013).

In „Was ist Philosophie?“ entwickeln Deleuze und Guattari (1996) die Vorstellung, „Karten“ (ebd., S. 44) zu zeichnen, um das Funktionieren von philosophischen Begriffswelten zu charakterisieren. Diese Karten zeigen, in welchen „Regionen“ Begriffe sich mit welchem „Wucherungsgrad“ und mit welchen „Rückkoppelungen, Verbindungen“ zu anderen sie sich bewegen (ebd., S. 48). Es liegt also nahe, solche Karten als (ungerichtete, nicht-hierarchische) Graphen zu visualisieren und dabei zu versuchen, datengeleitet vorzugehen, um sich nicht auf einen bestimmten Diskurs festzulegen, sondern das komplexe Geflecht von intertextuellen Bezügen zu berechnen und zu visualisieren.

### 3.2 Datengeleitete Verfahren und Visualisierung

Visualisierungen von Daten dienen einerseits der Illustration von gewonnener Erkenntnis, beispielsweise in der Form von Balken-, Streu- oder Liniendiagrammen, die Mess- oder Zählwerte repräsentieren. Solche Visualisierungen werden „presentation graphics“ (Präsentationsgrafiken) genannt (Chen et al. 2008, S. 4). Andererseits sind Visualisierungen aber auch eigenständige Mittel der Erkenntnisgewinnung, wenn andere Formen der Repräsentation von Wissen (wie Listen, Tabellen oder Texte) zu umfangreich oder zu komplex sind, um als Ganze erfasst und gedeutet werden zu können. Dies ist bei datengeleiteten Analysen der Normalfall. Visualisierungen dieser Art werden zur Gruppe der „exploratory graphics“ (explorative Visualisierungen) gezählt (Schumann/Müller 1999, S. 5; Chen et al. 2008, S. 5).

Explorative Visualisierungsmethoden werden insbesondere im Bereich der „Visual Analytics“ (Keim et al. 2010; Chen et al. 2008) eingesetzt. Visualisierungen transformieren, gewichten und filtern komplexe Daten und bringen sie dadurch in eine Form, die sie als Informationen erfassbar und interpretierbar machen. Visualisierungen sind damit keine Abbildungen der Wirklichkeit, sondern aufgrund von Relevanzkriterien geordnete und damit interpretative Reduktionen von Daten, die auf der Basis gestalterischer Vorgaben visuell repräsentiert werden.

Der Nutzen von explorativen Visualisierungsmethoden zeigt sich insbesondere bei datengeleiteten Analysen. Beispielgebend sind die technischen

und naturwissenschaftlichen Disziplinen, wo vor allem in Mathematik, Informatik, Naturwissenschaften und *life-sciences* bei der Analyse komplex vernetzter Daten mit Visualisierungen gearbeitet wird. Auch im Bereich der Sozialwissenschaften wird der Nutzen von Visualisierungen immer öfter entdeckt, etwa im Bereich der Netzwerkanalyse. In den Kultur- und Geisteswissenschaften steckt die Entwicklung von durch Visualisierung geleiteten Methoden noch am Anfang, zeigt aber z. B. im Paradigma der Visuellen Textanalyse (Risch et al. 2008; Rohrdantz et al. 2010) ihr enormes Potenzial.

### 3.3 Kollokationen und Kollokationsgraphen

Der Begriff der Kollokation ist in der Linguistik durchaus umstritten (vgl. Evert 2009, S. 1212 f.). Übereinstimmend wird jedoch in den meisten Definitionen davon ausgegangen, dass es sich bei Kollokationen um rekurrente Kookkurrenzen von mindestens zwei lexikalischen Einheiten in einer begrenzten Menge sprachlicher Daten handelt. Typische Beispiele für Kollokationen sind „Weg“ und „einschlagen“ oder „Geschichte“ und „erzählen“, die in sprachlichen Daten jeweils überzufällig häufig miteinander auftreten. Je nach Erkenntnisinteresse und Anwendungsgebiet der Kollokationsanalysen werden weitere Bestimmungen als konstitutive Bestandteile des Kollokationsbegriffs formuliert: etwa im Hinblick auf die Art der Relation zwischen den Wörtern oder im Hinblick auf die Menge der Sprachdaten, in denen die lexikalischen Einheiten gemeinsam vorkommen. So definiert Bartsch (2004, S. 76), Kollokationen seien „lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other“. Evert (2009, S. 1213 f.) dagegen, dem wir in unserem Projekt folgen, unterscheidet einen rein statistischen Kollokationsbegriff von dem Begriff der *multiword expression*. Ersterer bezieht sich auf ein in natürlichen Sprachen direkt beobachtbares gemeinsames Auftreten von Wörtern, letzterer basiert auf theoretischen Annahmen über die Semantizität und Pragmatizität der Wortverbindungen und ist somit in höherem Maße von theoriegeleiteten Interpretationen abhängig.

Die Grundlage für die Berechnung von Kollokationen ist die Bestimmung von Kollokatoren. Kollokatoren sind Wörter, die gemeinsam mit einem bestimmten Wort, der „Basis“, auftreten. Je nach linguistischem Erkenntnisinteresse legt man ein Fenster von  $n$  Wörtern links und  $m$  Wörtern rechts der Basis fest und wertet dieses Fenster bei jedem Auftreten der Basis im Korpus aus. So erhält man eine Liste von Wörtern, die gemeinsam mit der gesuchten Basis auftreten, sowie ihrer Frequenzen. Die Liste der Kollokatoren alleine ist jedoch noch nicht aussagekräftig. Funktionswörter wie Artikel oder Präpositionen treten als Angehörige von geschlossenen Wortklassen insgesamt sehr

viel häufiger auf als Adjektive, Verben oder Nomen. Das häufige gemeinsame Auftreten des Wortes „der“ mit dem Wort „Weg“ reicht nicht dafür aus, „der“ als Kollokation von „Weg“ zu klassifizieren. Um zu überprüfen, ob eine Basis und ein Kollokator überzufällig häufig miteinander auftreten, muss untersucht werden, ob der Kollokator in anderen Kontexten nicht ähnlich häufig auftritt wie im Kontext der Basis. Ein statistischer Signifikanztest gibt darüber Auskunft, ob der potenzielle Kollokator überzufällig häufig zusammen mit der Basis auftritt.<sup>15</sup>

Da es in unserem Projekt darum geht, Frames zu identifizieren, die in Zeitungstexten signifikant häufig zusammen vorkommen, ist das Konstrukt der Kollokation, verstanden als ein überzufälliges gemeinsames Auftreten zweier Phänomene, ein geeigneter Anknüpfungspunkt. Allerdings sind folgende Anpassungen nötig:

- Für die Berechnung von Kollokationen auf der Basis von Lexemen ist die Linearität sprachlicher Daten entscheidend, weshalb die Definition des Suchfensters links und rechts der Basis einen großen Einfluss auf die Berechnung hat. Im Fall der Frameanalyse fällt dieser Aspekt weg, da in einem Text ein Frame höchstens einmal vorkommen kann, unabhängig von der Länge des Textes.
- Die Größe des Fensters zur Bestimmung von Kollokatoren ist deswegen variabel und entspricht jeweils der Menge der im jeweiligen Text vorkommenden Frames.
- Als Vergleichsgröße für die Berechnung der Signifikanz des gemeinsamen Auftretens der Frames wurde nicht die Anzahl der laufenden Wortformen in den Teilkorpora gewählt, sondern die Anzahl der Texte mit bzw. ohne gemeinsames Auftreten.<sup>16</sup>

Während sich Informationen über die Kollokationen zu einer einzigen Basis leicht als Liste oder Tabelle überblicken lassen, ist die Analyse von Kollokationsnetzen, also Informationen über die Kollokationen mehrerer Basen, die selbst Kollokationen anderer Basen sein können, bei einer nennenswerten Menge an Verbindungen ohne Visualisierung nicht mehr möglich. Basis-Kollokationen-Verbindungen und Kollokationsnetze lassen sich aber beispielsweise als gerichtete Graphen visualisieren. Graphen sind – einfach ge-

---

15 Für eine weiterführende Darstellung in Frage kommender statistischen Verfahren vgl. Evert (2009, S. 1224–1236).

16 Frame- und Kollokationsanalyse wurde mittels selbst geschriebener Software durchgeführt, ebenso wie die Überführung der Einzelergebnisse in ein Kollokationsnetz. Das Setzen des Graphen erfolgte mit Hilfe des GraphViz-Software-Pakets (Ellson et al. o. J.) und mit Gephi ([www.gephi.org](http://www.gephi.org)).

sprochen – eine Menge von Punkten („Knoten“ / „vertices“ / „nodes“), zwischen denen Linien („Kanten“ / „edges“) verlaufen. In unserem Projekt repräsentieren die Punkte Frames, die Kanten signifikante Kookkurrenzen zwischen zwei Knoten.

Es gibt unzählige Möglichkeiten einen Graphen mit mehreren Knoten und Kanten zu zeichnen.<sup>17</sup> Insbesondere für komplexe Netze wurden verschiedene Algorithmen entwickelt, die automatisiert das Layout eines Graphen erzeugen. Dabei kommen auch ästhetische Kriterien zur Anwendung, damit die Lesbarkeit des Graphen gesichert und optimiert wird. Dies sind unter anderen:

1. die Minimierung der Überschneidungen von Kanten,
2. die Minimierung der Fläche,
3. die Minimierung der Länge der Kanten, (vgl. Fleischer/Hirsch 2001, S. 19 f.)
4. die Verhinderung des Überschreibens von Knoten in sehr großen Graphen.

Da über die Struktur der Graphen in unserem Projekt nichts vorgängig bekannt ist, wir beispielsweise keine Hierarchien oder Verwandtschaften voraussetzen können, bietet sich die Verwendung einer Layout-Methode an, die prinzipiell für alle Arten von Graphen geeignet ist. Die Wahl fiel auf die forced-based-Methode (Kräftediagramm), genauer das „Force Atlas 2“-Layout, wie es in Gephi<sup>18</sup> implementiert ist. Diese Methode modelliert einen Graphen als ein physikalisches System, das von Kräften (forces) zwischen den Knoten gebildet wird (Fleischer/Hirsch 2001, S. 20). Man kann sich dabei die Kanten als Sprungfedern und die Knoten als Ringe vorstellen, in die die Sprungfedern eingehakt sind. Erlaubt man einem solchen physischen Graphen, sich ohne Beschränkung zu bewegen, wird er in einen Zustand kommen, in dem die potentielle Energie in den Federn möglichst gering ist (Landgraf 2001, S. 173). Die für die Optimierung des Graphen-Layout hilfreiche Analogie dabei ist, dass ein Graph dann Ansprüche an seine Informativitäts- und die oben genannten Lesbarkeits-Kriterien am besten erfüllt, wenn er sich in einem Zustand der Entspannung, d.h. in einem Zustand mit geringer energetischer Ladung befindet (vgl. Brandes 2001, S. 71).

---

17 Eine Übersicht bieten die Standardwerke und Kaufmann/Wagner (2001) und Tamassia (2013).

18 Gephi ist eine OpenSource-Software zur Visualisierung und Manipulation von Graphen (vgl. [www.gephi.org](http://www.gephi.org)).

Die Implementierung von Algorithmen zur Sicherung der Lesbarkeit der Graphen hat zur Folge, dass nicht alle Aspekte eines Graphen als Visualisierung relevanter Informationen gelesen werden können. So sind beispielsweise die Kantenlängen und die Winkel zwischen zwei Kanten nicht bedeutungstragend. Der Algorithmus sorgt jedoch dafür, dass zwei oder mehr Knoten, die mit einer gleichen oder sehr ähnlichen Menge anderer Knoten verbunden sind, in relativer Nähe zueinander dargestellt werden. Verdichtungsbereiche im Graphen lassen also auf eine Gruppe von Knoten mit ähnlichen Vernetzungsmerkmalen schließen. Solche Verdichtungsbereiche können darüber hinaus über eine Netzwerkanalyse wie der „Louvain-Methode“ (Blondel et al. 2008) visualisiert werden: In den weiter unten verwendeten Graphen sind Knoten, die viele Verbindungen untereinander aufweisen und deshalb einen Verdichtungsbereich darstellen, durch gleiche Einfärbung codiert.

Kollokationsgraphen bieten damit die Möglichkeit, Frames mit ähnlichen Vernetzungsmerkmalen anhand von Verdichtungsräumen im Graphen visuell zu identifizieren. Zudem kann es interessant sein, jene Kanten zu untersuchen, durch die verschiedene Frame-Cluster miteinander verknüpft sind. Aufgrund der Komplexität der Daten wären diese Analysen ohne Visualisierung kaum möglich. Im Folgenden wollen wir zeigen, wie Kollokationsgraphen genutzt werden können, um die Qualität von Umbrüchen zu untersuchen.

#### **4. Datengeleitete Frameanalyse am Beispiel des SPIEGEL/ZEIT-Korpus II: Wandel von Framekollokationen in Umbruchzeiten**

Ausgehend von den in Abschnitt 2 berechneten Umbruchjahren haben wir berechnet, welche Framekonstellationen sich im Zuge der Umbrüche stark verändert haben.

Um die Qualität des Umbruchs, der sich in den Jahren 2001 bis 2008 vollzogen hat, genauer zu untersuchen, haben wir für die Jahre 2001 bis 2004 und 2005 bis 2008 je einen Kollokationsgraphen berechnet. Dabei haben wir sämtliche Texte aus den jeweiligen Untersuchungszeiträumen einer automatisierten Frameanalyse unterzogen und daran anschließend untersucht, welche Frames signifikant häufig im jeweiligen Zeitraum miteinander in Einzeltexten kookkurrieren. Um die Unterschiede der beiden Zeitabschnitte leichter vergleichen zu können, haben wir Knoten und Kanten beider Rhizome zu einem gemeinsamen Frame-Kollokationsgraphen vereinigt, wobei für die Jahre ab 2005 bzw. die Jahre bis und mit 2004 je die Kanten farblich hervorgehoben wurden.



schaftsthemen.<sup>19</sup> In der Grafik (und analog in allen Grafiken weiter unten) sind Verknüpfungen, die erst für die Zeit nach dem Umbruch typisch sind, schwarz hervorgehoben. Verbindungen, die für die Zeit vor dem Aufbruch typisch sind, sind gestrichelt. Auffallend ist die neue Verbindung zwischen den Frames „Verleihen“ (Lexeme wie „anlegen“, „verleihen“, „Darlehen“, „Kredit“ etc.) und „Tausch, Handel“ (Lexeme wie „kaufen“, „spekulieren“, „Umsatz“, „Börse“, „Markt“ etc.), während bis 2004 mit „Verleihen“ eher „Kaufen“ und „Tausch, Handel“ mit „Handel“ kookkurriert. Diese Veränderung könnte das wiedererwachte Vertrauen in die Märkte nach der sog. Dot-com-Blase im Jahr 2000 widerspiegeln. Die Berichterstattung in SPIEGEL und ZEIT scheinen zudem die Finanzwirtschaft verstärkt als Kredit- und Börsen-zentriert darzustellen – eine Welt, die ein paar Jahre später mit der globalen Finanzkrise an Attraktivität einbüßen wird.

Der Kampf zwischen Wirtschaft und Politik um das Primat in der Demokratie widerspiegelt sich auch im Cluster „Politik“ (Abbildung 10): Die neuen Verbindungen zwischen den Frames „Partei“, „Wille“ deuten eine politische Polarisierung an, bei der sich die Parteien im Bundestags-Wahljahr 2005 profilieren wollen. Der politische Gestaltungswille steht zudem möglicherweise im Konflikt mit dem Grundgesetz (Frame „Gesetz“). Zudem wird die Politik an die „Pflicht“ erinnert, zu handeln.

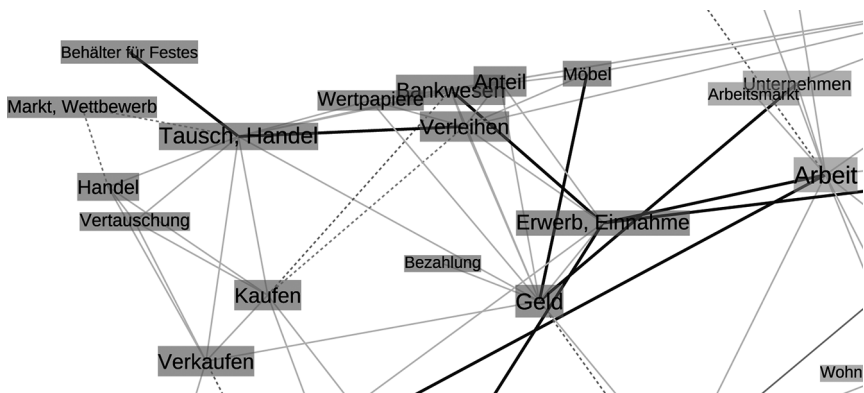


Abbildung 9: Cluster „Bankwesen“ Umbruchjahre 2004/2005.

19 Beim Frame „Möbel“ handelt es sich um eine Fehlannotation, da mit dem darin vorhandenen Lexem „Bank“ im Kontext das Finanzinstitut gemeint ist.



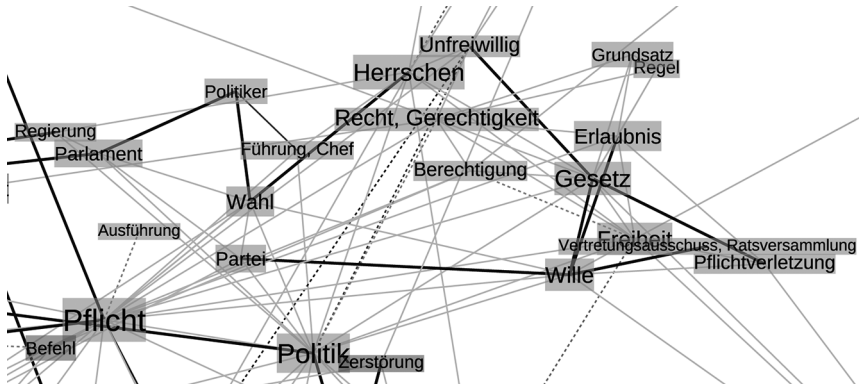


Abbildung 10: Cluster „Politik“ Umbruchjahre 2004/2005

Ein weiterer Cluster dreht sich um Frames im Bereich Familie, Schule, Kindheit (Abbildung 11). Hinter dem Frame „Fruchtbarkeit“ verbergen sich die Lexeme „Eltern“, „Elterngeneration“ und „Kind“, „Neugeborenes“ und der Frame „Vorbereitung“ steht unter anderem für eine Reihe von Lexemen im Bereich der Ausbildung. Mit dem Frame „Studium, Universität“ wird mit den Umbruchjahren 2004/2005 also ein neuer Fokus auf Bildungsthemen deutlich.

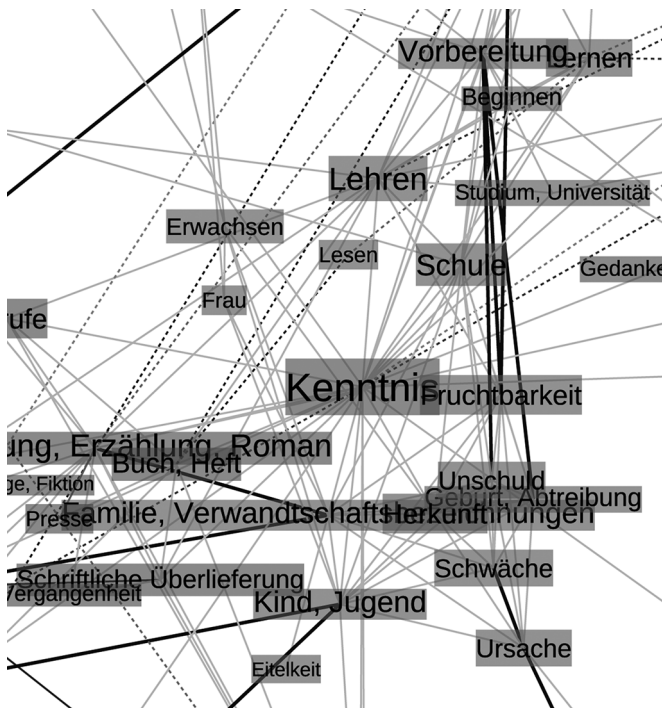


Abbildung 11: Cluster „Familie“ Umbruchjahre 2004/2005





ist die neue Kookkurrenz der Frames „Geld“ und „Kaufen“ sichtbar, eigentlich keine überraschende Verbindung, die jedoch zehn Jahre später nicht mehr typisch ist (stattdessen korreliert „Geld“ mit „Verkaufen“). Dies könnte eine Wirtschaftsberichterstattung andeuten, die das Investitionsvermögen der Wirtschaft thematisiert, gerade auch im Hinblick auf die wirtschaftlichen Herausforderungen im wiedervereinigten Deutschland.

## 5. Fazit

Unser Beitrag sollte zeigen, welche Chancen die digitale Welt auch für diskurslinguistische Fragestellungen bietet: Noch nie war es so einfach, an sehr große Datenmengen für die Analyse zu kommen. Dies alleine ist jedoch nur von Nutzen, wenn sich die Analysewerkzeuge anpassen. Einerseits kann aus einem reichen Fundus von Methoden aus der Computer- und Korpuslinguistik, aber auch aus der Informatik (Data Mining), Statistik etc. geschöpft werden. Andererseits ermöglichen die Datenmenge und die verfügbaren Methoden aber auch einen neuen Zugang zu Daten, nämlich datengeleitete Verfahren, die weniger stark mittels traditioneller Analysekatégorien bestimmte Hypothesen überprüfen, sondern strukturentdeckend funktionieren und eher der Hypothesengewinnung dienen. Es ist unsere Überzeugung, dass solche Ansätze bestimmten Prämissen der Diskursanalyse entgegen kommen: Das rhizomartige Wuchern von Diskursen, über Textgrenzen hinweg, jeglicher Hierarchie trotzend, kann mit datengeleiteten Verfahren, die sich nicht von vornherein bezüglich Datengrundlage und Themen beschränken, eher nachgebildet werden. Es ist nicht mehr notwendig, sich aus forschungspraktischen Gründen auf eine bestimmte Auswahl von Texten zu einem gesetzten Thema zu beschränken. Dies ermöglicht es, neue Forschungsfragen zu verfolgen, wie z. B. die unsrige, datengeleitet zeitgeschichtliche Umbrüche durch Frameanalysen zu entdecken und genauer zu beschreiben.

Die vorgestellten Methoden sind ein erster Schritt, der zeigen soll, in welche Richtung es gehen könnte. Eine Reihe weiterer Methoden für den Umgang mit Big Data könnten für diskurslinguistische Zwecke genutzt werden und wir hoffen, dass viele weitere Wissenschaftlerinnen und Wissenschaftler damit beginnen, in diese Richtung zu experimentieren. Eine Bereitschaft, sich auf technische Fragestellungen einzulassen ist dabei unumgänglich. Idealerweise lernt man dafür programmieren, denn der autonome Umgang mit Technik ist die Grundbedingung für eine selbstbestimmte Forschungstätigkeit im digitalen Zeitalter.

## Literatur

- Anderson, C. (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. In: Wired Magazine Band 16.07, [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory/](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/) [Abruf: 14. Mai 2014].
- Angermüller, J. (2007): Diskurs als Aussage und Äußerung – die enunziative Dimension in den Diskurstheorien Michel Foucaults und Jacques Lacans. In: Warnke, I. (Hrsg.): Diskurslinguistik nach Foucault, Linguistik – Impulse & Tendenzen. Berlin und New York: de Gruyter, S. 53–80.
- Bartsch, S. (2004): Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence: Gunter Narr Verlag.
- Blondel, V. D./Guillaume, J.-L./Lambiotte, R./Lefebvre, E. (2008): Fast unfolding of communities in large networks. In: Journal of Statistical Mechanics: Theory and Experiment, Nr. 10, S. P10008.
- Böke, K./Jung, M./Niehr, T./Wengeler, M. (2000): Vergleichende Diskurslinguistik. Überlegungen zur Analyse national heterogener Textkorpora. In: Niehr, T./Böke, K. (Hrsg.): Einwanderungsdiskurse: vergleichende diskurslinguistische Studien. Wiesbaden: Westdeutscher Verlag, S. 11–36.
- Brandes, U. (2001): Drawing on Physical Analogies. In: Kaufmann, M./Wagner, D. (Hrsg.): Drawing Graphs. Methods and Models. Berlin, Heidelberg und New York et al.: Springer, S. 71–86.
- Bubenhof, N. (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse, Sprache und Wissen. Berlin und New York: de Gruyter.
- Bubenhof, N./Scharloth, J. (2013a): Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren. In: Warnke, I./Meinhof, U./Reisigl, M. (Hrsg.): Diskurslinguistik im Spannungsfeld von Deskription und Kritik, Diskursmuster – Discourse Patterns. Berlin: Akademie-Verlag, S. 147–168.
- Bubenhof, N./Scharloth, J. (2013b): Korpuspragmatische Methoden für kulturanalytische Fragestellungen. In: Benitt, N./Koch, C./Müller, K./Schüler, L./Saage, S. (Hrsg.): Kommunikation Korpus Kultur: Ansätze und Konzepte einer kulturwissenschaftlichen Linguistik, Giessen Contributions to the Study of Culture. Trier: WVT, S. 47–66.
- Bublitz, H./Bührmann, A. D./Hanke, C./Seier, A. (1999): Das Wuchern der Diskurse. Frankfurt und New York: Campus Verlag.
- Busse, D. (1987): Historische Semantik. Analyse eines Programms, Sprache und Geschichte. Stuttgart: Klett-Cotta.
- Busse, D. (2006): Text – Sprache – Wissen. Perspektiven einer linguistischen Epistemologie als Beitrag zur Historischen Semantik. In: Scientia Poetica 10, S. 101–137.
- Busse, D./Teubert, W. (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, D./Hermanns, F./Teubert, W. (Hrsg.): Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik. Opladen: Westdeutscher Verlag, S. 10–28.
- Chen, C./Härdle, W./Unwin, A. (2008): Handbook of data visualization, Springer handbooks of computational statistics. Heidelberg und Berlin: Springer.
- Dahinden, U. (2006): Framing: Eine integrative Theorie der Massenkommunikation. Konstanz: UVK.
- Deleuze, G./Guattari, F. (1992): Tausend Plateaus: Kapitalismus und Schizophrenie. Berlin: Merve.
- Deleuze, G./Guattari, F. (1996): Was ist Philosophie? Frankfurt am Main: Suhrkamp.

- Deleuze, G./Parnet, C. (1980): *Dialoge*. Frankfurt am Main: Suhrkamp.
- Dornseiff, F. (2004): *Der deutsche Wortschatz nach Sachgruppen: Mit einer lexikographisch-historischen Einführung und einer ausführlichen Bibliographie zur Lexikographie und Onomasiologie*. Berlin und New York: Walter de Gruyter.
- Ellson, J./Gansner, E./Hu, Y./Bilgin, A.: *Graphviz – Graph Visualization Software*. <http://www.graphviz.org> [Abruf: 14. Mai 2014].
- Evert, S. (2009): *Corpora and Collocations*. In: Lüdeling, A./Kytö, M. (Hrsg.): *Corpus Linguistics: An International Handbook*, *Handbooks of linguistics and communication science*. Band 2. Berlin und New York: Walter de Gruyter, S. 1212–1248.
- Fleischer, R./Hirsch, C. (2001): *Graph Drawing and Its Applications*. In: Kaufmann, M./Wagner, D. (Hrsg.): *Drawing Graphs. Methods and Models*. Berlin, Heidelberg und New York et al.: Springer, S. 1–22.
- Goffman, E. (1974): *Frame Analysis*. Garden City and New York: Doubleday.
- Hermanns, F. (1995): *Sprachgeschichte als Mentalitätsgeschichte. Überlegungen zu Sinn und Form und Gegenstand historischer Semantik*. In: Gardt, A./Mattheier, K./Reichmann, O. (Hrsg.): *Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien*. Tübingen: Niemeyer, S. 69–101.
- Kämper, H. (2007): *Linguistik als Kulturwissenschaft. Am Beispiel einer Geschichte des sprachlichen Umbruchs im 20. Jahrhundert*. In: Kämper, H./Eichinger, L. M. (Hrsg.): *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache, Studien zur deutschen Sprache*. Tübingen: Narr, S. 419–439.
- Kaufmann, M./Wagner, D. (2001): *Drawing Graphs: Methods and Models*. Berlin und New York: Springer.
- Keim, D. A./Kohlhammer, J./Ellis, G./Mansmann, F. (2010): *Mastering the Information Age – Solving Problems with Visual Analytics*. Goslar: Eurographics Association.
- Konerding, K.-P. (1993): *Frames und lexikalisches Bedeutungswissen: Untersuchungen zur linguistischen Grundlegung einer Frametheorie und zu ihrer Anwendung in der Lexikographie*. Berlin und New York: Walter de Gruyter.
- Landgraf, B. (2001): *3D Graph Drawing*. In: Kaufmann, M./Wagner, D. (Hrsg.): *Drawing Graphs. Methods and Models*. Berlin, Heidelberg und New York et al.: Springer, S. 172–192.
- Neidhardt, F./Rucht, D. (1993): *Auf dem Weg in die „Bewegungsgesellschaft“?* In: *Soziale Welt* 44, S. 305–326.
- Nelson, T. E./Clawson, R. A./Oxley, Z. M. (1997): *Media Framing of a Civil Liberties Conflict and its Effect on Tolerance*. In: *The American Political Science Review* 91(3), S. 567–583.
- Perkuhn, R./Belica, C. (2006): *Korpuslinguistik – Das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik*. In: *Sprachreport* 22(1), S. 2–8.
- Rettie, R. (2004): *Using Goffman’s frameworks to explain presence and reality*. In: *7th Annual International Workshop on Presence*. Valencia: ISPR, S. 117–124.
- Risch, J./Kao, A./Poteet, S./Wu, J. Y.-J. (2008): *Text Visualization for Visual Text Analytics*. In: Simoff, S./Böhlen, M./Mazeika, A. (Hrsg.): *Visual Data Mining, Lecture Notes in Computer Science*. Band 4404. Berlin und Heidelberg: Springer, S. 154–171.
- Rohrdantz, C./Koch, S./Jochim, C./Heyer, G./Scheuermann, G./Ertl, T./Schütze, H./Keim, D. A. (2010): *Visuelle Textanalyse*. In: *Informatik-Spektrum* 33(6), S. 601–611.
- Scharloth, J./Bubenhof, N. (2011): *Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse*. In: Felder, E./Müller, M./Vogel, F. (Hrsg.): *Korpusprag-*

- matik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen. Berlin und New York: de Gruyter, S. 195–230.
- Scharloth, J./Eugster, D./Bubenhof, N. (2013): Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn. In: Busse, D./Teubert, W. (Hrsg.): Linguistische Diskursanalyse. Neue Perspektiven. Wiesbaden: VS, S. 345–380.
- Schiller, A./Teufel, S./Thielen, C. (1995): Guidelines für das Tagging deutscher Textcorpora mit STTS. Stuttgart: Universität Stuttgart.
- Schmid, H. (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. Stuttgart: Universität Stuttgart.
- Schumann, H./Müller, W. (1999): Visualisierung: Grundlagen und allgemeine Methoden. Heidelberg und Berlin: Springer.
- Spitzmüller, J./Warnke, I. H. (2011): Diskurslinguistik: eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse. Berlin und New York: Walter de Gruyter.
- Tamassia, R. (2013): Handbook of Graph Drawing and Visualization. London: Chapman & Hall/CRC.
- Teubert, W./Cermáková, A. (2007): Corpus Linguistics: A Short Introduction. London: Bloomsbury Academic.
- Ziem, A. (2008): Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz., Sprache und Wissen. Berlin und New York: de Gruyter.