Unverbindlicher Preprint von: Bubenhofer, Noah/Scharloth, Joachim: 95. Kulturwissenschaftliche Orientierung in der Computer- und Korpuslinguistik. In: Sprache - Kultur - Kommunikation / Language - Culture - Communication Ein internationales Handbuch zu Linguistik als Kulturwissenschaft / An International Handbook of Linguistics as a Cultural Discipline. Bd. 43. Berlin, Boston : De Gruyter, 2016, S. 924–933

95. Kulturwissenschaftliche Orientierung in der Computer- und Korpuslinguistik

Noah Bubenhofer, Joachim Scharloth

1. Einleitung

Das im Zuge der Digitalisierung gewachsene Interesse an empirischen, quantitativen Analysen von authentischen Sprachdaten hat auch jene Bereiche der Linguistik erfasst, die sozial- und kulturwissenschaftlich orientiert sind. Ihr Methodenrepertoire speist sich aus zwei Forschungstraditionen, die mehr und mehr zusammenwachsen: dem *Information Retrieval* und *Text Mining* einerseits und der Computer- und Korpuslinguistik andererseits. Digitalisierung bedeutet zählbar machen und so ist beiden Forschungsrichtungen gemeinsam, dass sie sprachliche Merkmale als Zahlen darstellen und mit Hilfe mathematischer Methoden untersuchen. Sprachliche Daten variieren im Gegensatz zu vielen anderen Datentypen in sehr vielen Dimensionen, sind daher auch eher niederfrequent verteilt und können auf unterschiedlichen Ebenen (Syntax, Lexik, Semantik etc.) analysiert werden. Texte sind Merkmalsvektoren, die die Distribution von Texteigenschaften repräsentieren. Eine Sammlung mehrerer Texte (Korpus) bildet eine Matrix.

Ansätze aus dem Bereich des *Text Mining* zielen häufig darauf, Modelle zu finden, die das Auffinden bestimmter Informationen in großen Textmengen ermöglichen. Sie lassen sich oft als Klassifikationsprobleme beschreiben, die mit Hilfe maschinellen Lernens gelöst werden. Dabei werden bereits klassifizierte Dokumente auf ihre linguistischen Eigenschaften hin analysiert (z.B. Distribution von Lemmata, Wortkombinationen, Buchstaben-n-Grammen etc.) und ein Klassifikator aus Merkmalsausprägungen bestimmt, der die vorher definierten Klassen möglichst gut trennt. Dieser Klassifikator kann dann dazu genutzt werden, künftige Klassifikationsaufgaben zu lösen, etwa um Korpora nach Textsorten zu sortierten oder Textstellen mit besonders hohem Informationsgehalt zu identifizieren. Hier wird deutlich, dass *Text Mining* eine starke Orientierung zur angewandten Forschung hat. Entsprechend ist die Frage, welche linguistischen Merkmale für die Klassifikation von besonderer Bedeutung sind und ob diese soziokulturelle Korrelate haben, von geringem Interesse.

Im Folgenden beschränken wir unsere Darstellung daher auf die sozial- und kulturwissenschaftlich interessierte Computer- und Korpuslinguistik. Hierfür werden wir zunächst die theoretischen Grundlagen (Kap. 2), dann unterschiedliche Forschungsparadigmen (Kap. 3) vorstellen; im Anschluss werden wir einen groben Überblick über grundlegende Typen von Analysekategorien geben (Kap. 4), ehe wir im letzten Kapitel die Bedeutung von Visualisierungen für die Analyse skizzieren (Kap.5).

2. Theoretische Grundlage: Pragmatische Wende und Korpuslinguistik

Folgt man Feilke (2003: 217ff), so lässt sich die pragmatische Wende in zwei Phasen einteilen. Die erste Phase erweiterte zwar den Bereich linguistischer Gegenstände und Kategorien, wurde jedoch vom systemlinguistisch interessierten Zweig der Disziplin umgedeutet: An die Pragmatik als Kind der Wende wurde der Anspruch herangetragen, parallel zum universalgrammatischen ein universalpragmatisches Erkenntnisinteresse zu verfolgen, das sich auf die Suche nach der Universalität von Sprechakten und deren tiefenstrukturellen Gemeinsamkeiten machte (vgl. Nerlich 1995: 311). Die zweite Phase der pragmatischen Wende, die in den letzten 30 Jahren die Linguistik verändert hat, setzte ihre Positionen in einigen Bereichen neu (Feilke, 2003, 217ff.). Besonders relevant sind dabei zwei neue Sichtweisen: (1) Die Formelhaftigkeit der sprachlichen Oberfläche rückt zuungunsten sprachlicher Universalien der Tiefenstruktur ins Zentrum der Theoriebildung. (2) Statt den Kontext einer sprachlichen Äußerung als gegeben zu betrachten, geht man davon aus, dass der Sprachgebrauch den Kontext (mit-)herstellt, d.h. sprachliche Äußerungen kontextualisiert. Aus diesen beiden Neuakztentuierungen folgt, dass idiomatische Prägungen das Resultat von konventionalisierten Interpretationen sind. Signifikant häufig auftretende sprachliche Muster können deshalb als das Ergebnis rekurrenter Sprachhandlungen der Sprecherinnen und Sprecher gedeutet werden, in die typische Verwendungskontexte, Handlungsziele und Interpretationsrahmen eingeschrieben sind.

Die Korpuslinguistik bietet für eine so bestimmte Pragmatik sowohl einen theoretischen Rahmen als auch Methoden, um nicht nur Belege zur Illustration eines Phänomens („corpus-illustrated linguistics“, (Tummers et al. 2005)), sondern sprachliche Regelmäßigkeiten herauszuarbeiten. Dabei profitiert die Korpuslinguistik von Methoden der quantitativen Linguistik und der Computerlinguistik, etwa um sprachliche Daten maschinell mit linguistischen Informationen auszuzeichnen oder statistische Auswertungen über auffällige Strukturen zu ermöglichen.

1) So bietet sich vor dem Hintergrund des britischen Kontextualismus das Konzept der „Kollokation“ (Firth 1957) verstanden als überzufällige Kookkurrenz von Wörtern an, um pragmatische Konzepte zu operationalisieren. Und Kollokationen zeigen aus einer semantischen Perspektive, dass Bedeutungen nicht an Einzelwörtern festmachbar sind, sondern erst durch den Kontext entstehen und damit komplexere Zeichen sind, denen ein Gebrauchswert eingeschrieben ist.

2) Weiter trifft sich das Interesse der Textlinguistik an Oberflächeneigenschaften mit Ansätzen der Computerlinguistik, Textklassifizierungen anhand von Merkmalen der Textoberfläche maschinell durchführen zu können. Zudem führte die Überzeugung, dass Textsortenwissen kulturspezifisch geprägt ist, zu kontrastiven Untersuchungen, die Korpora als Basis für die empirische Überprüfung der Hypothesen verwenden.

3) Konstruktionsgrammatische Überlegungen, die von Kriterien wie Nicht-Kompositionalität und deswegen Konventionalität von Konstruktionen ausgehen, werden zwar erst in jüngerer Zeit durch korpuslinguistische Methoden auf eine breite empirische Basis gestellt (Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2010), teilen aber die Grundannahme einer gebrauchsorientierten Sprachtheorie und sind damit anschließbar an korpuslinguistische Methoden.

4) In der Phraseologie ist eine Ausweitung des Phraseologismenbegriffs zu beobachten, wo das Interesse nicht nur stark idiomatischen Wendungen, sondern auch schwach oder

überhaupt nicht idiomatischen Wortverbindungen gilt (Burger 1998, p.36ff.) und computerlinguistische Methoden oder aber zumindest korpusgestützte Analysen immer bedeutender werden, um potenzielle Phraseologismen in großen Korpora aufzufinden oder aber deren Verbreitung und Verwendung zu überprüfen (Steyer 2003; Stubbs 2003; Ptashnyk et al. 2010; Bubenhofer 2008).

5) Für die von Foucault inspirierte Diskurslinguistik sind korpuslinguistische Methoden attraktiv, um Serien von Aussagen auch quantitativ zu untersuchen. So lassen sich in Korpora, die bestimmte Diskurse repräsentieren, musterhaft auftretende Elemente wie Argumentationsfiguren, Topoi oder allgemeiner Sprachgebrauchsmuster entdecken und deren Verwendung diskurslinguistisch deuten (Wengeler 2003; Scharloth 2005; Jung 1996; Bubenhofer 2009a; Spitzmüller 2005; Teubert 2006).

Diese knapp gehaltene Aufzählung macht deutlich, dass eine ganze Reihe von Forschungsfragen im Nachgang zur pragmatischen Wende die kultur- und gesellschaftsanalytisch interessierte Linguistik offen machten für Methoden der empirischen Textanalyse und damit die Genese der Korpuspragmatik forcierten. Gerade die Korpuslinguistik ist dabei nicht einfach eine Hilfswissenschaft, sondern ebenso ein Kind der pragmatischen Wende und begünstigt durch technische Fortschritte (steigende Rechenleistung und Speicherkapazität von Computern, Verfügbarkeit von elektronischen Korpora, computerlinguistische Methoden der Textanalyse) immer stärker eine eigene Disziplin mit eigener Erkenntnislogik.

3. Datenbasiertes vs. Datengeleitetes Paradigma

Korpuslinguistik ist heute nicht mehr nur eine Methode, sondern generierte einen Denkstil, der viele Bereiche der Sprachwissenschaft nachhaltig verändert. Am ehesten der Vorstellung von Korpuslinguistik als einer Methode entspricht das Paradigma der datenbasierten korpuslinguistischen Analyse („corpus-based“), die nicht erst seit der Verfügbarkeit von digitalen Korpora erfolgreich angewandt wird. Korpora dienen demnach der Überprüfung von Forschungshypothesen. Die Hypothesen, die unabhängig von der Analyse des Korpus entwickelt wurden, formulieren Annahmen über interpretative Konstrukte, die mittels bereits bewährter interpretativer linguistischer Analysekategorien an einem Korpus überprüft werden sollen.

Diesem deduktiven Vorgehen steht die Möglichkeit eines induktiven Vorgehens zur Seite, das die Grundlage des Paradigmas der datengeleiteten Analyse („corpus-driven“) bildet. Dieses Paradigma wird von Tognini-Bonelli (2001, p.84ff.) vor dem Hintergrund der Arbeiten von Sinclair (1991) expliziert und im deutschen Sprachraum in Arbeiten von (Perkuhn et al. 2005; Belica & Steyer 2006; Steyer 2004; Bubenhofer 2009b) u.a. verbreitet. Digitale Korpora sind hier nicht nur „Belegsammlungen oder Zettelkästen in elektronischer Form“, sondern ermöglichen eine eigene „korpuslinguistische Perspektive“ (Perkuhn & Belica 2006, p.2). Statt eine Hypothese mit vorher festgelegten Analysekategorien zu überprüfen, werden in einem Korpus sämtliche Muster berechnet, die sich bei der Anwendung vorher festgelegter Algorithmen ergeben. Diese Muster werden im Anschluss kategorisiert. Damit geraten häufig Evidenzen in den Fokus, die entweder quer zu den vorher existierenden Erwartungen stehen und die Grundlage für neue Hypothesen sind, oder im besten Fall sogar solche Evidenzen, die die Bildung neuer interpretativer linguistischer Analysekategorien nahelegen. Es ist dieses Potenzial datengeleiteter Analysen, das es erlaubt, der Korpuslinguistik über eine empirische Methode hinaus den Status eines Denkstils zuzuschreiben. Denn das Ernstnehmen

empirischer Widerständigkeiten, die sich mit traditionellen linguistischen Kategorien nicht beschreiben lassen, birgt die Möglichkeit eines neuen Zugangs zu Sprache und den Kategorien ihrer Beschreibung.

Zwar verzichtet das datengeleitete Paradigma auf das Formulieren von Hypothesen und auf eine Festlegung auf bestimmte Analysekategorien, es ist jedoch offensichtlich, dass auch beim datengeleiteten Verfahren vorgängiges Wissen in den Forschungsprozess einfließt (vgl. Scharloth & Bubenhofer 2011), und zwar:

1. durch die Wahl der Korpora,
2. hinsichtlich der Gestaltung der Algorithmen zur Musterberechnung,
3. bei der Festlegung dessen, was als 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 (vgl. McEnery u. a. 2006: 8-11)

Häufig ergänzen sich beide Ansätze jedoch auch: Der erste Zugriff auf die Daten erfolgt datengeleitet, um sprachliche Auffälligkeiten zu ermitteln, die dann für die Operationalisierung zentraler Konzept benutzt werden. Die eigentliche Analyse erfolgt dann datenbasiert mit Hilfe des datengeleitet ermittelten Messinstruments.

4. Typen von Analysekategorien

Grundsätzlich haben alle computerlinguistisch erfassbaren Einheiten das Potenzial, als sozial oder kulturell signifikant gedeutet zu werden. In der bisherigen Forschung haben sich aber einige Kategorientypen als besonders fruchtbar erwiesen, deren Berechnung teilweise auch in korpuslinguistische Standardsoftware als Feature integriert sind. Die Analyse zielt dann auf das Auffinden statistisch signifikanter Unterschiede in der Distribution der Einheiten zwischen (Teil-)Korpora. Diese signifkanten Muster werden dann als sozial oder kulturell signifikant gedeutet.

4.1. Lemmata und Wortklassen

Der Fokus auf Einzellexeme ist der korpuslinguistisch naheliegendste Zugang zum Sprachgebrauch. Dabei müssen die Ebenen Wortform und Lemma unterschieden werden, wobei letztere erst nach der Lemmatisierung des Korpus verfügbar ist, die entweder manuell oder maschinell (z.B. über die Lemmatisierungskomponente eines Wortarten-Taggers) erstellt werden kann. Weiter können morphologische Aspekte mit einbezogen werden, indem z.B. der Fokus auf Komposita oder bestimmten Derivationen liegt.

Die Historische Semantik und die Begriffsgeschichte (Busse 1987; Hermanns 1995) zeigen das Deutungspotenzial von solchen Analysen auf, indem die diskursive, historische oder kulturelle Prägung von Wortfeldern und Begriffen korpuslinguistisch ergründet werden. Neben hypothesenprüfenden Verfahren, bei denen apriori definierte

Lexeme untersucht werden, gibt es eine Vielzahl von Methoden, um hypothesengenerierend vorzugehen und für bestimmte Korpora oder Teilkorpora im Vergleich zu Referenzkorpora typische Lexeme zu berechnen.

Statt Lemmata können auch formal oder semantisch bestimmte Wortklassen (z.B. Gradpartikel, Indefinitpronomen, Negationswörter) als Indikatoren für kulturelle oder soziale Phänomene gedeutet werden. Besonders im *Opinion Mining* und in der *Sentiment Analysis* wird häufig mit Wortlisten als Indikatoren für die Tonalität bzw. die semantische Prägung von Texten gearbeitet (Pang & Lee 2008).

4.2. Kollokationen

Zu jenen Typen von Analysekategorien, die nicht nur die Distribution von Lemmata oder Wortklassen, sondern die Verwendungskontexte von Wörtern in den Blick nehmen, zählen Kollokationen. Bei ihnen handelt es sich um "recurrent and predictable word combinations, which are a directly observable property of natural language" (Evert 2008: 1214) Für kultur- und sozialwissenschaftlich interessierte Linguistik sind Unterschiede im Kollokationsprofil einzelner Lemmata in verschiedenen (Sub-)Korpora von besonderem Interesse, verweisen sie doch auf unterschiedliche idiomatische Prägungen. Daneben können Kollokationsgraphen berechnet werden, die mehrere oder alle Lemmata in einem Korpus und deren typische Verbindungen zu anderen Lemmata abbilden (Scharloth et al. 2013).

4.3. n-Gramme

N-Gramme sind Einheiten einer durch *n* bestimmten Anzahl aufeinander folgender Wörter (Manning & Schütze 2002, p.192ff.). Für n=2 wird von Bigrammen, n=3 von Trigrammen etc. gesprochen. Normalerweise werden n-Gramme als kontinuierliche Wortfolgen verstanden. Weiter gefasst fallen aber auch diskontinuierliche Wortgruppen (Bubenhofer 2009b, p.118) oder Wortgruppen mit freier Wortreihenfolge („Concgram“; (Cheng et al. 2006)) unter den Terminus. N-Gramme können nur Wortformen oder aber Kombinationen von Wortformen, Wortarten und Lemmata („Collostructions“, (Stefanowitsch & Gries 2003); „komplexe n-Gramme“, (Scharloth & Bubenhofer 2011)) umfassen. Verwandt mit den n-Grammen sind die „syntagmatischen Muster“ von (Belica 2001; Keibel & Belica 2007), die ausgehend von der Berechnung von Kollokatoren zu diesen wiederum Kollokatoren berechnen und so in der Folge die typische Reihenfolge der Kollokatoren und die typischen Filler der Lücken syntagmatisch beschreiben können.

Für die Berechnung von n-Grammen können kombinatorische Zählverfahren von statistischen Verfahren unterschieden werden: Erstere zählen für eine bestimmte Länge von n alle möglichen unterschiedlichen n-Gramme in einem Korpus. Letztere verwenden analog zur Berechnung von Kollokationen statistische Assoziationsmaße, um die Bindungsstärke zwischen den Gliedern des n-Gramms zu berechnen.

Das Konzept der n-Gramme kommt den Bemühungen einiger Theorien, bei Analysen den Blick über das Einzelwort hinaus zu weiten, entgegen: In der Phraseologie werden n-Gramme (ebenso wie Kollokationen) genutzt, um Phrasen oder Idiome zu entdecken.

4.4. Syntaktische Kategorien

Auch syntaktische Kategorien können dann für maschinelle kultur- oder gesellschaftsanalytische Untersuchungen fruchtbar gemacht werden, wenn sie eine semantische Ladung haben. So kann beispielsweise die Distribution von bestimmten

Nebensatztypen als ein Indikator für den Gebrauch argumentativer Muster dienen, die Füllung semantischer Rollen als Indikator für Agency gedeutet werden oder in historischer Perspektive der Wandel im Valenzgefüge eines Verbs als Verweis auf einen Mentalitätswandel (vgl. Linke 2003a und Scharloth 2010).

4.5. semantische Taxonomien

In sozial- oder kulturwissenschaftlichen Korpusanalysen kommen zudem unterschiedliche Verfahren zum Einsatz, die die Komplexität der Texte auf eine überschaubare Menge semantischer Merkmale reduzieren sollen. Für diesen Zweck werden geordnete Wortschätze (z.B. Dornseiff), semantische Taxonomien (z.B. WordNet) oder spezialisierte Ontologien (Stuckenschmidt 2009) zur automatischen Annotation eingesetzt. Sie werden dann beispielsweise im *Opinion Mining* (Esuli & Sebastiani n.d.) oder zur Bestimmung von Frames in mentalitätsgeschichtlich orientierten Studien (Scharloth et al. 2013) eingesetzt.

5. Visualisierung

Datengeleitete Analysen großer Korpora resultieren in einer großen Menge von Ergebnisdaten. Wenn andere Formen der Repräsentation von Wissen wie Listen, Tabellen oder Texte zu umfangreich oder zu komplex sind, um in ihrer Gesamtheit erfasst und gedeutet werden zu können (Chen et al. 2008, p.5), unterstützen Visualisierungen den Forschungsprozess. Visualisierungen sind dann nicht nur Illustrationen („presentation graphics“), sondern eigenständige Mittel der Erkenntnisgewinnung („exploratory graphics“) (Schumann & Müller 1999, p.5). Die Entwicklung von Methoden zur Visualisierung ist daher ein integraler Bestandteil datenintensiver Forschungsprozesse in der kultur- und gesellschaftsanalytisch interessierten Korpuslinguistik.

Beispielgebend sind die technischen Disziplinen, wo vor allem in Mathematik, Informatik, Naturwissenschaften und *life-sciences* bei der Analyse komplex vernetzter Daten mit Visualisierungen gearbeitet wird, die in der Medizin den Namen „bildgebende Verfahren“ tragen. Solche Verfahren folgen dem Paradigma der „Visual Analytics“ (Keim et al. 2010; Chen et al. 2008): 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, d.h. bildlich repräsentiert werden.

Als Visualisierungen mit explorarivem Wert, die in der kultur- und gesellschaftsanalytisch interessierten Linguistik zum Einsatz kommen, sind vor allem Graphen, etwa zur Visualisierung komplexer Kollokationsnetze oder narrativer Muster, und Karten, beispielsweise zur Visualisierung der räumlichen Distribution sozial signifikanter sprachlicher Phänomene, zu nennen.

6. Literatur (in Auswahl)

Belica, C., 2001. *Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des*

Sprachgebrauchs,

Belica, C. & Steyer, K., 2006. Korpusanalytische Zugänge zu sprachlichem Usus. *AUC (Acta Universitatis Carolinae), Germanistica Pragensia*, XX.

Bubenhofer, N., 2008. "Es liegt in der Natur der Sache…". Korpuslinguistische Untersuchungen zu Kollokationen in Argumentationsfiguren. In C. Mellado Blanco, ed. *Studien zur Phraseologie aus textueller Sicht*. Philologia – Sprachwissenschaftliche Forschungsergebnisse. Hamburg: Kova\v c, pp. 53–72.

Bubenhofer, N., 2009a. *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*, Berlin, New York: de Gruyter.

Bubenhofer, N., 2009b. *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*, Berlin, New York: de Gruyter.

Burger, H., 1998. *Phraseologie. Eine Einführung am Beispiel des Deutschen*, Berlin: Erich Schmidt.

Busse, D., 1987. *Historische Semantik. Analyse eines Programms*, Stuttgart: Klett-Cotta.

Chen, C., Härdle, W. & Unwin, A. eds., 2008. *Handbook of data visualization*, Springer. Available at: http://sfx.ethz.ch/sfx_locater?sid=ALEPH:EBI01&genre=book&isbn=9783540330370&id=doi:10.1007/978-3-540-33037-0 Online via SFX.

Cheng, W., Greaves, C. & Warren, M., 2006. From N-Gram to Skipgram to Concgram. *International Journal of Corpus Linguistics*, 11, pp.411–433(23).

Esuli, A. & Sebastiani, F., SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining.

Evert, S., 2008. Corpora and Collocations. In A. Lüdeling & M. Kytö, eds. *Corpus linguistics: an international handbook*. Mouton de Gruyter.

Feilke, H., 2000. Die pragmatische Wende in der Textlinguistik. In K. Brinker, ed. *Text- und Gesprächslinguistik/Linguistics of Text and Conversation*. Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science. Berlin/New York: de Gruyter, pp. 64–82.

Feilke, H., 2003. Textroutine, Textsemantik und sprachliches Wissen. In A. Linke, H. Ortner, & P. R. Portmann-Tselikas, eds. *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*. Reihe Germanistische Linguistik. Tübingen: Niemeyer, pp. 209–230.

Firth, J.R., 1957. Modes of Meaning. In *Papers in Linguistics 1934–1951*. London: Oxford University Press, pp. 190–215.

Gries, S.T. & Stefanowitsch, A., 2010. Cluster analysis and the identification of collexeme classes. In S. Rice & J. Newman, eds. *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI, pp. 59–72.

Hermanns, F., 1995. Sprachgeschichte als Mentalitätsgeschichte. Überlegungen zu Sinn und Form und Gegenstand historischer Semantik. In A. Gardt, K. Mattheier, & O. Reichmann, eds. *Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien*. Tübingen: Niemeyer, pp. 69–101.

Jung, M., 1996. Linguistische Diskursgeschichte. In K. Böke, M. Jung, & M. Wengeler, eds. *Öffentlicher Sprachgebrauch. Praktische, theoretische und historische Perspektiven. Georg Stötzel zum 60. Geburtstag gewidmet*. Opladen: Westdeutscher Verlag, pp. 453–472.

Keibel, H. & Belica, C., 2007. CCDB: A Corpus-Linguistic Research and Development Workbench. In *Proceedings of the 4th Corpus Linguistics Conference*. Birmingham.

Keim, D.A. et al., 2010. *Mastering the Information Age - Solving Problems with Visual Analytics*, Goslar: Eurographics Association. Available at: http://www.vismaster.eu/book/.

Linke, A., 2003. Spaß haben. Ein Zeitgefühl. In J. K. Androutsopoulos & E. Ziegler, eds. *Standardfragen. Soziolinguistische Perspektiven auf Sprachgeschichte, Sprachkontakt und Sprachvariation*. Variolingua. Frankfurt am Main: Lang, pp. 63–79.

Manning, C.D. & Schütze, H., 2002. *Foundations of Statistical Natural Language Processing* 5. ed., Cambridge, Massachusetts: The MIT Press.

McEnery, T., Richard, X. & Yukio, T., 2006. *Corpus-Based Language Studies. An advanced Resource Book*, London/New York: Routledge.

Nerlich, B., 1995. The 1930s – At the Birth of a Pragmatic Conception of Language. *Historiographica Linguistica*, XXII(3), pp.311–334.

Pang, B. & Lee, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), pp.1–135.

Perkuhn, R. et al., 2005. Korpustechnologie am Institut für Deutsche Sprache. In J. Schwitalla & W. Wegstein, eds. *Korpuslinguistik deutsch: synchron – diachron – kontrastiv. Würzburger Kolloquium 2003*. Tübingen: Niemeyer, pp. 57–70.

Perkuhn, R. & Belica, C., 2006. Korpuslinguistik – Das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik. *Sprachreport*, 22(1), pp.2–8.

Ptashnyk, S., Hallsteinsdóttir, E. & Bubenhofer, N. eds., 2010. *Korpora, Web und Datenbanken. Computergestützte Methoden in der modernen Phraseologie und Lexikographie/Corpora, Web and Databases. Computer-Based Methods in Modern Phraseology and Lexicography.*, Baltmannsweiler: Schneider Verlag Hohengehren.

Scharloth, J., 2010. *1968 - Eine Kommunikationsgeschichte*, Fink Wilhelm GmbH + Co.KG.

Scharloth, J., 2005. Die Semantik der Kulturen. Diskurssemantische Grundfiguren als Kategorien einer linguistischen Kulturanalyse. In D. Busse, T. Niehr, & M. Wengeler, eds. *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*. Reihe Germanistische Linguistik. Tübingen: Niemeyer, pp. 133–148.

Scharloth, J. & Bubenhofer, N., 2011. Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse. In E. Felder, M. Müller, & F. Vogel, eds. *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen*. Berlin, New York: de Gruyter, pp. 195–230.

Scharloth, J., Eugster, D. & Bubenhofer, N., 2013. Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn. In D. Busse & W. Teubert, eds. *Linguistische Diskursanalyse: neue Perspektiven*. Interdisziplinäre Diskursforschung. Springer Fachmedien Wiesbaden, pp. 345–380. Available at: http://dx.doi.org/10.1007/978-3-531-18910-9_11.

Schumann, H. & Müller, W., 1999. *Visualisierung: Grundlagen und allgemeine Methoden*, Springer DE.

Sinclair, J., 1991. *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

Spitzmüller, J., 2005. *Metasprachdiskurse. Einstellungen zu Anglizismen und ihre wissenschaftliche Rezeption*, Berlin: de Gruyter.

Stefanowitsch, A. & Gries, S.T., 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics*, 8(2), pp.209–243.

Steyer, K., 2004. Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In K. Steyer, ed. *Wortverbindungen – mehr oder weniger fest*. Institut für Deutsche Sprache. Jahrbuch 2003. Berlin, New York: de Gruyter, pp. 87–116.

Steyer, K., 2003. Korpus, Statistik, Kookkurrenz. Lässt sich Idiomatisches "berechnen"? In H. Burger, A. Häcki-Buhofer, & G. Gréciano, eds. *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifik der Phraseologie*. Phraseologie und Parömiologie. Baltmannsweiler: Schneider, pp. 33–46.

Stubbs, M., 2003. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), pp.215–244.

Stuckenschmidt, H., 2009. *Ontologien: Konzepte, Technologien und Anwendungen*, Berlin, Heidelberg: Springer.

Teubert, W., 2006. Korpuslinguistik, Hermeneutik und die soziale Konstruktion von Wirklichkeit. *Linguistik online*, 28(3), pp.41–60.

Tognini-Bonelli, E., 2001. *Corpus Linguistics at Work*, Amsterdam: Benjamins.

Tummers, J., Heylen, K. & Geeraerts, D., 2005. Usage-Based Approaches in Cognitive Linguistics: A Technical State of the Art. *Corpus Linguistics and Linguistic Theory*, 1(2), pp.225–261.

Wengeler, M., 2003. *Topos und Diskurs: Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960 – 1985)*, Tübingen: Niemeyer.