

Noah Bubenhofer / Marc Kupietz

Einleitung

1. Funktionen von Visualisierungen in den Wissenschaften

Visualisierungen von Daten spielen in den Wissenschaften eine wichtige Rolle im Forschungsprozess. Einerseits dienen sie 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 u. a. 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 Ganzes erfasst und gedeutet werden zu können. Visualisierungen dieser Art werden zur Gruppe der „Exploratory Graphics“ (explorativen Visualisierungen) gezählt (Chen u. a. 2008, S. 5; Schumann und Müller 1999, S. 5).

Explorative Visualisierungsmethoden werden insbesondere im Bereich der Visual Analytics (Keim u. a. 2010; Chen u. a. 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. Visualisierungen könnten demnach auch als Scharnier zwischen quantitativ-maschinellen und qualitativ-interpretierenden Analysen angesehen werden, da durch einen iterativen Prozess der interpretativen Interaktion mit den Daten Modelle generiert werden (vgl. den Visual Analytics Process nach Keim u. a. 2010, S.10): „Visualisation becomes the medium of a semi-automated analytical process, where humans and machines cooperate using their respective, distinct capabilities for the most effective results“ (Keim u. a. 2010, S.14). Insbesondere explorative Visualisierungen sind entsprechend nicht Endprodukt, sondern typischerweise Zwischenprodukt und Mittel, die (1) quantitativ-maschinelle mit der (2) qualitativ-interpretierenden Analyse in einem iterativen,

empirisch-hermeneutischen Erkenntnisprozess zu kombinieren. Dadurch wird etwa die Abduktion neuer, vielversprechender Hypothesen möglich (vgl. Jockers 2013; Kupietz/Keibel 2009, S. 48).

Der Wert visueller Methoden für die Analyse großer Datenmengen wird zwar in einigen Publikationen zur statistischen Theorie und Methode erkannt, trotzdem mahnen Chen et al. (2008, S. 4) die fehlende Reflexion über die gängigen Methoden der Visualisierung statistisch gewonnener Daten an: “Examples abound in almost every issue of every scientific journal concerned with quantitative analysis. There are occasionally articles published in a more theoretical vein about specific graphical forms, but little else” (Chen u. a. 2008, S. 4). Allerdings bleibt anzumerken, dass es eine Reihe von Arbeiten gibt, die die semiotischen und kognitiven Grundlagen der Visualisierung reflektieren und Regeln für die Erstellung von Grafiken formulieren (Bertin 1967; Tufte 1983; Tufte 1997; Schumann und Müller 1999; Unwin u. a. 2006), die Algorithmen zur Erstellung von Visualisierungen, etwa von Graphen, diskutieren (Tamassia 2013; Brandes u. a. 2013) oder den Wert von Visualisierungen aus Sicht der Benutzerinnen und Benutzer, z. B. im Web, darlegen (Hearst 2009; Hearst und Rosner 2008).

Insbesondere existiert eine Reihe von Arbeiten, die grundlegende semiotische und kulturelle Aspekte des Diagramms reflektieren: Welcher Art ist dieses Zeichen? Wie werden Diagramme eingesetzt, um Wissen zu ordnen, zu generieren oder zu kommunizieren? Welche historischen Grundfiguren des Diagramms gibt es und wie hängen diese mit ideengeschichtlichen Entwicklungen zusammen? Hierzu sind in den letzten Jahren einige Arbeiten entstanden, die eine Theorie der Diagrammatik skizzieren und elaborieren – oft im Rückgriff auf Charles Sanders Peirce (Bauer und Ernst 2010; Bender und Marrinan 2010; Bredekamp u. a. 2008; Krämer 2016; Liebsch und Mößner 2012; Reichert 2013; Siegel 2009; Stetter 2005; Stjernfelt 2007).

Die Wurzeln der explorativen Datenanalyse (“Exploratory Data Analysis”, “EDA”) gehen auf Tukey (1977) und Benzécri (1973b; 1973a) zurück, wobei die Geschichte der Visualisierung statistisch gewonnener Daten viel älter ist (Friendly 2005; Tufte 1983). So gilt Michael Florent van Langren 1644 erstellte Grafik der Schätzungen verschiedener Astronomen zur Longitudinal-Differenz zwischen Toledo und Rom als erste visuelle Repräsentation statistischer Daten (Tufte 1997, S. 15). In der Folge wurden immer häufiger Visualisierungen verwendet, nicht nur, um komplexe statistische Zusammenhänge zu präsentieren, sondern auch, um sich einen Überblick über die Daten zu verschaffen und überhaupt die immer größer werdenden Datenbestände und darin auftretende Zusammenhänge analysieren zu können (Zhang 2008, S. IX). Die Grundlagen der visuellen Datenanalyse werden in einer Reihe von Werken erarbeitet und beispielhaft angewandt (Zhang 2008; Dill u. a. 2012; Chen u. a. 2008; Burkhart und Eppler 2004; Mazza 2009; Tufte 1983; Keim u. a. 2010; Arnold 2008).

2. Traditionelle Visualisierungen in der Sprachwissenschaft

In der Linguistik sind besonders in der Dialektologie Visualisierungen in Form von Karten schon lange gebräuchlich und sind sowohl „Dokumentations-“ als auch „Forschungsmittel“ (Naumann 1982) – dienen also sowohl der Präsentation von Ergebnissen als auch der Exploration von Daten. Es entwickelten sich verschiedene Typen von Themenkarten (Originalformkarten, Punktsymbolkarten, Flächenkarten, kombinierte Karten etc.), die sich zwischen Dokumentation und Interpretation bewegen. Auch in der Phonetik sind Visualisierungen für explorative Zwecke wichtig, z. B. Spektrogramme, um die Lage der Formanten zu erkennen und damit beispielsweise die Stimmqualität oder prosodische Eigenschaften zu messen (Reetz 2003).

In der strukturalen Syntax sind Baumgraphen und andere Visualisierungsmöglichkeiten von Strukturen (vgl. z. B. die syntagmatische Verkettung als Spirale bei Mikuš 1952; zit. nach Thümmel 1993a, S. 271) weit mehr als nur eine Darstellungshilfe, sie sind vielmehr Ausdruck strukturalistischer Theoriebildung (Thümmel 1993b; Heringer 1993). Ähnlich verhält es sich mit zahlreichen weiteren Visualisierungen von Modellen, die Sprachwirklichkeit beschreiben wollen. Exemplarisch sei auf die im germanistischen Raum relativ bekannte Visualisierung des soziolinguistischen Varietätenmodells von Löffler (1994) verwiesen. Es folgt dem Forschungsparadigma der Varietätenlinguistik, die Sprache als komplexe Menge von sprachlichen Varietäten und nicht als „unmittelbar gegebene[n] (homogene[n]) Gegenstand“ ansieht (Bußmann 2002, S. 729), was in der Visualisierung durch sich überlagernde Vektoren widerspiegelt wird. Ein weiterer wichtiger Anwendungsbereich von Visualisierungen sind Stammbäume in der vergleichenden Sprachwissenschaft. Schleicher (1860, S. 28) gilt als erster Sprachwissenschaftler, der eine Stammbaumdarstellung für den Sprachvergleich eingesetzt hat (Sutrop 2012, S. 299). Der Stammbaum reproduziert als gerichteter Graph bestimmte Ordnungsprinzipien, die sich nur bedingt mit modernen Auffassungen von Sprachfamilien vereinbaren lassen (Sutrop 2012, S. 320). Alternative Darstellungsformen ergeben sich dabei auch aus neuen methodischen, statistischen Zugängen (Fox 1995; Jäger 2014).

Weniger offensichtlich sind die Formen der Visualisierung in der Gesprächsanalyse: Dort müssen die flüchtigen Daten des Gesprächs dokumentiert werden, üblicherweise in Form einer Transkription (Redder 2001; Deppermann 2001). Erst die Transkription erlaubt anschließend die Exploration der Daten (Sager 2001). Die Art der Transkription richtet sich nach dem Erkenntnisinteresse, sodass verschiedene Transkriptionsstandards existieren, die als eine Form von Visualisierung je unterschiedliche Aspekte der komplexen Daten (Transliteration, Normalisierung, Intonation, Betonung etc.) hervorheben.

Interessant ist auch ein Blick über die engen disziplinären Grenzen zu den Literaturwissenschaften: Hier gilt Moretti (2000; 2009) als einer der Wegbereiter für eine neue, visuelle Sicht auf Literatur, die seinem Paradigma des Distant Reading folgt. Voraussetzung dafür ist die Computerphilologie, die beispielsweise die kritische Edition von Texten mit den Möglichkeiten der digitalen Aufbereitung (Annotation, dynamische Textdarstellung etc.) und Analyse verbindet (Jannidis 1999; Jannidis et al. 2017; Lauer 2011).

Einen erhellenden Überblick über Visualisierungen in der Linguistik bietet Harleman Stewart (1976). Sie reflektiert den Einfluss von grafischen Repräsentationen auf die Theoriebildung und analysiert Visualisierungen wie Baumgraphen in der vergleichenden Sprachwissenschaft, Phonetik, Syntax etc. Dabei wird die Komplexität des Visualisierungsprozesses deutlich, durch den nicht nur Daten interpretiert, sondern heuristisch Theorien modelliert werden. Diese ältere Publikation reflektiert jedoch noch nicht die neueren Entwicklungen im Bereich der Visualisierungen in der Sprachwissenschaft, die sich zudem seit den 1970er-Jahren in vielen Bereichen stark gewandelt hat.

3. Visuelle Textanalyse: Visualisierungen in der Korpuslinguistik und den datenintensiven Digital Humanities

In der Sprachwissenschaft und den Digital Humanities ist es insbesondere die Korpuslinguistik, bei der der Bedarf für neue Formen der visuellen Analyse stark ansteigt. Bei hypothesengeleiteten Ansätzen entstehen quantitative Analyseergebnisse, die visualisiert werden können („presentation graphics“). Doch die Verfügbarkeit großer Textmengen erlaubt es auch, datengeleitete Analyseverfahren anzuwenden, die der Hypothesengenerierung dienen. Im größeren Kontext der Digital Humanities zeigt sich zudem die Chance, nicht nur mit Textdaten zu arbeiten, sondern verschiedene Datentypen (Bilder, Daten historischer Ereignisse, geografische Informationen – GIS, soziodemografische Daten etc.) integrieren zu können. In der Korpuslinguistik werden deshalb vermehrt Methoden der analytischen Statistik und des Data Minings angewandt, um die verfügbaren Daten auswerten zu können (Manning und Schütze 2002; Baayen 2008; Gries 2009b).

Allerdings bringt die Analyse solcher Daten eine Reihe von Herausforderungen mit sich: 1) Textdaten gehören zu den unstrukturierten Datentypen und unterscheiden sich von anderen Daten, die dem Data Mining normalerweise zugrunde liegen. 2) Die Daten sind oft heterogen, da sie unterschiedliche Datentypen vereinen. 3) Die Daten sind oft komplex, da die einzelnen Datentypen wiederum eine Vielzahl von Ebenen umfassen können: Bei Textdaten sind das verschiedene Annotationsebenen, wie sie typischerweise in Korpora, die mit Methoden des Natural Language Processing aufbereitet worden sind, auftreten

(Wortartkategorien, Lemma, syntaktische Struktur), aber auch (halb-)manuell erzeugte Annotationen.

In anderen Disziplinen, die mit Big Data dieser Art arbeiten, erwiesen sich visuelle Analysemethoden, eben „exploratory graphics“, als besonders fruchtbar (Tukey 1977; Thomas und Cook 2005; Unwin u. a. 2006; Chen u. a. 2008; Dill u. a. 2012). Als Teilgebiet der visuellen Analyse etablieren sich gegenwärtig die Visual Text Analytics (visuelle Textanalyse), die das Paradigma der Datenvisualisierung auf Textdaten anwendet (Risch u. a. 2008; Rohrdantz u. a. 2010). Erste Anwendungsbeispiele sind vielversprechend, doch fehlt noch weitgehend die theoretische und methodische Reflexion.

Visuelle Analysemethoden von Textdaten können dann gewinnbringend eingesetzt werden, wenn ein Analyseverständnis vorherrscht, bei dem nicht die Einzelbelege im Vordergrund stehen, sondern bei dem mit statistischen Mitteln Regularitäten im Sprachgebrauch aufgedeckt werden. Dazu stehen immer größere Korpora in Größenordnungen ab 1 Mia. Textwörter zur Verfügung, und es wird deutlich, dass ein Mehr an Daten auch ein Mehr an Analysemöglichkeiten bietet (Church und Mercer 1993). Die dafür nötigen statistischen Methoden werden gegenwärtig entwickelt und diskutiert, wie eine Vielzahl von methodologisch-statistischen Arbeiten (vgl. z. B. Kilgarriff 2005; Gries 2005; Gries 2008a; Gries 2009a; Gries 2010b; Hilpert & Gries 2009; Gries 2010a; Evert 2005; Rietveld & Hout 2005; Biber & Jones 2009) zeigt, wobei sich diese Forschungsrichtung auch bereits in Lehrbüchern niederschlägt (Gries 2008b; Baayen 2008). Fragen der Visualisierung generell, insbesondere auch der visuellen Analyse, scheinen dabei noch sekundär zu sein, obwohl z. B. Gries betont, dass die Visualisierung der Ergebnisse deskriptiver Analyse hilfreich ist (Gries 2008b, S.268) und die bereits verfügbaren statistischen Methoden, darunter auch Formen der Visualisierung, noch längst nicht ausgeschöpft sind (Gries 2010b, S. 24).

Korpus- und computerlinguistische Anwendungen visueller Textanalyse liegen z. B. für diskursive Daten vor (Luo u. a. 2012), um Sprachwandel oder semantische Variation zu analysieren (Hilpert 2011; Hao u. a. 2010; Rohrdantz, Hao u. a. 2012) oder die Lesbarkeit von Texten visuell auszudrücken (Oelke, Spretke u. a. 2012). Weiter gibt es Vorschläge, kontinuierlich entstehende Textdaten zu visualisieren (Rohrdantz u. a. 2011; Diakopoulos u. a. 2010) oder Ergebnisse von maschinellen semantischen Analysen, z. B. von Kundenrezensionen, zusammenfassend darzustellen (Alper u. a. 2011; Rohrdantz, Hao u. a. 2012; Shi u. a. 2010). Einige Arbeiten versuchen neue Visualisierungsformen für traditionelle, aber unbefriedigende Formate zu finden, wie z. B. Kollokationsgraphen oder Wortwolken (Gambette und Veronis 2009; Rockwell u. a. 1999; Wattenberg und Viegas 2008; Culy und Lyding 2010; Leblanc und Pérès 2010; Oelke, Eklund u. a. 2012; Collins u. a. 2009; Brandes u. a. 2006), diatopische Karten und andere Geotextdaten (Vriend u. a. 2011; Gregory und Hardie 2011), Netzwerke (Efer u. a.

2012), Syntax-Baumgraphen (Derrick & Archambault 2010) oder andere grammatische Muster (Elliott u. a. 2001; Säily u. a. 2011). Visualisierungen werden auch eingesetzt, um komplexe Korrelationen darzustellen, z. B. in Datensammlungen linguistischer Eigenschaften von Vornamen (Wattenberg 2005) oder bei der Berechnung von Ähnlichkeiten zwischen Sprachen oder Dialekten (Rohrdantz, Hund u. a. 2012; Zastrow 2011). Bei der maschinellen Textklassifikation dienen Visualisierungen auch dazu, die Auswahl der Variablen für die Modellierung zu unterstützen (May u. a. 2010; Oelke u. a. 2008; Chuang u. a. 2012).

Der Forschungsstand zeigt, dass in der Sprachwissenschaft sowohl für Darstellungszwecke als auch für die Datenexploration häufig Visualisierungstechniken eingesetzt werden, die Reflexion darüber jedoch oft fehlt. So werden beispielsweise Wortwolken zur Darstellung von häufigem Vokabular eingesetzt – insbesondere auch außerhalb der Wissenschaften scheint diese Darstellung, z. B. als Tag-Clouds, sehr attraktiv zu sein –, die Visualisierung weist aber den gravierenden Mangel auf, dass die Position des Wortes in der Wolke nicht semantisiert ist (die Wörter also zufällig angeordnet sind) oder die Semantisierung unwichtige Kriterien abbildet (bei alphabetischer oder gestalterisch optimierter Anordnung). Darüber hinaus sind die statistischen Berechnungsmethoden meist nicht transparent oder entsprechen nicht dem State of the Art, wenn z. B. die Größe des abgebildeten Wortes in ikonischem Verhältnis zur absoluten Auftretenshäufigkeit (unter Ausschluss von bestimmten Stoppwörtern) statt im Verhältnis zur statistischen Signifikanz der Frequenz im Vergleich zu einem Referenzkorpus steht.

4. Konzeption des Buches

Das vorliegende Buch ist aus dem im November 2014 veranstalteten Herrenhäuser Symposium *Visuelle Linguistik – Theorie und Anwendung von Visualisierungen in der Sprachwissenschaft*¹ entstanden. Ziel des Symposiums war es, vor dem oben dargelegten historischen und theoretischen Hintergrund die aktuellen und zukünftigen Herausforderungen und Chancen im Hinblick auf Visualisierungen in der Linguistik zu diskutieren. Ein wesentliches Anliegen des Symposiums war es dabei, ein sowohl horizontal breites Spektrum verschiedener Ansätze in der Linguistik und den angrenzenden Disziplinen in den Digital Humanities aufzugreifen als auch vertikal das Spektrum zwischen meta-theoretischen Überlegungen und konkreten Anwendungen abzubilden, um verschiedene Ansatzalternativen und ihre Integration in Methodologien und Arbeitsprozesse im Hinblick auf den möglichen Erkenntnisgewinn in einem interdisziplinären Teilnehmerfeld zu diskutieren.

1 Gefördert von der VolkswagenStiftung (Az. 88445).

Das Buch enthält schriftliche Ausarbeitungen ausgewählter Beiträge des Symposiums und gliedert sich in die drei großen Bereiche I Visual Linguistics, II Praxis und III Tools. Im ersten Teil sind Beiträge versammelt, die die theoretischen und methodologischen Grundlagen von Visualisierungen in der Linguistik diskutieren. Noah Bubenhofers Beitrag zur Eröffnung ist gleichzeitig als Einführung ins Thema und starkes Plädoyer für eine neue Sicht auf das Feld gedacht, er dient zudem auch dazu, die folgenden Beiträge zu kontextualisieren. Ebenfalls grundlegenden Charakter hat der Beitrag von Rainer Perkuhn und Marc Kupietz, in dem insbesondere die theoretisch-methodologischen Grundlagen für Visualisierungen im gegenwärtig wichtigsten Bereich der Linguistik, der Korpuslinguistik, dargelegt werden. Die weiteren Beiträge des ersten Teils adressieren weitere Bausteine einer Methode der visuellen Linguistik (Lauersdorf), wobei Jana Pflaegings Beitrag auch formal das Thema aufgreift, indem ihr Standpunkt zur Ästhetisierung linguistischer Wissensvermittlung auch als visuelles Statement gestaltet ist.

Im zweiten Teil folgen Beiträge, die einen Einblick vermitteln, wie in die Forschungspraxis mit visuellen Analysemöglichkeiten umgegangen wird. Einerseits werden darin linguistisch, literaturwissenschaftlich und psycholinguistisch interessante Erkenntnisse zu verschiedenen Themen – Referenzen in der Lyrik (Hoenen), Toponyme in literarischen Texten (Barbatesi) und Story Complexity (Flekova / Stoffel / Gurevych / Keim) – präsentiert, andererseits die Funktion und Bedeutung der dafür benutzten visuellen Analyseinstrumente kritisch reflektiert.

Der dritte Teil „Tools“ fokussiert stärker die Ebene des Analysewerkzeugs und zeigt anhand verschiedener Anwendungsgebiete Methoden der visuellen Analyse. Den Auftakt bildet ein Beitrag zur sehr universell einsetzbaren Programmiersprache R, die häufig für visuelle Analysen unter Einsatz verschiedenster Methoden eingesetzt wird (Wolfer, Hansen). Es folgen Beiträge über Werkzeuge, zunächst zur explorativen Analyse von Korpora mit einer dreistufigen Realisierung der Visualisierungsprozesse (Rüdiger), dann zur Darstellung und Interpretation von Topic Models (Hinneburg / Oberländer) und zur Signifikanzschätzung von Frequenzunterschieden bei N-Grammen (Todorova / Chinkina). Im letzten Beitrag wird der Einfluss von Visualisierungen auf die Usability von Annotationstools diskutiert (Burghardt), sodass ein breites Feld von visuellen Methoden und Anwendungsbereichen in der Linguistik skizziert werden kann.

Die Herausgeber bedanken sich bei der VolkswagenStiftung für die großzügige Finanzierung des Symposiums und der vorliegenden Publikation. Ebenfalls möchten wir uns beim Verlag für die umsichtige Betreuung und das sorgfältige Lektorat und die Produktion des Buches bedanken. Und schließlich geht der Dank an alle Autorinnen und Autoren dieses Bandes, ohne deren Engagement weder das Symposium erfolgreich durchgeführt noch das Buch hätte entstehen können.

Bibliografie

- Alper, Basak, Huahai Yang, Eben Haber und Eser Kandogan. 2011. "Opinion-Blocks: Visualizing Consumer Reviews." In *IEEE Workshop on Interactive Visual Text Analytics for Decision Making*. Providence, RI. <http://vialab.science.uoit.ca/textvis2011/papers/textvis%202011-alper.pdf>.
- Arnold, Claus. 2008. *Visualisierung im Information Retrieval*. Saarbrücken: Dr. Müller.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bauer, Matthias, Ernst, Christoph. 2010. *Diagrammatik: Einführung in ein kultur- und medienwissenschaftliches Forschungsfeld*. Bielefeld: transcript.
- Bender, John und Michael Marrinan. 2010. *The Culture of Diagram*. Stanford, Calif: Stanford University Press.
- Benzécri, Jean-Paul. 1976-1980. *L'analyse des données*. Paris: Dunod.
- Benzécri, Jean-Paul. 1973a. *L'Analyse des correspondants: introduction, théorie, applications diverses notamment à l'analyse des questionnaires, programmes de calcul*. [S.l.]: Bordas.
- Benzécri, Jean-Paul. 1973b. *L'analyse des données : leçons sur l'analyse factorielle et la reconnaissance des formes et travaux du laboratoire de statistique de l'Université de Paris VI*. Paris: Dunod.
- Bertin, Jacques. 1967. *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*. Paris: Mouton.
- Biber, Douglas und James K. Jones. 2009. „Quantitative methods in corpus linguistics.“ In *Corpus Linguistics*, herausgegeben von Anke Lüdeling und Merja Kytö. Berlin: Mouton de Gruyter, 1286–1304.
- Brandes, Ulrik, Linton C. Freeman und Dorothea Wagner. 2013. Social Networks. In *Handbook of Graph Drawing and Visualization*, herausgegeben von Roberto Tamassia. London: Boca Raton.
- Brandes, Ulrik, Martin Hoefer und Jürgen Lerner. 2006. *WordSpace: Visual Summary of Text Corpora*. <https://dx.doi.org/10.1117/12.647867>.
- Bredenkamp, Horst, Birgit Schneider, und Vera Dünkel, Hrsg. 2008. *Das Technische Bild: Kompendium zu einer Stilgeschichte wissenschaftlicher Bilder*. Berlin: Akademie-Verlag.
- Burkhard, Remo A. und Martin J. Eppler. 2004. *Knowledge Visualization*. In: *Encyclopedia of Knowledge Management*. Hershey, PA: Idea Group Reference, 551–560.
- Bußmann, Hadumod. 2002. *Lexikon der Sprachwissenschaft*. 3., aktual. und erw. Aufl.. Stuttgart: Kröner.

- Chen, Chun-houh, Wolfgang Härdle und Antony Unwin, Hrsg. 2008. *Handbook of Data Visualization*. Berlin: Springer (Springer Handbooks of Computational Statistics).
- Chuang, Jason, Christopher D. Manning und Jeffrey Heer. 2012. "Termite: Visualization Techniques for Assessing Textual Topic Models." In: *Advanced Visual Interfaces*. <http://vis.stanford.edu/papers/termite> (letzter Zugriff am 27. November 2017).
- Church, KW und RL Mercer. 1993. "Introduction to the special issue on computational linguistics using large corpora." *Computational Linguistics*. 19 (1): 1–24.
- Collins, Christopher, Fernanda B. Viegas und Martin Wattenberg. 2009. "Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora." In *2009 IEEE Symposium on Visual Analytics Science and Technology*, Atlantic City. 91–98. <https://doi.org/10.1109/VAST.2009.5333443>.
- Culy, Chris und Verena Lyding. 2010. "Double Tree: An Advanced KWIC Visualization for Expert Users." In *2010 14th International Conference Information Visualisation*, London, 98–103. <https://doi.org/10.1109/IV.2010.24>.
- Deppermann, Arnulf. 2001. *Gespräche analysieren*. Opladen: Leske + Budrich.
- Derrick, Donald und Daniel Archambault. 2010. „TreeForm: Explaining and Exploring Grammar Through Syntax Trees.“ *Literary and Linguistic Computing* 25 (1): 53–66. <https://doi.org/10.1093/lc/fqp031>.
- Diakopoulos, Nicholas, Mor Naaman und Funda Kivran-Swaine. 2010. „Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry.“ In *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 115–122. doi: <https://doi.org/10.1109/VAST.2010.5652922>.
- Dill, John, Rae Earnshaw, David Kasik, John Vince und Pak Chung Wong, Hrsg. 2012. *Expanding the Frontiers of Visual Analytics and Visualization*. 2012. London: Springer.
- Efer, Thomas, Jens Blecher und Gerhard Heyer. 2012. *Leipziger Rektoratsreden 1871–1933 Insights into Six Decades of Scientific Practice*. In: *International Conference on Historical Corpora*. <http://asv.informatik.uni-leipzig.de/publication/file/239/HistCorp2012-EferBlecherHeyer-Rektoratsreden.pdf> (letzter Zugriff am 18. Januar 2018).
- Elliott, J., E. Atwell und B. Whyte. 2001. *Visualisation of long distance grammatical collocation patterns in language*. In: *Proceedings. Fifth International Conference on Information Visualisation*, 297–302. <https://doi.org/10.1109/IV.2001.942073>.
- Evert, Stefan. 2005. *The Statistics of Word Cooccurrences. Word Pairs and Collocations*. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. <http://dx.doi.org/10.18419/opus-2556>.
- Finnegan, Ruth. 1992. *Oral Poetry*. Bloomington, Ind.: Indiana University Press
- Fisher, Ronald A. 1950. *Statistical Methods for Research Workers*. 11. ed. London: Oliver and Boyd

- Fox, Anthony. 1995. *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press.
- Friendly, Michael. 2005. „Milestones in the History of Data Visualization: A Case Study in Statistical Historiography.“ In *Classification: The Ubiquitous Challenge*, herausgegeben von Claus Weihs und Wolfgang Gaul. New York: Springer, 34–52.
- Gambette, Philippe und Jean Veronis. 2010. „Visualising a Text with a Tree Cloud.“ In: IFCS’09: International Federation of Classification Societies Conference, March 2009, Dresden, Germany. <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00373643v2> (letzter Zugriff am 28. November 2017).
- Gregory, Ian N. und Andrew Hardie. 2011. „Visual GISTing: bringing together corpus linguistics and Geographical Information Systems.“ *Literary and Linguistic Computing* 26 (3): 297–314. <https://doi.org/10.1093/lc/fqro22>.
- Gries, Stefan Thomas. 2010a. „Corpus linguistics and theoretical linguistics: A lovehate Relationship?“ Not necessarily ... *International Journal of Corpus Linguistics* 15 (17): 327–343.
- Gries, Stefan Thomas. 2008a. „Dispersions and adjusted frequencies in corpora.“ *International Journal of Corpus Linguistics* 13 (35): 403–437.
- Gries, Stefan Thomas. 2009a. „Dispersions and adjusted frequencies in corpora: further explorations.“ *Language and Computers* 71 (1): 197–212.
- Gries, Stefan Thomas. 2005. „Null-hypothesis significance testing of word frequencies: a follow-up on Kilgarriff.“ *Corpus Linguistics and Linguistic Theory*. 1 (2): 277–294.
- Gries, Stefan Thomas. 2009b. *Quantitative corpus linguistics with R: a practical introduction*. New York: Routledge.
- Gries, Stefan Thomas. 2008b. *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht (Studienbücher zur Linguistik).
- Gries, Stefan Thomas. 2010b. „Useful statistics for corpus linguistics.“ In: *A mosaic of corpus linguistics Selected approaches*, herausgegeben von Aquilino Sánchez und Moisés Almela. Frankfurt am Main: Lang, 269–291.
- Hao, M.C., M. Marwah, H. Janetzko u. a. 2010. „Visual analysis of frequent patterns in large time series“. In *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 227–228, <https://doi.org/10.1109/VAST.2010.5650766>.
- Harleman Stewart, Ann. 1976. *Graphic representation of models in linguistic theory*. Bloomington: Indiana University Press.
- Hearst, M.A., Rosner, D. 2008. Tag Clouds: „Data Analysis Tool or Social Signaller?“ In: *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 160–160, doi: <https://doi.org/10.1109/HICSS.2008.422>.
- Hearst, Marti. 2009. *Search user interfaces*. Cambridge: Cambridge University Press.

- Heringer, Hans Jürgen. 1993. „Basic Ideas and the Classical Model.“ In *Syntax*. Berlin: de Gruyter, 298–316 (Handbücher zur Sprach- und Kommunikationswissenschaft 9, 1).
- Hilpert, Martin. 2011. „Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora.“ *International Journal of Corpus Linguistics*. 16 (4), 435–461, doi: <https://doi.org/10.1075/ijcl.16.4.01hil>.
- Hilpert, Martin und Stefan Thomas Gries. 2009. „Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition.“ *Literary and Linguistic Computing*. 24 (4), 385–401, doi: <https://doi.org/10.1093/lc/fqn012>.
- Jäger, Gerhard. 2014. „Lexikostatistik 2.0.“ In *Sprachverfall? Dynamik – Wandel – Variation*. Berlin: de Gruyter (Jahrbuch 2013).
- Jannidis, Fotis. 1999. „Was ist Computerphilologie?“ In *Jahrbuch für Computerphilologie* (1): 39–60.
- Jannidis, Fotis, Hubertus Kohle, Malte Rehbein (Hrsg.). 2017. *Digital Humanities: Eine Einführung*. Stuttgart: Metzler.
- Jockers, Mathew L. 2013. *Macroanalysis: Digital Methods and Literary History*. Champaign, IL: University of Illinois Press.
- Keim, Daniel A., Jörn Kohlhammer, Geoffrey Ellis und Florian Mansmann. 2010. *Mastering the Information Age - Solving Problems with Visual Analytics*. Goslar: Eurographics Association. <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf>.
- Kilgariff, Adam. 2005. „Language is Never, Ever, Random.“ *Corpus Linguistics and Linguistic Theory*. 1 (2): 263–276.
- Krämer, Sybille. 2016. *Figuration, Anschauung, Erkenntnis: Grundlinien einer Diagrammatologie*. Berlin: Suhrkamp.
- Kupietz, Marc und Holger Keibel. 2009. „Gebrauchsbasierte Grammatik: Statistische Regelmäßigkeit.“ In: *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*, herausgegeben von Marek Konopka und Bruno Strecker. Berlin: de Gruyter, 33–50.
- Lauer, Gerhard. 2011. „Bibliothek aus Daten.“ In *Die digitale Bibliothek*, herausgegeben von Christine Haug und Vincent Kaufmann. Wiesbaden: Harrassowitz, 79–86 (Kodex. Jahrbuch der Internationalen Buchwissenschaftlichen Gesellschaft 1).
- Leblanc, Jean-Marc und Marie Pérès. 2010. „Visualiser les données textuelles : Propositions de fonctionnalités pour une modélisation tridimensionnelle du discours constructeur d’espaces.“ *Transeo Review* 2 (3): 16–25. <https://halshs.archives-ouvertes.fr/halshs-01147433> (letzter Zugriff am 27. November 2017).

- Liebsch, Dimitri, Mößner, Nikola. 2012. *Visualisierung und Erkenntnis. Bildverstehen und Bildverwenden in Natur- und Geisteswissenschaften*. Köln: Herbert von Halem.
- Löffler, Heinrich. 1994. *Germanistische Soziolinguistik*. Berlin: E. Schmidt.
- Lord, Albert Bates. 1960. *The Singer of Tales*. Cambridge, Mass.: Harvard University Press
- Luo, Dongning, Jing Yang, Milos Krstajic, William Ribarsky und Daniel Keim. 2012. "EventRiver: Visually Exploring Text Collections with Temporal References." *IEEE Transactions on Visualization and Computer Graphics* 18 (1): 93–105, doi: <https://doi.org/10.1109/TVCG.2010.225>.
- Manning, Christopher D. und Hinrich Schütze. 2002. *Foundations of Statistical Natural Language Processing*. 5. Aufl. Cambridge, Massachusetts: The MIT Press.
- May, Thorsten, James Davey und Jörn Kohlhammer. 2010. "Combining statistical independence testing, visual attribute selection and automated analysis to find relevant attributes for classification." *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*. 239–240, doi: <https://doi.org/10.1109/VAST.2010.5654445>.
- Mazza, Riccardo. 2009. *Introduction to Information Visualization*. London: Springer.
- Mikuš, Radivoj Francis. 1952. «Quelle est en fin de compte la structure-type du language." *Lingua* (3): 430–470.
- Moretti, Franco. 2000. "Conjectures on World Literature." *New Left Review* (1): 54–68.
- Moretti, Franco. 2009. *Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte*. Frankfurt am Main: Suhrkamp (edition suhrkamp).
- Naumann, Carl Ludwig. 1982. „Kartographische Datendarstellung.“ In: *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Berlin: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft), 667–692.
- Oelke, Daniela, Peter Bak, Daniel A. Keim, Mark Last und Guy Danon 2008. „Visual evaluation of text features for document summarization and analysis.“ In: *2008 IEEE Symposium on Visual Analytics Science and Technology*. 75–82. <https://doi.org/10.1109/VAST.2008.4677359>.
- Daniela Oelke, Ann-Marie Eklund, Svetoslav Marinov und Dimitrios Kokkinakis. 2012a. "Visual Analytics and the Language of Web Query Logs – A Terminology Perspective." In *Proceedings of the 15th EURALEX International Congress*, 541–548.
- Daniela Oelke; David Spretke; Andreas Stoffel; Daniel A. Keim. 2012b. „Visual Readability Analysis: How to Make Your Writings Easier to Read.“ In *IEEE Transactions on Visualization and Computer Graphics* 18 (5): 662–674.

- Parry, Milman. 1971. *The making of Homeric verse: the collected papers of Milman Parry*. Oxford: Clarendon Press.
- Redder, Angelika. 2001. „Aufbau und Gestaltung von Transkriptionssystemen.“ In: *Text- und Gesprächslinguistik / Linguistics of Text and Conversation*. Berlin: de Gruyter, 1038–1059 (Handbücher zur Sprach- und Kommunikationswissenschaft 16, 2).
- Reetz, Henning. 2003. *Artikulatorische und akustische Phonetik*. Trier: WVT, Wiss. Verlag.
- Reichert, André. 2013. *Diagrammatik des Denkens Descartes und Deleuze*. Berlin: de Gruyter.
- Rietveld, Toni und Roeland van Hout. 2005. *Statistics in Language Research: Analysis of Variance*. Berlin: de Gruyter.
- Risch, John, Anne Kao, Stephen Poteet und Y.-J. Jason Wu. 2008. “Text Visualization for Visual Text Analytics.” In *Visual Data Mining*, herausgegeben von Simeon J. Simoff, Michael H. Böhlen und Arturas Mazeika. Berlin: Springer (Lecture Notes in Computer Science), 154–171.
- Rockwell, G., J. Bradley und P. Monger. 1999. “Seeing the text through the trees: visualization and interactivity in text applications.” *Literary and Linguistic Computing*. 14 (1): 115–130. <https://doi.org/10.1093/lc/14.1.115>.
- Rohrdantz, Christian, Ming C. Hao, Umeshwar Dayal, Lars-Erik Haug, Daniel A. Keim. 2012a. “Feature-based Visual Sentiment Analysis of Text Document Streams.” In *ACM Transactions on Intelligent Systems and Technology, Special Issue on Intelligent Visual Interfaces for Text Analysis* 3, issue 2, article 26.
- Rohrdantz, Christian, Michael Hund, Thomas Mayer, Bernhard Wälchli und Daniel A. Keim. 2012b. “The World’s Languages Explorer: Visual Analysis of Language Features in Genealogical and Areal Contexts.” *Computer Graphics Forum*, 31, issue 3pt1: 935–944, doi: 10.1111/j.1467-8659.2012.03086.x.
- Rohrdantz, Christian, Steffen Koch, Charles Jochim, Gerhard Heyer, Geric Scheuermann, Thomas Ertl, Hinrich Schütze und Daniel A. Keim. 2010. “Visuelle Textanalyse.” *Informatik-Spektrum* 33 (6): 601–611, <https://doi.org/10.1007/s00287-010-0483-x>.
- Rohrdantz, Christian, Daniela Oelka, Miloš Krstajić und Fabian Fischer. 2011. „Real-Time Visualization of Streaming Text Data: Tasks and Challenges.“ In *IEEE Workshop on Interactive Visual Text Analytics for Decision Making*. Providence, RI. <http://vialab.science.uoit.ca/textvis2011/papers/textvis%202011-rohrdantz.pdf>.
- Sager, Sven F. 2001. „Formen und Probleme der technischen Dokumentation von Gesprächen.“ In *Text- und Gesprächslinguistik / Linguistics of Text and Conversation*. Berlin: de Gruyter, 1022–1033 (Handbücher zur Sprach- und Kommunikationswissenschaft, 16,2).

- Säily, Tanja, Terttu Nevalainen, und Harri Siirtola. 2011. "Variation in noun and pronoun frequencies in a sociohistorical corpus of English." In *Literary and Linguistic Computing*. 26 (2): 167–188, doi: <https://doi.org/10.1093/llc/fq004>.
- Schleicher, August. 1860. *Die deutsche Sprache*. Stuttgart: Cotta.
- Schumann, Heidrun und Wolfgang Müller. 1999. *Visualisierung: Grundlagen und allgemeine Methoden*. Berlin: Springer.
- Shi, Lei, Furu Wei, Shixia Liu, Li Tan, Xiaoxiao Lian und Michelle X. Zhou. 2010. "Understanding text corpora with multiple facets." In *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 99 –106, doi: <https://doi.org/10.1109/VAST.2010.5652931>.
- Siegel, Steffen. 2009. *Tabula: Figuren der Ordnung um 1600*. Berlin: Akademie-Verlag.
- Stetter, Christian. 2005. *Bild, Diagramm, Schrift*. In *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*. München: Fink (Kulturtechnik).
- Stjernfelt, Frederik. 2007. *Diagrammatology: an investigation on the borderlines of phenomenology, ontology, and semiotics*. Dordrecht: Springer.
- Sutrop, Urmars. 2012. „Estonian Traces in the Tree of Life Concept and in the Language Family Tree Theory.“ *ESUKA – JEFUL*. 1 (3): 297–326.
- Tamassia, Roberto, Hrsg. 2013. *Handbook of Graph Drawing and Visualization*. Boca Raton: CRC Press.
- Thomas, James J. und Kristin A. Cook, Hrsg. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society.
- Thümmel, Wolfgang. 1993a. „Der europäische Strukturalismus.“ In *Syntax*. Berlin: de Gruyter, 257–280 (Handbücher zur Sprach- und Kommunikationswissenschaft 9,1).
- Thümmel, Wolfgang. 1993b. „Geschichte der Syntaxforschung.“ In: *Syntax*. Berlin: de Gruyter, 130–199 (Handbücher zur Sprach- und Kommunikationswissenschaft 9,1).
- Tufte, Edward R. 1983. *The visual display of quantitative information*. Cheshire, Conn.: Graphics Press.
- Tufte, Edward R. 1997. *Visual explanations: images and quantities, evidence and narrative*. Cheshire, Conn.: Graphics Press.
- Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley (Addison Wesley Series in Behavioral Science. Quantitative Methods).
- Unwin, Antony, Martin Theus und Heike Hofmann. 2006. *Graphics of Large Datasets. Visualizing a Million*. Berlin: Springer (Statistics and Computing).
- Vriend, Folkert de, Lou Boves und Roeland van Hout. 2011. "Visualization as a research tool for dialect geography using a geo-browser." *Literary and Linguistic Computing* 26 (1): 17–34, doi: <https://doi.org/10.1093/llc/fq027>.

- Wattenberg, M. und F. B. Viegas. 2008. „The Word Tree, an Interactive Visual Concordance.” In: *IEEE Transactions on Visualization and Computer Graphics* 14 (6): 1221 –1228, <https://doi.org/10.1109/TVCG.2008.172>.
- Wattenberg, Martin. 2005. *Baby Names, Visualization, and Social Data Analysis*. In: *INFOVIS '05 Proceedings of the 2005 IEEE Symposium on Information Visualization*, October 23–25, 2005. <https://doi.org/10.1109/INFVIS.2005.1532122>.
- Zastrow, Thomas. 2011. „Neue Analyse- und Visualisierungsmethoden in der Dialektometrie.“ Dissertation, Universität Tübingen.
- Zhang, Jin. 2008. *Visualization for Information Retrieval*. Berlin: Springer.

