

Daniel Burckhardt/Alexander Geyken/  
Achim Saupe/Thomas Werneke

## DISTANT READING IN DER ZEITGESCHICHTE

Möglichkeiten und Grenzen einer  
computergestützten Historischen Semantik  
am Beispiel der DDR-Presse

Elektronisches Publizieren und online verfügbare Quellenbestände haben die Geschichtswissenschaft verändert. Gesunkene Zugangshürden und einfache Suchmöglichkeiten beschleunigen den Blick in Quellen sowie Literatur und verbreitern die empirische Basis. Programme zum Bibliographieren, Recherchieren und Exzerpieren ersetzen den Zettelkasten, strukturieren unser Wissen neu und vereinfachen den Blick über geographische und disziplinäre Grenzen hinaus.<sup>1</sup> In der Geschichtswissenschaft viel weniger bekannt sind dagegen Möglichkeiten zur Untersuchung großer Textkorpora mittels computerlinguistischer Methoden.<sup>2</sup> Solche Forschungen versuchen qualitative und quantitative Analyseschritte miteinander zu verbinden. Dieser Sprung von der »digitalisierten« zu einer »digitalen« Geschichtswissenschaft, die algorithmische Analyseverfahren methodisch innovativ nutzt und reflektiert, ist längst noch nicht vollzogen.<sup>3</sup>

- 
- 1 Siehe Lara Putnam, The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast, in: *American Historical Review* 121 (2016), S. 377-402.
  - 2 Vgl. Wolfgang Schmale, Big Data in den Kulturwissenschaften, in: ders. (Hg.), *Digital Humanities. Praktiken der Digitalisierung, der Dissemination und der Selbstreflexivität*, Stuttgart 2015, S. 125-137; Martin Schaller, Arbeiten mit digital(isiert)en Quellen: Herausforderungen und Chancen, in: ebd., S. 15-30; Silke Schwandt, Digitale Methoden für die Historische Semantik. Auf den Spuren von Begriffen in digitalen Korpora, in: *Geschichte und Gesellschaft* 44 (2018), S. 107-134.
  - 3 So auch für die Kunstgeschichte: Johanna Drucker, Is There a »Digital« Art History?, in: *Visual Resources* 29 (2013), S. 5-13, hier S. 7.

Im Zuge der Digitalisierung von Quellen, des Aufbaus größerer thematischer Korpora und einer gestiegenen Attraktivität (und oft Unergründlichkeit) bekannter Tools wie dem Ngram Viewer von Google<sup>4</sup> oder auch Visualisierungsmöglichkeiten wie Word Clouds ist zwar ein wachsendes Interesse für digitale Werkzeuge zur Auswertung ausgedehnter Text- und Bildbestände festzustellen. Zugleich bestehen weiterhin methodische Vorbehalte, die aus dem verbreiteten hermeneutischen Zugriff der Geschichtswissenschaft oder aber wissenschaftshistorisch noch aus der Diskussion um Methoden und Ergebnisse quantifizierender Forschungen in den 1970er- und 1980er-Jahren resultieren.<sup>5</sup>

Die mit digitalen Hilfsmitteln vorgehende Analyse von Sprachgebrauch, Sprachwandel und Wortverwendungszusammenhängen – oder vielleicht noch besser: von »Thematisierungsereignissen« und »Thematisierungskonjunkturen«<sup>6</sup> – schließt an Arbeiten der Begriffsgeschichte, der Historischen Semantik und Diskursgeschichte an. Mit der verstärkten Digitalisierung von Quellenbeständen und den Möglichkeiten automatisierter Auszählung gerieten begriffs- und diskursgeschichtliche Ansätze aber in eine latente, wenn auch selten offen thematisierte Krise: Müsste sich der bisher weitgehend hermeneutisch nachvollzogene Begriffs- und Bedeutungswandel bzw. auch diskursive Wandel nicht statistisch erhärten (oder korrigieren) lassen? Müssten sich »semantische Netze«, die Polysemie von Begriffen (oder aber Aussageereignisse und diskursive Formationen) nicht anhand eines Wortumfeldes bzw. anhand von Thematisierungszusammenhängen besser beschreiben lassen? Eine digital informierte Historische Semantik kann das Methodenspektrum erweitern und verändern, indem breitere empirische Belege herangezogen sowie neue Evidenz- und Plausibilisierungsstrategien genutzt werden.

Mit Blick auf die bisherige Forschung lässt sich zunächst feststellen, dass die digitalisierte Geschichtswissenschaft, welche methodischen Vorannahmen sie auch immer impliziert, weiterhin auf der Favorisierung des individuellen *Close Reading* gegenüber maschinellen Methoden des *Distant Reading* beruht.<sup>7</sup> Dieser letztere

4 Siehe <<https://books.google.com/ngrams>>. Dazu Tobias Hodel, Das kleine Digitale. Ein Plädoyer für Kleinkorpora und gegen Großprojekte wie Googles Ngram-Viewer, in: Michael Hagner/Caspar Hirschi (Hg.), *Digital Humanities*, Zürich 2013, S. 103-119.

5 Siehe u.a. Konrad H. Jarausch (Hg.), *Quantifizierung in der Geschichtswissenschaft: Probleme und Möglichkeiten*, Düsseldorf 1976.

6 Philipp Sarasin, Sozialgeschichte vs. Foucault im Google Books Ngram Viewer. Ein alter Streitfall in einem neuen Tool, in: Pascal Maeder/Barbara Lüthi/Thomas Mergel (Hg.), *Wozu noch Sozialgeschichte? Eine Disziplin im Umbruch. Festschrift für Josef Mooser zum 65. Geburtstag*, Göttingen 2012, S. 151-174, hier S. 163.

7 Franco Moretti, *Distant Reading*, London 2013 (dt.: *Distant Reading. Aus dem Englischen übersetzt von Christine Pries*, Konstanz 2016); ders., Conjectures on World Literature, in: *New Left Review* 1/2000, S. 54-68; sowie Stefan Jänicke u.a., On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges (2015), URL: <<http://www.etrapp.eu/wp-content/uploads/2015/07/paper.pdf>>; Barbara Herrnstein Smith, What Was »Close Reading«? A Century of Method in Literary Studies, in: *Minnesota Review* 87 (2016), S. 57-75.

Begriff, von Franco Moretti bereits im Jahr 2000 eingeführt, hat sich seit einiger Zeit zu einem Leitbegriff der Digital Humanities entwickelt. Er betont den Unterschied einer quantitativen und statistischen Auswertung großer Textmengen mit Hilfe des Computers gegenüber einer qualitativen, individuellen Lektüre weniger Texte. Diese Ansätze schließen sich aber gegenseitig nicht aus: Überzeugend wird die Analyse meist erst dann, wenn *Distant* und *Close Reading* eng miteinander verschränkt werden. In den Sozialwissenschaften wurde dafür der Begriff »Blended Reading« ins Spiel gebracht.<sup>8</sup> Für die Geschichtswissenschaft schlagen wir den Begriff »Digitale Hermeneutik« vor.

Bisher gibt es trotz vielfältiger Anstrengungen, spezielle Software für die Geisteswissenschaften zu entwickeln,<sup>9</sup> nur wenige Studien, die aufzeigen, wie sich Arbeitsweisen in der Geschichtswissenschaft – und besonders in der Zeitgeschichte – durch digitale Werkzeuge verändern. Als Beispiel, wie sich mittels lexikometrischer Verfahren neue Analysemöglichkeiten und neue Erkenntnisse ergeben, soll hier ein DDR-Presserkorpus genutzt werden (»Neues Deutschland« und »Berliner Zeitung«), wobei als westliches Vergleichskorpus die Hamburger »ZEIT« herangezogen wird.<sup>10</sup> Als digitales Tool verwenden wir DiaCollo, ein Browser-basiertes Werkzeug zur Analyse und Visualisierung von Worthäufigkeiten und Kollokationen im zeitlichen Verlauf.<sup>11</sup> Auch wenn das DDR-Presserkorpus nur einen Einblick in die weitgehend gelenkte Pressesprache der DDR gewährt und die »ZEIT« den westdeutschen Pressesprachegebrauch

8 Alexander Stulpe/Matthias Lemke, Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung, in: Matthias Lemke/Gregor Wiedemann (Hg.), *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Wiesbaden 2016, S. 17-61. Ähnlich argumentieren Friedrich Alexander/Chris Biemann, Digitale Begriffsgeschichte? Methodologische Überlegungen und exemplarische Versuche am Beispiel moderner Netzsemantik, in: *Forum Interdisziplinäre Begriffsgeschichte* 5 (2016) H. 2, S. 78-96, hier S. 96. Bei Thomas Weitin wird daraus das »Scalable Reading«: Thomas Weitin, Scalable Reading, in: *Zeitschrift für Literaturwissenschaft und Linguistik* 47 (2017), S. 1-6.

9 Für den deutschsprachigen Raum siehe beispielsweise das eHumanities-Förderprogramm des Bundesministeriums für Bildung und Forschung (2011): <<https://www.bmbf.de/foerderungen/bekanntmachung-643.html>>.

10 Das von der Staatsbibliothek zu Berlin erstellte DDR-Presserkorpus umfasst ca. 4 Millionen Artikel aus über 40 Jahren aus den Tageszeitungen »Neues Deutschland«, »Berliner Zeitung« und »Neue Zeit« vom Beginn ihres Erscheinens bis in die Zeit nach 1989/90. Der Bestand wurde komplett digitalisiert und bis in die Einzelartikel segmentiert im Volltext erschlossen. Vgl. <<http://zefys.staatsbibliothek-berlin.de/ddr-presse/>>. Als Recherche-Möglichkeiten stehen Kalenderfunktion und Volltextsuche bereit. Ergänzt werden diese Inhalte durch das vom Leibniz-Zentrum für Zeithistorische Forschung Potsdam aufgebaute Begleitportal »Presse in der DDR: Beiträge und Materialien«, <<http://pressegeschichte.docupedia.de>>. 2015 wurde das DDR-Presserkorpus als »Kurationsprojekt« der Facharbeitsgruppe Zeitgeschichte vom Zentrum Sprache an der Berlin-Brandenburgischen Akademie der Wissenschaften in die Infrastruktur von CLARIN-D eingebunden. Der Zugang erfolgt über <<https://clarin.bbaw.de/de/corpus/>>. CLARIN-D – »Common Language Resources and Technology Infrastructure« in Deutschland – ist eine vom Bundesministerium für Bildung und Forschung geförderte Infrastrukturmaßnahme für Sprachressourcen in den Geistes- und Sozialwissenschaften (<<https://www.clarin-d.net/de/>>).

11 Bryan Jurish, DiaCollo: On the Trail of Diachronic Collocations, in: Koenraad De Smedt (Hg.), *CLARIN Annual Conference 2015. Book of Abstracts. Wrocław, Poland, October 14-16*, Bergen 2015, S. 28-31.

nur sehr eingeschränkt repräsentiert, steht damit für die deutsch-deutsche Zeitgeschichte ein lückenloser, frei zugänglicher Bestand von Zeitungstexten zur Verfügung, der die Jahre von 1946 bis 1990 in Ost und West abdeckt.

Der Sprachgebrauch in der DDR wurde seit den späten 1960er-Jahren auf vielfältige methodische Weise, in unterschiedlichen Disziplinen und oft vergleichend in deutsch-deutscher Perspektive erforscht.<sup>12</sup> Neben solchen komparativen Untersuchungen hat speziell die Erforschung der DDR-Pressesprache sowie der kontrollierten öffentlichen Diskurse in der DDR zu aufschlussreichen Ergebnissen geführt. Mehrfach wurde dabei auf die spezifische Entdifferenzierung der offiziell verlautbarten Sprache hingewiesen, die über einen »hegemonialen Sprachstil« sowie das »Definitions- und Deutungsmonopol« der SED zu einer Ritualisierung der Sprache geführt habe, in der die »Bekräftigung von Werten, die Demonstration von Loyalität und die Gestaltung von Herrschaftsbeziehungen gegenüber den manifesten Inhalten der Rede an Gewicht gewannen« und die »chronische Legitimationskrise« sowie Stagnation der SED-Diktatur widerspiegeln.<sup>13</sup>

Auch in Gesellschaften mit einem freien Journalismus zeigt sich die soziale Wirklichkeit nicht direkt im Sprachgebrauch der Presse, »denn die Massenmedien sagen uns nicht, wie die Welt ist, sondern wie sie wahrgenommen wird«.<sup>14</sup> Für die staatlich gelenkte Presse in der DDR<sup>15</sup> mit einer von der Ausbildung bis in den Redaktionsalltag systematisch operationalisierten »sozialistischen Journalistik«<sup>16</sup> kommt jedoch eine weitere Ebene hinzu: Häufig erfahren wir hier weniger darüber, wie die Welt

- 
- 12 Manfred W. Hellmann, Schrittmacher. Untersuchungen zum Sinnbezirk des vorbildlichen Werkstätigen in der Zeitungssprache der DDR, in: *Muttersprache* 80 (1970), S. 5-15 (Teil 1), S. 127-135 (Teil 2); ders. (Hg.), *Zum öffentlichen Sprachgebrauch in der Bundesrepublik Deutschland und in der DDR. Methoden und Probleme seiner Erforschung*, Düsseldorf 1973; Gotthard Lerchner (Hg.), *Sprachgebrauch im Wandel. Anmerkungen zur Kommunikationskultur in der DDR vor und nach der Wende*, Frankfurt a.M. 1992, 2., durchges. Aufl. 1996; Bettina Bock/Ulla Fix/Steffen Pappert (Hg.), *Politische Wechsel – sprachliche Umbrüche*, Berlin 2011.
- 13 Ralph Jessen, Diktatorische Herrschaft als kommunikative Praxis. Überlegungen zum Zusammenhang von »Bürokratie« und Sprachnormierung in der DDR-Geschichte, in: *Presse in der DDR: Beiträge und Materialien*, 1.6.2011, URL: <[http://pressegeschichte.docupedia.de/wiki/Diktatorische\\_Herrschaft\\_als\\_kommunikative\\_Praxis\\_Version\\_1.html](http://pressegeschichte.docupedia.de/wiki/Diktatorische_Herrschaft_als_kommunikative_Praxis_Version_1.html)>; Ulla Fix, Entdifferenzierung und Ritualisierung von Textsorten im öffentlichen Sprachgebrauch der DDR – das Beispiel Presstexte, in: Stéphanie Benoist/Laurent Gautier/Marie Genevieve Gerrer (Hg.), *Politische Konzepte in der DDR. Zwischen Diskurs und Wirklichkeit*, Frankfurt a.M. 2013, S. 37-68, hier S. 42ff.; dies., *Ritualität in der Kommunikation der DDR*, Frankfurt a.M. 1998; Steffen Pappert, *Politische Sprachspiele in der DDR. Kommunikative Entdifferenzierungsprozesse und ihre Auswirkungen auf den öffentlichen Sprachgebrauch*, Frankfurt a.M. 2003.
- 14 Christian Geulen, Plädoyer für eine Geschichte der Grundbegriffe des 20. Jahrhunderts, in: *Zeit-historische Forschungen/Studies in Contemporary History* 7 (2010), S. 79-97, hier S. 83. Vgl. auch Matthias Lemke/Alexander Stulpe, Text und soziale Wirklichkeit, in: *Zeitschrift für germanistische Linguistik* 43 (2015), S. 52-83.
- 15 Jürgen Wilke, Presseanweisungen. Organisation, Themen, Akteure, Sprechakte, in: *Presse in der DDR: Beiträge und Materialien*, 23.5.2011, URL: <[http://pressegeschichte.docupedia.de/wiki/Presseanweisungen\\_Version\\_1.html](http://pressegeschichte.docupedia.de/wiki/Presseanweisungen_Version_1.html)>.
- 16 Anschaulich etwa: *Wörterbuch der sozialistischen Journalistik*, Leipzig 1979.

wahrgenommen wurde, als darüber, wie sie wahrgenommen werden sollte. Da die Kenntnis dieser offiziellen Sicht und der entsprechenden Formulierungen besonders für Menschen im Bildungswesen, in öffentlichen Verwaltungen und der Armee für die Sicherung der eigenen Position und Karriere hohe Bedeutung hatte, war die DDR nicht etwa trotz, sondern gerade wegen der gelenkten Medien ein ausgesprochenes »Zeitungsleseland«.<sup>17</sup>

Unsere Leitfrage im Hinblick auf die DDR-Pressesprache ist nun einerseits, inwieweit sich anhand von korpusstatistischen Frequenz- und Kollokationsanalysen eine Veränderung der Zukunfts- und Fortschrittssemantik abzeichnet und sich andererseits die These von der Entdifferenzierung und Ritualisierung der DDR-Pressesprache argumentativ stützen lässt. Beides kann im Rahmen dieses Beitrags nur explorativ erfolgen.

## 1. Operationalisierungsmöglichkeiten

Für eine computerlinguistisch informierte Digitale Hermeneutik, Historische Semantik und Diskursgeschichte bietet es sich an, mit Hilfe systematisch erstellter Korpora die Verwendung von Wörtern in ihren jeweiligen Kontexten zu analysieren: »Dieses Verfahren erlaubt einmal die Ermittlung begrifflicher Äquivalenz (oder Teil-Äquivalenz) zwischen Wörtern in verschiedenen Texten des Korpus (z.B. wenn alternative Bezeichnungen in lexikalisch vergleichbarer Umgebung vorkommen); zum anderen ermöglicht der Zugriff auf Wörter die Feststellung von Bedeutungswandel (wenn dasselbe Wort zunehmend in anderen Kontexten erscheint).«<sup>18</sup> Auf diese Weise wurden in den letzten Jahren – meist in sprachwissenschaftlichen Zusammenhängen und der Nutzung digitaler Tools vorausgehend – politisch und gesellschaftlich »kontroverse Begriffe« untersucht und der Sprachwandel im Umfeld politischer Zäsuren analysiert.<sup>19</sup> Zeithistorische Studien versuchen auf der Basis von Volltextanalysen quantitative und

17 Anke Fiedler/Michael Meyen, Zeitunglesen in der DDR, in: *Presse in der DDR: Beiträge und Materialien*, 1.4.2012, URL: <[http://pressegeschichte.docupedia.de/wiki/Zeitunglesen\\_in\\_der\\_DDR](http://pressegeschichte.docupedia.de/wiki/Zeitunglesen_in_der_DDR)>.

18 Dietrich Busse/Wolfgang Teubert, Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik, in: dies. (Hg.), *Linguistische Diskursanalyse: neue Perspektiven*, Wiesbaden 2013, S. 13-30. Eine prägnante und verständliche Einführung in Terminologie und Methodik der Korpuslinguistik bieten Rainer Perkuhn/Holger Keibel/Marc Kupietz, *Korpuslinguistik*, Paderborn 2012.

19 Georg Stötzel/Martin Wengeler (Hg.), *Kontroverse Begriffe, Geschichte des öffentlichen Sprachgebrauchs in der Bundesrepublik Deutschland*, Berlin 1995; Heidrun Kämper, *Opfer – Täter – Nichttäter. Ein Wörterbuch zum Schulddiskurs 1945–1955*, Berlin 2007; dies., *Wörterbuch zum Demokratiediskurs 1967/68*, Berlin 2013; Bock/Fix/Pappert, *Politische Wechsel* (Anm. 12); Dieter Herberg/Doris Steffens/Elke Tellenbach, *Schlüsselwörter der Wendezeit. Wörter-Buch zum öffentlichen Sprachgebrauch 1989/90*, Berlin 1997; Philipp Dreesen, *Diskursgrenzen. Typen und Funktionen sprachlichen Widerstands auf den Straßen der DDR*, Berlin 2015.

qualitative Methoden zu verbinden.<sup>20</sup> Für die computergestützte Auswertung von Texten gemäß diesen Ansätzen stehen – bei entsprechender Aufbereitung bzw. Annotierung der Korpora – verschiedene Distant-Reading-Methoden zur Verfügung, die man in vier Kategorien unterteilen kann.

*Volltextsuchen* gewährleisten den schnellen und umfassenden Zugriff auf Belegstellen. Neben der üblichen Stichwortsuche können durch linguistische Annotation sowohl die Wortart (*part of speech*, POS) und nach einer Lemmatisierung auch alternative Formen des Suchbegriffs in die Volltextsuche einbezogen werden – dann findet eine Abfrage nach »ist« auch »sein« oder »gewesen«. Komplexere Abfragen finden etwa all jene Adjektive im Korpus, die mit »anti-« beginnen oder auf »-bar« enden.

*Lexikometrische Werkzeuge* unterstützen die Bestimmung der absoluten oder relativen Häufigkeit eines Lemmas in den einzelnen Dokumenten. Nach Gruppierung und Summierung gemäß dem Veröffentlichungsdatum können diese Häufigkeiten in Form von Wortverlaufskurven über Jahre oder Jahrzehnte visualisiert werden.<sup>21</sup> Ein weiteres lexikometrisches Werkzeug ist die auf statistischen Korrelationen beruhende Kookkurrenzanalyse – wie dies im Folgenden anhand des Werkzeugs DiaCollo vorgeführt wird. Sie erlaubt die Suche nach empirisch typischen Wortverbindungen in unmittelbarer oder enger Nachbarschaft, sogenannten Kollokationen. Dies können feste Begriffspaare sein – »kommunistische Partei« – oder auf den ersten Blick überraschende Kombinationen wie »Genosse« und »sterben«. »Dieses ›typische Ko-Vorkommen« wird als ›Assoziation« gemessen und statistisch mit einem Signifikanztest modelliert: Wenn die beiden Einheiten in den Daten signifikant häufiger zusammen vorkommen, als es bei einer zufälligen Verteilung der Wörter im Korpus anzunehmen wäre, handelt es sich um eine Kollokation.«<sup>22</sup>

*Methoden des maschinellen Lernens* können zum algorithmischen Gruppieren (*Clustering*) von Texten oder Themenfeldern in Texten gemäß vorgegebener oder automatisch bestimmter Kategorien angewendet werden. Die Art der Kategorisierung wird durch die konkrete Forschungsfrage bestimmt und erfordert in der Regel eine manuelle Annotation, die anhand einer kleinen Teilmenge der Textsammlung vorgenommen wird. Dieses Vorgehen – in den Sozialwissenschaften als qualitative Inhaltsanalyse bekannt – wird in einem Folgeschritt maschinell auf einen deutlich umfassenderen

20 Florian Greiner, *Wege nach Europa. Deutungen eines imaginierten Kontinents in deutschen, britischen und amerikanischen Printmedien, 1914–1945*, Göttingen 2014; Ariane Brill, *Abgrenzung und Hoffnung. »Europa« in der deutschen, britischen und amerikanischen Presse, 1945–1980*, Göttingen 2014.

21 Solche einfachen Wortzählungen lassen sich in verschiedene Richtungen erweitern. Wird etwa über ein Wörterbuch bestimmten Einträgen ein positiver, negativer oder neutraler Gefühlswert zugewiesen, lassen sich die einzelnen Wortverlaufskurven zu einer kumulativen Stimmungskurve aufsummieren (*sentiment analysis*). Vgl. Alberto Acerbi u.a., The Expression of Emotions in 20th Century Books, in: *PLOS ONE* 8 (2013) H. 3, e59030. Die darin gezeigte Stimmungskurve basierend auf dem Google-Books-Korpus zeigt sowohl die »roaring twenties« als auch die »vingt glorieuses«.

22 Noah Bubenhofer/Joachim Scharloth, Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate, in: *Zeitschrift für germanistische Linguistik* 43 (2015), S. 1-26, hier S. 6.

Bestand ausgeweitet.<sup>23</sup> Im Kontext der Digital Humanities werden zum *Clustering* häufig Topic-Modeling-Methoden benutzt.<sup>24</sup> Bei diesen wird nicht die Art, sondern die Zahl der Kategorien festgelegt; eine manuelle Eingruppierung entfällt. Statt etwa die konkrete Rubrikenstruktur einer Zeitung vorzugeben (Politik, Wirtschaft, Kultur, Sport), liefern diese Algorithmen eine Reihe von Begriffen, die für einen – vom Menschen nicht immer eindeutig zu erkennenden – Themenbereich charakteristisch sind.<sup>25</sup>

*Netzwerkanalysen* ermöglichen es, die Beziehungen zwischen Texten innerhalb des Korpus zu untersuchen, wobei jedes Dokument als Knoten im Netz verstanden wird. Zwischen solchen Knoten (*nodes*) wird jeweils eine Kante (*edge*) gezogen, wenn ein bestimmter Begriff in mehreren Dokumenten auftaucht. Dies eignet sich für Studien zur Verbreitung neu geprägter Fachbegriffe oder für die Analyse von Personen- oder Ortsnetzwerken.<sup>26</sup>

In der folgenden Analyse zeigen wir, dass bereits die beiden ersten Methoden wertvolle Einsichten liefern können, d.h. eine Volltextsuche sowie die mathematisch relativ einfachen lexikometrischen Methoden der Frequenz- und Kollokationsanalyse. Dabei sind die Einstiegshürden, die bei vergleichbaren Vorhaben bislang eine hohe

23 Im Rahmen des Projekts »e-pol« wurden auf diese Weise die Konjunkturen neoliberaler Argumentationszusammenhänge in verschiedenen Zeitungen (»Frankfurter Allgemeine Zeitung«, »ZEIT«) und Zeitungsressorts (Politik, Wirtschaft, Kultur) visualisiert. Vgl. Andreas Niekler/Gregor Wiedemann/Gerhard Heyer, Leipzig Corpus Miner – A Text Mining Infrastructure for Qualitative Data Analysis. Terminology and Knowledge Engineering 2014, June 2014, Berlin, URL: <<https://hal.archives-ouvertes.fr/hal-01005878>>.

24 Tze-I.Yang/Andrew J. Torget/Rada Mihalcea, Topic Modeling on Historical Newspapers, in: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Association for Computational Linguistics, 24.6.2011, URL: <<http://dl.acm.org/citation.cfm?id=2107649>>. Eine Einführung in Topic Modeling für Historiker\*innen bieten Shawn Graham/Scott Weingart/Ian Milligan, Getting Started with Topic Modeling and MALLET, in: Programming Historian, 2.9.2012, URL: <<http://programminghistorian.org/lessons/topic-modeling-and-mallet>>.

25 Die »Neue Zürcher Zeitung« hat für den Jahresrückblick 2014 ein Topic Model mit 70 Begriffen erstellt: <<https://storytelling.nzz.ch/2014/chronik/>>. Gut lesbare Begriffsfelder sind in diesem Fall im Bereich Wirtschaft erkennbar (»mio fr mrd umsatz jahr quartal gewinn vorjahr unternehmen konzern...«) sowie für Politik (»partei regierung prozent parlament wahl wahlen stimmen parteien politischen mehrheit ministerpraesident...«); andere Begriffsfelder etwa zu Katastrophen decken sich nicht direkt mit der Rubrikenstruktur der Zeitung.

26 Daniel Burckhardt, Comparing Disciplinary Patterns: Exploring the Humanities through the Lens of Scholarly Communication, in: Digital Humanities Quarterly 11 (2017) H. 2, URL: <<http://www.digitalhumanities.org/dhq/vol/11/2/000298/000298.html>>. Für computerlinguistische Relationsextraktion aus biographischen Texten vgl. André Blessing/Jonas Kuhn, Textual Emigration Analysis (TEA), in: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-14). European Language Resources Association (ELRA), 2014, Reykjavik, S. 2089-2093, URL: <[http://www.lrec-conf.org/proceedings/lrec2014/pdf/1009\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/1009_Paper.pdf)>. Zum Potential einer systematischen Extraktion und Disambiguierung von Personen- und Ortsentitäten in Zeitungskorpora vgl. Thomas Lansdall-Welfare u.a., Content Analysis of 150 Years of British Periodicals, in: *Proceedings of the National Academy of Sciences of the United States of America* 114 (2017), S. E457-E465.

digitale Expertise erforderten, relativ gering gehalten. Besonders hilfreich für historische Untersuchungen ist die mit DiaCollo verbundene Möglichkeit zur diachronen Analyse und Visualisierung von Kollokationen direkt im Webbrowser.<sup>27</sup> Die Echtzeitdarstellung der Ergebnisse ermöglicht interaktive Experimente. Durch die Verknüpfung jeder Kollokation mit der entsprechenden Belegstelle im Artikelarchiv wird ein stetiger Wechsel zwischen *Distant* und *Close Reading* möglich.

## 2. Frequenzanalysen: »Zukunft« und »Fortschritt«, »Kritik« und »Sorge«

Die folgenden Beispiele sollen das Potential von Frequenzanalysen im Hinblick auf den Wandel der Zukunftsemantik des staatssozialistischen Systems beleuchten, die in den letzten Jahren in mehreren Studien analysiert worden ist.<sup>28</sup> Solche Frequenzanalysen, die dank des Ngram Viewer von Google in den Geschichtswissenschaften seit einigen Jahren vermehrt Beachtung finden,<sup>29</sup> zeigen auf, wie häufig (absolut oder relativ) eine spezifische Form in einem bestimmten Segment des Korpus auftritt. Während die Verwendung von »Zukunft« im »Neuen Deutschland« (ND) über den gesamten Zeitraum bemerkenswert stabil blieb, finden wir für den in der sozialistischen Sinnwelt eng damit verknüpften »Fortschritt« eine starke Konjunktur. Dieser Begriff erreichte im ND um 1977 eine relationale Frequenz von 400 pro 1 Million Wörtern (*Token*), bevor er signifikant abfiel.<sup>30</sup>

27 Eine Einführung in diese Analyse-Umgebung mit Beispielen aus dem frei zugänglichen Bestand der »ZEIT« (1946–2015) bieten <<https://clarin-d.de/de/kollokationsanalyse-in-diachroner-perspektive>> sowie <<https://kaskade.dwds.de/diacollo-tutorial/>>.

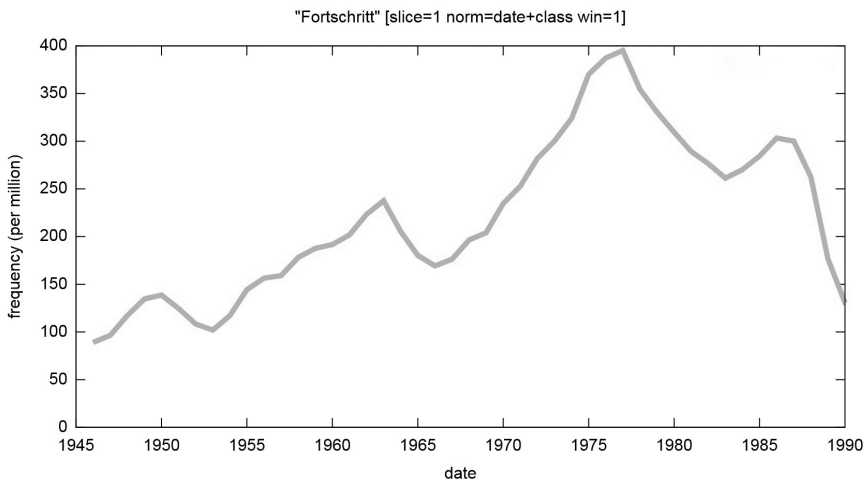
28 Vgl. u.a. unter Einbezug ausgewählter Artikel des DDR-Pressekorpus Martin Sabrow, Chronos als Fortschrittsheld: Zeitvorstellungen und Zeitverständnis im kommunistischen Zukunftsdiskurs, in: Igor Polianski/Matthias Schwartz (Hg.), *Die Spur des Sputnik. Kulturhistorische Expeditionen ins kosmische Zeitalter*, Frankfurt a.M. 2009, S. 117-134; ders., Zukunftspathos als Legitimationsressource. Zu Charakter und Wandel des Fortschrittsparadigmas in der DDR, in: Heinz-Gerhard Haupt/Jörg Requate (Hg.), *Aufbruch in die Zukunft. Die 1960er-Jahre zwischen Planungseuphorie und kulturellem Wandel. DDR, ČSSR und Bundesrepublik Deutschland im Vergleich*, Weilerswist 2004, S. 165-184; Joachim Radkau, *Geschichte der Zukunft. Prognosen, Visionen, Irrungen in Deutschland von 1945 bis heute*, München 2017, S. 297-316.

29 Beispielhaft Sarasin, Sozialgeschichte vs. Foucault (Anm. 6), sowie Franco Moretti/Dominique Pestre, Bankspeak: The Language of World Bank Reports, 1946–2012. Pamphlets of the Stanford Literary Lab 9 (March 2015), URL: <<https://litlab.stanford.edu/LiteraryLabPamphlet9.pdf>>. Da Google im Rahmen seines Buch-Digitalisierungsprogramms primär mit Universitätsbibliotheken kooperiert, sind wissenschaftliche Publikationen mit der entsprechenden Fachterminologie in diesem Datenbestand überdurchschnittlich stark vertreten.

30 Im Vergleichskorpus der »ZEIT« überschreitet »Fortschritt« demgegenüber niemals einen Wert von 90 pro 1 Million *Token* in einem Jahr. Zum »Sinnverlust des Fortschrittsbegriffs« in der Spätzeit der DDR siehe Sabrow, Zukunftspathos als Legitimationsressource (Anm. 28), S. 183.



Prüft man, welche fünf häufigsten Kollokationen der Fortschrittsbegriff mit sich führte und welchen Änderungen diese unterlagen,<sup>31</sup> so zeigt sich zunächst, dass der Fortschrittsbegriff im ND von 1946 bis 1989 ein relativ stabiles semantisches Umfeld hatte. Während der »wissenschaftlich-technische Fortschritt« von Anbeginn der zentrale Terminus war, kam ab ca. 1965 noch »technologisch« hinzu. Der »soziale Fortschritt« überzog gegenüber dem »wirtschaftlichen«. Der Kollokationsrahmen kann erweitert werden, um feinere Veränderungen zu verfolgen und bestimmte Phrasen nachzuvollziehen, etwa die ab Mitte der 1970er-Jahre inflationär verwendete Formel der »Beschleunigung des wissenschaftlich-technischen Fortschritts«.



Häufigkeit des Begriffs »Fortschritt« im »Neuen Deutschland«

Auffällige Häufungen oder Schwächungen eines Begriffs können jedoch nicht immer als Belege für einen umfassenden sprachlichen – oder gar gesellschaftlichen – Wandel herangezogen werden. In der Tages- oder Wochenpresse sind starke Veränderungen im Frequenzverlauf häufig eng an tagespolitische Diskussionen gebunden. So finden wir im ND deutliche Hinweise, dass der Begriff »Jahr 2000«, der einen konkreten Zukunftshorizont eröffnete, Mitte der 1980er-Jahre recht abrupt eine

31 Um festzustellen, ob solche Verbindungen typisch oder zufällig sind, erfordert der statistische Signifikanztest eine Mindesthäufigkeit der beiden Lemmata. Bei deutlich weniger als 1.000 Einzelnennungen in einem vorgegebenen Zeitfenster liefert der Test daher nur selten ein signifikantes Ergebnis. Bei weniger häufigen Begriffen muss deshalb oft die Auflösung der zeitlichen Analyse verringert werden. Umgekehrt können hochfrequente Begriffe auch in engeren Intervallen untersucht werden, zum Beispiel in Zwei- oder Fünf-Jahres-Schnitten. Damit lassen sich auch schnelle Änderungen im Sprachgebrauch zeitlich präzise lokalisieren.



»Kurs 2000«. XI. Parteitag der SED, 1986.  
 Farbiges Bildplakat, 40 x 29 cm  
 (Gestaltung: Hajo Schüler,  
 Foto: Andreas Mußmann;  
 Dokumentationszentrum  
 Alltagskultur der DDR,  
 Inventarnummer 2600/03-DDR)

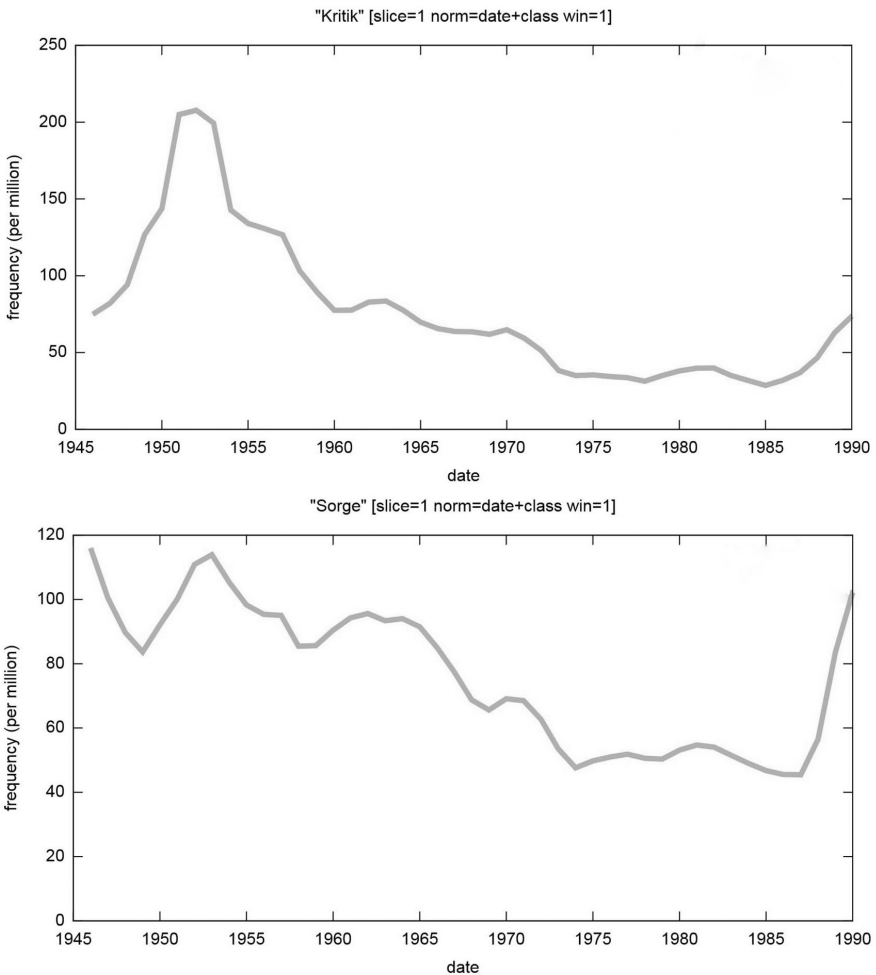
massive Konjunktur erfuhr. Ebenso wie die relative Stabilität des Zukunftsbegriffs mag dies zunächst Thesen widersprechen, die über den Verlust des Fortschritthorizonts in der DDR zugleich auch den der Zukunft insgesamt behaupten.<sup>32</sup> Jede anhand von Verlaufskurven formulierte Hypothese sollte jedoch stets mittels der »Keyword-in-Context«-Funktion (KWIC) anhand konkreter Belegstellen überprüft und eingeordnet werden. Hier zeigen die Einzelbelege recht schnell, dass das »Jahr 2000« weniger einen abstrakten Zukunftshorizont meinte, sondern anlässlich des XI. Parteitages der SED im Jahr 1986 zur realen Bezugsgröße politischer Planung wurde.

Statt nach einzelnen Begriffen aus einem konkreten Forschungsbereich zu suchen, kann auch vom Korpus ausgehend nach den insgesamt häufigsten Begriffen gefragt werden. Um diese Möglichkeiten der DiaCollo-Analyseumgebung sowie das heuristische Potential des DDR-Pressekorpus bei einer Korpus-geleiteten<sup>33</sup> Vorgehensweise

32 Vgl. etwa Radkau, *Geschichte der Zukunft* (Anm. 28), S. 316, sowie auch Hagen Schönrich, »Wir können uns das Tempo nicht aussuchen«. Die DDR-Reportagerie »Wettlauf mit der Zeit« (1986–1989), in: *Zeithistorische Forschungen/Studies in Contemporary History* 13 (2016), S. 179-186.

33 Zur Unterscheidung zwischen *corpus-based* und *corpus-driven* vgl. Noah Bubenhofer, Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse, in: Ingo H. Warnke/Jürgen Spitzmüller (Hg.), *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, Berlin 2008, S. 407-434.

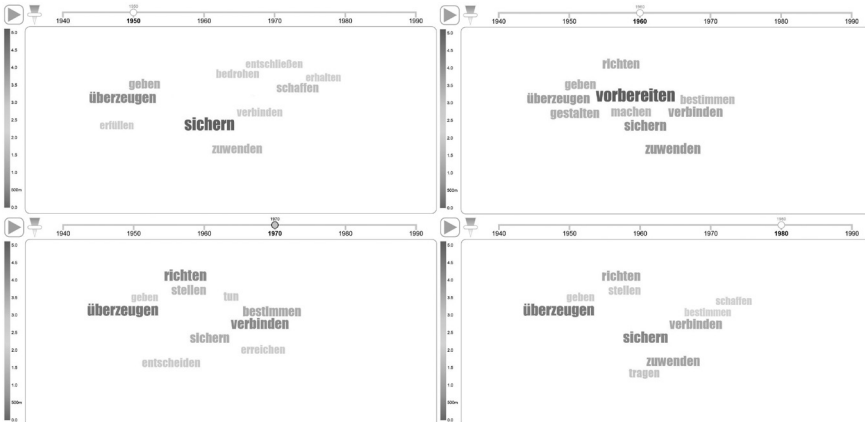
zu verdeutlichen, haben wir die häufigsten Begriffe des Korpus ermittelt und deren Beziehungen zueinander ausgewertet. Ein Blick auf eine vom Computer ausgezählte Liste im ND führt zunächst eher vorhersehbare Begriffe auf wie »Volk« (Platz 3), »Arbeit« (7) »Frieden« (8). Dazu gesellen sich aber auch überraschendere Begriffe wie »Gespräch« (Platz 60), »Bürger« (77), »Kritik« (487) und »Sorge« (485). Sucht man in einem nächsten Schritt nach den Konjunkturen dieser Begriffe, wird deutlich, dass »Kritik« und »Sorge« im ND erst immer seltener vorkamen, bevor sie seit Mitte der 1980er-Jahre wieder verstärkt Einzug in die Tagespresse fanden.



Häufigkeit der Begriffe »Kritik« und »Sorge« im »Neuen Deutschland«

### 3. Diachrone Kollokationsanalysen I: Zukunftssemantik

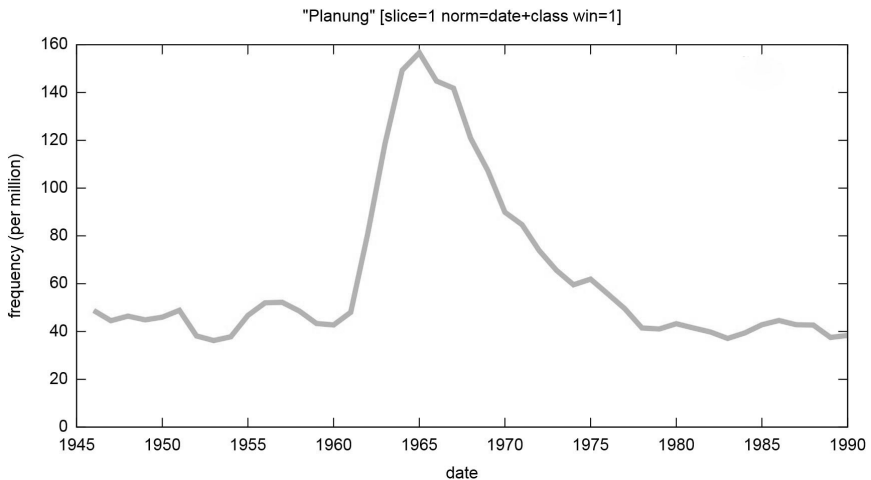
Während wir bisher Frequenzverläufe in den Vordergrund gestellt haben, schauen wir nun vertieft auf die Möglichkeiten der Kollokationsanalyse. Dies kann im Anschluss an aktuelle Forschungen zum historischen Wandel von Zukunftsvorstellungen Aufschluss darüber geben, wie in verschiedenen Modi – etwa der Erwartung, der Gestaltung, der Risikoabwägung und der Erhaltung – über mögliche Zukünfte geschrieben wurde.<sup>34</sup> So lassen sich durch DiaCollo die mit »Zukunft« einhergehenden Verben herausfiltern. Diese zeigen trotz einer relativen Stabilität einige signifikante Unterschiede zwischen den Dekaden: In den 1950er-Jahren galt es, die Zukunft zu »sichern«, sich ihr »zuzuwenden«, wobei hier der entsprechende Text der DDR-Staatshymne entweder direkt zitiert oder aber in vielen Formen adaptiert wurde. Hingegen »bereitete« man sich gerade in den 1960er-Jahren auf die Zukunft »vor«. Dabei handelte es sich um eine Formulierung, die vor allem zwischen 1966 und 1970 Konjunktur hatte. Die Zukunft konnte nach damaligem Verständnis einerseits vorbereitend gestaltet werden, andererseits barg sie trotz allem auch Gefahren. Deshalb startete das ND mit dem Deutschen Fernsehfunk und dem Deutschen Demokratischen Rundfunk 1968 die publizistische Gemeinschaftsaktion »Auf die Zukunft vorbereitet?«.



Mit dem Begriff »Zukunft« verbundene Verben im »Neuen Deutschland«  
(oben: 1950er- und 1960er-Jahre, unten: 1970er- und 1980er-Jahre. Eine Dekade bildet sich wie folgt ab: 1.1.1950–31.12.1959 zum Beispiel für die 1950er-Jahre.)

34 Vgl. Rüdiger Graf/Benjamin Herzog, Von der Geschichte der Zukunftsvorstellungen zur Geschichte ihrer Generierung. Probleme und Herausforderungen des Zukunftsbezugs im 20. Jahrhundert, in: *Geschichte und Gesellschaft* 42 (2016), S. 497-515, insb. S. 500.

Eine »offene Zukunft« kannte das ND während des gesamten Zeitraums der DDR nicht; erst aus dem Jahr 1990 finden wir zwei Beispiele. Im Vergleichskorpus der »ZEIT« gibt es immerhin 58 Erwähnungen. Eine »unsichere Zukunft« wurde im ND dem kapitalistischen Ausland zugeschrieben (16 Treffer). So bestätigte eine kolumbianische Delegation im Jahr 1959 den Bürgern der DDR: »Die Menschen sind hier insbesondere von der Angst vor einer unsicheren Zukunft frei.«<sup>35</sup> Zukunftsvorstellungen sind eng verwoben mit der Hoffnung, Generierung und Erhaltung von »Sicherheit«, einem Begriff, der zum zentralen Vokabular der politisch-sozialen Sprache in beiden Teilen Deutschlands gehörte.<sup>36</sup> Schauen wir also darauf, was in den vier Dekaden der DDR alles »gesichert« werden sollte: In der Aufbauphase der DDR galt es, »Ernährung« und »Versorgung« zu sichern, »Frieden«, »Zukunft« und »Existenz«. Während es den 1950er-Jahren vorbehalten blieb, den »Wohlstand« zu sichern, ging es in der zweiten Hälfte der 1960er-Jahre darum, die »Produktion« zu sichern. In den 1980er-Jahren mussten dann das »Lebensniveau«, genauer das »materielle und kulturelle Lebensniveau«, aber auch weiterhin die »Versorgung«, der »Frieden« und die »Zukunft« gesichert werden – »dauerhaft«, »materiell«, »zuverlässig« oder zumindest »schrittweise«. Wichtig erschien im Hinblick auf die Bewahrung des Lebensniveaus, das »Erreichte« und »Errungene« zu sichern. Hier spiegelt sich die Wirtschaftskrise der DDR. Anhand entsprechender Frequenzanalysen zeigt sich auch, dass man in der Endphase der DDR offensichtlich nicht mehr so viel »erreichen« und »erringen« konnte, wohingegen man ungebrochen »plante«. Die regelrechte Planungseuphorie Mitte der 1960er-Jahre war jedoch schon lange abgeklungen.

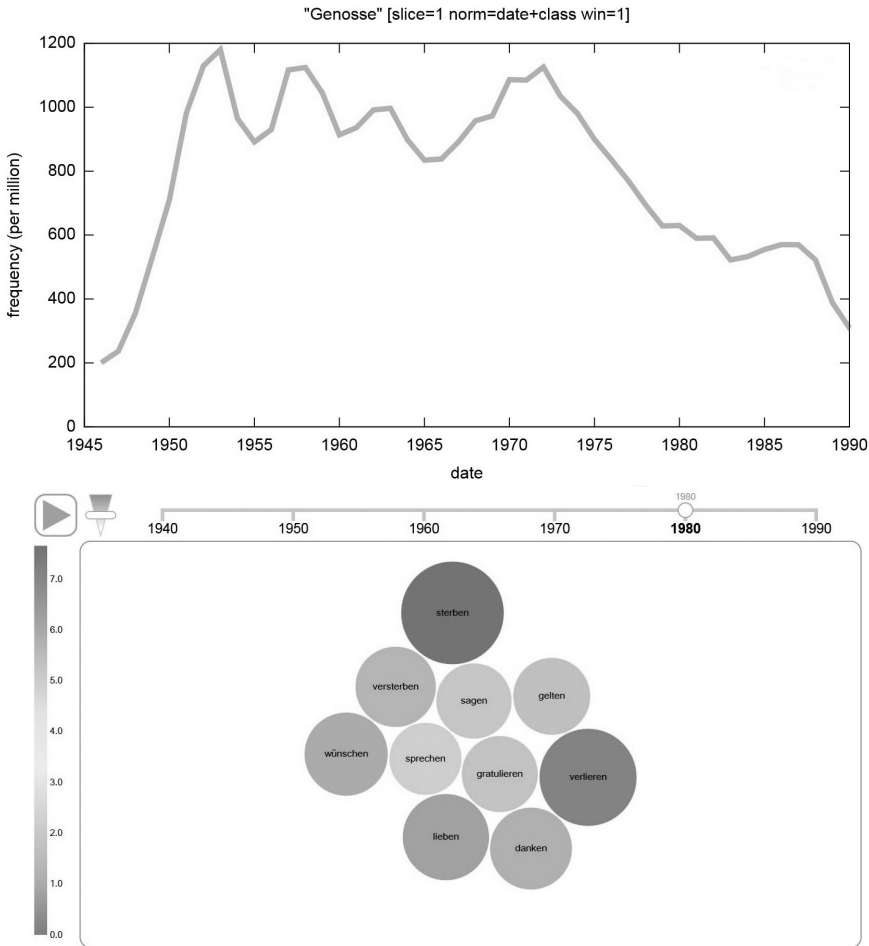


Häufigkeit des Begriffs »Planung« im »Neuen Deutschland«

35 *Neues Deutschland*, 24.5.1959.

36 Für die Bundesrepublik, aber auch mit einigen Hinweisen auf die DDR siehe Eckart Conze, *Die Suche nach Sicherheit. Eine Geschichte der Bundesrepublik Deutschland von 1949 bis in die Gegenwart*, München 2009.

Wie attribuierte man nun die Zukunft? Sie erschien durchgehend als »glückliche Zukunft«, oder verstärkt: als »friedliche und glückliche Zukunft«. Dies schwächte sich allenfalls moderat in den 1980er-Jahren ab, wobei gegen Ende der Dekade die »gemeinsame« Zukunft etwas stärker adressiert wurde (z.B. der »gemeinsame Kampf der Völker für eine friedliche Zukunft«). Während um 1980 der Frieden als »das A und das O einer glücklichen Zukunft der Menschheit« galt, war die Sicherung einer »Zukunft dauerhaften Friedens« später eine oft benutzte Floskel.



Häufigkeit des Begriffs »Genosse« (1945–1990) und mit diesem Begriff verbundene Verben (während der 1980er-Jahre) im »Neuen Deutschland«

Ab Mitte der 1970er-Jahre finden wir in der Presse – Stichwort Biermann-Ausbürgerung – auch eine moderate rhetorische Abkehr von der SED-eigenen Ideologie.<sup>37</sup> Beispielfhaft zeigt der Frequenzverlauf von »Genosse« diesen Bruch in der staatssozialistischen Rhetorik. Hier sind jedoch nicht allein die schieren Häufigkeiten aufschlussreich, sondern vor allem die Verschiebungen im Begriffskontext. Wenn wir mit DiaCollo auf die mit »Genosse[n]« verbundenen Verben schauen, so zeigt sich, dass die »Genossen« von 1945 bis 1975 »berichteten«, »betonten«, und »erklärten«; im schlimmsten Fall hatte man sie »verloren«. Ab 1976 war diese verbale Vielfalt verschwunden: Am charakteristischsten für die »Genossen« der Spätzeit der DDR war es, zu »sterben«.

#### 4. Diachrone Kollokationsanalysen II: Entdifferenzierung der DDR-Pressesprache

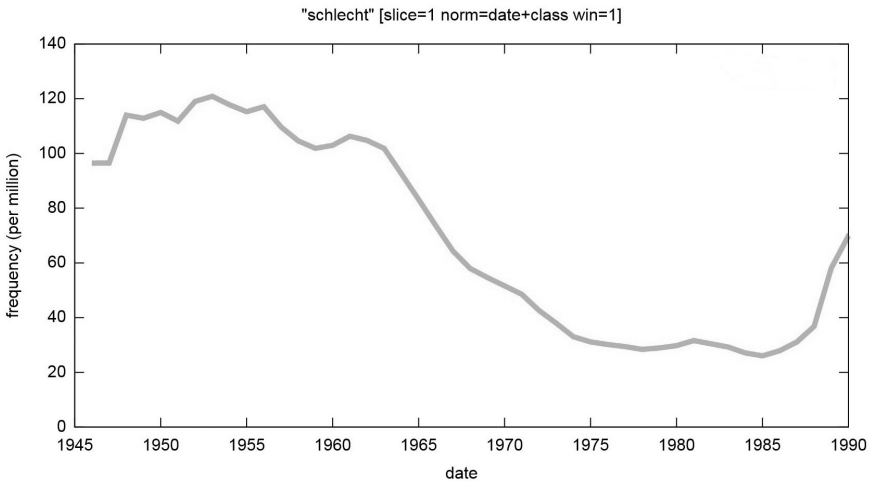
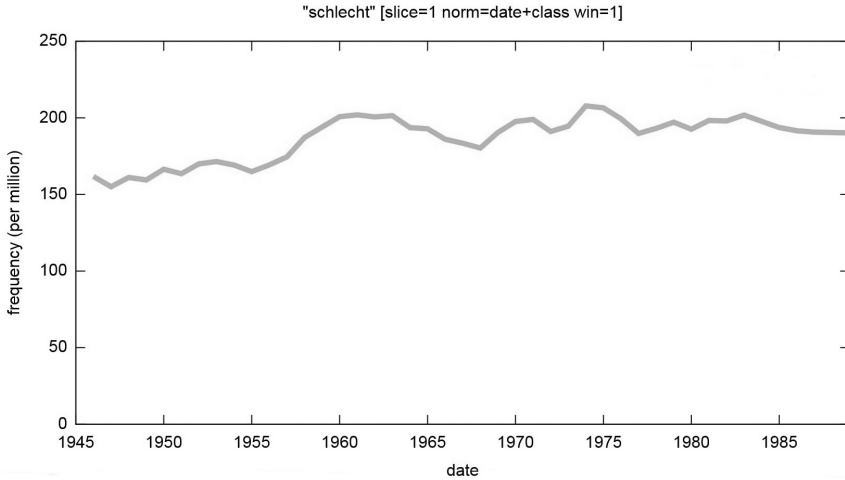
Wie verhält es sich nun mit der in den oben genannten Studien geäußerten These einer Entdifferenzierung der Sprache in den DDR-Printmedien? Die britische Historikerin Mary Fulbrook formulierte gestützt auf ihre Arbeiten zu »Informationsberichten« von Partei und Gewerkschaft 1996 die These, dass die zunehmende Formelhaftigkeit der offiziellen Sprache ab 1960/61 einem besseren Funktionieren des politischen Systems zuzuschreiben sei.<sup>38</sup> Parallel ging die Klassenkampf- bzw. Kalter-Krieg-Rhetorik im Zuge von Entstalinisierung und Mauerbau zurück, wie man an abnehmenden Frequenzverläufen vieler »giftiger Worte« der SED-Diktatur zeigen kann, aber auch anhand einer selteneren Negativberichterstattung über den Westen.<sup>39</sup> Eindrucksvoll zeigt sich dies anhand des Rückgangs des Wörtchens »schlecht«. Während es im Vergleichskorpus der »ZEIT« bemerkenswert konstant vorkam, verlor es im Osten zwischen 1961 und 1985 fast zwei Drittel seiner Nennungen – bevor es danach wieder deutlich anstieg (siehe die Grafiken auf der folgenden Seite).

Ein formelhafter Berichtsstil lässt sich mit DiaCollo schnell nachweisen: So liefert das bis Mitte der 1980er-Jahre zunehmend verwendete »Gespräch« als zugehörige Kollokate »freundschaftlich«, »Honecker«, »Zentralkomitee« und über die entsprechenden Belegstellen einen paradigmatischen Absatz für diese in den letzten beiden Jahrzehnten der DDR immer mehr Raum einnehmende Form der Berichterstattung: »Erich Honecker, Generalsekretär des Zentralkomitees der SED und Vorsitzender des Staatsrates der DDR, empfing gestern Kazimierz Barcikowski, Mitglied des Politbüros und Sekretär des Zentralkomitees der Polnischen Vereinigten Arbeiterpartei, zu einem

37 Vgl. dazu Fiedler/Meyen, *Zeitungen* (Anm. 17).

38 Mary Fulbrook, *Methodologische Überlegungen zu einer Gesellschaftsgeschichte der DDR*, in: Richard Bessel/Ralph Jessen (Hg.), *Die Grenzen der Diktatur. Staat und Gesellschaft in der DDR*, Göttingen 1996, S. 274-297, hier S. 279.

39 Ulrich Weißgerber, *Giftige Worte der SED-Diktatur. Sprache als Instrument von Machtausübung und Ausgrenzung in der SBZ und der DDR*, Berlin 2010.



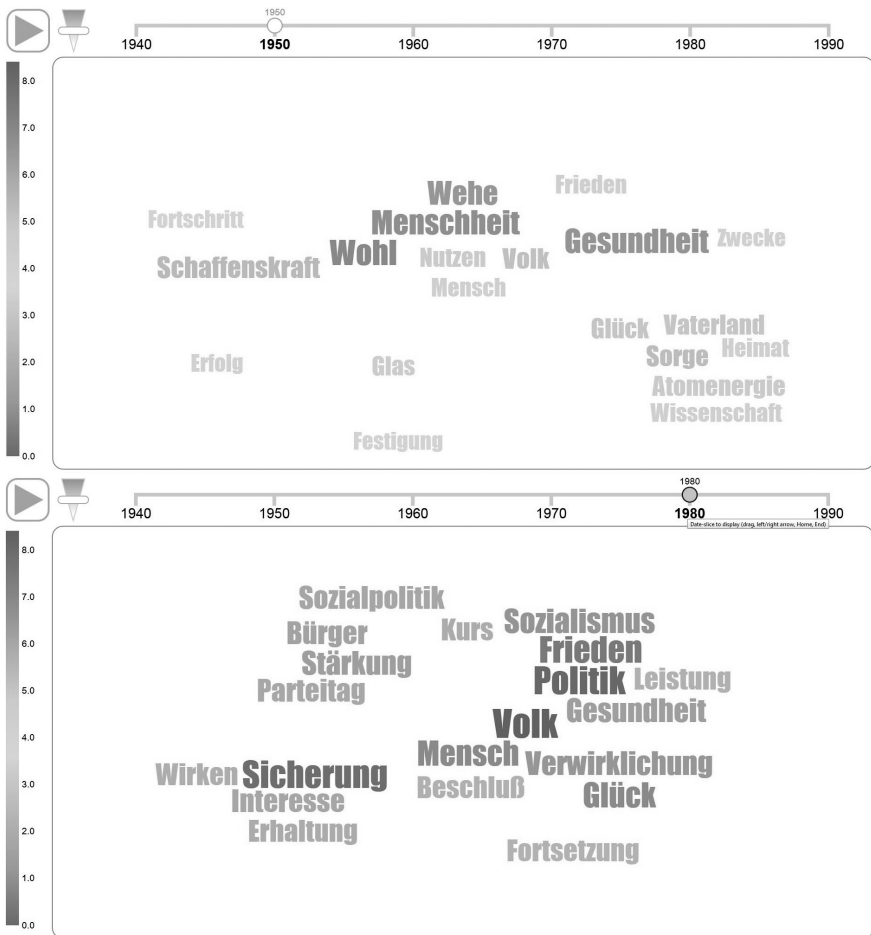
Häufigkeit des Wortes »schlecht« in der »ZEIT« (oben) und im »Neuen Deutschland« (unten)

freundschaftlichen Gespräch [...].«<sup>40</sup> Die »ideologisch motivierte Entdifferenzierung«<sup>41</sup> der DDR-Pressesprache ging darüber jedoch hinaus. Blickt man etwa auf die Kollokate des Begriffs »Wohl«, so kann man feststellen, dass mit Beginn der 1970er-Jahre etliche Substantive öfter in einem gemeinsamen Satzkontext auftraten. Ersichtlich

40 *Berliner Zeitung*, 8.2.1984.

41 Ulla Fix, Text- und diskurslinguistische Analyse der Zeitungssprache der DDR, in: dies., *Sprache, Sprachgebrauch und Diskurse in der DDR. Ausgewählte Aufsätze*, Berlin 2014, S. 349-396, hier S. 358.





Kollokationen für »Wohl« im »Neuen Deutschland«: 1950er-Jahre (oben) und 1980er-Jahre (unten)

wird dies in DiaCollo aus deren Färbung: Je dunkler die Kollokate in der Word Cloud von DiaCollo erscheinen, umso stärker ist die kontextuelle Nähe beider Begriffe im besagten Zeitraum, d.h. ihr Assoziationsgrad.

Für die 1980er-Jahre erscheinen viele Kollokate zu »Wohl« in dieser starken Färbung. »Politik«, »Volk«, »Sicherung«, »Frieden«, »Glück« und »Mensch« stechen hier besonders hervor.<sup>42</sup> An Hochwertwörtern wie »Wohl« sowie deren Wortumgebung

42 Entsprechende Abfragen zu den Begriffen »Fortschritt«, »Kollektiv« und »Verwirklichung« haben ähnliche Ergebnisse gezeigt.

kann man Charakteristika des Verlautbarungsjournalismus der DDR-Presse veranschaulichen. Dabei rückt eine Eigenschaft in den Blick, die man als »modular« bezeichnen kann. Zwei Beispielsätze: »Unsere gemeinsamen Ziele – und dafür setzen wir täglich unser Wissen und Können ein – sind das Wohl des Volkes, das Glück der Menschen, eine Zukunft in Frieden auf der Welt.«<sup>43</sup> »Wir schlugen zum Wohl der Menschen, für das Glück des Volkes, im Interesse der Arbeiterklasse und aller Werktätigen den Kurs der Einheit von Wirtschafts- und Sozialpolitik ein.«<sup>44</sup> Nach dem Baukastenprinzip wurden Hochwertwörter in einem vergleichbaren Rahmen immer wieder neu »aneinandergesteckt«. Der Kontext erscheint nachrangig, und die Begriffe verloren weitgehend ihre sprachliche Bindung an bestimmte Themen. Ob nun »Glück und Wohl für Mensch und Volk« (zum Thema Frieden), »Glück und Wohl für Soldaten und ihre Familien« (zum Thema Militärdienst), »Glück und Wohl für Arbeiter und Werktätige« (zum Thema Wirtschaftsentwicklung) – die modulare Verwendung dieser Hochwertwörter ermöglichte es, eine Vielfalt von Themen in rituell wiederkehrenden Wortzusammenhängen zu subsumieren.

## 5. Fazit und Ausblick

Abhängig vom konkreten Untersuchungsinteresse müssen in einer lexikometrischen Textanalyse generell die Abfragemethode und entsprechende Parameter passend gewählt und die Ergebnisse stichprobenweise sorgfältig überprüft werden. Oder wie es Justin Grimmer und Brandon M. Stewart mit Blick auf politikwissenschaftliche Analysen treffend formulierten:

- »(1) All quantitative models of language are wrong – but some are useful.
- (2) Quantitative methods for text amplify resources and augment humans.
- (3) There is no globally best method for automated text analysis.
- (4) Validate, Validate, Validate.«<sup>45</sup>

Korpusbasierte Analysen können zunächst einmal Annahmen und Ergebnisse der Forschung validieren, differenzieren oder falsifizieren. Mit den hier vorgestellten lexikographischen Werkzeugen, insbesondere der Untersuchung von Kollokationen, kann darüber hinaus gezeigt werden, wie im Rahmen der DDR-Pressesprache Begriffe semantisch in ihrer Bedeutungsvarianz beschränkt und verdichtend modular eingesetzt wurden. Anhand gleichbleibender Kollokationen kann die relative Stabilität von Semantiken visualisiert werden, Veränderungen geben Aufschluss über ihren Bedeutungswandel, der nun über Attribuierungen äußerst differenziert erschlossen und beschrieben werden kann. Und wie das Beispiel »Jahr 2000« zeigt, können auf der

43 *Neues Deutschland*, 8.10.1988.

44 *Neues Deutschland*, 12.8.1989.

45 Justin Grimmer/Brandon M. Stewart, Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, in: *Political Analysis* 21 (2013), S. 267-297, hier S. 269.

Basis von Frequenzanalysen erschlossene Wortkonjunkturen auch ohne umfassende Kenntnisse der DDR-Geschichte recht schnell mit politischen Ereignissen und sozio-kulturellen Zäsuren in Verbindung gebracht werden.

Mit DiaCollo und dem DDR-Pressekorpus steht ein nützliches Instrument für Frequenz- und Kollokationsanalysen zur Verfügung, das im Bereich der (vergleichenden) DDR-Forschung mit geringem Aufwand eingesetzt werden kann – gerade auch dort, wo das Hauptinteresse nicht der Historischen Semantik als solcher gilt. Ohne Zweifel ist es eher die Ausnahme, dass mit einer Ressource wie derjenigen zur DDR-Presse ein Korpus bereitsteht, das Forschungen in einem spezifischen Feld gut überprüfbar macht. Dennoch wäre zu diskutieren, ob sich die hier nur kurz angerissenen Ergebnisse in einigen Punkten verändern würden, wenn ein breiteres Korpus zur Sprache der SED-Diktatur verfügbar wäre, das auch Beschlüsse der Parteitage, Sitzungsprotokolle der DDR-Volkammer, Wörterbücher und weitere Aktenbestände mit einbezüge – im Hinblick auf die DDR-Zeitungen etwa die als »Donnerstag-Argus« bekannten Presseanweisungen. Dass die in der Pressesprache klar zu erkennenden Setzungen von Begriffen parteiintern zugleich umstritten, kontrovers und umkämpft sein konnten, müsste allerdings im Rahmen einer systematisch vergleichenden Analyse mit anderen Korpora oder im klassischen Quellenstudium erschlossen werden. Ein solches größeres Korpus müsste einerseits die Quellenbasis um herrschaftssprachliche Erzeugnisse erweitern, andererseits aber auch Äußerungen von Gegenbewegungen einbeziehen, wie wir sie beispielsweise im politischen Samisdat der DDR finden. Dies könnte darüber Auskunft geben, inwieweit auch die Sprache der Opposition von der Ritualisierung der Herrschaftssprache betroffen war oder umgekehrt nonkonforme Äußerungen und Begriffe Wege fanden, den publizierten politischen Diskurs sprachlich zu reformieren. Doch nicht für jede derartige Frage wird es sich lohnen, ein eigenes digitales Korpus aufzubauen.

Solche Herausforderungen bei der Korpusbildung deuten die Schwierigkeiten an, mit denen weit größere Projekte zu kämpfen hätten, etwa eine computerlinguistisch unterstützte Historische Semantik des 20. Jahrhunderts oder auch ein Diktaturenvergleich zwischen NS- und DDR-Sprache. Ein Referenzkorpus für die (deutschsprachige) Geschichte des 20. Jahrhunderts fehlt ebenso wie für andere Länder. Die Identifikation passender Quellen aus unterschiedlichsten Gattungen, also repräsentativen Teilkorpora, sowie die zahlreichen damit verbundenen rechtlichen Hürden lassen ein solches Vorhaben als herausragende, in näherer Zukunft vermutlich aber kaum lösbare Aufgabe erscheinen. Insofern dürfte es zielführender sein, zunächst weitere kleinere Korpora aufzubereiten, die in einem jeweils klar abgesteckten Bereich geschichtswissenschaftlichen Relevanz- und computerlinguistischen Qualitätskriterien genügen. Neben zusätzlichen Zeitungs- und Zeitschriftenarchiven könnten beispielsweise Parlaments- und Fraktionsprotokolle eine signifikante Menge an Texten bereitstellen, die für die Geschichtswissenschaften bedeutsam wären.

Mit den korpusgeleiteten Analysemöglichkeiten werden sich – so ist zu vermuten – weniger die Fragen der Hermeneutik und Historischen Semantik (Begriffs- und Diskursgeschichte) als vielmehr die Beschreibungsweisen verändern. »Massenquellen« wie Zeitungen oder Parlamentsprotokolle geraten in den Blick, während »Höhenkammliteratur«

(und das mit ihr individuell und intellektuell vorgebrachte Argument) zumindest nicht unmittelbar im Fokus des *Distant Reading* steht. Schwierig bleibt es, auf der Basis eines digitalisierten Korpus »Grundbegriffe«, »Pathosformeln« oder »Signaturen« herauszukristallisieren. Hier ist man – bei allen Fallstricken, die schon im analogen Zeitalter diskutiert wurden – weiterhin auf eine quellenbasierte Benennung charakteristischer Begriffe angewiesen, auch wenn Korpus-getriebene digitale Listen Forscher\*innen dabei unterstützen können und die computergestützte »Keyword-in-Context«-Funktion Fehldeutungen verhindern hilft. Was sich effizienter und zugleich konkreter als bisher beschreiben lässt, sind Konjunkturen und Ritualisierungen hochfrequenter Begriffe und Wortzusammenhänge, die ohne digitale Hilfsmittel kaum nachvollziehbar wären. Hinzu tritt die Möglichkeit, den Wandel semantischer Felder und Begriffsnetze zu visualisieren. Dies beinhaltet Attribuierungen, darunter etwa Verben, die zur Analyse von Sprechakten herangezogen werden können. Bei allem bleibt das Lesen der Quellen zwingend notwendig, zumal sich mit der hier angewandten Methode die Bedeutung eines Lemmas zunächst an seinem näheren Wortumfeld erschließt – nämlich im Rahmen eines Satzes. Eine weiterführende Diskussion über Grenzen und Möglichkeiten der Digitalen Hermeneutik und der *Distant-Reading*-Verfahren ist daher notwendig. Die neuen Verfahren schließen nicht aus, dass mehr gelesen wird als zuvor.

Für zusätzliche Grafiken siehe die Internet-Version unter <https://zeithistorische-forschungen.de/1-2019/id=5694>.

**Daniel Burckhardt**

Institut für die Geschichte der deutschen Juden  
Beim Schlump 83 | D-20144 Hamburg  
E-Mail: burckhardtd@geschichte.hu-berlin.de

**PD Dr. Alexander Geyken**

Berlin-Brandenburgische Akademie der Wissenschaften  
Jägerstr. 22/23 | D-10117 Berlin  
E-Mail: geyken@bbaw.de

**Dr. Achim Saupe**

Leibniz-Zentrum für Zeithistorische Forschung  
Am Neuen Markt 1 | D-14467 Potsdam  
E-Mail: saupe@zzf-potsdam.de

**Dr. Thomas Werneke**

Leibniz-Zentrum für Zeithistorische Forschung  
Am Neuen Markt 1 | D-14467 Potsdam  
E-Mail: werneke@zzf-potsdam.de