



DHd2020

SPIELRÄUME

DIGITAL HUMANITIES ZWISCHEN MODELLIERUNG UND INTERPRETATION

7. Jahrestagung des Verbands Digital Humanities im deutschsprachigen Raum



digital humanities im
deutschsprachigen raum

7. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V.

DHd 2020

Spielräume
Digital Humanities zwischen Modellierung und Interpretation

Konferenzabstracts

Universität Paderborn
02. bis 06. März 2020

Goldsponsoren



WAYS ayfie LYNX

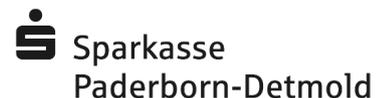
Silbersponsoren



Bronzesponsor



Sponsor des Social Events



Kooperationspartner



Die Abstracts wurden von den Autorinnen und Autoren in einem Template erstellt und mittels des von Marco Petris, Universität Hamburg, entwickelten DHConvalidators in eine TEI konforme XML-Datei konvertiert.

Herausgeber: Christof Schöch

Redaktion und Korrektur der Auszeichnungen:

Nina Seemann, Benjamin Bellgrau

Konvertierung TEI nach PDF: Nina Seemann

<https://github.com/NinaSeemann/DHd2020-BoA>

Historie der Autorinnen und Autoren sowie Versionen der Konversionskripte:

Attila Klett (2019)

<https://github.com/texttechnologylab/DHd2019BoA>

Claes Neuefeind (2018)

<https://github.com/GVogeler/DHd2018>

Aramís Concepción Durán (2016)

<https://github.com/aramiscd/dhd2016-boa.git>

Karin Dalziel (2013)

<https://github.com/karindalziel/TEI-to-PDF>

Konferenz-Logo: Benjamin Bellgrau

Online verfügbar: <https://doi.org/10.5281/zenodo.3666690>

ISBN 978-3-945437-07-0

[2] **Halliday, M. A. K.** (1985): *An Introduction to Functional Grammar*. London: Edward Arnold.

[3] **Han, S.**: Free Google Translate API for Python. Available on <https://pypi.org/project/googletrans/>. Last accessed at 15. August 2019.

[4] **Hosseini-Maasoum, S. M. / Mahdiyan, M.** (2012): Punctuation in Translation: The Unseen Side of the Coin. *Mediterranean journal of social sciences*, 3(11):25–32.

[5] **Kirkman, J.** (2006): *Punctuation Matters: Advice on Punctuation for Scientific and Technical Writing*. Routledge study guides. Routledge.

[6] **Kobus, N.** (2006): *Hard cover: Gedichte*. Ardey Verlag, Münster.

[7] **Lyrikline Literaturwerkstatt Berlin**: Lyrikline: listen to the poet. Available on www.lyrikline.org. Last accessed at 03. September 2019.

[8] **Newmark, P.** (1988): *A Textbook of Translation*. Prentice Hall.

[9] **Papineni, K. / Roukos, S. / Ward, T. / Zhu, W-J** (2002): BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.

[10] **Shiyab, S. M.** (2017): *Translation: Concepts and Critical Issues*. Garant Publishers.

[11] **Wesling, D.** (1996): *The Scissors of Meter: Grammetrics and Reading*. University of Michigan Press.

„Romantik“ im aktuellen parteipolitischen Diskurs auf Twitter

Duan, Tinghui

tinghui.duan@uni-jena.de
DFG Graduierten Kolleg "Modell Romantik"; Jena University Language & Information Engineering Lab (JULIE Lab)

Buechel, Sven

sven.buechel@uni-jena.de
Jena University Language & Information Engineering Lab (JULIE Lab)

Hahn, Udo

udo.hahn@uni-jena.de
Jena University Language & Information Engineering Lab (JULIE Lab)

Die vorliegende Arbeit versucht, im Rahmen einer empirisch fundierten Diskursanalyse von Texten sozialer Medien eine Brücke zwischen qualitativ-hermeneutischer Kulturwissenschaft (hier: Literatur- und Politikwissenschaft) und quantitativ-komputationeller digitaler Geisteswissenschaft zu bauen und beide Methodenlinien synergetisch miteinander zu verschränken. In diesem erweiterten Abstract beschreiben wir einen neuen Datensatz von Twitter-Beiträgen deutscher Parlamentarier des 19. Deutschen Bundestags als Da-

tengrundlage der Diskursanalyse und erste Teilergebnisse, die aus der Analyse dieses Datensatzes resultieren. Ein Fixpunkt dieses Vorgehens ist das historisch markierte Epochenkonstrukt der Romantik in seiner literarischen und sozialen Ausformung (Lebensform, Wertekanon usw.) und seine (Wieder-)Aufnahme bzw. Adaption im aktuellen parteipolitischen Diskurs in Deutschland.

Ausgangspunkt unserer Arbeiten waren Beobachtungen, die einen Bezug zwischen rechtspopulistischen Parteien und Symbolen der deutschen Romantik nahelegten. Während der AfD-Politiker Björn Höcke von seinem Parteikollegen beispielsweise als „romantischer Nationalist“ bezeichnet wurde, trug sein Parteigenosse Andreas Wild bei einem Auftritt im Bundestag eine blaue Kornblume an seinem Revers. Diese Blume, ein zentrales Symbol der Romantik, wurde in den 1930er Jahren sogar zu einem Erkennungszeichen der illegalen Nationalsozialisten in Österreich. Die semantische Doppelbesetzung der blauen Kornblume eröffnet folglich rechtspopulistischen Politikern einen diskursiven Spielraum, sich einerseits implizit an den Nationalsozialismus anzulehnen, andererseits diese Identifikation in der Öffentlichkeit nicht eindeutig zum Ausdruck bringen zu müssen.

Um diese Einzelbeobachtungen systematischer einordnen und die Hypothese von der auffälligen Verwendung von Konzepten der Romantik-Epoche im Diskursverhalten einer rechtspopulistischen Partei einer strengeren Prüfung unterziehen zu können, entwickelten wir ein Korpus von Twitter-Beiträgen aller Abgeordneten des (aktuellen) 19. Bundestags (es kann damit als Ergänzung der Redenkorpora des Bundestags von Barbaresi (2018) bzw. Blätte & Blessing (2018) betrachtet werden, die aber auch frühere Legislaturperioden umfassen). Dieses Korpus sollte Grundlage für eine computerlinguistische Diskursanalyse zur Prüfung der Hypothese sein (einen ähnlichen Studienansatz zur Überprüfung sprachlich markierter Stereotypen zwischen politischen Parteien beschreiben Sylwester & Purver (2015)).

Korpus: Für unsere Untersuchung haben wir DeBAC (*Deutscher BundestagsAbgeordnete-Corpus*), das nach unserem Kenntnisstand erste Twitter-Korpus deutscher Bundestags-abgeordneter für die laufende 19. Legislaturperiode, aufgebaut. Es umfasst zum Zeitpunkt der Abfassung dieses Abstracts (Januar 2020) 887.008 Tweets von 478 Parlamentariern über einen Zeitraum vom 21.11.2008 bis 2.1.2020; dieses Korpus wird fortlaufend aktualisiert. Es umfasst *alle* im Bundestag vertretenen Parteien sowie parteilose Abgeordnete.

Da dieser Datensatz natürlich nicht nur für Fragestellungen im Romantik-Kontext, sondern für die deutschsprachige politische Diskursanalyse generell wertvoll sein kann, stellen wir es der Fachöffentlichkeit zur Verfügung (<https://github.com/JULIElab/DeBAC>). Aus rechtlichen Gründen distribuieren wir dabei nur die Tweet-IDs und dazugehörigen Metadaten (u.a. Autor, Erstellungszeitpunkt und Parteizugehörigkeit), während die Rohtexte über ein ebenfalls mitgeliefertes Skript heruntergeladen werden können.

Analytik: Im ersten Anlauf suchten wir nach Stichwörtern, die Romantik-Konzepte indizieren. Hierzu wurde eine explorative Umfrage unter mehreren Literaturwissenschaftlern (allesamt Mitglieder des Graduiertenkollegs „Modell Romantik“ an der Friedrich-Schiller-Universität Jena)¹ durchgeführt, um gebräuchliche lexikalische Signale für diese Epoche zu bestimmen. Dabei stellte sich heraus, dass nicht nur direkte Lexikalisierungen wie „Romantik“, „Romantiker“, „romantisch“ romantikrelevant sind, sondern auch solche wie „Gemeinschaft“,

„Wesen“, „Glauben“, „Heimat“ (man denke an Friedrich Schlegels *Über den Republikanismus*, Novalis' *Glauben und Liebe* usw.). Das Suchergebnis wurde sowohl quantitativ analysiert als auch qualitativ interpretiert. Die folgende Tabelle zeigt die Häufigkeiten von Tweets mit diesen Stichwörtern und ihre Zuordnung zu Parteien:

Tabelle 1: Häufigkeit der Stichwörter mit Romantikbezug, gruppiert nach Parteien im Bundestag. Tweets der insgesamt vier fraktionslosen Abgeordneten (mit sehr niedrigen Belegzahlen) sind zur Übersichtlichkeit nicht aufgeführt

Suchwort (Regulärer Ausdruck)	CDU/CSU	SPD	AfD	FDP	LINKE	GRÜNE	S
/[Rr]oman- tik/	29	14	7	11	20	15	96
/[Rr]oman- tisch/	9	10	7	13	2	3	44
/[Rr]oman- tisier/	1	2	2	5	0	4	14
/[Gg]lau- ben/	375	298	350	252	198	277	1750
/[Gg]e- mein- schaft/	424	399	104	234	260	343	1764
/[Ww]e- sen/	925	844	504	700	688	835	4496
/[Hh]ei- mat/	1504	941	562	312	314	639	4272
Insgesamt	3267	2508	1536	1527	1478	2116	12436

Die Tabelle zeigt, dass die direkten Lexikalisierungen „*Romantik*“, „*Romantiker*“ und „*romantisch*“ vergleichsweise selten vorkommen und wenn, dann verweisen sie meist auf eine Lesart im Sinne von „*realitätsfern*“, z.B.:

#Grüne und #Linke wollen, dass #Karlsruhe die Patenschaft für ein Seenotrettungsschiff einer Nichtregierungsorganisation (NGO) im Mittelmeer übernimmt. Eine romantische, realitätsferne Weltansicht. (<https://twitter.com/MarcBernhardAfD/status/1062048613923201026>)

Dagegen kommen indirektere Lexeme wie „*Gemeinschaft*“ und „*Heimat*“ weitaus häufiger vor und werden im Sinne eines abgrenzenden und ausschließenden Charakters eingesetzt, z.B.:

Feste, Feiern, Schwimmbäder: Der Verlust öffentlicher Orte und von Gemeinschaftserlebnissen. Nicht alle haben private Pools. <https://t.co/jZsxnMfjCP> (https://twitter.com/Renner_AfD/status/1155441711105134592)

#Bayern gibt Unsummen für illegale Migranten aus. Geld, das vielen älteren Menschen fehlt, die Jahrzehnte für unsere Heimat und unsere Gesellschaft hart gearbeitet haben. Schützen Sie unser Sozialsystem gegen Armutsseinerwanderung und geben wir den Rentnern mehr. #AfD zur #LtwBayern <https://t.co/0imAQg3oCj> (<https://twitter.com/ProfMaier/status/1044102746411073536>)

Diese überwiegend qualitative inhaltsanalytische Vorgehensweise haben wir anschließend durch eine einfache quantitative Untersuchung im Rahmen einer automatischen Emotionsanalyse ergänzt (s.a. entsprechende Vorarbeiten von Hellrich et al. (2019) bzw. Buechel et al. (2017)). Hierzu haben wir sämtliche Tweets unseres Korpus mithilfe des Software-Werkzeugs JEmAS (Buechel & Hahn 2016) analysiert und ihnen so einen emotionalen Stimmungswert anhand der darin vorkommenden Lexeme zugewiesen.

Dieses Verfahren liefert für relativ häufige Wörter intuitiv plausible Ergebnisse. Das Lexem „*Heimat*“, das in insgesamt

4.325 Tweets vorkommt, wird etwa von CDU und CSU am positivsten verwendet und von Der Linken am wenigsten (aber immer noch) positiv. Demgegenüber mussten wir feststellen, dass für unsere Ausgangsforschungsfrage zentrale Begriffe („*Romantik*“, „*romantisch*“, „*romantisieren*“) in unserem derzeitigen Korpus zu selten vorkommen, um damit auf Grundlage von reinen Worthäufigkeiten zuverlässige Daten erheben zu können. Eine sinnvolle Erweiterung unserer bisherigen Arbeiten besteht daher in der Anwendung fortgeschrittenerer komputationaler Modelle zur Emotionserkennung, die etwa auf Deep Learning (Nay 2016) oder Topic Modeling (Nguyen et al., 2015) beruhen. Unsere Studie ist damit dem weiteren Kontext der Meinungsklima- und Emotionsanalytik im Umfeld parlamentarischer politischer Akteure zuzuordnen (vgl. a. Abercrombie & Batista-Navarro 2018, Green & Larasati 2018, Blätte 2018, van der Zwaan et al. 2016, Rheault et al. 2016, Nguyen et al. 2015, Zirn 2014, Lietz et al. 2014), ein aktueller Schwerpunkt im zur Zeit stark expandierenden Bereich *Computational Social Science*.

Danksagung. Tinghui Duan ist Doktorand des Graduiertenkollegs „Modell Romantik“, das von der DFG unter Fördernummer GRK 2041 gefördert wird; Sven Buechel ist Mitarbeiter eines unter der Förderlinie „Big Data in der makroökonomischen Analyse“ (Fachlos 2; GZ 23305/003#002) geförderten Projekts des Bundesministeriums für Wirtschaft; Udo Hahn ist PI in beiden Projekten. Die Autoren bedanken sich bei den zwei anonymen Gutachtern für Ihre kritische Anmerkungen und bei Christof Schöch für seine verständnisvolle Kommunikation.

Fußnoten

1. <http://modellromantik.uni-jena.de/>

Bibliographie

Abercrombie, Gavin / Batista-Navarro, Riza T. (2018): "Identifying opinion-topics and polarity of parliamentary debate motions", in: *WASSA 2018 – Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis @ EMNLP 2018* 280-285.

Barbatesi, Adrien (2018): "A corpus of German political speeches from the 21st century", in: *LREC 2018 – Proceedings of the 11th International Conference on Language Resources and Evaluation* 792-797.

Blätte, Andreas (2018): "Zum Verwechseln ähnlich? Eine Klassifikationsanalyse parlamentarischen Diskursverhaltens auf Basis des PolMine-Plenarprotokollkorpus", in: *Computational Social Science. Die Analyse von Big Data, Nomos* 139-162.

Blätte, Andreas / Blessing, André (2018): "The GermaParl corpus of parliamentary protocols", in: *LREC 2018 – Proceedings of the 11th International Conference on Language Resources and Evaluation* 810-816.

Buechel, Sven / Hahn, Udo (2016): "Emotion analysis as a regression problem: dimensional models and their implications on emotion representation and metrical evaluation", in: *ECAI 2016 – Proceedings of the 22nd European Conference on Artificial Intelligence* 1114-1122.

Buechel, Sven / Hellrich, Johannes / Hahn, Udo (2017): "The course of emotion in three centuries of German text: a methodological framework", in: *dh 2017 – Digital Humanities*

2017: *Conference Abstracts of the 2017 Conference of the Alliance of Digital Humanities Organizations (ADHO)*.

Green, Nathan / Larasati, Septina (2018): "The first 100 days: a corpus of political agendas on Twitter", in: *LREC 2018 – Proceedings of the 11 th International Conference on Language Resources and Evaluation* 2785-2789.

Hellrich, Johannes / Buechel, Sven / Hahn, Udo (2019): "Modeling word emotion in historical language: quantity beats supposed stability in seed word selection", in: *LaTeCH-CLfL 2019 – Proceedings of the 3 rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature @ NAACL-HLT 2019* 1-11.

Lietz, Haiko / Wagner, Claudia / Bleier, Arnim / Strohmaier, Markus (2014): "When politicians talk: assessing on-line conversational practices of political parties on Twitter", in: *ICWSM 2014 – Proceedings of the 8 th International AAAI Conference on Weblogs and Social Media* 285-294.

Nay, John J. (2016): "gov2vec: learning distributed representations of institutions and their legal text", in: *NLP + CSS 2016 – Proceedings of the [1 st] Workshop on Natural Language Processing and Computational Social Science @ EMNLP 2016* 49-54.

Nguyen, Viet-An / Boyd-Graber, Jordan / Resnik, Philip / Miler, Kristina (2015): "Tea Party in the House: a hierarchical ideal point topic model and its application to Republican legislators in the 112th Congress", in: *ACL-IJCNLP 2015 – Proceedings of the 53 rd Annual Meeting of the Association for Computational Linguistics & 7 th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing* 1438-1448.

Rheault, Ludovic / Beelen, Kaspar / Cochrane, Christopher / Hirst, Graeme (2016): "Measuring emotion in parliamentary debates with automated textual analysis", in: *PLoS ONE*, 11, e0168843.

Sylwester, Karolina / Purver, Matthew (2015): "Twitter language use reflects psychological differences between Democrats and Republicans", in: *PLoS ONE*, 10, e0137422.

van der Zwaan, Janneke M. / Marx, Maarten / Kamps, Jaap (2016): "Validating cross-perspective topic modeling for extracting political parties' positions from parliamentary proceedings", in: *ECAI 2016 – Proceedings of the 22 nd European Conference on Artificial Intelligence* 28-36.

Zirn, Căcilia (2014): "Analyzing positions and topics in political discussions of the German Bundestag", in: *Proceedings of the Student Research Workshop @ ACL 2014* 26-33.

Routinen, Ressourcen und Tools der digitalen Texterforschung. Ein einfacher Einstieg

Horstmann, Jan

jan.horstmann@uni-hamburg.de
Universität Hamburg, Deutschland

Flüh, Marie

marie.flueh@uni-hamburg.de
Universität Hamburg, Deutschland

Petris, Marco

marco.petris@uni-hamburg.de
Universität Hamburg, Deutschland

Traditionelle und digitale Arbeitsweisen

Die Anwendung computergestützter Verfahren in den Geistes- und Kulturwissenschaften prägt seit geraumer Zeit die Entwicklung unterschiedlicher Fachdisziplinen (vgl. Thaller 2012). Neue Methoden bahnen sich ihren Weg in den Methodenkanon ganz unterschiedlicher Domänen (vgl. Sahle 2015). Wie aber kann man Lehrenden – mit den unterschiedlichen Ansprüchen universitär Dozierender oder Lehrender an Schulen – einen möglichst niedrigschwelligen, aber dennoch wissenschaftlich seriösen Zugang zu dem Repertoire digitaler Methoden der Texterforschung eröffnen, das zum Spektrum der Digital Humanities zählt? Wie kann man sowohl Begeisterung wie kritische Kompetenz im konkreten Umgang mit Verfahren der digitalen Textanalyse so vermitteln, dass die Alltagspraxis des Lehrens und Forschens davon profitiert? Man muss nicht immer gleich einen theoretischen „Paradigmenwechsel“ ausrufen, sondern kann das „neue“ Feld besser zunächst im „hands-on“-Modus erschließbar machen. Durch einen niedrigschwelligen Disseminationsansatz entsteht die Möglichkeit, dass alte Fragen und neue Methoden sinnvoll aufeinander bezogen werden können (vgl. etwa Horstmann / Kleymann 2019).

Das im November 2017 an der Universität Hamburg gestartete DFG-Projekt forTEXT (<https://fortext.net>) entwickelt vor diesem Hintergrund Strategien zur Dissemination digitaler Verfahren für die Arbeit mit Texten (vgl. Horstmann / Jacke / Meister 2018). In den auf der projekteigenen Webseite als Open-Access-Publikationen bereitgestellten zitierfähigen Besprechungen von Routinen, Ressourcen und Tools werden sämtliche Phasen eines literaturwissenschaftlichen Forschungsprojekts abgedeckt. Das Projekt leistet damit die Übersetzungsarbeit zwischen literaturwissenschaftlichen Fragestellungen und technischem Know-how, die für die Vermittlung digital gestützten Arbeitens an traditionellere Geisteswissenschaftlerinnen notwendig ist.

Routinen

In der Rubrik Routinen stellen wir einführende Einträge zu digitalen *Methoden* der Textdigitalisierung, -annotation, -analyse, -visualisierung, -präsentation etc. zur Verfügung, in denen neben Definition, Diskussion und technischen Hintergründen stets auch die literaturwissenschaftliche Tradition der jeweiligen Methode betont wird. In *Lerneinheiten* zum Selberlernen werden Nutzerinnen schrittweise an die Umsetzung der vorgestellten Methode in Kombination mit der Anwendung eines konkreten Tools (vgl. Abschnitt 4) und ausgewählter Ressourcen (vgl. Abschnitt 3) herangeführt. Die