

Romantik im Wandel der Zeit – eine quantitative Untersuchung

johannes.hellrich@uni-jena.de
Jena University Language & Information Engineering
(JULIE) Lab, Friedrich-Schiller-Universität Jena,
Deutschland

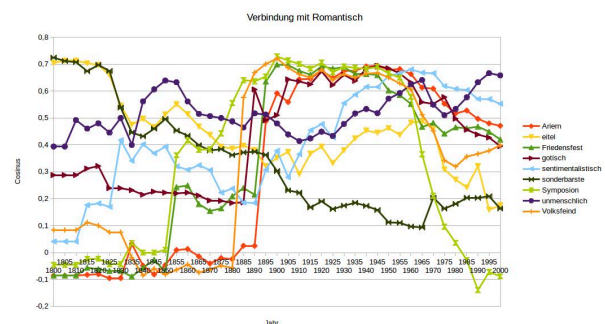
udo.hahn@uni-jena.de
Jena University Language & Information Engineering
(JULIE) Lab, Friedrich-Schiller-Universität Jena,
Deutschland

„Romantik“ und „romantisch“ bezeichnen heute Profanes, wie ein *privates Abendessen in edlem Ambiente bei Kerzenschein* oder ein *idyllisch gelegenes Hotel mit Butzenglasscheiben*, meinte aber ursprünglich *Ästhetisches, insbesondere Literarisches* (DWB). Dieser Bedeutungswandel kann durch die automatische Analyse eines großen Korpus' quantifiziert werden; etwa unter Verwendung des Google Books N-Gram Korpus, das 4% aller je gedruckten Bücher enthält (Michel et al. 2011). Methodische Grundlage für die Analyse sind Verfahren aus der Computerlinguistik (distributionelle Semantik), mittels derer die Bedeutung von Wörtern über den für sie typischen Kontext (also ihren direkten textuellen Kontext) approximiert wird. Ein aktuelles und mächtiges Verfahren ist word2vec (Mikolov et al. 2013), das auf Forschungsarbeiten zu künstlichen neuronalen Netzen basiert (LeCun et al. 2015). Damit gewonnene Repräsentationen können sowohl synchron als auch diachron auf ihre Ähnlichkeit hin verglichen werden, wodurch Bedeutungswandel quantifiziert werden kann. Dies wurde bereits am Beispiel des Englischen durch Kim et al. (2014) demonstriert, die für einen Zeitraum von 100 Jahren Wortrepräsentationen erzeugten und verglichen. Ein Ergebnis war die Quantifizierung des Bedeutungswandels von „gay“ hin zu einer Bezeichnung für (männliche) Homosexualität. Zwischen den word2vec-Repräsentationen einzelner Wörter sind zudem semantisch sinnvolle arithmetische Operationen möglich (Mikolov et al. 2013). Beim Vergleich der modernen und der historischen Bedeutung von „romantisch“ und „Romantik“ könnte somit ermittelt werden, ob sie einander ähnlicher sind, wenn beispielsweise sexuelle Aspekte ignoriert werden.

Alternative Verfahren zur Quantifizierung von Bedeutungswandel nutzen die Kookkurrenz von Wörtern in Bi-Grammen (Gulordava / Baroni 2011), oder einen auf Nachbarwörtern trainierten Klassifikator (Mihalcea /

Nastase 2012). Dabei kann der erste Ansatz lediglich lokale Zusammenhänge erfassen, während der zweite vordefinierte Zeiträume erfordert, zwischen denen ein Bedeutungsunterschied gesucht werden soll. Riedl, Steuer und Biemann (2014) entwickelten einen distributionellen Thesaurus, der für vordefinierte Zeiträume ähnliche Wörter gruppiert. Nachteil dieser Methoden ist wiederum die Notwendigkeit, den untersuchten Zeitraum in Abschnitte zu unterteilen. Vorteilhaft gegenüber word2vec ist die Möglichkeit, die einzelnen Bedeutungen polysemer Wörter getrennt zu erfassen – statt einer veränderlichen Gesamtbedeutung, liegen Teilbedeutungen mit unterschiedlicher Frequenz vor. Eine semantisch schwächere Form des quantitativen Zugangs, aber nützlich für spätere qualitative Interpretationen, ist die Visualisierung von Kollokationen im Zeitverlauf (Jurish 2015).

Die bei unseren Untersuchungen erwarteten Ergebnisse umfassen nicht nur die zunehmende Trivialisierung der Wörter „romantisch“ und „Romantik“ während der letzten 200 Jahre, sondern auch eine Reflexion des deutschen Nationalismus im 19. und 20. Jahrhundert, von dem die Epoche der Romantik instrumentalisiert wurde (Kremer 2007: 50-58). Dem entspricht etwa die erhöhte Frequenz von „Romantik“ in deutschen Texten der Zwischenkriegszeit und direkten Nachkriegszeit, die mit einer Verdrängung des Bi-Gramms „romantische Liebe“ einhergeht, das dafür Ende des 20. Jahrhunderts seine maximale Popularität erreicht. Auch erste Ergebnisse mit einer an Kim et al. (2014) angelehnten Untersuchung, für die ein word2vec-Modell auf dem deutschen Google Books 5-gram Korpus trainiert wurde, entsprechen dieser Erwartung – Abbildung 1 zeigt, dass „romantisch“ während der ersten Hälfte des 20. Jahrhunderts eine hohe Ähnlichkeit (als Cosinus im Vektorraum) zu „Arien“ und „Volksfeind“ hatte.



Wörter mit hoher Ähnlichkeit zu „romantisch“ im Zeitverlauf (hoher Cosinus entspricht hoher Ähnlichkeit).

Geplante Folgearbeiten beinhalten, neben der geisteswissenschaftlichen Einordnung der quantitativen Ergebnisse, den Vergleich über mehrere europäische Sprachen hinweg und die Einbeziehung von Sentiment

Analysis-Technologien, um die emotionale Ladung der Wörter im Verlauf der Zeit abzubilden (Acerbi et al. 2013).

Die beschriebenen Arbeiten sind Teil eines Promotionsvorhabens am von der Deutschen Forschungsgemeinschaft finanzierten Graduiertenkolleg „Modell 'Romantik'“ der Friedrich-Schiller-Universität Jena.

Association for Computational Linguistics. Human Language Technologies – NAACL-HLT 2013 746–751.

Riedl, Martin / Steuer, Richard / Biemann, Chris (2014): "Distributed Distributional Similarities of Google Books over the Centuries ", in: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)* 1401–1405.

Fußnoten

1. Am 24.8.2015 ermittelt über <https://books.google.com/ngrams/>.

Bibliographie

Acerbi, Alberto / Lamos, Vasileios / Garnett, Philip / Bentley, R. A (2013): "The Expression of Emotions in 20th Century Books", in: *PLoS ONE* 8, 3: e59030.

DWB: Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig 1854–1961. Quellenverzeichnis Leipzig 1971. <http://dwb.uni-trier.de/> [letzter Zugriff 30. Juli 2015].

Gulordava, Kristina / Baroni, Marco (2011): "A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus ", in: *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics @ EMNLP 2011* 67–71.

Jurish, Bryan (2015): "DiaCollo: On the trail of diachronic collocations", in: *Proceedings of the CLARIN Annual Conference 2015* 28–31.

Kim, Yoon / Chiu, Yi-I / Hanaki, Kentaro / Hegde, Darshan / Petrov, Slav (2014): "Temporal Analysis of Language through Neural Language Models ", in: *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science* 61–65.

Kremer, Detlef (2007): *Romantik*. Lehrbuch Germanistik. 3. Auflage. Stuttgart: Metzler.

LeCun, Yann / Bengio, Yoshua / Hinton, Geoffrey (2015): "Deep learning", in: *Nature* 521, 7553: 436–444.

Michel, Jean-Baptiste / Shen, Yuan K. / Aiden, Aviva P. / Veres, Adrian / Gray, Matthew K. / The Google Books Team / Pickett, Joseph P. / Hoiberg, Dale / Clancy, Dan / Norvig, Peter / Orwant, Jon / Pinker, Steven / Nowak, Martin A. / Aiden, Erez L. (2011): "Quantitative Analysis of Culture Using Millions of Digitized Books ", in: *Science* 331, 6014: 176–182.

Mihalcea, Rada / Nastase, Vivi (2012): "Word Epoch Disambiguation: Finding How Words Change over Time", in: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* 259–263.

Mikolov, Tomas / Yih, Wen-tau / Zweig, Geoffrey (2013): "Linguistic Regularities in Continuous Space Word Representations", in: *Proceedings of the 2013 Conference of the North American Chapter of the*