Wenn der Funke überspringt – Word Embeddings im Dienst der Wissenschaftsgeschichte

Hellrich, Johannes
johannes.hellrich@uni-jena.de
Graduiertenkolleg „Modell Romantik“, Friedrich-Schiller-Universität Jena, Jena, Deutschland; Jena University Language & Information Engineering Lab (JULIE Lab), Friedrich-Schiller-Universität Jena, Jena, Deutschland

Stöger, Alexander
alexander.stoeger@uni-jena.de
Graduiertenkolleg „Modell Romantik“, Friedrich-Schiller-Universität Jena, Jena, Deutschland

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

Einleitung

Das moderne Verständnis von Elektrizität fußt auf wissenschaftlichen Entdeckungen des 17. und 18. Jahrhunderts, die die technische Nutzbarmachung ab dem 19. Jahrhundert ermöglichten. Unsere Studie wendet state-of-the-art Methoden der distributionellen diachronen Semantik an, um dies anhand des damit einhergehenden Wandels der Wortsemantik von *Elektrizität*, *electricity*, *elektrisch*, *electrical*, *Funken* und *spark* nachzuvollziehen.[1] Im Fokus liegt die Entwicklung der menschlichen Wahrnehmung eines Phänomens, das zuerst als nur schwer begreifliche Naturerscheinung galt, bevor es durch wissenschaftliche Experimente zu einer erklärten Kraft und einem unverzichtbaren Teil des menschlichen Alltags wurde. Wir sind von anderen Studien zur Entwicklung wissenschaftlicher Konzepte (Hall u.a. 2008; Mimno 2012; Fankhauer u.a. 2016; Schumann 2016) inspiriert, die die Entwicklung von Wortgruppen beschreiben (vgl. etwa Topic Modelling nach Blei u.a. (2003)). Dagegen untersuchen wir einzelne Wörter, um differenzierte Aussagen zur Entwicklung des Konzepts Elektrizität machen zu können, das insbesondere in seinem frühen Untersuchungsstadium noch stark fluktuiert.

Methoden

Die von uns verwendeten Verfahren der distributionellen Semantik basieren auf der strukturalistischen Annahme, dass die Semantik von Wörtern über die typischerweise in ihrer Nähe stehenden Wörter approximiert werden kann, prägnant zusammengefasst: “You shall know a word by the company it keeps!” (Firth 1957: 11). Zur Durchführung unserer Experimente nutzen wir JeSemE[2] (Hellrich u. Hahn 2017b). JeSemE basiert auf Word Embeddings, einer state-of-the-art distributionellen Methode zur Messung von Wortähnlichkeit. Word Embeddings sind niedrigdimensionale[3] Vektorrepräsentationen, die die Semantik von Wörtern anhand aller Kontexte repräsentieren, in denen Wörter in einem Korpus beobachtet wurden. Somit benötigen sie keinerlei manuell erstellte Wissensbasis (Wörterbücher oder Ontologien), sondern nur Korpora. Der populärste Algorithmus zur Erzeugung von Word Embeddings ist word2vec (Mikolov u.a. 2013), jedoch kommt in JeSemE stattdessen SVD $_{\text{PPMI}}$ (Levy u.a. 2015) zur Anwendung. Grund für diese Entscheidung ist dessen hohe Reliabilität, wohingegen mit word2vec durchgeführte Experimente nur bedingt wiederholbar sind (Hellrich u. Hahn 2017a). Durch Word Embeddings, die mit Texten aufeinanderfolgender Zeitabschnitte trainiert wurden, kann die diachrone semantische Entwicklung von Wörtern nachvollzogen werden, indem man die zum jeweiligen Zeitpunkt ähnlichsten Wörter ermittelt (Kim u.a. 2014; Kulkarni u.a. 2015; Hamilton u.a. 2016; Hellrich u. Hahn 2016; Jo 2016). Außerdem kann JeSemE anhand der statistischen Maße PPMI (Bullinaria u.a. 2007) und # ^{2} (vgl. etwa Manning u. Schütze (1999)) die spezifischsten Kontextwörter eines Worts identifizieren, also Wörter, die häufiger in seiner Nähe stehen, als auf Grund zufälliger Prozesse zu erwarten wäre. Letzteres dient als Plausibilitätscheck für die mittels Word Embeddings ermittelte Ähnlichkeit.

Korpora

Die beiden von uns untersuchten Korpora sind das Kernkorpus des Deutschen Textarchivs (DTA; Geyken (2013)), das deutschsprachige Sachtexte und literarische Texte des 17. bis 19. Jahrhunderts sammelt, und das Royal Society Corpus (RSC; Kermes u.a. (2016)), das die ersten beiden Jahrhunderte der „Philosophical Transactions of the Royal Society of London“ (1665–1869, älteste wissenschaftliche Zeitschrift in englischer Sprache) beinhaltet. Tabelle 1 zeigt die Kerndaten der von JeSemE prozessierten Korpora, dabei wurden die Korpora in Zeitabschnitte mit ähnlichen Textmengen unterteilt.[4] Beide Korpora sind mit Informationen zur orthographischen Normalisierung und Lemmatisierung annotiert, die von uns genutzt wurde.

Korpus	Zeitraum	Zeitabschnitte	Wörter im Korpus	Modellierte Wörter
DTA	1751–1900	jeweils 30 Jahre	$79{,}6 \times 10^6$	5.388
RSC	1750–1869	50, 50 und 29 Jahre	$24{,}7 \times 10^6$	3.080

Tabelle 1: Kerndaten der verwendeten Korpora.

Ergebnisse

Ausgehend vom wissenschaftshistorischen Forschungsstand ist mit einer semantischen Entwicklung von *Elektrizität* und *electricity* zu rechnen, die weg von beobachteten Naturphänomenen und hin zu kontrollierten Experimenten und systematisch erfassten Grundsätzen führt (Home 2016: 368–71). Wir testeten diese Hypothese, indem wir mittels JeSemE die Ähnlichkeit der beiden Wörter zu anderen bestimmten, die im historischen naturwissenschaftlichen Diskurs relevant waren.

Die Abbildungen[5] 1 und 2 zeigen diese Entwicklung für *electricity* im Englischen und *Elektrizität* im Deutschen. Die Ergebnisse entsprechen der Hypothese, *electricity* wird den mit Experimentalaufbauten verbundenen Wörtern *spark* [‚Funke‘] und *conductor* [‚Leiter‘] im Lauf der Zeit immer ähnlicher, während die Ähnlichkeit zu *lightning* [‚Blitz‘] abnimmt. Analoge Entwicklungen sind auch für *Elektrizität* im DTA erkennbar, besonders deutlich für die Ähnlichkeit zu *Blitz*. Dem zunehmenden theoretischen Verständnis entspricht die steigende Ähnlichkeit zu *magnetism* [‚Magnetismus‘] beziehungsweise *Magnet* — die Entdeckung des Elektromagnetismus im 19. Jahrhundert führte zu einer Vielzahl wissenschaftlicher Publikationen durch Forscher wie Michael Faraday, einem Mitglied der Royal Society.

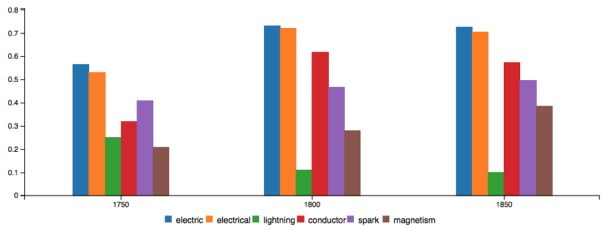

Abbildung 1: Ähnlichkeit ausgewählter Wörter zu ‚electricity‘ im RSC.

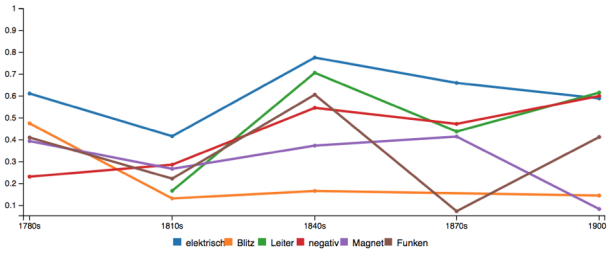

Abbildung 2: Ähnlichkeit ausgewählter Wörter zu ‚Elektrizität‘ im DTA.

Die Analyse spezifischer Kontextwörter (mittels normalisiertem x^2) weist auf einen engen Zusammenhang mit elektrischen Ladungen in DTA und RSC hin, wie in Abbildungen 3 für das RSC gezeigt. Dabei werden *vitreous* [‚gläsern‘] und *resinous* [‚harzig‘] in den historischen Texten entsprechend ihrer elektrischen Eigenschaften synonym zu *negative* und *positive* verwendet (siehe etwa Lichtenberg u. Erxleben (1787: 434)).

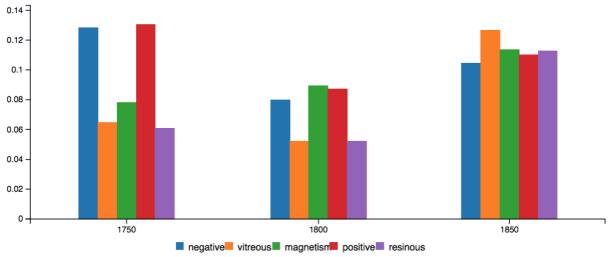

Abbildung 3: Spezifische Kontextwörter für ‚electricity‘ im RSC.

Ein weiterer klarer Hinweis auf das steigende Verständnis elektrischer Phänomene ist die Entwicklung der Adjektive *electrical* und *elektrisch*. Für diese ist ein starker Rückgang der Spezifität von *Materie*, respektive *matter*, erkennbar, wie in Abbildung 4 für das DTA gezeigt. Dies entspricht der bis ins 19. Jahrhundert verbreiteten Idee einer *electrical matter* [‚Elektrisches Fluidum‘], einer abstrakten und unspezifizierten Kraft, die auch als Erklärung für Lebensvorgänge verwendet wurde (Steigerwald 2013). Auffällig ist die unterschiedliche Entwicklung von *Funken* und *spark* – während die Spezifizität von ersterem stetig sinkt, ist die von letzterem vergleichsweise konstant. Funkenbildung war Teil vieler populärer Schauexperimente zur Elektrizität, so beispielsweise bei der „Elektrischen Apotheose“, bei der ein künstlicher Heiligenschein erzeugt wurde (Hochadel 2006: 528).

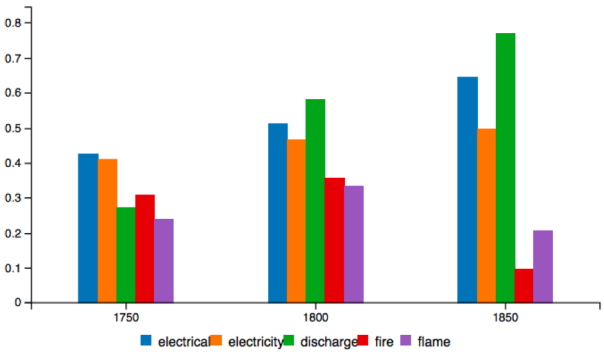

Abbildung 4: Ähnlichkeit ausgewählter Wörter zu ‚spark‘ im RSC.

Diesem Unterschied zwischen DTA und RSC entspricht, dass das Wort *Funken* eine vergleichsweise konstante Ähnlichkeit zu *Feuer* und *Flamme* hat, während die Ähnlichkeit von *spark* und *fire* bzw. *flame* ab Mitte des 19. Jahrhunderts abnimmt, wie in Abbildung 5 gezeigt. Mögliche Erklärungen sind sowohl sprachspezifische semantische Unterschiede, als auch die unterschiedliche Zusammensetzung der beiden Korpora, wie etwa eine Wechselwirkung der im DTA enthaltenen literarischen Texte (vgl. etwa „Wenn er dich mit seinem Auge electrisirt, fühl es, daß es ein königlicher Funke sey" (Hippel 1781: 28)). Wissenschaftshistorische Studien legen aber auch nahe, dass die im 18. Jahrhundert mit Elektrizität durchgeführten Experimente bis ins 19. Jahrhundert länger und öfter wiederholt und erweitert wurden, als das im deutschen Sprachraum der Fall war. Während die Erscheinung dort zum Ausgang des 18. Jahrhunderts abnahm und rasch zum Elektromagnetismus umgedeutet wurde, verlieren die Elektrizitätsexperimente in Großbritannien nicht an Beliebtheit (vgl. Morus (1998)).

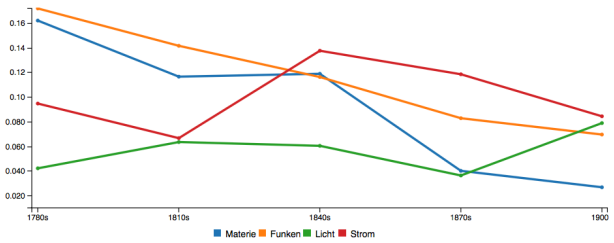

Abbildung 5: Spezifische Kontextwörter für ‚elektrisch' im DTA.

Schluss

Durch die Anwendung von state-of-the-art Methoden der distributionellen diachronen Semantik konnten wir unsere Hypothese zur Entwicklung des Verständnisses von Elektrizität, weg von einem in der Natur beobachteten Phänomen und hin zu etwas im Labor Erzeugtem, das mit anderen Phänomenen wie Magnetismus kombiniert untersucht wird, bestätigen. Dazu erkundeten wir die semantische Entwicklung der Wörter *Elektrizität*, *electricity*, *elektrisch*, *electrical*, *Funken* und *spark* mittels des SVD $_{PPMI}$ Word Embedding Verfahrens und des statistischen Maßes x^2. Die Bestätigung unserer Erwartung, die auf dem wissenschaftsgeschichtlichen Forschungsstand basiert, sehen wir als Hinweis für die methodologische Eignung unserer Herangehensweise.

Wir konnten JeSemE erfolgreich zur Bestätigung bereits bestehender Thesen, die auf close-reading-Methoden aufbauen, einsetzen. Wir hoffen, darauf aufbauend Arbeitsabläufe zu entwickeln, die einen gezielten Zugang zu forschungsrelevanten Texten innerhalb umfangreicher Textkorpora wie historische Fachzeitschriften ermöglichen und somit close und distant reading (vgl. Moretti (2013)) verbinden. Die Analyse der Bedeutungsentwicklung über lange Zeiträume und mehrere internationale Korpora hinweg eröffnet neue Forschungsperspektiven, durch die bisher unzugängliche Eigenheiten oder Veränderungen erkennbar werden. Offen bleibt die Frage, wie unsere fokussierte Betrachtung von Einzelwörtern am besten mit der abstrakteren Analyse von Konzepten — dem in der aktuellen Forschung dominierenden Ansatz — verbunden werden kann.

Danksagung

Die beschriebenen Arbeiten wurden von der Deutschen Forschungsgemeinschaft im Rahmen des Graduiertenkollegs „Modell 'Romantik'" (GRK 2041/1) gefördert.

Fußnoten

1. Im Folgenden wird *kursiv* verwendet um untersuchte Wörter vom sonstigen Text abzusetzen.
2. www.jeseme.org
3. Typischerweise einige wenige hundert Dimensionen, in JeSemE kommen 500 zum Einsatz.
4. Grundsätzlich decken beide Korpora längere Zeiträume ab als in den in JeSemE genutzten Versionen, das ein Mindestmaß an Text für die angemessene Modellierung der Wortsemantik benötigt.
5. Die Darstellungsarten für DTA und RSC wurden entsprechend der jeweils für die Analyse nutzbaren Zeitabschnitte gewählt. Alternative Dartellungen wie in Kulkarni u.a. (2015) oder Hamilton u.a. (2016), die Bewegungen innerhalb einer Art Wordcloud nutzen, wurden verworfen, da sie fälschlicherweise eine statische Semantik der Vergleichswörter implizieren.

Bibliographie

Blei, David M. / Ng, Andrew Y. / Jordan, Michael I. (2003): "Latent Dirichlet Allocation", in: *The Journal of Machine Learning Research* 3: 993–1022.

Bullinaria, John A. / Levy, Joseph P. (2007): "Extracting semantic representations from word co-occurrence statistics: A computational study", in: *Behavior Research Methods* 39(3):510–26.

Fankhauser, Peter / Knappen, Jörg / Teich, Elke (2016): "Topical diversification over time in the royal society corpus", in: *Digital Humanities 2016* 496–500.

Firth, John Rupert (1957): "A synopsis of Linguistic Theory, 1930–1955", in: *Studies in linguistic analysis* 1–32.

Geyken, Alexander (2013): "Wege zu einem historischen Referenzkorpus des Deutschen: das Projekt Deutsches Textarchiv", in: *Perspektiven einer corpusbasierten historischen Linguistik und Philologie* 221–34.

Hall, David / Jurafsky, Daniel / Manning, Christopher D. (2008): "Studying the History of Ideas Using Topic Models", in: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* 363–71.

Hamilton, William L. / Leskovec, Jure / Jurafsky, Dan (2016): "Diachronic Word Embeddings reveal statistical laws of semantic change", in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* 1489–501.

Hellrich, Johannes / Hahn, Udo (2016): "Measuring the Dynamics of Lexico-Semantic Change Since the German Romantic Period", in: *Digital Humanities 2016* 545–7.

Hellrich, Johannes / Hahn, Udo (2017a): "Don't Get Fooled by Word Embeddings-Better Watch their Neighborhood", in: *Digital Humanities 2017* 250–2.

Hellrich, Johannes / Hahn, Udo (2017b): "Exploring Diachronic Lexical Semantics with JeSemE", in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics – System Demonstrations* 31–6.

Hippel (1781): *Lebensläufe nach Aufsteigender Linie* . Meines Lebenslaufs Dritter Theil. Zweyter Band.

Hochadel, Oliver (2006): „The Business of Experimental Physics: Instrument Makers and Itinerant Lecturers in the German Enlightenment“, in: *Science & Education 2007* 525–37.

Home, R. W. (2016): „Experimental Physics“, in: Porter, Roy: *The Cambridge History of Science, Volume 4: Eighteenth-Century Science* 363–71.

Jo, Eun Seo (2016): "Diplomatic history by data. Understanding Cold War foreign policy ideology using networks and NLP", in: *Digital Humanities 2016* 582–5.

Kermes, Hannah / Degaetano-Ortlieb, Stefania / Khamis, Ashraf / Knappen, Jörg / Teich, Elke (2016): "The royal society corpus: From uncharted data to corpus", in: *Proceedings of the Tenth International Conference on Language Resources and Evaluation* 1928–31.

Kim, Yoon / Chiu, Yi-I / Hanaki, Kentaro / Hegde, Darshan / Petrov, Slav (2014): "Temporal analysis of language through neural language models", in: *Proceedings of the Workshop on Language Technologies and Computational Social Science @ ACL 2014* 61–5.

Kulkarni, Vivek / Al-Rfou, Rami / Perozzi, Bryan / Skiena, Steven (2015): "Statistically significant detection of linguistic change", in: *Proceedings of the 24th International Conference on World Wide Web – Technical Papers* 625–35.

Levy, Omer / Goldberg, Yoav / Dagan, Ido (2015): "Improving distributional similarity with lessons learned from Word Embeddings", in: *Transactions of the Association for Computational Linguistics* 3:211–25.

Lichtenberg, Georg Christoph / Erxleben, Johann Christian Polykarp (1787): *Anfangsgründe der Naturlehre* . Dietrich, 4. Auflage.

Manning, Chris / Schütze, Hinrich (1999): *Foundations of Statistical Natural Language Processing.* MIT Press, Kapitel 5: Collocations.

Mikolov, Tomas / Chen, Kai / Corrado, Gregory S. / Dean, Jeffrey (2013): "Efficient estimation of word representations in vector space", in: *Workshop Proceedings of the International Conference on Learning Representations* https://arxiv.org/abs/1301.3781

Mimno, David (2012): "Computational historiography: Data mining in a century of classics journals", in: *Journal on Computing and Cultural Heritage* 5(1): Article 3.

Moretti, Franco (2013): *Distant Reading* . Verso.

Morus, Iwan Rhys (1998): *Frankenstein's Children. Electricity, Exhibition, And Experiment in Early-Nineteenth-Century London.* Princeton University Press.

Schumann, Anne-Kathrin (2016): "Brave new world: Uncovering topical dynamics in the ACL anthology reference corpus using term life cycle information", in: *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities @ ACL 2016* 1–11.

Steigerwald, Joan (2013): "Rethinking organic vitality in Germany at the turn of the nineteenth century", in: Normandin, Sebastian / Wolfe, Charles T.: *Vitalism and the Scientific Image in Post Enlightenment Life Science, 1800-2010* , Springer, 51-75.