

Textexplorationen in der digitalen Literaturwissenschaft Eine kritische und angewandte Auseinandersetzung mit Repräsentations- und Interpretationsansätzen von Text

Brandl, Stephanie

stephanie.brandl@tu-berlin.de
Københavns Universitet; Technische Universität Berlin

Lassner, David

lassner@tu-berlin.de
Technische Universität Berlin

Krömer, Cora

cora.kroemer@fau.de
Universitätsbibliothek Erlangen-Nürnberg

Baillet, Anne

anne.baillot@univ-lemans.fr
Le Mans Université

Anforderungen an den Ort

Maximalanzahl Teilnehmende: 25
Räumliche Anforderungen:

- Beamer
- Whiteboard/Tafel
- Stromversorgung für Laptops der Teilnehmenden
- Wifi

Anforderungen an die Teilnehmenden

Idealerweise bringen Teilnehmende ihre eigenen Laptops mit, die bestenfalls schon die nötige Software vorinstalliert haben. Wir werden kurz vor der Konferenz eine Willkommens-E-Mail mit einer Liste der relevanten Software verschicken. Die praktischen Sitzungen werden mithilfe von Jupyter Notebooks (Python3, Jupyter) abgehalten. Wir planen zusätzlich als Absicherung einen Online-Zugang zu einem JupyterHub Server mit vorinstallierten Paketen für Teilnehmende, bei denen die Installation Schwierigkeiten macht.

Wir ermutigen die Teilnehmenden ausdrücklich einen eigenen Datensatz mitzubringen, an Beispiel dessen die Aufgaben ausge-

führt werden können. Wir gehen davon aus, dass dies aufschlussreich für die jeweiligen Teilnehmenden ist und zusätzlich den Workshop bereichert. Wir haben außerdem bereits mit zwei Dateninstitutionen Kontakt aufgenommen. Beide haben Interesse bekundet sowohl Daten für den Workshop zur Verfügung zu stellen als auch am Workshop teilzunehmen. Damit stellen wir sicher, dass auch für Teilnehmende, die keine Daten mitbringen, Material vorhanden ist, und im Zweifel auch Expert:innen vor Ort sind, die etwaige Zwischenergebnisse der verwendeten Methoden evaluieren und in Kontext setzen können.

Beschreibung

Traditionelle Methoden und Theorien der Literaturwissenschaft können in vier Typen unterteilt werden: autor-, text-, leser- und kontextorientiert [Köppe & Winko, 2013]. Die Literaturwissenschaft zeichnet sich durch einen Methodenpluralismus aus [Nünning & Nünning 2020], der in den letzten Jahren vor allem durch kognitionswissenschaftliche [Zunshine, 2015] und empirische Ansätze [Kuiken & Jacobs, 2020] weiter ausdifferenziert wurde. Gerade diese haben zu neuen Erkenntnissen in leserorientierten Ansätzen geführt. Umgekehrt lässt sich auch vermuten, dass neue Lektürewesen, –in unserem Fall maschinengenerierte Lektüren– literaturwissenschaftliche Innovationen direkt oder mittelbar als Folgewirkung hervorbringen [Parr & Honold, 2018]. Nicht nur die Literaturwissenschaft an sich, sondern auch ihre leserorientierten Ansätze zeichnen sich durch Theorien- und Methodenpluralismus aus [Rautenberg & Schneider, 2015], z. B. rezeptionstheoretische Ansätze [Willand, 2014].

Die Verbreitung von digitalen Lesemedien führt zu einer neuen Beschäftigung mit Repräsentationsformen von Texten und Textorganisation [Saemmer, 2015], da der Sinn eines Textes nicht davon unabhängig gedacht werden kann. Dies führt zu neuen und zu erlernenden Lesestrategien, die es zu erforschen gilt.

Auf der anderen Seite prägen visuelle Repräsentationen von Literatur schon lange ihre Institutionalisierung: Literarische Bewegungen werden in Schulbüchern in Form von Zeitleisten dargestellt, dramatische Handlungen mit der Freytagschen Pyramide abgebildet, und dies schon seit dem 19. Jahrhundert. Bereits diese Ansätze setzen sich zum Ziel, übergeordnete Strukturen anschaulich zu machen, die sich aus dem literarischen Text selbst ableiten lassen. Mit der bourdieuschen Literatursoziologie etablierte sich in den 60er Jahren die Vorstellung der Literatur als ein Feld, das von diversen Dynamiken durchzogen wird. Die digitalen Methoden, spätestens seit Morettis *Graphs Maps Trees* (2005), gehen in der räumlich-visuellen Exploration von Literatur weiter. Der Einsatz von Machine Learning-Methoden seit den 2010er Jahren führte einerseits zur Diversifizierung der Visualisierungsmethoden (WordClouds als Visualisierung von Topic Models, Netzwerke als Ergebnis von Named Entity Recognition oder Distanzgraphen aus stilometrischen Analysen), andererseits zur Repräsentation von immer abstrakteren Literaturphänomenen, bei denen der unmittelbare Zusammenhang mit dem literarischen Text nur mittels komplexer Rechenmethoden nachzuvollziehen ist. In diesem Zusammenhang haben sich Word Embeddings als eine mögliche vektorielle Repräsentationsmethode etabliert, die sich unmittelbar an die Semantik des literarischen Textes anlehnt. Wörter werden hier, basierend auf ihrem Kontext in Bezug auf andere Wörter im Text, als Vektoren dargestellt. Der resultierende Vektorraum bildet so semantische wie syntaktische Beziehungen zwischen Wörtern ab.

Word Embeddings bilden häufig die Brücke zwischen dem Text als Zeichenfolge und der Darstellung, mit der digitale Literatur-

wissenschaftler:innen ihr Korpus lesen, also bspw. als Input für ein ML-Modell.

In *Reading Machines* rückt Stephen Ramsay eine *Radical Transformation* [Ramsay 2011] der Lesart in den Fokus: kein sequentieller Leseprozess sondern z.B. Kapitel in zufälliger Reihenfolge, aber auch etwa 'nur die häufigsten 100 Worte lesen'. Damit stellt sich in hermeneutischer Perspektive die Frage der Angemessenheit der gewählten Transformation. Wenn wir die Überführung in Word Embeddings ebenfalls als solch eine Radical Transformation betrachten, wie wirkt sich das auf das Lesen aus? Ist einem Word Embedding, das auf dem Werk eines Autors oder einer Autorin trainiert wurde, die Angemessenheit für sie oder ihn in gewisser Weise durch den Lesehorizont dessen einprogrammiert? Zieht man Erkenntnisse aus neurowissenschaftlicher Studien heran, lässt sich sogar feststellen, dass bereits eine kleine Veränderung der visuellen Darstellung eines Gedichts einen messbaren Einfluss auf das Lesen hat [Fechino et al., 2020].

Eine weitere große Herausforderungen bei der Verwendung von Word Embeddings ist die zuverlässige Evaluation ihrer Qualität, das heißt, wie sehr sie mit der gewünschten Bedeutung übereinstimmen. Es gibt zwar eine Menge von Evaluationsmethoden, wie Analogy Tests oder die Applikation von Downstream Tasks, häufig klopfen diese allerdings nur Teile der zugrundeliegenden Repräsentation ab und geben kein umfassendes Bild über ihre Qualität.

Ein in diesem Zusammenhang viel diskutierter Aspekt von Word Embeddings sind 'Biases', der zugrundeliegenden Daten, die durch die Modellarchitekturen reproduziert, oder sogar verstärkt werden können. Dies ist insbesondere bei Systemen ein Problem, die auf so großen und heterogenen Datenmengen trainiert wurden, dass es sich im Nachhinein schwer dokumentieren lässt, welche Biases dem jeweiligen Modell innewohnen (Bender et al., 2021). In Bezug auf die häufig kleineren, gut mit Metadaten versehenen Korpora in den DH, steckt in diesen reproduzierten Biases aber tatsächlich ein interessantes Forschungsfeld (Gebru et al., 2018). So lässt sich beispielsweise sehen, in welchen Zeiträumen welche Biases besonders vorherrschend waren, bzw. wann diese möglicherweise wieder verschwinden. Hierfür könnten mehrere Word Embeddings für verschiedene Zeiträume trainiert, und dann miteinander verglichen werden.

Diese Art von Vergleich ist aber nicht nur auf Biases beschränkt, man kann sie ebenfalls als eine Form des 'Realitätschecks' durchführen, wenn man Dynamiken zwischen Word Embeddings betrachtet: Bei einem Korpus, das einen größeren Zeitraum umfasst, verändern sich darin allgemein die Repräsentationen der Worte so, wie man es von der Bedeutung der Worte auch erwarten würde?

In dem Workshop möchten wir uns der beschriebenen Thematik von zwei Seiten nähern, wir geben (1) einen Überblick über traditionelle Theorien und Methoden der Literaturwissenschaft, insbesondere auch der Leseforschung und kontextualisieren letztere mit Erkenntnissen aus neurowissenschaftlichen Laborstudien. Wir geben (2) eine Zusammenfassung der aktuellen Methoden zu Textrepräsentation [z.B. Glove, Bert] und Beispiele für DH-Projekte, in denen diese bereits verwendet werden, insbesondere fokussieren wir uns hier auch auf gängige Visualisierungsmethoden und die damit zusammenhängenden Limitationen. Beides wollen wir durch praktische Übungen an eigenen oder bereitgestellten Datensätzen näherbringen. Ziel ist es, dass im Laufe des Workshops Verknüpfungen und Schnittmengen der unterschiedlichen Lesebegriffe aus den verschiedenen Fachrichtungen hergestellt und gefunden werden.

Zum Abschluss bieten wir konkrete Möglichkeiten an, welche Visualisierungen auf welche Weise ihren Platz als literaturwissenschaftliche Methode finden können.

Ablauf

Der Ablauf des Workshops gliedert sich in Vorträge sowie praktische Sitzungen, bei denen das im Vortrag zuvor besprochene direkt von den Teilnehmenden ausprobiert werden soll. Im folgenden soll ein kurzer Überblick über die einzelnen Vortragsteile gegeben werden.

Traditionelle Theorien und Methoden der Literaturwissenschaft

(Cora Krömer)

Theorien und Methoden der Literaturwissenschaft können allgemein in vier Typen unterteilt werden: text-, autor-, leser-, kontextorientiert [Köppe & Winko, 2013]. Als Einführung in den Workshop werden wir uns insbesondere einen Überblick über text- und leserorientierte Theorien und Methoden verschaffen, um die nachfolgenden Einheiten besser einordnen zu können. Digitale Texte (z. B. Hypertexte), neue empirische Quellen (z. B. Online-Leserkommentare) und digitale Methoden laden dazu ein, über das Lesen und Interpretieren von Texten und ihren Visualisierungen nachzudenken.

Einführung Textrepräsentationen ML

(David Lassner)

Hier werden verschiedene für DH relevante Textrepräsentationen besprochen. Dabei wird zum einen projektbezogen der typische Ablauf (Buch-Scan-OCR-TEI-NLP) durchgegangen und die verschiedenen verwendeten Repräsentationen verglichen, zum anderen die Entwicklung von Textrepräsentationen in NLP der letzten Jahren nachvollzogen, also von Bag-of-Words, über Word2Vec bis hin zu Bytepair Encoding und kontextualisierten Embeddings und der Frage, wie man aus diesen wieder generelle Wortrepräsentationen erhält.

Neuroscience

(Stephanie Brandl)

In welcher Form und Geschwindigkeit Menschen Texte lesen wird bereits seit Jahrzehnten untersucht [Rayner, 1998] und modelliert [Reichle et al., 2003]. Oberflächliche Eigenschaften wie Worthäufigkeiten oder Wortlängen beeinflussen unsere Lesegeohnheiten. Neuere Forschung zeigt zudem, dass auch die Darstellung des Textes eine Rolle spielt, ob ein Gedicht beispielsweise in der ursprünglichen Form oder als Prosatext gezeigt wird [Fechino et al., 2020]. Dementsprechend ist es auch wichtig, Word Embeddings so darzustellen, dass der sehr dichte Informationsgehalt verarbeitet und evaluiert werden kann.

Limitationen von Word Embeddings

(Stephanie Brandl)

Word Embeddings sind stark abhängig vom zu Grunde liegenden Datensatz, beispielsweise lernen statische Methoden wie GloVe genau eine Repräsentation pro Wort, so dass zeitliche oder andere Entwicklungen im Datensatz nicht beachtet werden und in einem Vektor verschmelzen. Diese Abhängigkeit führt auch dazu, dass Verzerrungen (Biases) im Datensatz möglicherweise in

den Wort-Vektoren auftauchen. Wenn beispielsweise alle Personen des medizinischen Personals im Datensatz weiblich sind, wird ein Algorithmus einen männlichen Bewerber möglicherweise für eher ungeeignet halten, um als Arzt zu arbeiten. Verschiedene Methoden wurden bereits veröffentlicht, um solche problematischen Strukturen (zeitliche Abhängigkeit, Biases uvm) in Word Embeddings aufzuzeigen [Basta et al., 2019; Brandl & Lassner, 2019; Gonen et al., 2020] und auch aufzulösen [Gonen et al., 2019; Manzini et al., 2019]. Wir werden einige davon besprechen und im Anschluss auch an den gelernten Word Embeddings anwenden.

Lesen von Textrepräsentationen und Visualisierungen

(Anne Baillet)

Zum Schluss wird auf die Entwicklung literaturwissenschaftlichen Umgangs mit Analyse bzw. Interpretation von Text eingegangen: zuerst auf das Faszinosum Netzwerk, dann auf die Herausforderung der Definition von Beziehungen zwischen Textelementen und ihrer Bedeutung, schließlich auf veröffentlichungstechnische Fragen, die mit der Einbettung dieser graphischen Repräsentationen einhergehen.

Bibliographie

- Basta, C. R. S., Ruiz Costa-Jussà, M., & Casas Manzanares, N.** (2019). "Evaluating the underlying gender bias in contextualized word embeddings." In *The 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 33-39).
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S.** (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).
- Brandl, S., & Lassner, D.** (2019). "Times Are Changing: Investigating the Pace of Language Change in Diachronic Word Embeddings". In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (pp. 146-150).
- Fechino M., Jacobs A.M., Lüdtke J. Following** (2020). "Jakobson and Lévi-Strauss' footsteps: A neurocognitive poetics investigation of eye movements during the reading of Baudelaire's 'Les Chats'". In *Journal of Eye Movement Research*, ff10.16910/jemr.13.3.4ff. Ffhal-02749759f
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K.** (2018). *Datasheets for datasets*. arXiv preprint arXiv:1803.09010.
- Gonen, H., & Goldberg, Y.** (2019). "Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them". In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 609-614).
- Gonen, H., Jawahar, G., Seddah, D., & Goldberg, Y.** (2020). "Simple, interpretable and stable method for detecting words with usage change across corpora". In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 538-555).
- Köppe, S. & Winko, S.** (2013). "Theorien und Methoden der Literaturwissenschaft". In Anz, T. (Hrsg.) *Handbuch Literaturwissenschaft*. Bd. 2. Stuttgart: J.B. Metzler. DOI 10.1007/978-3-476-01271-5. (pp. 285-371).
- Kuiken, D. & Jacobs, A.** (2021). *Handbook of Empirical Literary Studies*. Boston: De Gruyter.
- Manzini, T., Lim, Y. C., Tsvetkov, Y., & Black, A. W.** (2019). "Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings". In *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
- Moretti, F.** (2005). *Graphs, Maps, Trees. Abstract Models for a Literary History*. London/New York.
- Nünning, V. & Nünning, A.** (2020). *Methods of Textual Analysis in Literary Studies Approaches, Basics, Model Interpretations*. Trier: Wissenschaftlicher Verlag Trier.
- Parr, R. & Honold A.** (2018). *Grundthemen der Literaturwissenschaft: Lesen*. Berlin; Boston: De Gruyter.
- Ramsay, S.** (2011). *Reading Machines: Toward and Algorithmic Criticism*. University of Illinois Press.
- Rautenberg, U. & Schneider, U.** (2015). *Lesen: ein interdisziplinäres Handbuch*. Berlin; Boston: De Gruyter.
- Rayner, K.** (1998). "Eye movements in reading and information processing: 20 years of research". *Psychological bulletin*, 124(3), 372.
- Reichle, E. D., Rayner, K., & Pollatsek, A.** (2003). "The EZ Reader model of eye-movement control in reading: Comparisons to other models". *Behavioral and brain sciences*, 26 (4), 445-476.
- Saemmer, A.** (2015). *Rhétorique du texte numérique: Figures de la lecture, anticipations de pratique*. Villeurbanne.
- Willand, M.** (2014). *Lesermodelle und Lesertheorien. Historische und systematische Perspektiven*. Berlin und Boston.
- Zunshine, L.** (2015). *The Oxford Handbook of Cognitive Literature Studies*. Oxford.