

Peter Wulff
 David Buschhüter
 Andrea Westphal
 Andreas Borowski

Universität Potsdam

Potentiale automatischer Sprachverarbeitung für die Fachdidaktik

Lehren und Lernen in den Naturwissenschaften sind eng verknüpft mit Lernen von Sprache, da Bedeutungskonstruktion naturwissenschaftlicher Inhalte durch Sprache vermittelt wird (Schnotz, 2006). Sprache hängt unter anderem mit fachlichem Lernen (Rincke, 2007), mit konzeptuellem Wissen (Nehm & Härtig, 2012) und mit reflexivem Denken (Poldner, van der Schaaf, Simons, van Tartwijk & Wijngaards, 2014) zusammen. Zugleich ist Sprache so komplex, dass sich einfache Zusammenhänge den Analysen entziehen (Härtig, Fraser, Bernholt & Retelsdorf, 2019) und Fragestellungen stark elementarisiert werden müssen, um Effekte zu erzeugen (Schnotz, 2006), was die Validität der Ergebnisse für Lehr- und Lernprozesse in authentischen Lernumgebungen gefährdet. Sprache ist unter anderem deshalb komplex zu analysieren, da Bedeutung im Zusammenspiel von Oberflächen- und Tiefenstruktur erzeugt wird, Kohärenz von lokaler und globaler Abhängigkeit determiniert ist, und Worte als Begriffe in verschiedenen Kontexten anders zu lesen sind (Schnotz, 2006). Zur Stärkung einer sprachbasierten Forschung und Lehre in den Naturwissenschaftsdidaktiken sind deshalb Methoden und Instrumente notwendig, die flexibel, reliabel und valide verschiedene Ebenen von Sprache analysieren können.

Ein möglicher Ansatz in dieser Richtung ist Natural Language Processing (NLP), das als Schnittelelement von Computerlinguistik, Kognitionsforschung und zahlreichen weiteren Domänen (Bird, Klein & Loper, 2009) bisher sehr mächtige Anwendungen hervorgebracht hat, die natürliche Sprachnutzung von Menschen abstrahieren, sodass ein systematischerer Zugriff auf Wissen und Sprachnutzung möglich wird. Beispielsweise konnten in Lehr-Lern-Kontexten sowohl in der Forschung als auch in der Lehre zielführende und effektive Anwendungen umgesetzt werden. Im Bereich Expertiseforschung wurde für lexikonbasierte und regelbasierte Classifier gezeigt, dass sich Unterschiede zwischen Expert*innen und Novizen*innen auch in deren Sprachverwendung zeigen, beispielsweise durch den Gebrauch einer kausalen Sprache oder bestimmter Konzepte (Nehm & Härtig, 2012; Pennebaker, Boyd, Jordan & Blackburn, 2015). Ziel dieses Beitrages ist es, anhand zweier spezifischer Anwendungsfälle Potentiale, Herausforderungen und Grenzen von NLP für die naturwissenschaftlichen Fachdidaktiken exemplarisch aufzuzeigen. Adressiert sind Forschende und Lehrende in den Naturwissenschaftsdidaktiken, die einen ersten Einblick in NLP-Anwendungen bekommen wollen.

Schriftliche Reflexionen und NLP

In unserer Forschung beschäftigen wir uns mit Reflexionstexten von Lehramtsstudierenden in der Physik im Praxissemester. Schriftliche Reflexionen bieten sich für die Anwendung von NLP aus mehrfacher Sicht an: Zum einen stellen Texte geordnete, längere Einheiten dar, die von zahlreichen Abhängigkeiten auf Wort-, Satz- und Paragraphenebene gekennzeichnet sind (Schnotz, 2006). Das Verfassen von Texten ist damit eine externe Repräsentation von Bedeutung, die es aus forschungs- als auch aus lehrspezifischer Perspektive zu rekonstruieren gilt. Zum anderen rekurrieren Reflexionstexte auf Erfahrungen von Lehrkräften, sodass es sich insgesamt um eine wohldefinierte Textsorte handelt, für die die Varianz in neuen Daten eingeschränkt ist und zu der bereits Forschungsbefunde vorliegen (Poldner et al., 2014).

Auf der Basis vorheriger Forschung (Nowak, Liepertz & Borowski, 2018) sehen wir Reflexion als Denkprozess, der zum Ziel hat, Erfahrungen und Wissen einer Lehrkraft zu strukturieren und zu abstrahieren. Hierbei wird häufig das Ideal des reflektierten Praktikers oder der reflektierten Praktikerin herangezogen, wobei Reflexion als wichtiges Verbindungsglied von Theorie (Wissen) und Praxis (Handeln) konzeptualisiert wird (Carlson et al., 2019). Um Wissen und Handeln verknüpfen zu können, müssen angehende Lehrkräfte Gelegenheit erhalten, sich kritisch und distanzierend mit ihren eigenen Erfahrungen auseinanderzusetzen. Insbesondere die schriftliche Reflexion wurde hierbei vielfach auch im Bereich der Naturwissenschaftsdidaktik als wirksames diagnostisches Tool implementiert (Hume, 2009; Poldner et al., 2014). Um Studierenden eine Orientierung zu ermöglichen, was reflexives Denken bedeutet, wurde deshalb ein Reflexionsmodell validiert (Nowak et al., 2018), welches zentrale Elemente einer Reflexion identifiziert. Hiernach sollten Studierende für eine Reflexion zunächst die Rahmenbedingungen einer reflektierten Sequenz angeben und anschließend die Unterrichtsvorgänge beschreiben. Danach bewerten sie das Geschehene und leiten auf dieser Basis Alternativen für weitere Handlungsoptionen ab. Zuletzt adressieren die Reflektierenden Konsequenzen für ihre eigene professionelle Entwicklung.

Empirische Studien deuten allerdings darauf hin, dass angehende Lehrkräfte eher auf einem niedrigen Niveau und auf oberflächlicher Ebene reflektieren (Korthagen, 2005). Poldner et al. (2014) mussten feststellen, dass der Großteil der Reflexionstexte ihrer Studierenden auf der Ebene der Beschreibung und Bewertung stattfand. Um Studierende dabei zu unterstützen, in ihren Reflexionstexten auch stärker Elemente zu berücksichtigen, in welchen sie Alternativen abwägen und Konsequenzen für die eigene professionelle Entwicklung ziehen, sollen im ersten Anwendungsfall die Elemente der Reflexion auf der Basis eines NLP-Classifiers automatisch kodiert werden, was dann zur Umsetzung eines automatischen Feedbacktools genutzt werden kann. Im zweiten Anwendungsfall sollen dann explorativ Themen in den Studierendenreflexionen identifiziert werden. Daraus ließe sich z. B. ableiten, welche Inhalte im Rahmen der Begleitveranstaltungen zum Praxissemester genauer diskutiert werden sollten.

Anwendungsfall 1: Automatisiertes Feedback für schriftliche Reflexionstexte

Um Reflexionstexte automatisiert zu klassifizieren, ist es notwendig zu definieren, welche Einheiten kodiert werden sollen und welche Aspekte diese Einheiten am besten charakterisieren, sodass ein maschineller Lernalgorithmus das Auftreten bestimmter Features mit den Elementen der Reflexion verbinden kann. Der Lernalgorithmus wird dann an einem Trainingsdatensatz trainiert und an einem Entwicklungsdatensatz optimiert, sodass dieser neue Daten (zurückgehaltener Testdatensatz) klassifizieren kann (Jurafsky & Martin, 2014). Die Validität der Kodierung kann dann auf der Basis einer Mensch-Computer-Übereinstimmung bestimmt werden. In dieser Studie wurden dazu die Texte von N=17 Studierenden im Praxissemester (insgesamt N=81 schriftliche Reflexionen) händisch kodiert. Als Kodiereinheit wurde eine thematisch zusammenhängende Einheit im Text genutzt, die jeweils einer der Elemente der Reflexion (Rahmenbedingung, Beschreibung, Bewertung, Alternativen, Konsequenzen) zugeordnet wurde. Zwei unabhängige Kodierer erreichten hierbei eine substantielle Übereinstimmung, sodass auf der Basis dieser Kodierungen ein Classifier trainiert werden konnte. Als Feature wurde wordcount herangezogen, der ein sehr einfach umzusetzendes Verfahren darstellt, indem für jedes Dokument (d.h., Kodiereinheit) ein ca. 5000-dimensionaler Vektor erzeugt wurde, der angibt, ob ein Wort aus dem Vokabular im Dokument enthalten ist oder nicht (one-hot encoding). Da die resultierende Datenmatrix sehr viele leere Zellen hat, kann zur Verbesserung der Performance eine Datentransformation (word2vec) vorgenommen werden, die zur Reduzierung der Dimensionen führt (Mikolov, Sutskever, Chen, Corrado & Dean, 2013). Eine Intuition für diese Transformation ist in Abb. 1 dargestellt. In dieser Abbildung wird deutlich, dass die Distanz zweier beliebiger Wörter im

Falle des one-hot encoding (linke Seite) stets gleich ist. Dies ist allerdings eine eher unpassende Annahme für die Arbeit mit Texten, da verschiedene Worte im Kontext eines Textes ähnliche semantische Information kodieren können. Daher ist eine Transformation empfehlenswert. Effektiv in diesem Zusammenhang ist beispielsweise die word2vec-Transformation, da diese auch Bedeutung abbildet und beispielsweise Wortanalogien über Vektoradditionen gefunden werden können (Jurafsky & Martin, 2014). Betrachtet man Abb. 1, kann man im rechten Graphen nachvollziehen, dass der eingebettete Vektor für Frau minus dem Vektor für Mann, angewendet auf Königin als Ergebnis König gibt.

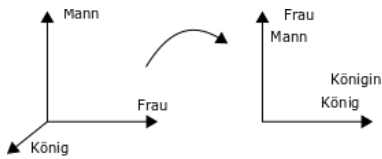


Abb. 1: Word2vec Intuition.

Um nun die Datenmatrix mit der abhängigen Variablen (hier: Elemente der Reflexion) zu verknüpfen, können verschiedene Classifier verwendet werden. Beispielsweise ist der Naive Bayes Classifier auch für kleine Datensätze geeignet, macht allerdings unter anderem die (naive) Annahme bedingter Unabhängigkeit der Features. Das führt dazu, dass der Naive Bayes Classifier hochkorrelierte Features doppelt zählt und damit die Evidenz verfälscht, wo beispielsweise der Logistic Regression Classifier robuster reagiert (Jurafsky & Martin, 2014). Viele dieser Algorithmen sind im scikit-learn-Paket (Pedregosa et al., 2011) der frei verfügbaren Software Python enthalten und einfach zu implementieren. Um dann die Mensch-Computer-Übereinstimmung eines Classifiers zu überprüfen, wird in der Regel der F1-score, als harmonisches Mittel aus Genauigkeit (Anteil der korrekt als positiv klassifizierten Fälle aus der Gesamtheit aller als positiv klassifizierten Fälle) und Trefferquote (Anteil korrekt als positiv klassifizierten Fälle aus der Gesamtheit der tatsächlich positiven Fälle) verwendet, der ein Maß für die Übereinstimmung darstellt.

Im vorliegenden Anwendungsfall konnten mit einem Logistic Regression Classifier auf der Basis von ca. 1700 Kodierungen nach entsprechender Optimierung von Hyperparametern, sowie Feature-Transformation nach word2vec ein F1-score von knapp .90 erreicht werden, was auf eine sehr gute Performance hindeutet. Damit ist dieser Classifier in der Lage, mit recht hoher Sicherheit Sätze in Studierendenreflexionen den entsprechenden Elementen zuzuordnen. Auf dieser Basis kann der Classifier nun eingesetzt werden, um beispielsweise Studierenden eine Rückmeldung zur Strukturiertheit und Vollständigkeit ihrer Reflexion zu geben.

Anwendungsfall 2: Thematische Analyse von schriftlichen Reflexionstexten

Aus methodischer Sicht der empirischen Bildungs- und Sozialforschung folgt Anwendungsfall 1 der Logik der deduktiven Kategorienbildung, bei der aus der Theorie bestimmte Kategorien abgeleitet und genutzt werden, um entsprechende Zuordnungen zu ermöglichen (Mayring, 2010). Der im folgenden dargestellte Ansatz folgt einer eher induktiven Herangehensweise (Bengfort, Bilbro & Ojeda, 2018). Kategorien am Textmaterial selbst zu entwickeln, kann in Bezug auf Reflexionstexte sinnvoll sein, da die theoretische Perspektive (Nowak et al., 2018) aus Sicht der Studierenden selbst einen Mangel an Relevanz besitzen könnte.

Ziel des folgenden Anwendungsfalls ist es deshalb, ein Kategoriensystem von Reflexionsthemen aus dem Textmaterial heraus zu entwickeln (s. auch Chen, Yu, Zhang &

Yu, 2016), ohne dabei die theoretische Perspektive apriori zu wählen. Diese Themen könnten z. B. von Dozierenden der Begleitveranstaltungen genutzt werden, um den Reflexionsbedarf aus Studierendenperspektive zu diagnostizieren. Der hier verwendete Ansatz, die Latent Dirichlet Allocation (LDA), gehört zu den Methoden des Topic Modeling (Blei, Ng & Jordan, 2003). Topic Modeling meint meist Methoden des unsupervised Machine Learning, um Themen (topics) aus Texten zu extrahieren (Bengfort et al., 2018). Unsupervised ist hier gleichbedeutend damit, dass apriori kein kategorisierter Datensatz vorhanden ist, der zum Training des Modells verwendet wird (Bengfort et al., 2018). In Abb. 2 wird die grundlegende Idee des Modells einer LDA visualisiert (angelehnt an Blei, Carin & Dunson, 2012). In diesem Modell ist jedem Dokument eine Verteilung über die Themen zugeordnet. Dokumente bestehen also aus einer Mischung von Themen zu bestimmten Anteilen (Abb 2. rechts). Die Themen wiederum entsprechen einer Verteilung über alle verschiedenen Begriffe (Abb. 2 links). Ähnlich wie bei der Bestimmung von Personen- und Itemparametern bei einer Rasch-Analyse werden mittels einer LDA aus den Texten (Daten) die entsprechenden Verteilungen für alle Dokumente über die Themen und für alle Themen über die verschiedenen Begriffe berechnet (genauer hierzu Blei et al., 2012; Blei et al., 2003).

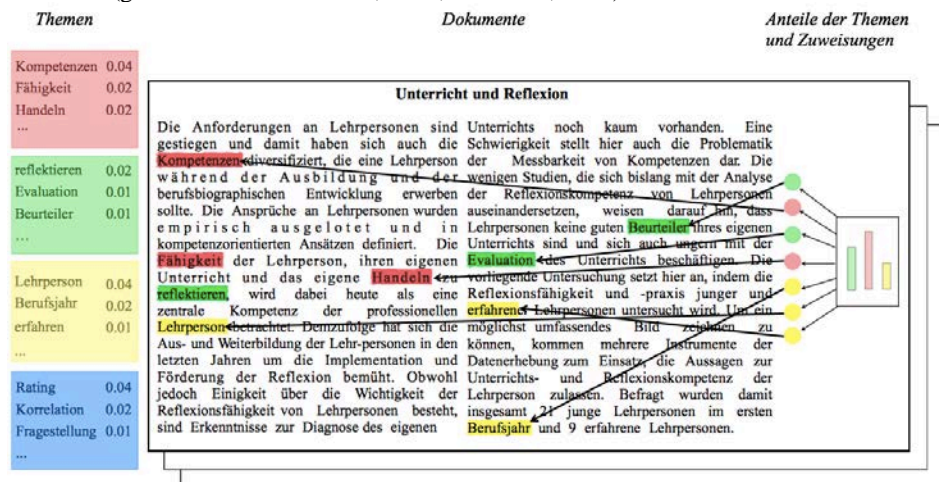


Abb. 2: Visualisierung der Grundidee einer LDA (Beispieltext wörtlich aus Wyss, 2013)

Auch bei diesem Verfahren müssen manuell bestimmte Hyperparameter gewählt werden. So ist auch die Anzahl der Themen manuell festzulegen. Um die entsprechende Entscheidung gibt es eine breite Diskussion (z. B. Greene, O'Callaghan & Cunningham, 2014; Chang, Boyd-Graber, Gerrish, Wang & Blei, 2009). Bei der LDA spielt die Reihenfolge der Wörter keine Rolle. Sie gehört zu den Bag-of-Words-Modellen. Auch sind streng genommen nicht einmal Wörter notwendig. Damit muss auch die Vergabe der Namen der Themen von der interpretierenden Person durchgeführt werden. Dazu verwendet die Person häufig die Rangfolge der Worte hinsichtlich ihrer Auftretenswahrscheinlichkeit bzw. -häufigkeit innerhalb eines Themas. Mittels des R-Paket LDAvis (Sievert & Shirley, 2014) lässt sich zudem berücksichtigen, dass bestimmte Begriffe in allen Themen hohe Auftretenswahrscheinlichkeiten haben. Mithilfe eines Koeffizienten Lambda (hier bei ca. 3) kann konfiguriert werden, zu welchem Maß die Überzufälligkeit berücksichtigt werden soll. Die resultierende Rangfolge wird dann zur Interpretation verwendet (Sievert & Shirley, 2014). Im Rahmen der vorliegenden 164 Reflexionstexte wurden folgende zehn Themen extrahiert: Kraft, Rechenaufgaben, Arbeitsphasen, Mechanik, Auswertung, E-Lehre, Wärmelehre, Allgemeines 1, Allgemeines 2, Unklar (letzteres war nicht kohärent genug, um es inhaltlich

interpretieren zu können). Im Folgenden sind für drei Themen, die entsprechend der oben erklärten Rangfolge auftauchen, die ersten sieben Begriffe kleingeschrieben dargestellt.

- Kraft: kräfte, wechselwirkungsprinzip, wirkt, kraft, pfeil, kräften, pfeile
- Rechenaufgaben: formel, s, test, sagte, aufgabe, schrieb, film
- Arbeitsphase: aufgaben, gruppen, übung, bearbeiten, bearbeitet, schülerversuch, aufgabenstellung

Nicht alle Themen hatten diese offensichtliche Kohärenz (z. B. E-Lehre: simulation, spannung, stromstärke, video, meinung, innerhalb, leiter), so dass hier mehr als nur die ersten sieben Wörter zur Interpretation und Vergabe des Kategoriennamens betrachtet werden müssen. Inwiefern die entsprechenden Kategorien hilfreich für eine gemeinsame Reflexion des Praxissemesters im Rahmen des Begleitseminars sein können, muss noch geklärt werden.

Chancen, Herausforderungen und Grenzen von NLP in der naturwissenschaftlichen Fachdidaktik

In den Anwendungsfällen standen jeweils eine supervised NLP (Klassifikation von Reflexionstexten) und eine unsupervised NLP Umsetzung (thematische Analyse von Reflexionstexten) im Zentrum. Aus Perspektive der fachdidaktischen Forschung bietet Anwendungsfall 1 Möglichkeiten, sprachliche Aspekte eines Kodiermanuals wie Wortbedeutung oder Syntax stärker explizit zu machen. Einmal trainiert, können selbst einfache NLP-basierte Classifier Expertise speichern und abrufen. Für die fachdidaktische Lehre hat das den Vorteil, ressourcenökonomisch zu sein (Burstein, 2009; Nehm & Härtig, 2012).

Gegenüber diesem Verfahren hat die Anwendung des Topic Modeling zunächst den Nachteil, dass der theoretische Blickwinkel hier nicht gewählt werden kann. Deshalb ist nicht evident, dass die entstehenden Kategorien einen apriori gewählten Zweck erfüllen. Dennoch bietet dieser Ansatz nach unserer Einschätzung Potential. Im Gegensatz zu vielen qualitativen Verfahren ermöglicht er eine enorme Reduktion der Theoriegeladenheit bei der Kategorienbildung. Wie im oben dargestellten Fall kann es sinnvoll sein, einen Text möglichst „theoriefrei“ zu interpretieren (auch wenn dieses Ideal nicht vollständig erreicht werden kann). Neben ihrer Ökonomie ist unsupervised Topic Modelling insbesondere dadurch attraktiv, dass es angewandt werden kann, wenn zu einer Textsorte noch keine oder nicht hinreichend ausgeschärfte theoretische Modelle vorliegen. Insofern kann die LDA für die Fachdidaktik ein exploratives Verfahren der Textanalyse über einfache deskriptivere Verfahren wie Worthäufigkeiten und Wortwolken hinaus darstellen.

Herausforderungen für die Adaption von NLP in der Fachdidaktik ergeben sich insbesondere im Zusammenhang mit der Notwendigkeit, repräsentative Korpora aufzubauen, an denen Forschungsfragen untersucht werden können. In diesem Zusammenhang sollten Modelle geteilt und gegenseitig geprüft werden, sodass Wissensaufbau in der Fachdidaktik systematisiert wird. Die Infrastrukturen und die Anwendungen für diesen Transfer sind noch nicht vorhanden, sollten aber in entsprechenden Anträgen mitgedacht werden (DFG, 2019). NLP scheint aufgrund seiner grundlegenden Bedeutung dafür eine geeignete Plattform.

Literatur

- Bengfort, B., Bilbro, R. & Ojeda, T. (2018). *Applied Text Analysis with Python. Enabling Language-Aware Data Products with Machine Learning*. Sebastopol: O'Reilly.
- Bird, S., Klein, E. & Loper, E. (2009). *Natural Language Processing with Python*. Beijing: O'Reilly.
- Blei, D. M., Carin, L. & Dunson, D. (2012). Probabilistic topic models. *Communication of the ACM*, 55(4), 77–84.
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5), 993–1022.
- Burstein, J. (2009). Opportunities for Natural Language Processing Research in Education. In A. Gebulkh (Hrsg.), *Springer lecture notes in computer science* (Bd. 5449, S. 6–27). New York, NY: Springer.
- Carlson, J., Daehler, K., Alonzo, A., Barendsen, E., Berry, A., Borowski, A. et al. (2019). The Refined Consensus Model of Pedagogical Content Knowledge. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge*. Singapore: Springer.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C. & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference* (S. 288–296).
- Chen, Y., Yu, B., Zhang, X. & Yu, Y. (2016). Topic modeling for evaluating students' reflective writing: A case study of pre-Service teachers' journals. In *ACM International Conference Proceeding Series* (S. 1–5).
- DFG. (2019). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*.
- Greene, D., O'Callaghan, D. & Cunningham, P. (2014). How Many Topics? Stability Analysis for Topic Models. In *European Conference on Machine Learning*. Verfügbar unter <http://derekgreene.com/papers/greene14topics.pdf>
- Härtig, H., Fraser, N., Bernholt, S. & Retelsdorf, J. (2019). Kann man Sachtexte vereinfachen? – Ergebnisse einer Generalisierungsstudie zum Textverständnis. *Zeitschrift für Didaktik der Naturwissenschaften*, 8(3), 231. <https://doi.org/10.1007/s40573-019-00105-7>
- Hume, A. (2009). Promoting higher levels of reflective writing in student journals. *Higher Education Research & Development*, 28(3), 247–260.
- Jurafsky, D. & Martin, J. H. (2014). *Speech and language processing* (Always learning, 2. ed., Pearson new internat. ed.). Harlow: Pearson Education.
- Korthagen, F. A. (2005). Levels in reflection. core reflection as a means to enhance professional growth. *Teachers and Teaching*, 11(1), 47–71.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Aufl.). Weinheim, Basel: Beltz.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *NIPS*, 13, 3111–3119.
- Nehm, R. H. & Härtig, H. (2012). Human vs. Computer Diagnosis of Students' Natural Selection Knowledge: Testing the Efficacy of Text Analytic Software. *Journal of Science Education and Technology*, 21(1), 56–73. <https://doi.org/10.1007/s10956-011-9282-7>
- Nowak, A., Liepertz, S. & Borowski, A. (2018). Reflexionskompetenz von Praxissemesterstudierenden im Fach Physik. In C. Maurer (Hrsg.), *Qualitätvoller Chemie- und Physikunterricht- normative und empirische Dimensionen. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Regensburg 2017*. Universität Regensburg.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pennebaker, J. W., Boyd, R. L., Jordan, K. & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. Austin, TX: The University of Texas at Austin.
- Poldner, E., van der Schaaf, M., Simons, P. R.-J., van Tartwijk, J. & Wijngaards, G. (2014). Assessing student teachers' reflective writing through quantitative content analysis. *European Journal of Teacher Education*, 37(3), 348–373. <https://doi.org/10.1080/02619768.2014.892479>
- Rinke, K. (2007). *Sprachentwicklung und Fachlernen im Mechanikunterricht. Sprache und Kommunikation bei der Einführung in den Kraftbegriff*. Berlin: Logos.
- Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In H. Blühdorn, E. Breindl & U. W. Waßner (Hrsg.), *Text - Verstehen. Grammatik und darüber hinaus* (S. 222–238). Berlin: de Gruyter.
- Sievert, C. & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Association for Computational Linguistics (Hrsg.), *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (S. 63–70).
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften*. Münster: Waxmann.