

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

Individuelle linguistische Variabilität in sozialen Medien: Ein multimediales Korpus

Tatjana Scheffler, Hannah Seemann, Lesley-Ann Kern

Soziale Medien wie Blogs oder Netzwerke wie Twitter stellen heute einen wichtigen Ort der informellen Textproduktion dar und eignen sich aufgrund ihres hohen Grades an Digitalisierung hervorragend als Grundlage für korpuslinguistisches Arbeiten. Digitale Korpora sozialer Medien sind als Schnittmenge zwischen konzeptioneller Mündlichkeit und schriftsprachlicher Ausarbeitung (Koch & Oesterreicher, 1985, 2007) von besonderer Bedeutung für die Untersuchung von Fragestellungen diverser linguistischer Teildisziplinen. Wir berichten über ein neues Social-Media-Korpus, das es erlaubt, sprachliche Daten derselben Sprecher:innen über zwei soziale Medien hinweg miteinander zu vergleichen und somit Rückschlüsse auf die individuelle linguistische Variabilität innerhalb sozialer Medien zu ziehen.

Zur Datensammlung wurde eine Liste von Elternblogger:innen verwendet, die sowohl auf Twitter aktiv sind als auch eigene Blogs pflegen. Zu dieser Liste wurden drei weitere bekannte Elternblogger:innen hinzugefügt. Die Tweets dieser Nutzer:innen zwischen Oktober 2016 und Februar 2017 wurden mit Hilfe des Python-Pakets *tweepy* über die Twitter API extrahiert. Über die Twitter-Biographie wurde auf den Blog der jeweiligen Person zugegriffen und der Text via RSS-Feed oder über FeedReader ausgelesen. Es entsteht jeweils ein Dokument mit Blogposts der Nutzer:innen und ein Dokument mit den gesammelten Tweets.

Im Anschluss an die Datensammlung wurden alle Blogger:innen kontaktiert und hatten die Möglichkeit, der Verwendung der Daten zu widersprechen (Opt-Out-Verfahren). Erhielten wir keine Rückmeldung, wurde von Zustimmung ausgegangen und die Daten wurden verwendet, nicht kontaktierbare Personen wurden mit ihren Daten entfernt. Alle Texte wurden manuell anonymisiert, sodass Rückschlüsse auf Einzelpersonen nicht länger möglich sind. Dafür wurden Personen- und Ortsnamen, Kontaktdaten, Usernamen sowie Links durch Platzhalter ersetzt. Um sicherzustellen, dass sowohl Blogposts als auch Tweets einer Person zugeordnet werden können, erhielten alle Nutzer:innen eine anonymisierte Zahlenkennung. Das Korpus beinhaltet Blogposts und Tweets von 44 Elternblogger:innen, insgesamt 468 Blogposts (ca. 360.000 Token) und 81.440 Tweets (ca. 1,2 Mio Token). Das Korpus liegt als Textdokument, tokenisiert

und in xml-Format vor. Es kann für wissenschaftliche Forschungen genutzt werden und ist online verfügbar¹.

Die parallele Sammlung von Blogposts und Tweets ermöglicht eine Untersuchung der intra-Autor:innen- und inter-Medien-Variation. Unsere These ist es, dass die Autor:innen ihre Sprachverwendung an das jeweilige Medium anpassen und sich somit Unterschiede in der Nutzung linguistischer Merkmale zwischen Blogs und Tweets einer Person finden lassen. Begründet wird die Wahl linguistischer Merkmale abgesehen von der persönlichen Präferenz durch die Affordanzen eines Mediums, die bestimmte Formen sprachlichen Ausdrucks überhaupt erst ermöglichen (Gibson, 2014). Dieser Aspekt spielt, genau wie weitere nicht-linguistische Faktoren, bei der Wahl des entsprechenden Registers bei der Kommunikation über soziale Medien eine essenzielle Rolle.

Diese These haben wir bisher v.a. anhand der Nutzung von Modalpartikeln und Intensivierern untersucht. Die Analyse dieser beiden linguistischen Phänomene ist im Kontext unseres Korpus von Interesse, da Intensivierer und Partikeln häufig der konzeptionellen Mündlichkeit zugeordnet werden (Lütten, 1979), Blogs und Tweets allerdings schriftliche Medien darstellen. Im Rahmen unserer bisherigen Untersuchungen wurden verschiedene Modalpartikeln und der Intensivierer so annotiert. Wir stellen fest, dass sowohl ein Unterschied in der Häufigkeit und Art der Nutzung dieser Ausdrücke zwischen den Medien erkennbar ist, aber dass einzelne Personen auch bestimmte idiosynkratische Nutzungspräferenzen haben.

Literatur:

Gibson, J. J. (2014). The Theory of Affordances. In *The Ecological Approach to Visual Perception*. Taylor & Francis Group.

Koch, P., & Oesterreicher, W. (2007). Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik*, 35(3). <https://doi.org/10.1515/zgl.2007.024>

Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe—Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36, 15–43. <https://doi.org/10.15496/publikation-20410>

Lütten, J. (1979). Die Rolle der Partikeln doch, eben und ja als Konsensus-Konstitutiva in gesprochener Sprache. In H. Weydt (Hrsg.), *Die Partikeln der deutschen Sprache* (S. 30–38). de Gruyter.

¹ <http://staff.germanistik.rub.de/digitale-forensische-linguistik/forschung/textkorpus-sprachliche-variation-in-sozialen-medien/>.