

Berfin Aktaş/Yulia Clausen/Tatjana Scheffler/Manfred Stede
(Potsdam)

Diskursstrategien in Sozialen Medien

Die Sprache in Sozialen Medien zeigt auf allen Ebenen eine hohe Variabilität und wurde daher als eine Mischung verschiedener Register (Tagliamonte/Denis 2008) analysiert, die sowohl informelle als auch formelle Formen umfassen. Im Gegensatz zu herkömmlichen schriftlichen Medien, wie z.B. Zeitungstexten, sind Soziale Medien wie Chat, Twitter, Forumdiskussionen, Facebook oder Blogs für diverse Autor/innen zugänglich, sind spontaner, und unterliegen weniger den sprachlichen Standards. Ein Teil der in diesen Kommunikationskanälen gefundenen Variabilität wird häufig auf Fehler zurückgeführt, es gibt jedoch viele Phänomene, die eigenen Regeln folgen. In Bezug auf das Kontinuum von konzeptueller/medialer Mündlichkeit/Schriftlichkeit (Koch/Oesterreicher 1985) befinden sich die meisten Social-Media-Beiträge in der Mitte des konzeptuellen Oralitätskontinuums, obwohl deren genaue Position bisher weitgehend unbestimmt ist. Soweit Nicht-Standard-Merkmale untersucht wurden, wurde meist die Orthographie, Morphologie, das Lexikon und die Syntax (z.B. Abkürzungen, Emoticons, Ellipsis) berücksichtigt.

Im Gegensatz dazu wählen wir einen pragmatischen, diskursorientierten Standpunkt: Welche Diskursstrategien wählen Sprecher/innen in Sozialen Medien und wie unterscheiden sich diese von bisher untersuchten Medienformen? Als Ausgangspunkt dient uns der Begriff des „Common Ground“ (CG; Stalnaker 1978, der das gemeinsame Wissen der Konversationsteilnehmer/innen bezeichnet, und dessen Maximierung als Ziel von Konversationen angesehen werden kann. Verschiedene Medien stellen nun verschiedene Mittel bereit, wie diskurspragmatische Bedeutung enkodiert werden kann. Im Allgemeinen stellen wir die Hypothese auf, dass die beobachtete Variabilität in den Sozialen Medien nicht aus einer „alles geht“-Einstellung der Sprecher/innen resultiert, sondern aus der Interaktion des internen grammatikalischen Systems eines jeden Einzelnen mit den sozialen und kontextuellen Beschränkungen eines bestimmten sprachlichen Beitrags.

Die spezifischen Phänomene, die wir untersuchen, tragen alle zum Common-Ground-Management im Diskurs bei. Wir untersuchen erstens die Funktionen ausgewählter Diskurspartikel (Frageanhängsel und Modalpartikeln), die sich auf den CG beziehen oder diesen verändern. Diese Partikeln sind häufig in gesprochener Konversation, aber nur wenig analysiert, weil sie in geschriebenen Texten selten vorkommen. Soziale Medien bieten hier ein interessantes Medium, welches spontane, gleichwohl geschriebene Sprache enthält. Zweitens adressieren wir Koreferenzphänomene wie die Referenzauflösung bei Personalpronomen. Aus

der Literatur ist bekannt, dass gesprochene Gespräche und geschriebene Texte sich signifikant in Bezug auf die Länge und Form der Koreferenzketten unterscheiden (z.B. Fox 1987; Amoia/Kunz/Lapshinova-Koltunski 2012). Drittens betrachten wir die Diskursstruktur durch Kohärenzbeziehungen, die einen Diskursabschnitt an einen Teil des vorherigen Diskurses anhängen. Dieser Schritt wird durch die Verwendung von Konnektoren der Sprecher/innen erleichtert, die die Art der Beziehung signalisieren können. Auch hier sind große Unterschiede zwischen Kohärenzrelationen in gesprochener Sprache und Text in der Literatur belegt, und wir richten die Aufmerksamkeit insbesondere auf Konversationen in Twitter, wo relativ wenige Personen miteinander interagieren und auf vorangehende Beiträge Bezug nehmen.

In der ersten Projektphase stand die Konstruktion eines Korpus für die Untersuchung der genannten Fragestellungen im Vordergrund. Für die Koreferenz konzentrieren wir uns auf das Englische und nutzen mit OntoNotes (Hovy et al. 2006) und ARRAU (Poesio/Artstein 2008) zwei bereits annotierte Korpora, in denen sowohl geschriebene als auch gesprochene Sprache vertreten ist (beide sind über das Linguistic Data Consortium¹ erhältlich). Ergänzend dazu haben wir ein Korpus von Twitter-Konversationen zusammengestellt und hinsichtlich der Koreferenzbeziehungen annotiert. Die gleichen Daten nutzen wir auch für die Untersuchung von Kohärenzrelationen, damit auch die Zusammenhänge zwischen beiden Annotationsebenen ausgelotet werden können. Die Kohärenz wird nach dem Muster der Penn Discourse Treebank (Prasad et al. 2008) analysiert, wo Konnektoren, konnexionanzeigende Phrasen sowie implizite Diskursrelationen mit ihren Argumenten annotiert werden, dabei aber zunächst keine Annahmen über eine resultierende übergeordnete Struktur getroffen werden. Für die Analyse der Diskurspartikeln vergleichen wir Konversationen aus dem deutschen Twitterkorpus von Scheffler (2014) mit Telefongesprächen aus CALLHOME (Karins et al. 1997).

Unsere ersten Analysen zeigen deutliche Unterschiede in der Verwendung der diskurspragmatischen Mittel in den Sozialen Medien im Vergleich zur geschriebenen oder gesprochenen Sprache. So werden z.B. Frageanhängsel wie „ne?“ in Gesprächen häufig verwendet, um sicherzustellen, dass Hörer/innen eines Faktus gewahr sind, während diese Verwendung auf Twitter fast völlig fehlt. Hier dienen Frageanhängsel zur Einforderung von Zustimmungen (1a) oder Bestätigungen (1b). Im Bereich der Koreferenz zeigen sich auch völlig neue Phänomene wie z.B. die exophore Referenz auf Objekte innerhalb von Bildern in Tweets wie in (2) – übertragen aus dem Englischen –, wo das Pronomen auf eine im Text nicht erwähnte Person auf dem Bild verweist.

¹ <https://www ldc.upenn.edu/>.

- (1) a. Total crazy, oder?
 b. [über ein Produktfoto] Rewe, oder?
- (2) Tweet1: [...] <eingebettetes Bild, Fahrgäste im Zug>
 Tweet2: die aussicht hätte mir gefallen ! zwinkere **ih**r zu !

Literatur

- Amoia, Marilisa/Kunz, Kerstin/Lapshinova-Koltunski, Ekaterina (2012): Coreference in spoken vs. written texts: A corpus-based analysis. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Doğan, Mehmet Uğur/Maegaard, Bente/Mariani, Joseph/Moreno, Asuncion /Odijk, Jan/Piperidis, Stelios (Hg.): Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, May 2012. Paris, S. 158–164.
- Calzolari, Nicoletta/Choukri, Khalid/Maegaard, Bente/Mariani, Joseph/Odijk, Jan/Piperidis, Stelios/Tapias, Daniel (Hg.) (2008): Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. Paris.
- Fox, Barbara A. (1987): Discourse structure and anaphora: Written and conversational English. (= Cambridge Studies in Linguistics 48). Cambridge.
- Hovy, Eduard/Marcus, Mitchell/Palmer, Martha/Ramshaw, Lance/Weischedel, Ralph (2006): OntoNotes: The 90% Solution. In: Proceedings of the Human Language Technology Conference of the NAACL, New York June 04–09 2006. Companion Volume: Short Papers. Stroudsburg, PA, S. 57–60.
- Karins, Krisjanis/MacIntyre, Robert/Brandmair, Monika/Lauscher, Susanne/McLemore, Cynthia (1997): CALLHOME German transcripts. Philadelphia.
- Koch, Peter/Oesterreicher, Wulf (1985): Sprache der Nähe – Sprache der Distanz. Muendlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Romanistisches Jahrbuch 36, S. 15–43.
- Poesio, Massimo/Artstein, Ron (2008): Anaphoric Annotation in the ARRAU Corpus. In: Calzolari et al. (Hg.), S. 1170–1174.
- Prasad, Rashmi/Dinesh, Nikhil/Lee, Alan/Miltsakaki, Eleni/Robaldo, Livio/Joshi, Aravind/Webber, Bonnie (2008): The Penn Discourse Treebank 2.0. In: Calzolari et al. (Hg.), S. 2961–2968.
- Scheffler, Tatjana (2014): A German Twitter snapshot. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Loftsson, Hrafn/Maegaard, Bente/Mariani, Joseph/Moreno, Asuncion/Odijk, Jan/Piperidis, Stelios (Hg.): Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland. Paris, S. 2284–2289.
- Stalnaker, Robert (1978): Assertion. In: Cole, Peter (Hg.): Pragmatics. (= Syntax and Semantics 9). New York, S. 315–332.
- Tagliamonte, Sali A./Denis, Derek (2008): Linguistic ruin? LOL! Instant messaging and teen language. In: American Speech 83, 1, S. 3–34.