

Gesprochenes Deutsch: Struktur, Variation, Interaktion

METHODENMESSE

Mittwoch, 06. März 2024, 16:15 Uhr bis 18:00 Uhr

„Alter, sprich mich bloß nicht an!“ Modalpartikeln in Popsongs – Kontextabhängige Detektion mit selbstlernenden Algorithmen

Sandra Hansen, Christian Lang, Roman Schneider, Ngoc Duyen Tanja Tu (alle IDS Mannheim)

Modalpartikeln (auch: Abtönungspartikeln) sind typisch für den gesprochenen Diskurs im Deutschen. Sie bringen – ohne die Geltung einer Proposition zu modifizieren – Erwartungen und Bewertungen der Kommunikationspartner zu ausgedrückten Sachverhalten zum Ausdruck, beispielsweise Resignation/Gleichgültigkeit in *Das ist eben so!* oder Bestärkung eines Urteils in *Das ist vielleicht ärgerlich!* Je nach theoretischem Ansatz und damit syntaktischer bzw. semantischer Charakterisierung fällt die Bestandsaufnahme unterschiedlich aus (vgl. z. B. Diewald 2007, Blühdorn 2019). Eine Identifikation wird zudem durch den Umstand erschwert, dass sie zumeist mehrfach heterosem sind, d. h. formgleiche Dubletten in anderen Wortarten haben und ohne aufwändige Kontextanalysen oft nicht von anderen Partikel-Subklassen, Adverbien, Adjektiven, Konnektoren usw. zu unterscheiden sind. Tagsets, die aktuellen Wortart-Klassifikatoren zu Grunde liegen (z. B. STTS), besitzen häufig gar kein eigenes Label für Modalpartikeln und selbst falls ein solches vorliegt (STTS 2.0; Westpfahl et al. 2017), bleibt die automatische Zuordnung aus den o.g. Gründen problematisch.

Vor diesem Hintergrund präsentiert unser Beitrag ein maschinelles Lernverfahren zur Identifikation ausgewählter Modalpartikeln in deutschsprachigen Songtexten. Die Textsorte eignet sich für das Vorhaben in besonderem Maße, da sie im konzeptionellen Kontinuum zwischen Nähe/Mündlichkeit und Distanz/Schriftlichkeit nachweisbar Merkmale beider Bereiche vereint (Broll/Schneider 2022) und die für Modalpartikeln typische dialogische Kommunikationssituation (vgl. Diewald 2007, 130) zumindest simuliert (Werner 2021). Darüber hinaus kombiniert die Datenbasis abwechslungsreiche grammatische und lexikalische Repertoire (Schneider/Lang/Hansen 2022). Dadurch scheint die Wahrscheinlichkeit hoch, dass darauf trainierte Modelle künftig gewinnbringend für die Detektion von Modalpartikeln auch in anderen Textsorten angewendet werden können. Mit dem Songkorpus (Schneider 2022; *songkorpus.de*) nutzen wir eine breit stratifizierte und mehrfach linguistisch annotierte empirische Ressource, die zum einen die erfolgreichsten Titel der ost-, west- und gesamtdeutschen Hitparaden seit 1970 enthält sowie thematische Archive für Sub- und Quergenres wie Hiphop oder Neue Deutsche Welle. Weiterhin umfasst das Korpus Archive mit den kompletten Werken ausgewählter Interpreten. Insgesamt deckt die Datenbasis mit derzeit ca. 9.000 Songtexten bzw. 2,5 Millionen Token mehr als ein halbes Jahrhundert mündlich kommunizierter Popkultur ab.

Die automatische Detektion der Modalpartikeln wird als Sequence Labeling Binary Classification Task modelliert: Zunächst ordnen wir einem Teil der Datenbasis tokenweise das Label „ist-Modalpartikel“ oder „ist-keine-Modalpartikel“ manuell zu, unter Dokumentation von

Interrater-Reliabilitäten. Diese Daten dienen verschiedenartigen maschinellen Lernverfahren als Trainings-, Test- und Validierungsset. Empirische Ergebnisse stellen wir gemeinsam mit den trainierten Modellen wissenschaftsöffentlich zur Verfügung, ebenso wie Verteilungsstatistiken der untersuchten Partikeln und ihrer Heterosemie in Popsongs und anderen Textsorten.

Literatur

- Broll, Sarah / Schneider, Roman (2023): Empirische Verortung konzeptioneller Nähe/Mündlichkeit inner- und außerhalb schriftsprachlicher Korpora. In: *Journal for Language Technology and Computational Linguistics* 36(1). 113-150.
- Blühdorn, Hardarik (2019): Modalpartikeln und Akzent im Deutschen. In: *Linguistische Berichte* 259. Hamburg: Buske. 275-317.
- Diewald, Gabriele (2007): Abtönungspartikel. In: Hoffmann, Ludger (Hg.): *Handbuch der deutschen Wortarten*. Berlin/New York: de Gruyter. 117-142.
- Schneider, Roman / Lang, Christian / Hansen, Sandra (2022): Das Vokabular von Songtexten im gesellschaftlichen Kontext – ein diachron-empirischer Beitrag. In: Kämper, Heidrun/Plewnia, Albrecht (Hrsg.): *Sprache in Politik und Gesellschaft. Perspektiven und Zugänge*. Berlin/Boston: de Gruyter, 2022. 295-304.
- Schneider, Roman (2022): Zwischen Schriftlichkeit und Mündlichkeit: Songtexte in der deskriptiven Sprachforschung. In: *Sprachreport* 1/2022. 38-50.
- Werner, Valentin (2021): Catchy and conversational? A register analysis of pop lyrics. *Corpora*. 16. 237-270.
- Westpfahl, Swantje / Schmidt, Thomas / Jonietz, Jasmin / Borlinghaus, Anton (2017): STTS 2.0. Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS). Mannheim: Institut für Deutsche Sprache. urn:nbn:de:bsz:mh39-60634