

58. Jahrestagung des Leibniz-Instituts für Deutsche Sprache vom 15. bis 17. März 2022
(als Online-Konferenz)

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

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

Entwicklungen in der Erforschung der Jugendsprache im digitalen Raum: Das NottDeuYTSch Korpus

Louis Cotgrove

In diesem Vortrag wird das *Nottinghamer Korpus Deutscher YouTube-Sprache* (Das *Nott-DeuYTSch*-Korpus) vorgestellt. Das Korpus wurde für meine Dissertation aufgebaut und besteht aus 3,1 Mio. *YouTube*-Kommentaren und 33,7 Mio. Token aus dem Zeitraum zwischen 2008 und 2018. Sie wurden von 112 deutschsprachigen *YouTube*-Kanälen gesammelt, die auf Jugendliche ausgerichtet sind. Das *NottDeuYTSch*-Korpus ist eines der ersten großen Korpora, das deutschsprachige Daten von Jugendlichen aus *You Tube*-Kommentaren sammelt. Diese bilden einen wichtigen Teil der Jugendkultur und des Jugenddiskurses ab, denn über 85% der 12- bis 19-Jährigen „schauen sich regelmäßig *YouTube*-Videos an“ (Bahlo et al. 2019: 80).

Das Korpus wurde entsprechend der Ansprüche an speziellere Web-Korpora aufgebaut (siehe Barbaresi 2019), um korpuslinguistische Arbeiten zur weiteren Untersuchung der Jugendsprache im digitalen Raum zu ermöglichen und die vorhandenen Korpora zu ergänzen, die sich auf andere Bereiche der Sprache im digitalen Raum konzentrieren. Hier wären zum Beispiel das MoCoDa2-Korpus von *WhatsApp*-Nachrichten (Beißwenger et al. 2020), das DiDi-Korpus von *Facebook*-Texten (Glaznieks and Frey 2020) und das IBK-Korpus mit mehreren Online-Quellen, z.B. Emails, *IRC*-Chats und Blogs (Lüngen and Kupietz 2020) zu nennen. Ab Frühling 2022 wird das *NottDeuYTSch*-Korpus über *KorAP* and *COSMAS II* als Teil des Deutschen Referenzkorpus (DeReKo) verfügbar sein und so den Zugriff für zukünftige Forschung verfügbar machen.

Im Vortrag wird die Methodik vorgestellt, die der Datenauswahl für das *NottDeuYTSch*-Korpus zugrunde liegt, einschließlich der Leitprinzipien für den Aufbau des Korpus und des Identifikationsverfahrens der *YouTube*-Kanäle, von denen die Kommentare gesammelt wurden. Darüber hinaus werden die Prozesse zur Erstellung des *NottDeuYTSch*-Korpus skizziert, methodische Aspekte wie die Ausgewogenheit und der Umfang des Korpus reflektiert und das Stichprobenverfahren eingehend erläutert.

Dieser Vortrag wird die vielen neuen Untersuchungsmöglichkeiten zeigen, die durch die sich aus den zentralen Eigenschaften des *NottDeuYTSch*-Korpus ergeben:

Ausführliche Metadaten: Das Korpus enthält viele Metadaten zu den Videos und Kommentaren. Diese ermöglichen sowohl die Analyse von Videogenres, Zeitrahmen, Nutzer*innen und

*YouTuber*innen* als auch Interaktions- und Diskursanalysen von Gesprächen zwischen Kommentator*innen, da die Kommentarstruktur der Threads erhalten bleibt.

Großer Datenbestand: Die hohe Anzahl an Token im *NottDeuYTSch*-Korpus ermöglicht eine umfangreiche quantitative Forschung, z.B.: längsschnittstudien für lexikalische, orthografische, morphologische und syntaktische Merkmale, die Untersuchung von linguistischen Merkmalen, die nicht häufig vorkommen, wie z. B. einige syntaktische Konstruktionen, und die Erforschung von linguistischen Merkmalen, die für die digital vermittelte Kommunikation spezifisch sind, wie z. B. Emojis und Hashtags.

Multilinguale Daten: Die Kommentare im *NottDeuYTSch*-Korpus sind überwiegend deutschsprachig (einschließlich dialektaler Verwendungen), aber es gibt auch eine nennenswert Präsenz anderer Kommentarsprachen wie Englisch, Türkisch und Russisch sowie sprachliche Elemente aus mehreren Sprachen innerhalb desselben Kommentars (das sogenannte *Translanguaging*, siehe García and Li 2014). Im Vortrag werden auf quantitativen und qualitativen Analysen basierende Beispiele der Mehrsprachigkeit in der Jugendsprache gezeigt.

Der Vortrag illustriert die breiten Anwendungsmöglichkeiten des *NottDeuYTSch*-Korpus zur Analyse der internetbasierten schriftlichen Kommunikation von Jugendlichen auf der Basis von Daten aus bisher noch wenig erforschten Social Media-Quellen. Damit kann er einen wichtigen Beitrag zu zukünftigen korpuslinguistischen Studien sowohl über digital-vermittelte Kommunikation als auch über Jugendsprache selbst leisten.

Literatur:

Bahlo, Nils et al. 2019. *Jugendsprache: Eine Einführung* (Berlin: J.B. Metzler)

Barbaresi, Adrien. 2019. 'The Vast and the Focused: On the Need for Thematic Web and Blog Corpora', in *Proceedings of the Workshop on Challenges in the Management of Large Corpora*, ed. by Piotr Bański et al. (Mannheim: Leibniz-Institut für Deutsche Sprache), pp. 29–32

Beißwenger, Michael et al. 2020. 'Die Mobile Communication Database 2 (MoCoDa 2)', in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, and Axel Schmidt (Berlin: De Gruyter), pp. 349–52

García, Ofelia, and Wei Li. 2014. *Translanguaging: Language, Bilingualism and Education* (London: Palgrave Macmillan)

Glaznieks, Aivars, and Jennifer-Carmen Frey. 2020. 'Das DiDi-Korpus: Internetbasierte Kommunikation Aus Südtirol', in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, and Axel Schmidt (Berlin: De Gruyter), pp. 353–54

Lüngen, Harald, and Marc Kupietz. 2020. 'IBK- Und Social Media-Korpora Am Leibniz-Institut Für Deutsche Sprache', in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, and Axel Schmidt (Berlin: De Gruyter), pp. 319–42 <<https://doi.org/10.1515/9783110679885-016>>