

## KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

Mittwoch, 16. März 2022, 11:45 Uhr

### Korpora für die Diskursanalyse

Marcus Müller (TU Darmstadt)

Für die Diskursforschung sind Korpora konstitutiv, seit Michel Foucault (1972: 10) die Bildung „kohärenter und homogener“ Textkorpora als eines der methodologischen Probleme der „neuen Geschichtswissenschaft“ ausgewiesen hat. In der Germanistischen Linguistik hat man sich oft auf den Operationalisierungsvorschlag Dietrich Busses und Wolfgang Teuberts (1994: 4) berufen, Diskurse „im forschungspraktischen Sinn“ als „virtuelle Textkorpora“ zu verstehen. Ebenfalls konstitutiv für die Diskursforschung ist es aber, Kontexte des Sprachgebrauchs in die Analyse einzubeziehen, z.B. thematische, situationale oder soziale (Müller 2012). Es geht also gleichermaßen um Serialität und Kontextsensitivität von Daten – zwei operationale Prinzipien, die in der Methodologie eher auseinanderstreben und in der praktischen Arbeit jeweils aufs Neue zusammenzubringen sind: Da Textkorpora bekanntlich nichts anderes als distributionelle Information über die Ausdrucksseite sprachlicher Zeichen (Gries 2009: 1226) bieten, müssen sie zusätzlich mit Kontextinformation angereichert werden. Daher ist die Frage nach angemessenen und notwendigen Repräsentationsformaten von Kontexten beim Korpusdesign, bei der Datenmodellierung und beim Preprocessing besonders wichtig. Eine große Rolle bei der korpuslinguistischen Kontextualisierung spielen auch Annotationen, insbesondere solche, die auf eigenen Tagsets beruhen, in einem ersten Schritt manuell durchgeführt werden und dann ggf. auf größere Datensätze angewendet werden sollen (z.B. Becker, Bender & Müller 2020; Müller, Bartsch & Zinn 2021).

Vor diesem Hintergrund thematisiert der Vortrag an ausgewählten Beispielen den Zusammenhang von Korpusaufbereitung, Datenanreicherung und Nutzungsszenarien im Kontext des Discourse Lab (<https://www.discourselab.de>), das an der TU Darmstadt und der Universität Heidelberg betrieben und in linguistischen und interdisziplinären Forschungs- und Lehrprojekten genutzt wird.

### Literatur:

- Becker, M., Bender, M. & Müller, M. (2020): Classifying Heuristic Textual Practices in Academic Discourse: A Deep Learning Approach to Pragmatics. In: International Journal of Corpus Linguistics (4)2020, pp. 426–460. <https://doi.org/10.1075/ijcl.19097.bec>
- Busse, D. & Teubert, W. (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, D., Hermanns, F. & Teubert, W. (Hgg.), Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik, 10–28. Opladen: VS Verlag für Sozialwissenschaften.

- Foucault, M. (1972): *The Archaeology of Knowledge*. New York: Pantheon Books. [L'Archéologie du savoir, Paris : Gallimard, 1969].
- Gries, S. Th. (2009): What is Corpus Linguistics?. *Language and Linguistics Compass*, 3: 1225–1241. doi:10.1111/j.1749-818X.2009.00149.x.
- Müller, M. (2012): Vom Wort zur Gesellschaft: Kontexte in Korpora: Ein Beitrag zur Methodologie der Korpuspragmatik. In: Felder, E., Müller, M. & Vogel, F. (Hgg.), *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*, 33–82.
- Müller, M., Bartsch, S. & Zinn, J.O. (2021): Communicating the unknown. An interdisciplinary annotation study of uncertainty in the coronavirus pandemic. In: *International Journal of Corpus Linguistics* 26/4, 498–531. <https://doi.org/10.1075/ijcl.21096.mul>