

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

Regionalsprachliche Korpora des Deutschen - Repositorien des Forschungszentrums Deutscher Sprachatlas

Hanna Fischer, Brigitte Ganswindt, Georg Oberdorfer

Um Forschungsdaten einerseits für die wissenschaftliche Community sowie die interessierte Öffentlichkeit verfügbar zu machen und sie andererseits in eine Langzeitarchivierung zu überführen, wurde in den letzten Jahren das Management dieser Ressourcen und damit auch der Umgang mit Daten im wissenschaftlichen Alltag stärker fokussiert. Darunter fällt auch der Aufbau von Repositorien, der als wichtige Aufgabe von empirisch arbeitenden Forschungseinrichtungen gelten kann.

Auch das Forschungszentrum Deutscher Sprachatlas (DSA) hat sich dieser Aufgabe verpflichtet und setzt neue Impulse im Bereich des sprachbezogenen Forschungsdatenmanagements: Das Forschungszentrum verfügt über eine umfangreiche Dokumentation regionalsprachlicher Datensätze aus Ton, Bild und Text, die digital aufgearbeitet (digitalisiert, indiziert, getaggt) und online in einem lebendigen Repositorium publiziert sind – u. a. in Form von digitalen Katalogen für Audiokorpora, einer Online-Bibliografie und digitalen Sprachkarten (Ganswindt/Kehrein/Lameli 2015). Die Audiokorpora umfassen circa 4.000 Übertragungen der „Wenkersätze“ in Dialekt und intendierte Standardsprache (vgl. z. B. Göschel 1977). Zum anderen enthält das Repositorium zahlreiche Sprachaufnahmen unterschiedlicher Erhebungssituationen aus internen und externen Projekten (z. B. Aufnahmen zum Mittelrheinischen Sprachatlas (Bellmann 1994) oder Tonkorpora des Preußischen Wörterbuchs (Goltz 2005) sowie flächendeckende Aufnahmen von Vorleseausssprache). Insgesamt stammen die Sprachaufnahmen aus mehr als 70 Jahren Dialektologie, so dass sie auch als Datengrundlage für kurzzeitdiachrone Forschungen geeignet sind.

Die verschiedenartigen Datenbestände sind zum gegenwärtigen Zeitpunkt in dem sprachgeografischen Informationssystem „REDE SprachGIS“ zusammengeführt, dessen Basis die geografische Informationsverarbeitung und -vernetzung ist. Es verfügt über spezifische Suchwerkzeuge, die Filterungen nach verschiedenen Korpora oder Dialekträumen erlauben und zudem einen text- und lemmabasierten Zugriff ermöglichen.

Aktuell wird weiterhin ein „Korpus der Regionalsprachen“ aufgebaut. Dieses Korpus speist sich vorwiegend aus Aufnahmen freier Gesprächssituationen, die momentan transkribiert und für Annotation und Analyse aufbereitet werden. In einer modularen Vorgehensweise wird das Korpus sukzessive erweitert mit dem Ziel, ein Referenzkorpus der deutschen Regionalsprachen anbieten zu können. In der Korpusgenerierung stehen neben grammatischen Aspekten (Morphologie, Syntax) auch gesprächs- und diskurslinguistische Fragestellungen im Erkenntnisinteresse (Diskursmarker und -organisation, sprachreflexive Handlungen), welche die bisherigen Datenaufbereitungen am DSA ergänzen. Das „Korpus der Regionalsprachen“

ist als Teil des frei zugänglichen Repositoriums des DSA mit einem eigenständigen Webinterface für Recherche und Export angelegt. Wert wird dabei auf das in der Korpuslinguistik als momentaner Best-practice-Ansatz angesehene Vorgehen gelegt. Dazu gehört die Darstellung in KWIC (Keyword in Context) und verschiedene Exportformate für eine optimale Les- und Publizierbarkeit sowie multifunktionale Suchoptionen, die Operatoren wie jene der RegEx sowie auch kategoriale Bestimmungen wie jene des POS-Taggings kennen. Mit der Weiterentwicklung des Repositoriums in den nächsten Jahren werden zentrale Zugänge zu dialektalen und regiolektalen Sprachdaten unterschiedlicher Provenienz aus dem gesamten deutschen Sprachgebiet der wissenschaftlichen Community und interessierten Öffentlichkeit zur Verfügung gestellt. Die Sprachaufnahmen und ihre teils mehrschichtigen Annotationslayer bieten dabei eine einmalige Forschungsressource für die germanistische Variationslinguistik und darüber hinaus.

Auf der Methodenmesse bekommen Sie einen Überblick über die Fülle und die Qualität der verfügbaren Daten und Annotationen sowie den benutzerfreundlichen Zugriff darauf. Des Weiteren werden erste Einblicke in das im Entstehen befindliche „Korpus der Regionalsprachen“ ermöglicht.

Literatur:

Bellmann, Günter (1994): Einführung in den Mittelrheinischen Sprachatlas (MRhSA). Tübingen.

Ganswindt, Brigitte / Roland Kehrein / Alfred Lameli (2015): Regionalsprache.de (REDE). In: Kehrein, Roland / Alfred Lameli / Stefan Rabanus (Hrsg.): Regionale Variation des Deutschen – Projekte und Perspektiven. Berlin/Boston: De Gruyter, 421–453.

Goltz, Reinhard (2005): Preussisches Wörterbuch: deutsche Mundarten Ost- und Westpreußens. Akademie der Wissenschaften und der Literatur Mainz. Begr. von Erhard Riemann. Fortgef. von Ulrich Tolksdorf. Hrsg. von Reinhard Goltz. Neumünster: Wachholtz

Göschel, Joachim (Hrsg.) (1977): Die Schallaufnahme deutscher Dialekte im Forschungsinstitut für deutsche Sprache. Marburg.