

Historische Korpora

Resch, Claudia; claudia.resch@oeaw.ac.at

Historische Korpora sind strukturierte Sammlungen von Daten älterer Sprachstufen, wobei der Begriff ‘historisch’ vage bleibt. Jost Gippert (2015, S. 9) bezieht ihn etwa auf altertümliche, alte oder mittelalterliche, jedenfalls aber auf nicht zeitgemäße Sprachstadien. In der Unterscheidung zu gegenwartssprachlichen Korpora ist, so Carmen Scherer (2014, S. 26), die zeitliche Nähe oder Distanz zur Gegenwart relevant: „Je größer der zeitliche Abstand zwischen der Entstehung eines Textes und der Gegenwart wird, umso eher ist ein Text und das Korpus, in dem er enthalten ist, als historisch einzustufen.“ Verallgemeinernd spricht Claudia Claridge (2008, S. 242) von „periods before the present-day language“ und ergänzt, dass diese etwa eine Generation vor der gegenwärtigen enden.

Die Ausgewogenheit von historischen Korpora ist immer von der Überlieferung abhängig und dadurch oftmals eingeschränkt. Im Gegensatz zu gegenwartssprachlichen Sprachdatensammlungen sind historische Korpora meist von geringerem Umfang. Nicht nur muss das Material für den Korpusaufbau meist erst in maschinenlesbare Form gebracht werden, auch bei der Annotation der historischen Daten stehen Wissenschaftlerinnen und Wissenschaftler vor nicht geringen Herausforderungen: Die variantenreiche und -tolerante Schreibweise vieler älterer Sprachstufen, verbunden mit anderen Flexionsmustern und syntaktischen Konventionen, erschwert die automatische Verarbeitung und Erschließung des Materials, das aus heutiger Sicht häufig zu den *non-standard*-Varietäten gezählt wird. In den letzten Jahren ist deshalb verstärkt versucht worden, spezifische Tagsets zu entwickeln bzw. bestehende Kategoriensysteme zu erweitern und automatische Annotationsverfahren an ältere Sprachstufen anzupassen.

Anhand verlässlich annotierter historischer Korpora lassen sich einerseits bestehende Erkenntnisse überprüfen und andererseits neue syntaktische, semantische, pragmatische, lexikologische oder andere sprachliche Phänomene (auch im diachronen Verlauf) untersuchen. Das Potential historischer Korpora bleibt allerdings bei weitem nicht nur auf die historische Sprachwissenschaft beschränkt, sondern könnte in den kommenden Jahren noch viel stärker für korpusbasierte Ansätze ihrer Nachbardisziplinen genützt werden, wie etwa der Literaturwissenschaft, der Geschichte, der Philosophie oder der Theologie.

Einen sehr guten Überblick über die derzeit verfügbaren historischen Korpora bietet die von CLARIN kuratierte "Historical CorporaListe.

Literatur:

- New Methods in Historical Corpora. Hrsg. von Paul Bennett, Richard Whitt, Silke Scheible und Martin Durrell. Tübingen: 2013.
- Claridge, Claudia: Historical Corpora. In: Corpus Linguistics. An International Handbook 1. Berlin, Boston: 2008, S. 242–259.
- Gippert, Jost: Preface. In: Historical Corpora. Challenges and Perspectives. Tübingen: 2015, S. 9-12.
- Kroymann, Emil; Thiebes, Sebastian; Lüdeling, Anke; Leser, Ulf: Eine vergleichende Analyse von historischen und diachronen digitalen Korpora. Technical Report 174 des Instituts für Informatik der Humboldt-Universität zu Berlin: 2004. URL: <https://www2.informatik.hu-berlin.de/sam/preprint/kroymann174.pdf>.
- Pettersson, Eva; Megyesi, Beáta: The HistCorp Collection of Historical Corpora and Resources. In: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference. University of Helsinki: 2018, S. 306–320.
- Scherer, Carmen: Korpuslinguistik. Heidelberg: 2014.
- Historische Korpuslinguistik. Hrsg. von Renata Szczepaniak, Stefan Hartmann und Lisa Dücker. Berlin, Boston: 2019.

Verweise:

Annotationsstandards, Data Mining, Digitalisierung, Konkordanz, Lemmatisierung, Markup, Metadaten, NLP, Normalisierung, Part-of-Speech-Tagging, Tagsets, Transkriptionsrichtlinien, Volltextsuche

Projekte:

Deutsches Textarchiv, DigiPress, Cosmas II, Austrian Baroque Corpus (ABaC:us), Mittelhochdeutsche Begriffsdatenbank (MHDBDB), Deutsch Diachron Digital (DDD), Referenzkorpus Mittelhochdeutsch (ReM);, RIGDES-Projekt (Register in Diachronic German Science), Bonner Frühneuhochdeutsch Korpus (FnhdC), CLARIN: Historische Korpora

Themen:

Einführung, Natural Language Processing, Digitale Editionswissenschaft

Zitiervorschlag:

Resch, Claudia. 2021. Historische Korpora. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.94>