

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

Die österreichische Presselandschaft digital: Das Austrian Media Corpus (AMC) – Aufbau, Bedienung und Möglichkeiten

Amelie Dorn¹, Jan Höll¹, Wolfgang Koppensteiner¹, Hannes Pirker² & Theresa Ziegler¹
(¹Universität Wien, Institut für Germanistik; ²ACDH-CH ÖAW)

Der Einsatz von Korpora in diversen Gebieten der Sprachforschung verzeichnete in den letzten Jahren einen deutlichen Anstieg, nicht zuletzt bedingt durch die voranschreitenden und umfassenden Digitalisierungsmaßnahmen u. A. in Geistes- und Sozialwissenschaften, aber auch durch die stetig neue Entwicklung unterschiedlicher Methoden und Werkzeuge, um digitalisierte Korpora zu erschließen oder digital erzeugte („born-digital“) Korpora systematisch zu analysieren und dadurch Muster, Gebrauch und Veränderungen in Sprache und Kommunikation aufzuzeigen und nachzuweisen. Das Austrian Media Corpus (AMC) (siehe Ransmayr 2014; Ransmayr, Mörth & Ďurčo 2013; 2017) stellt ein wertvolles Beispiel eines digital erzeugten Textkorpus dar, das für weitreichende Analysen und auch als Vergleichskorpus herangezogen werden kann. Das AMC wird durch eine Kooperation der Austria Presse Agentur (APA) und der Österreichischen Akademie der Wissenschaften (ÖAW), konkret dem Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) zu Forschungszwecken bereitgestellt. Es ist derzeit mit ca. 45 Mio. Artikeln und 11 Milliarden Token eines der größten deutschen Textkorpora, das inhaltlich für unterschiedlichste sprachwissenschaftliche Fragestellungen zum (österreichischen) Deutschen herangezogen werden kann. Das AMC umfasst neben allen österreichischen Tageszeitungen auch die gängigsten Wochen- und Monatszeitschriften. So macht das AMC nahezu die gesamte Printmedienproduktion der letzten 25 Jahre eines Landes für die wissenschaftliche Forschung zugänglich. Diese Tatsache macht das AMC einzigartig in seinem inhaltlichen Umfang.

Der vorliegende Beitrag leistet zweierlei: Einerseits erfolgt eine extensive Darstellung des AMC, welche neben theoretisch-technischen Aspekten auch eine praktische, interaktive Auseinandersetzung mit dem Korpus bietet (NoSketch Engine (Kilgarriff et al. 2014; Rychlý 2007)). BesucherInnen der Methodenmesse erhalten die Möglichkeit über das NoSketch Engine-Interface selbst Abfragen im Korpus vorzunehmen. Andererseits wird an einer Auswahl an bereits durchgeführten Studien der Nutzen des AMC für die korpuslinguistisch orientierte Forschung demonstriert (vgl. Höll 2020, Koppensteiner 2015, Ziegler 2021). Unter anderem wird der Einsatz des AMC in den Kontext des Spezialforschungsbereichs „Deutsch in Österreich. Variation – Kontakt – Perzeption“ (Lenz 2018) gestellt, der dieses Korpus bereits zur Analyse des Gebrauches von sogenannten „Austriazismen“ aus variationslinguistischer Perspektive einsetzt.

Literatur:

- Austrian Media Corpus (amc) <https://amc.acdh.oeaw.ac.at/> [letzter Zugriff: 11.11.2021]
- Höll, Jan (2020): Wissensstrukturen im Kontext der deutschen Rechtschreibreform. Framebasierte Analyse eines medialen Diskurses in Österreich. Masterarbeit, Universität Wien. Wien.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: Ten Years on. In: *Lexicography* 1, 7–36.
- Koppensteiner, Wolfgang (2015): Das österreichische Deutsch im plurizentrischen Kontext. Diplomarbeit, Universität Wien. Wien.
- Lenz, Alexandra N. (2018): The Special Research Programme „German in Austria. Variation – Contact – Perception“. In: Ammon, Ulrich & Marcella Costa (Hgg.): *Sprachwahl im Tourismus – mit Schwerpunkt Europa*. (Sociolinguistica. Internationales Jahrbuch für europäische Soziolinguistik 32). Berlin/Boston, 269–277.
- NoSketch Engine <https://nlp.fi.muni.cz/trac/noske> [letzter Zugriff: 24.11.2021]
- Ransmayr, Jutta (2014): Neue Forschungsmöglichkeiten zum österreichischen Deutsch mit dem Austrian Media Corpus (AMC). In: Ransmayr, Jutta, Andrea Moser-Pacher & Ilona Elisabeth Fink: *Österreichisches Deutsch und Plurizentrik*. (Informationen zur Deutschdidaktik. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule 3/2014). Innsbruck: StudienVerlag, 63–68.
- Ransmayr, Jutta, Karlheinz Mörth & Matej Ďurčo (2013): Linguistic Variation in the Austrian Media Corpus. Dealing with the challenges of large amounts of data. In: *Procedia - Social and Behavioral Sciences* 95, 111–115.
- Ransmayr, Jutta, Karlheinz Mörth & Matej Ďurčo (2017): AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum österreichischen Deutsch. In: Resch, Claudia & Wolfgang U. Dressler (Hgg.): *Digitale Methoden der Korpusforschung in Österreich* (Veröffentlichungen zur Linguistik und Kommunikationsforschung 30). Wien: Verlag der Österreichischen Akademie der Wissenschaften, 27–38.
- Rychlý, Pavel (2007): Manatee/Bonito-A Modular Corpus Manager. In: *Proceedings of Recent Advances in Slavonic Natural Language Processing 2007 (RASLAN)*, 65–70.
- Ziegler, Theresa (2021): Über Geschenk-s-körbe und Schokolade-Ø-torten (zu runden Geburtstagen). Sneak Peek auf eine Abschlussarbeit über areal-horizontale Verfungstendenzen bei NN-Komposita mit schwankenden Fugenelementen in der österreichischen Standard(schrift)sprache. In: Kim, Agnes, Katharina Korecky-Kröll, Ludwig M. Breuer, Jan Höll & Wolfgang Koppensteiner: *Vom Tun nicht lassen können. Historische und rezente Perspektiven auf sprachliche Variation (in Österreich und darüber hinaus)*. Festgabe für Alexandra N. Lenz zum runden Geburtstag. (WLG 89) Wien: Institut für Sprachwissenschaft, Universität Wien.