

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

Dienstag, 15. März 2022, 14:40 Uhr

Das ZDL-Regionalkorpus: Design, lexikographische Nutzung und Korpusergänzungen

Andreas Nolda, Alexander Geyken, Adrien Barbaresi (BBAW Berlin)

In der 2015 erschienenen Publikation über das Variantenwörterbuch formulieren Bickel, Hofer und Suter (2015: 544) fünf Desiderata, die ein Korpus als Datengrundlage für die lexikographische Beschreibung diatopischer Variation im Standarddeutschen idealerweise erfüllen sollte:

1. Die Textbasis müsse gezielt nach national und regional zuordenbaren Texten abgesehen werden können, und die potenziellen Varianten müssten räumlich zugeordnet werden können.
2. Das Korpus sollte möglichst nur neuere und neueste standardsprachliche Texte enthalten.
3. Das Korpus sollte groß genug sein, um auch bei selteneren Varianten aussagekräftige Treffermengen zu liefern.
4. Das Korpus sollte linguistisch getaggt sein.
5. Absolute und relative Frequenzen einer Variante in den Vollzentren des Deutschen sollten automatisiert erhoben werden können.

„Ein linguistisches Korpus, das alle diese Wünsche erfüllt,“ stellen Bickel, Hofer und Suter fest, „gibt es zur Zeit [2015] noch nicht.“

In unserem Beitrag stellen wir das *ZDL-Regionalkorpus* (Nolda, Barbaresi und Geyken 2021) vor, das 2020 am Zentrum für digitale Lexikographie der deutschen Sprache (ZDL) erstmals verfügbar gemacht wurde und seitdem fortlaufend aktualisiert und ausgebaut wird. Aufgrund seines Designs und seiner Funktionalität strebt das ZDL-Regionalkorpus die Erfüllung der o.g. Desiderata an. Das ZDL-Regionalkorpus umfasst über 31 Millionen Zeitungsartikel aus Lokal- und Regionalressorts deutschsprachiger Zeitungen (Stand Dez. 2021). Es deckt den Zeitraum seit 1993 ab und wird monatlich aktualisiert. Alle Zeitungsartikel sind gemäß dem Arealkonzept (<https://www.dwds.de/d/regionalangaben>) annotiert. Die Texte sind lemmatisiert und mit Wortarten versehen. Die über die Webseite des „Digitalen Wörterbuch der deutschen Sprache“ (DWDS, www.dwds.de) bereitgestellte Benutzerschnittstelle ermöglicht die spezifische Suche in Arealen und stellt Werkzeuge zur frequenzbasierten tabellarischen und kartographischen Darstellung zur Verfügung.

Die Daten des ZDL-Regionalkorpus sind der Ausgangspunkt für die lexikographische Beschreibung der diatopischen Variation im Digitalen Wörterbuch der deutschen Sprache. Bislang wurden etwa 1000 regional markierte Stichwörter mit Hilfe des ZDL-Regionalkorpus neu beschrieben. Wir zeigen anhand ausgewählter Beispiele, für welche Phänomene das ZDL-Regionalkorpus gut geeignet ist, gehen aber ebenso auf die Grenzen des Korpus ein.

Im dritten Teil des Beitrags zeigen wir anhand einiger ausgewählter Beispiele, wie das ebenfalls im ZDL erstellte Web-Monitor Korpus als ergänzende Ressource für ZDL-Regionalkorpus

eingesetzt werden kann. Das Web-Monitor Korpus vereinigt über 500 reichweitenstarke Quellen des deutschen Sprachraums. Er wird täglich aktualisiert und weist derzeit eine Größe von über 1,1 Milliarden Tokens auf (<https://www.dwds.de/d/korpora/webmonitor>).

Literatur:

- Barbaresi, Adrien und Alexander Geyken (2020). Die Webkorpora im DWDS – Strategien des Korpusaufbaus und Nutzungsmöglichkeiten. In *Deutsch in Sozialen Medien: Interaktiv – multimedial – vielfältig*, hg. v. Konstanze Marx, Henning Lobin und Axel Schmidt, Institut für Deutsche Sprache: Jahrbuch 2019, Berlin: de Gruyter, 345–348.
- Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – neu: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.
- Geyken, Alexander, Adrien Barbaresi, Jörg Didakowski, Bryan Jurish, Frank Wiegand und Lothar Lemnitzer (2017). Die Korpusplattform des „Digitalen Wörterbuchs der deutschen Sprache“ (DWDS). *Zeitschrift für germanistische Linguistik*, 45(2), 327–344.
- Nolda, Andreas, Adrien Barbaresi und Alexander Geyken (2021). Das ZDL-Regionalkorpus: Ein Korpus für die lexikografische Beschreibung der diatopischen Variation im Standarddeutschen. In *Deutsch in Europa: Sprachpolitisch, grammatisch, methodisch*, hg. v. Henning Lobin, Andreas Witt und Angelika Wöllstein, Institut für Deutsche Sprache: Jahrbuch 2020, Berlin: de Gruyter, 317–322.