

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

Korpora zur Fußballlinguistik – Eine multilinguale Forschungsressource zur Sprache der Fußballberichterstattung

Simon Meier-Vieracker

Im Projekt fussballlinguistik.de werden seit einigen Jahren multilinguale Korpora mit Texten der internetbasierten Fußballberichterstattung wie etwa Spielberichte, Taktikanalysen und v.a. Liveticker aufgebaut und über eine webbasierte Plattform (www.fussballlinguistik.de/korpora) der Forschungsöffentlichkeit zugänglich gemacht (Meier 2017, 2020). Inzwischen kann in vollständig morphosyntaktisch annotierten Texten in vierzehn europäischen Sprachen (deutsch, Schweizer Standarddeutsch, englisch, norwegisch, niederländisch, französisch, italienisch, spanisch, portugiesisch, polnisch, tschechisch, russisch, griechisch, ungarisch) im Gesamtumfang von rund 78 Millionen Tokens recherchiert werden. Damit dokumentiert die Ressource in bislang einmaliger Weise den Schriftsprachgebrauch in der thematischen Domäne der Fußballberichterstattung. Vor allem die durch konzeptionelle Mündlichkeit gekennzeichnete, aber gleichwohl redaktionelle Textsorte Liveticker ist damit zum ersten Mal in größerem Umfang für korpuslinguistische Zwecke dokumentiert. Auch wenn die Fußballsprache seit vielen Jahren intensiv beforscht wird, mussten Forschende bislang entweder eigene Datensammlungen erstellen oder auf die thematisch unspezifischen Pressekorpora zurückgreifen, die natürlich auch Fußballtexte enthalten. Mit der neuen Ressource können aber viel fokussierter und zugleich auf einer breiten empirischen Basis die Eigenschaften dieser Sondersprache auf unterschiedlichen sprachlichen Ebenen untersucht werden.

Insbesondere die lexikologisch und phraseologisch orientierte Forschung zur Fußballsprache (Schmidt 2009; Burkhardt 2014), aber auch an syntaktischen Eigenheiten interessierte Forschung (Jürgens 2009), findet durch die Korpora eine nützliche empirische Basis, auch und gerade in sprachvergleichender Perspektive. Die Analysesoftware CQPweb (Hardie 2012) erlaubt nicht nur die bequeme Suche nach Belegen und ihrer kontextuellen Einbettung, sondern ermöglicht mit den implementierten statistischen Auswertungsmethoden wie Kollokationen und Keywords auch datengeleitete Explorationen. Die verfügbaren Metadaten, etwa zu Textsorte, Datum oder Wettbewerb, erlauben zudem die Erhebung von bedingten Verteilungen (Stefanowitsch 2020), so dass auch diachrone Entwicklungen oder mediale Prägungen (etwa live vs. non-live) untersucht werden können.

Die Korpora zur Fußballlinguistik bieten aber nicht nur für korpuslinguistisch versierte Forschende eine nützliche Ressource, sondern können aufgrund der globalen Popularität des Fußballs auch in didaktischen Settings eingesetzt werden. In Frage kommen einerseits (fremd)sprachdidaktische Anwendungen, um die Besonderheiten der medialen Fußballsprache auch sprachkontrastiv zu vermitteln (Burkhardt 2010). Andererseits sind methodendidaktische Zugänge möglich, um am Beispiel der populären und lebensweltlich relevanten Fußballsprache korpuslinguistische Fragestellungen und Techniken zu vermitteln und corpus literacy (Mukerjee 2004) zu fördern. Ergänzt wird die Korpusressource durch ein Online-Tutorial

(<https://ali.gsw.tu-dresden.de/tutorial.html>). Auf dem Blog fussballlinguistik.de werden zudem kleinere Forschungsergebnisse in popularisierender Form publiziert, welche die Nutzungsmöglichkeiten der Korpora veranschaulichen.

Aus Urheberrechtsgründen können in der Analysesoftware nur Ausschnitte der Texte angezeigt werden, ein Download der ganzen Datensätze ist nicht möglich. In einem GitHub-Repository werden jedoch Scraping-Skripte angeboten, mit denen ein Teil der Korpora from the scratch selbst erstellt werden kann, so dass dann auch avanciertere korpuslinguistische Zugänge wie etwa die Berechnung von Word Embedding-Modellen möglich sind.

Literatur:

- Burkhardt, Armin (2010): Der deutsche Fußball und seine Sprache. Ein Beitrag zur Fach- und Sondersprachendidaktik im DaF/Unterricht. In: Brünner Hefte zu Deutsch als Fremdsprache 3 (2), S. 149–166.
- Burkhardt, Armin (2014): Vom Runden und vom Eckigen, in das es hinein muss. Zu den Bezeichnungen von Ball und Tor in der deutschen Fußballsprache. In: Muttersprache 2014 (4), S. 290–301.
- Hardie, Andrew (2012): CQPweb — combining power, flexibility and usability in a corpus analysis tool. In: International Journal of Corpus Linguistics 17 (3), S. 380–409. doi:10.1075/ijcl.17.3.04har.
- Jürgens, Frank (2009): Syntaktische Formen bei der Fußballberichterstattung. In: Burkhardt, Armin/Schlobinski, Peter (Hg.): Flickflack, Foul und Tsukahara: der Sport und seine Sprache. Mannheim: Dudenverlag. S. 160–174.
- Meier, Simon (2017): Korpora zur Fußballlinguistik – eine mehrsprachige Forschungsressource zur Sprache der Fußballberichterstattung. In: Zeitschrift für germanistische Linguistik 45 (2), S. 345–349. doi:10.1515/zgl-2017-0018.
- Meier, Simon (2020): Blogs, Bots & Co. – Public Humanities in den Sozialen Medien. In: Marx, Konstanze/Lobin, Henning/Schmidt, Axel (Hg.): Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig. (= Jahrbuch des Instituts für Deutsche Sprache 2019). Berlin, Boston: De Gruyter. S. 359–362. doi:10.1515/9783110679885-021.
- Mukherjee, Joybrato (2004): Bridging the Gap between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany. In: Connor, Ulla/Upton, Thomas A. (Hg.): Applied Corpus Linguistics: A Multidimensional Perspective. Leiden: Brill. S. 239–250. doi:10.1163/9789004333772.
- Schmidt, Thomas (2009): The Kicktionary – a multilingual lexical resource of football language. In: Multilingual FrameNets in Computational Lexicography Methods and Applications. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110212976.1.101.
- Stefanowitsch, Anatol (2020): Corpus linguistics: A guide to the methodology. (= Textbooks in Language Sciences). Berlin: Language Science Press.