

Korpuslinguistische Ermittlung von Neologismuskandidaten in DeReKo

Harald Längen, Leibniz-Institut für Deutsche Sprache, Mannheim

Abstract

Dieser Beitrag stellt die aktuelle quantitative Methode zur Ermittlung von Neologismuskandidaten in aktuellen Jahrgängen von Korpusdaten im Deutschen Referenzkorpus DeReKo vor, welche auf Keibel et al. (2010) basiert. Die Methode zielt, dem Szenario des Neologismenwörterbuch des IDS folgend, darauf ab, solche neuen Wörter einer Dekade zu ermitteln, die Teil der Allgemeinsprache betrachtet werden können und im Lexikalierungsprozess bereits fortgeschritten sind. Die Methode zielt also z.B. nicht auf Okkasionalismen oder neue Wörter, die nur in einer Fachdomäne verwendet werden (Terminologie), ab.

Die Strategie ist (ähnlich zu Quasthoff 2007), ein Zeitungskorpus aus DeReKo in jahresbezogene Teilkorpora zu partitionieren, und darin Wortformen mit typischen zeitlichen Frequenzverläufen von Neologismen zu ermitteln, d.h. mit einem Ansteigen der Frequenz in der Gegenwart und mit Mindesthäufigkeiten als Bedingungen, um ihre Etablierung in der Allgemeinsprache zu modellieren und um Okkasionalismen auszuschließen. In unserem Ansatz werden insbesondere Zeitverläufe von Wortformen in einer Periode B (z.B. die laufende Dekade seit 2011) mit ihren Zeitverläufen in einer davorliegenden Periode A (z.B. die davorliegende Dekade 2001-2010) verglichen. Auf die resultierende, sehr große Kandidatenliste werden danach verschiedene Filter angewandt, um z.B. verbliebene Regionalismen und Namen zu entfernen. Die immer noch sehr große gefilterte Liste wird dann von den Mitarbeiterinnen des Projekts Neuer Wortschatz lexikografisch analysiert.

Die Methode wird mindestens jährlich angewendet, jeweils mit bestimmten Anpassungen, an, um die neuesten Kandidaten zu ermitteln. Der Beitrag diskutiert die Ergebnisse der aktuellen Anwendung und geht anhand einer Evaluation der Ergebnisse der letzten Anwendungen auf Stärken und Schwächen des Ansatzes ein.

Literatur

- Falk, I., Dernhard, D. & Gérard, C. (2014). From Non Word to New Word: Automatically identifying Neologisms in French Newspapers. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*. Reykjavik, Iceland.
- Keibel, H., Hennig, S. & Perkuhn, R. (2010). *Effiziente halbautomatische Detektion von Neologismuskandidaten*. Technical Report IDS-KL-2010-01. Mannheim: Institut für Deutsche Sprache.
- Klosa, A. & Längen, H. (2018): New German Words: Detection and Description. In: Čibej, Jaka/Gorjanc, Vojko/Kosem, Iztok/Krek, Simon (Hg.): *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, 17-21 July, Ljubljana. Ljubljana: Znanstvena založba, 2018. S. 559-569
- Quasthoff, U. (2007). *Deutsches Neologismenwörterbuch. Neue Wörter und Wortbedeutungen in der Gegenwartssprache*. de Gruyter: Berlin.