

NEOCRAWLER. ERKENNEN UND BEOBACHTEN LEXIKALISCHER INNOVATIONEN IM WEB

Quirin Würschinger / Daphné Kerremans / Jelena Prokic / Hans-Jörg Schmid (München)

Der Wortschatz einer Sprache ist dynamisch und wird von ihren Sprechern ständig erweitert und verändert. Ziel unseres Projekts ist es, die Diffusion von englischen Neologismen im Web zu beobachten, beteiligte Faktoren und deren Einflussgrößen zu ermitteln und zugrundeliegende Mechanismen lexikalischer Innovation zu modellieren. Wir bauen dabei auf Längsschnittdaten aus früheren Phasen des Projekts (Kerremans 2015) sowie auf Fallstudien zum Einfluss von sozialen Medien und medienpezifischen Verwendungsweisen (Würschinger et al. 2016) auf.

Der NeoCrawler (Kerremans et al. 2012) dient dazu, Daten zum erstmaligen Auftreten und zur weiteren Verbreitung von englischen Neologismen im World Wide Web zu erheben. Zum einen durchsucht der Crawler das Web nach Kandidaten für lexikalische Innovation und speichert diese in einer Datenbank ab. Zum anderen sammelt er mithilfe der Google Custom Search API wöchentlich Daten zu deren Verbreitung und erstellt ein Korpus der Gebrauchskontexte. Dabei wird eine Reihe von Pre- und Post-Processing-Schritten zur Bereinigung, Vorbereitung und Analyse der Daten durchgeführt. Dadurch können Daten über die Gebrauchshäufigkeiten der (potenziellen) Neologismen, über spezifische Verwendungsweisen sowie über deren Verwendung in verschiedenen Texttypen und semantischen Domänen gewonnen werden, welche für die Diffusion lexikalischer Innovationen ausschlaggebend sind.

Literatur:

- Kerremans, Daphné/Stegmayr, Susanne/Schmid, Hans-Jörg (2012): The Neocrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. In: Allan, Kathryn/Robinson, Justyna A. (Hg.): *Current Methods in Historical Semantics*. Berlin: de Gruyter Mouton, 59-96.
- Kerremans, Daphné (2015): *A Web of New Words: A corpus-based study of the conventionalization process of English neologisms*. Frankfurt am Main: Peter Lang.
- Würschinger, Quirin/Elahi, Mohammad Fazli/Zhekova, Desislava/Schmid, Hans-Jörg (2016): Using the web and social media as corpora for monitoring the spread of neologisms: The case of 'rapefugee', 'rapeugee', and 'rapugee'. In: *Proceedings of the 10th Web as Corpus Workshop (WaC-10)*. Berlin: Association for Computational Linguistics, 35-43.