

# NeoCrawler: Erkennen und Beobachten lexikalischer Innovationen im Web

Prof. Dr. Hans-Jörg Schmid, Dr. Daphne Kerremans, Dr. Jelena Prokic, Quirin Würschinger  
Ludwig-Maximilians-Universität München

## 1 Forschungsgegenstand

Untersuchung der Diffusion lexikalischer Innovationen:

- Welche Konventionalisierungspfade weisen lexikalische Innovationen auf?
- Welche strukturellen und sozio-kognitiven Faktoren haben auf deren Verbreitung Einfluss und in welchem Ausmaß?

Methodische Herausforderungen:

- Wie lassen sich potenzielle Neologismen im frühen Stadium erkennen?
- Wie lässt sich deren Diffusion systematisch messen und beschreiben?
- Wie lassen sich mögliche Einflussfaktoren (semi-)automatisch erheben?

## 2 Methodik und Korpus

Der NeoCrawler (Kerremans, Stegmayr und Schmid 2012) nutzt Googles Custom Search API, um in wöchentlichen Intervallen das World Wide Web nach Neologismen zu durchsuchen. Wie in Abbildung 1 dargestellt, ermittelt dabei das Modul **Discoverer** (linksseitig) neue Kandidaten für Neologie, wohingegen der **Observer** (rechtsseitig) wöchentlich deren Auftreten im Web beobachtet.

Durch dieses Verfahren entsteht ein **dynamisches Korpus**. Durch Post-Processing-Verfahren und Klassifikatoren kann neben der reinen Betrachtung der Gebrauchshäufigkeiten auch die Rolle von Variablen wie des Auftretens in unterschiedlichen Diskurstypen (z.B. Online-Zeitungen, Blogs, Foren) oder Diskursdomänen (z.B. Technologie, Politik, Sport) untersucht werden.

- Zeitraum der Beobachtung: 12/2009 bis 01/2011; seit 07/2016
- Anzahl aktuell beobachteter Kandidaten: 569

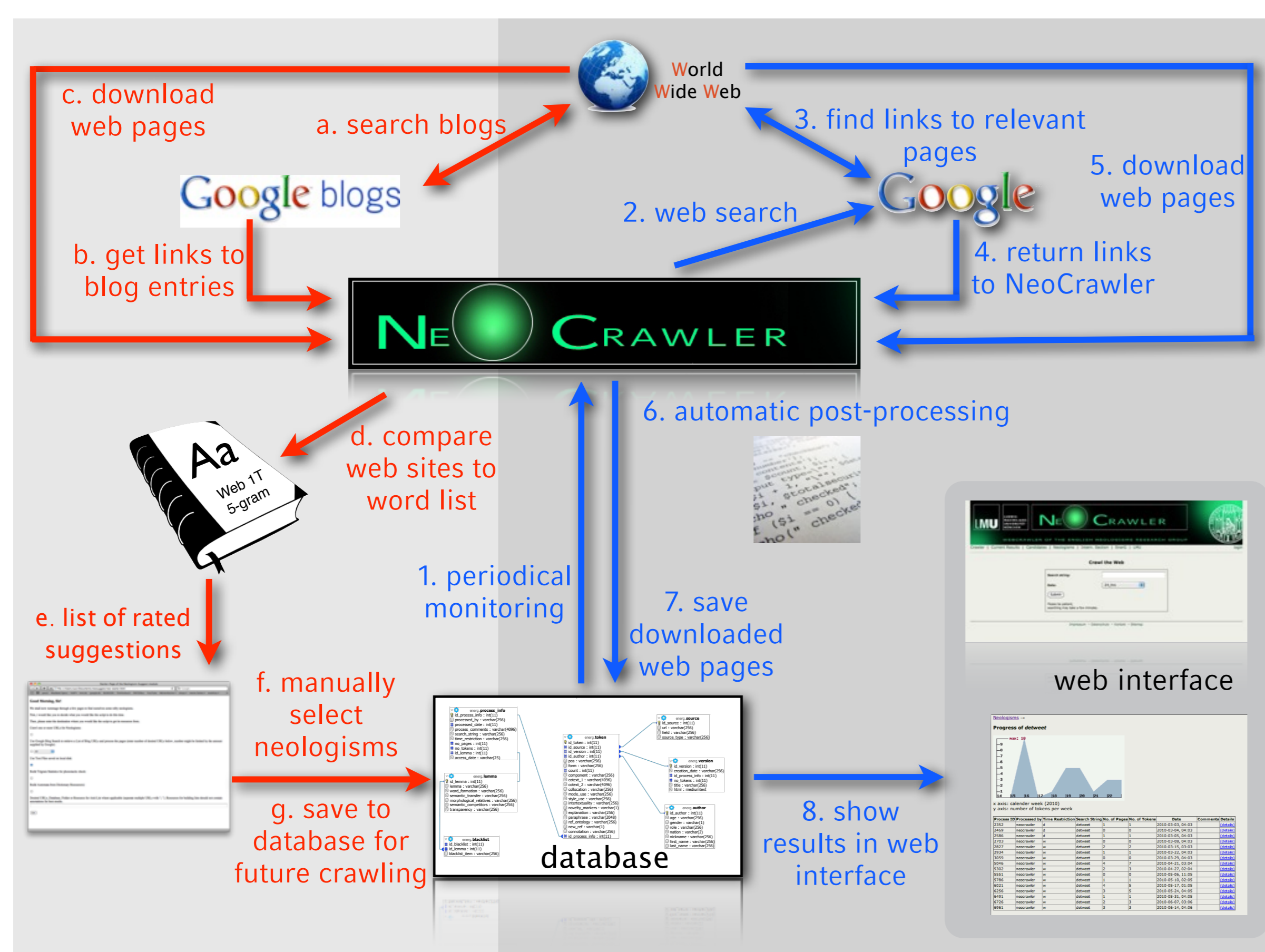


Abbildung 1: Architektur des NeoCrawlers

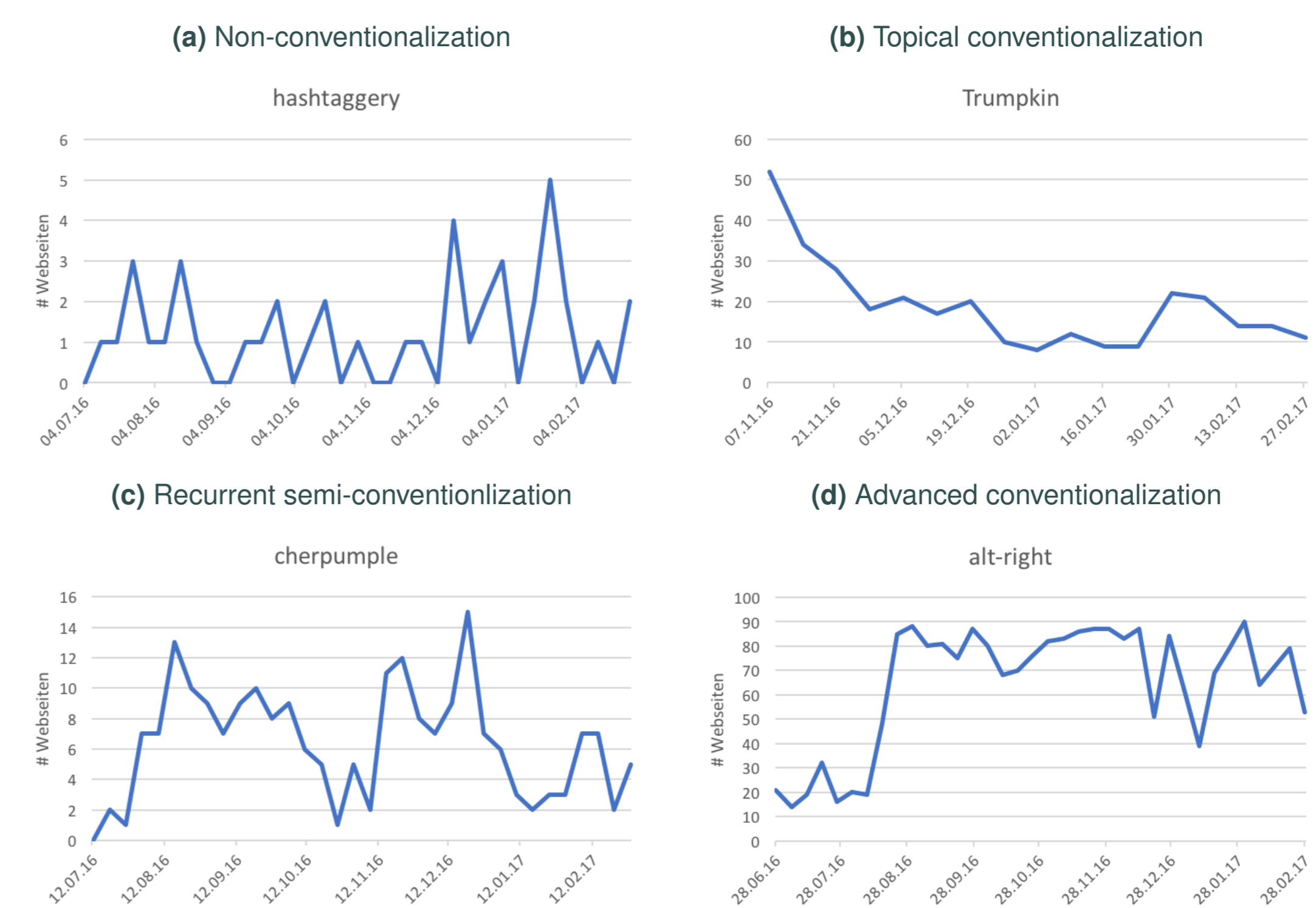
Ein **Web-Interface** ([www.anglistik.uni-muenchen.de/neocrawler](http://www.anglistik.uni-muenchen.de/neocrawler)) ermöglicht die Abfrage der gesammelten Daten, zeigt Kontexte an, bietet Möglichkeiten zur manuellen Klassifikation und einen quantitativen Überblick durch Visualisierung der beobachteten Gebrauchshäufigkeiten. Zugang zum Web-Interface kann bei Interesse auf der Webseite beantragt werden.

## 3 Ergebnisse

### 3.1 Konventionalisierungsgrade

Diffusion nimmt die zentrale Rolle im Prozess der Institutionalisierung (Lipka 2002) und Konventionalisierung (Schmid 2015) von lexikalischen Innovationen ein. Aus der Beobachtung unserer Daten ergibt sich ein vierstufiges Kontinuum. (Kerremans 2015, S. 115–147) Abbildung 2 zeigt vier prototypische Vertreter aus unserem Datensatz.

Abbildung 2: Konventionalisierungsgrade



### 3.2 Phasen und Einflussfaktoren der Konventionalisierung anhand von *alt-right*

**Innovation:** Prozess der Einführung neuer lexikalischer Einheit

- *formale Ökonomie*, *morpho-semantische Transparenz* und *etablierte Synonyme*: Effizienz gegenüber Langform „Alternative Right“ (älterer Terminus)
- *soziales Prestige*: maßgebliche Prägung durch 'White Supremacist' Richard Spencer
- *nameworthiness*: Salienz des Konzepts im Diskurs der jeweiligen Sprechergemeinschaft

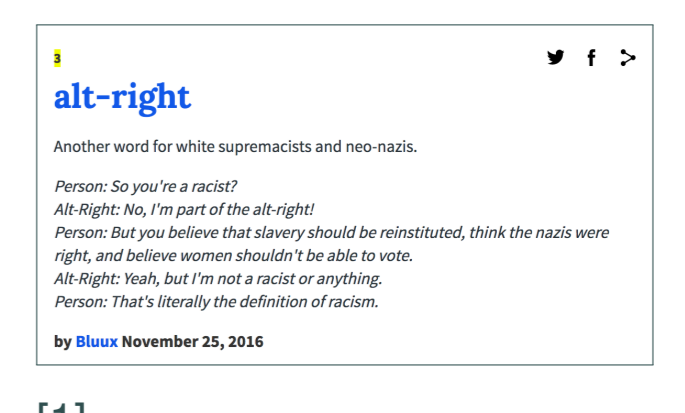
**Usualisierung** (Blank 2001): Gebräuchlichkeit der Verwendung des neuen Wortes innerhalb eines spezifischen Gebrauchskontexts oder innerhalb einer spez. Sprechergemeinschaft

- *sozio-emotionaler Wert*: kommunikative Motivation: Gruppenidentität und Ausgrenzung nach außen
- *Gebrauchsfrequenz*: Steigerung der „usage intensity“ (Stefanowitsch und Flach 2017) innerhalb der Gebrauchsdomäne
- *Ko- und Kontextualisierung*: Gebrauch in prototypischen Gebrauchskontexten mit kontingenter De- und Konnotation
- *soziolinguistisches Netzwerk*: enges Netz an Beziehungen, Gruppenidentität und Abgrenzung nach außen



**Diffusion:** Verbreitung des neuen Wortes in erweiterte Sprechergemeinschaft, weitere Diskursdomänen oder Diskurstypen

- *sozialer Wert*: Salienz und Wert des Konzepts im gesamtgesellschaftlichen Diskurs
- *Diskursdomänen*: Ausweitung auf Diskurs um den US-Wahlkampf von Donald Trump
- *Diskurstypen*: Verbreitung von Foren und privaten Blogs hin zu formelleren Quellen wie Online-Zeitungen
- *sozio-linguistisches Netzwerk*: erweiterte Sprechergemeinschaft aus, nicht mehr auf rechtsradikale Kreise beschränkt



**Normierung:** Festlegung der Form-Bedeutungs-Relation, sowohl implizit als auch explizit (zu gegebenem Zeitpunkt bei *alt-right* noch vergleichsweise hohe Divergenz, s. [1])

### Literatur

Blank, Andreas (2001). „Pathways of lexicalization“. In: *Language Typology and Language Universals/ La Typologie des Langues et les Universaux Linguistiques/ Sprachtypologie und Sprachliche Universalien. An International Handbook/ Manuel International/Ein Internationales Handbuch*. Hrsg. von Martin Haspelmath et al. Bd. 2. Mouton, S. 1596–1608.

Kerremans, Daphné (2015). *A Web of New Words. A Corpus-Based Study of the Conventionalization Process of English Neologisms*. Bd. 15. English corpus linguistics. Frankfurt a. M.: Lang.

Kerremans, Daphné, Susanne Stegmayr und Hans-Jörg Schmid (2012). „The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change“. In: *Current Methods in Historical Semantics*. Hrsg. von Kathryn Allan und Justyna Robinson. Berlin: Mouton de Gruyter, S. 59–96.

Lipka, Leonhard (2002). *English Lexicology: Lexical Structure, Word Semantics and Word-Formation*. Narr Studienbücher. Tübingen: Narr.

Schmid, Hans-Jörg (2015). „A blueprint of the Entrenchment-and-Conventionalization Model“. In: *Yearbook of the German Cognitive Linguistics Association* 3, S. 1–27.

Stefanowitsch, Anatol und Susanne Flach (2017). „The corpus-based perspective on entrenchment“. In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Hrsg. von Hans-Jörg Schmid. Boston, USA: American Psychology Association und de Gruyter Mouton, S. 101–128.

Würschinger, Quirin et al. (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, Germany: Association for Computational Linguistics, S. 35–43.

[1] Urban Dictionary, Definition [1], <http://www.urbandictionary.com/define.php?term=alt-right>, 05.03.2017

### Förderung

Dieses Projekt wird durch Mittel der Deutschen Forschungsgemeinschaft (DFG), Projekt SCHM 1232/5-1, gefördert.