Diffusion of Lexical Innovations

Investigating the Spread of English Neologisms on the Web and on Twitter

Quirin Würschinger, LMU Munich

FJUEL 2018
Munich
14 September, 2018

Research questions

- Which new words enter the English language?
- How do they diffuse?
- Which factors affect how they diffuse?

Which new words enter the English language?

Urban Dictionary

TRENDING RN - SEPTEMBER 13, 2018

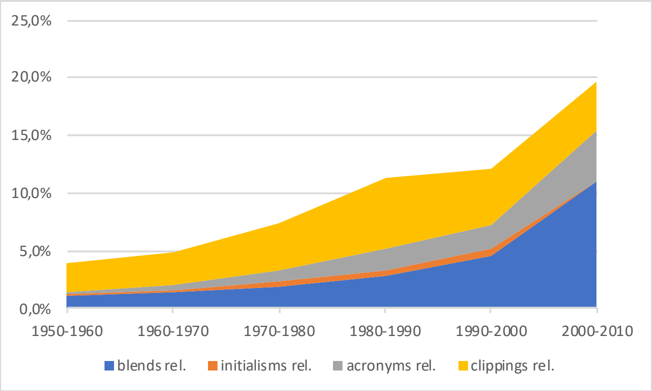
1. fo fo fo
2. gammin
3. Solider
4. Adele Syndrome
5. sleepyboner
6. ingenuine
7. I like your shoes
8. shekels
9. FWB
10. 7/11 Was a Part Time Job
11. kissogram
12. tampico
13. Boin
14. nobber
15. sheckles
16. Throwing Neck
17. sponcon
18. kante
19. Amala
20. sit on it and rotate it
21. Cartier
22. the spongebob
23. Flossed
24. 24/7/365
25. The Crab
26. Dominatrix
27. throw neck
28. Happy 9/11
29. GRU
30. Forelsket

What is a ‘new word’?

- **nonce-formations**: used once, but have not diffused
- **neologisms**: have diffused to some degree, but are still perceived to be ‘new’
- **conventional words**: have successfully diffused and are known to the majority of the speech community

Which words enter the English lexicon?

Morphological productivity



Which words are entering the English language?

NeoCrawler: Discoverer module (Kerremans and Prokic 2018)

- goal: investigating *incipient* diffusion
- method:
  - retrieve sample of web pages
  - dictionary matching
  - semi-manual selection of candidates
  - store in database ($\approx$ 1,000 lemmas)

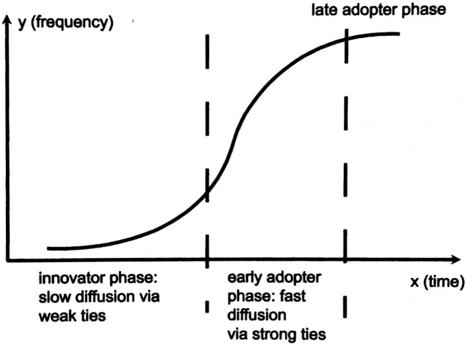
How do new words diffuse and become conventional?

Previous work

- **cultural innovation**: S-curves (Rogers 1962; Rogers and Shoemaker 1971), big data (Kim, McFarland and Leskovec 2017)
- **sociolinguistics and language change**: mainly phonology and syntax, diffusion, early and late adopters (Labov 1980; J. Milroy and L. Milroy 1985; Croft 2000)
- **structural**: lexicalization, institutionalization, establishment (Bauer 1983; Lipka 1992)
- **corpus linguistics**:
  - recent work: large-scale studies, bigger samples (Eisenstein et al. 2014; Grieve, Nini and Guo 2016)
  - tools: *NeoCrawler* (Kerremans, Stegmayr and Schmid 2012), *Wortwarte* (Lemnitzer 2018), *Logoscope* (Bernhard et al. 2015), *Neoveille* (Cartier 2017)

S-curve

Figure 1: Integration of Milroy's and Rogers' model of diffusion stages into an S-curve (Kerremans 2015, p. 65)



How can we model diffusion?

The EC model (Schmid 2015) – a simplified account:

- **coining**: first use
- **usualization**: agreement over communicative function
- **diffusion**: spread to new usage contexts and speakers
- **normation**: establishment of norms about how to use new words

Which factors influence diffusion?

lemma-inherent (type level)

- form
  - transparency
  - productivity of word-formation pattern
  - formal appeal
- meaning
  - semantic domain
  - existing near-synonyms
  - nameworthiness

in usage (token level)

- sociolinguistic
  - density of social network
  - speakers' prestige
- cognitive
  - formal salience in use
  - metalinguistic uses
- pragmatic
  - type of source
- emotive-affective
  - sentiment

Dimensions of diffusion

new uses bring about ...

- spread across speakers
- spread across usage contexts

		usage contexts	
		low	high
speakers	low	*hypostatization*	*alt-left*
	high	*electron*	*DNA*

How can we measure diffusion empirically?

- detecting candidates: Discoverer
- investigating diffusion
  - on the web: NeoCrawler
  - on social media: Twitter

How do new words spread on the web?

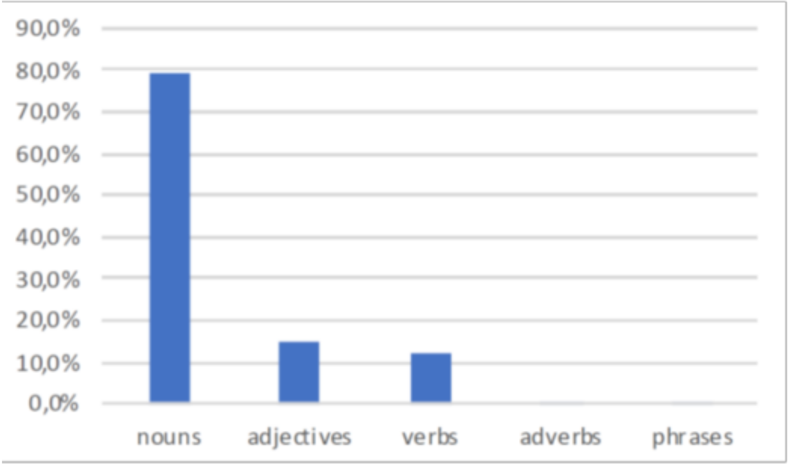
NeoCrawler (Kerremans, Stegmayr and Schmid 2012)

- weekly Google Searches[1] (about 1,000 lemmas)
- download all html pages found
- pre- and post-processing
- corpus compilation

[1]Google Custom Search API

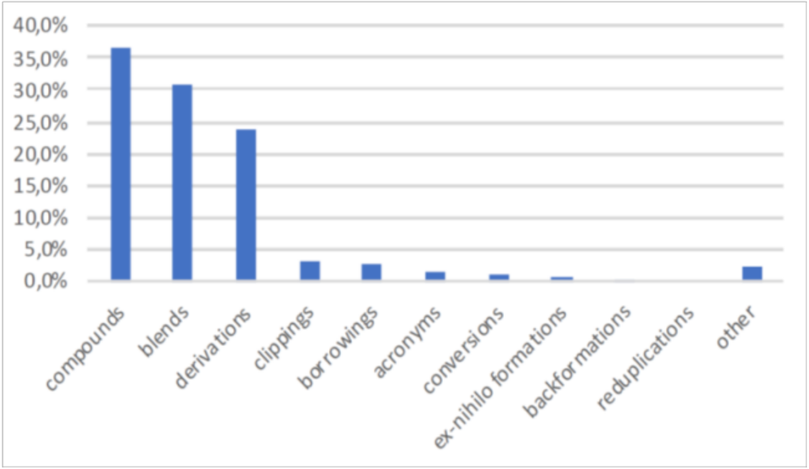
Results

Word classes



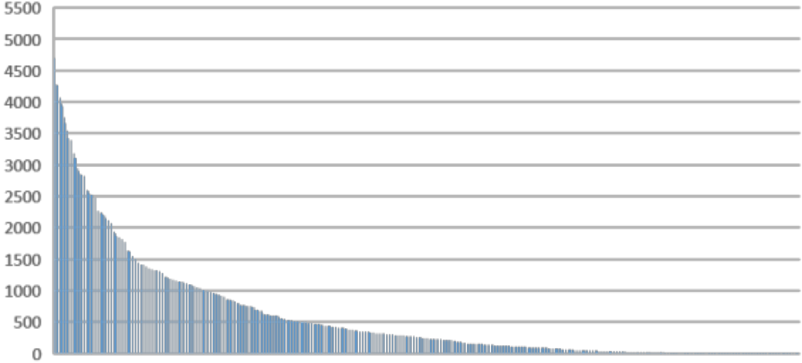
Results

Word-formation processes



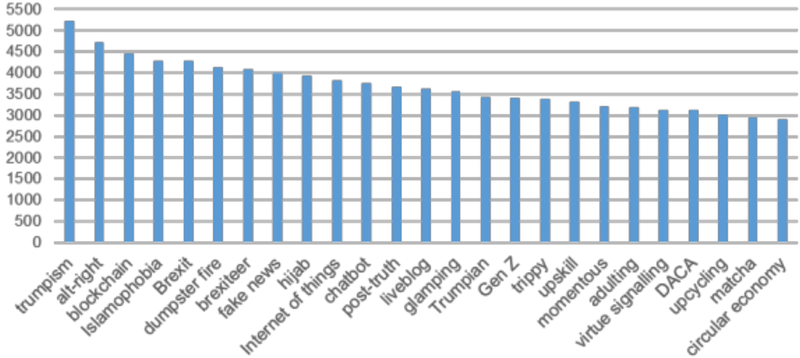
Results

Diffusion of all candidates



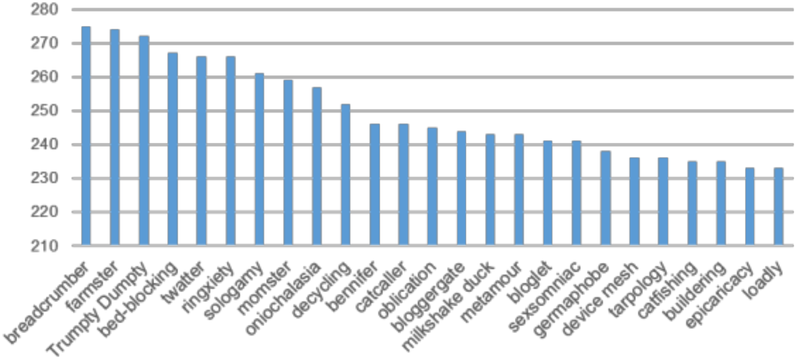
Results

Top 25 items



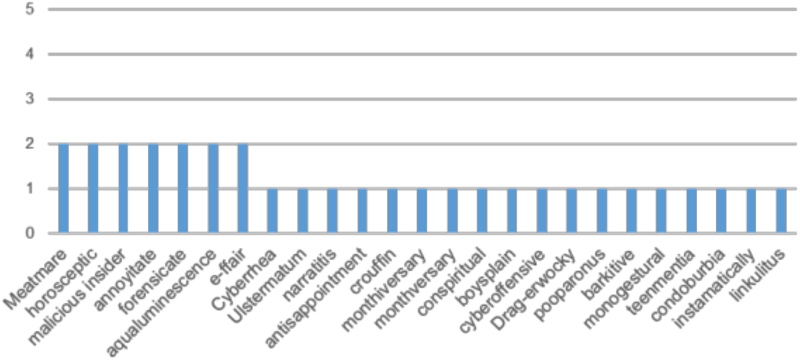
Results

Items around median



Results

Bottom 25 items



How do new words spread on Twitter?

Methodology

- advantages
    - going back in time
    - high temporal resolution
    - user metadata (social, geographic)
    - social network data
- tools
    - ongoing Twitter mining: *TAGS*
    - web scraping: *Twitter Scraper*

clipped form of earlier term *Alternative Right*, coined by White Supremacist Richard Spencer

formed in analogy (and opposition) to pre-existing *alt-right*

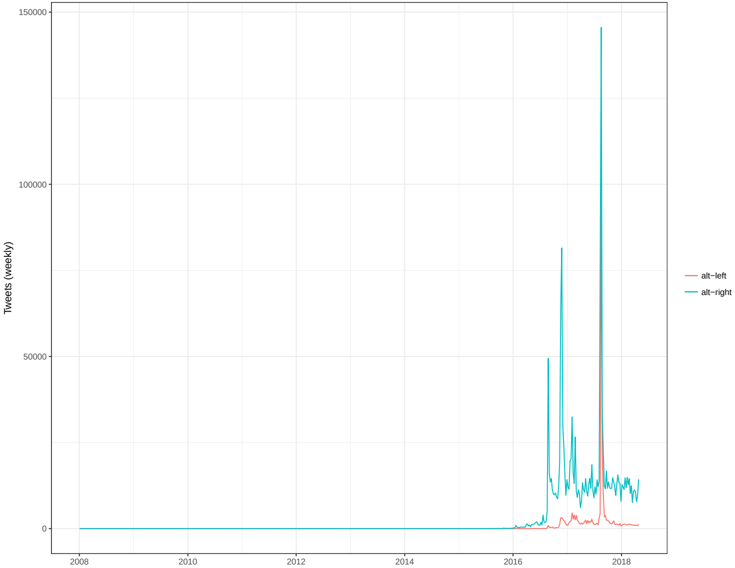
Corpus examples

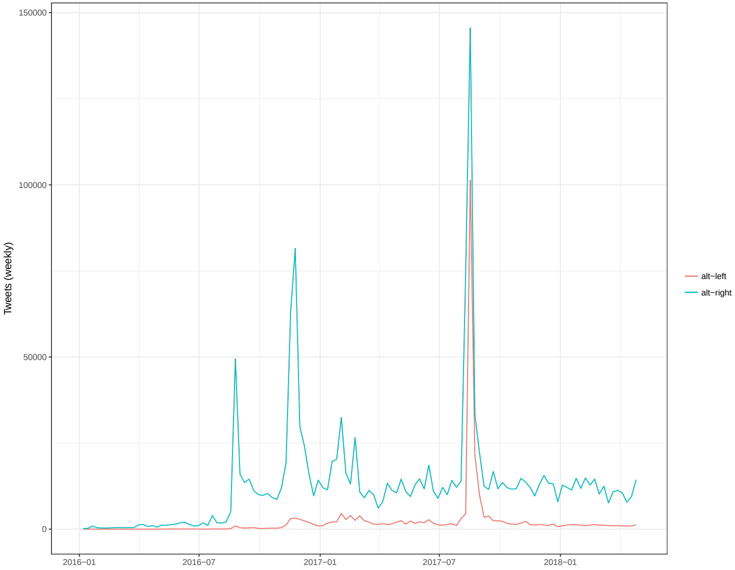
use of *alt-left* in 2016

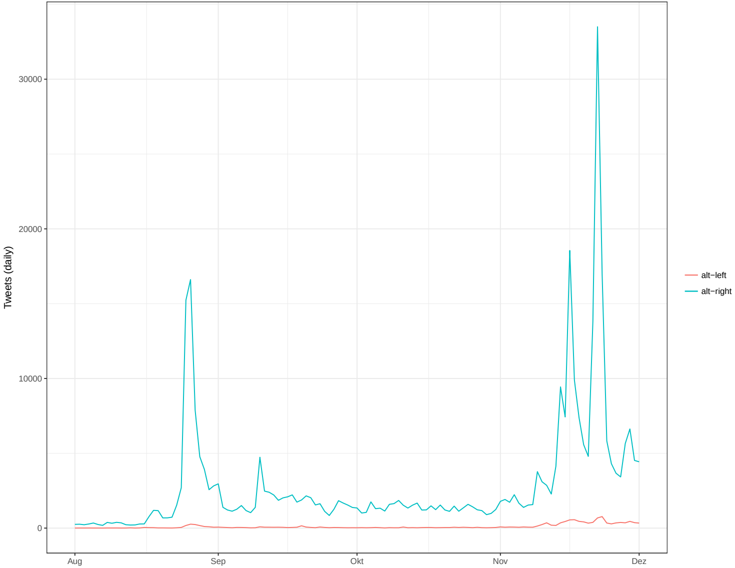
Corpus examples

use of *alt-left* in 2017

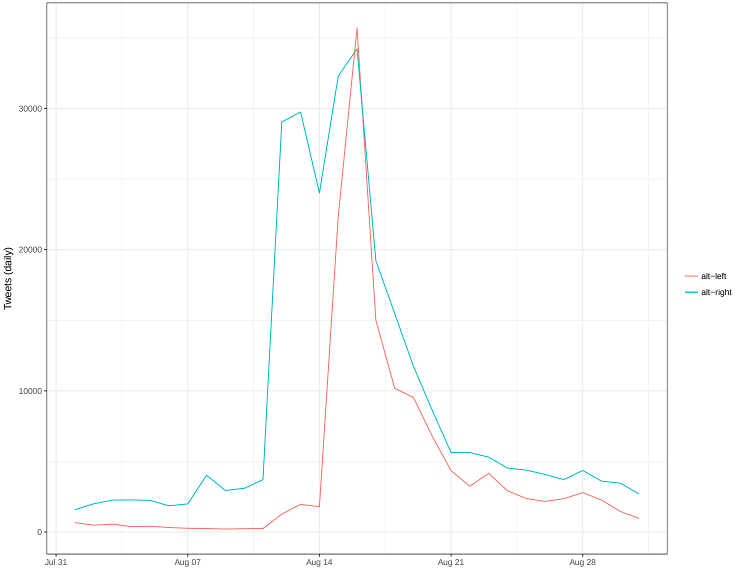
Zooming in on diffusion




150000
100000
50000
0
Tweets (weekly)
2016−01
2016−07
2017−01
2017−07
2018−01
alt−left
alt−right



August 25, 2016: Hillary Clinton's speech against alt-right
November 22, 2016: Trump publicly defends Steven Bannon



August 12, 2017: Charlottesville Rally
August 16, 2017: Trump attacking ‘alt-left’

Zooming in on diffusion

the ‘social’ network

Social network analysis

	alt-left	alt-right
number of tweets	295,968	1,760,777
number of individual speakers	117,607	550,798
avg. weighted degree	0.855	1.044
modularity	0.937	0.877

→ *alt-right* shows a high degree of diffusion over an extended time window

→ *alt-left* shows some diffusion, but remains to be used by smaller pockets of the speech community

Implications

- S-curves not to be expected due to effects of topicality[2]
- differentiated view on diffusion: sub-communities
- 'influencers' drive innovation
- social network characteristics influence diffusion

[2] and for other reasons that we could discuss ...

Thanks!

[3]

[3] OED Word of the Year 2015