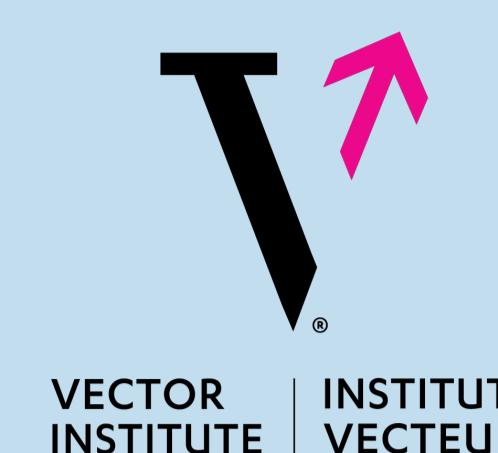


From *sunblock* to *softblock*: Analyzing the correlates of neology in published writing and on social media

Maria Ryskina, Matthew R. Gormley, Kyle Mahowald, David R. Mortensen, Taylor Berg-Kirkpatrick, Vivek Kulkarni



Are mechanisms of word emergence (neology) domain-dependent?

Published writing (books, magazines):
e-mail, sunblock, sitcom, ...

Social media (Twitter):
oomf, softblock, 5ever, ...

Data: COHA (1800–1989)
COCA (1990–2012)
 31M tweets (2007–2010)
229M tweets (2011–2021)

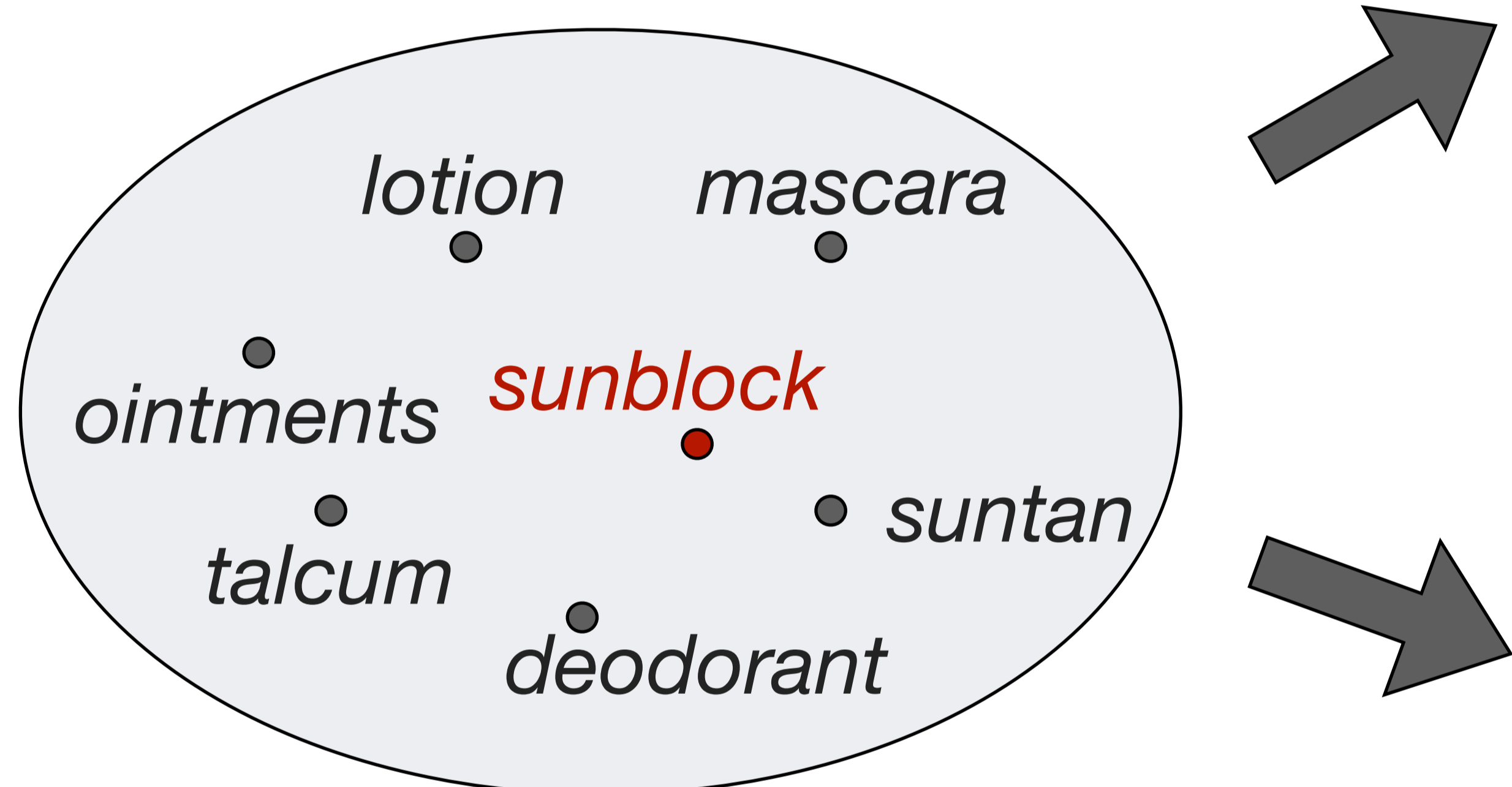
Neologisms extracted by frequency pattern and paired with control words:
 video ↔ *radio*, *oomfs* ↔ *buddies*

Hypotheses (Ryskina et al., 2020):

Supply: new words emerge in sparser areas of the semantic space

Demand: new words emerge in areas where words grow in popularity

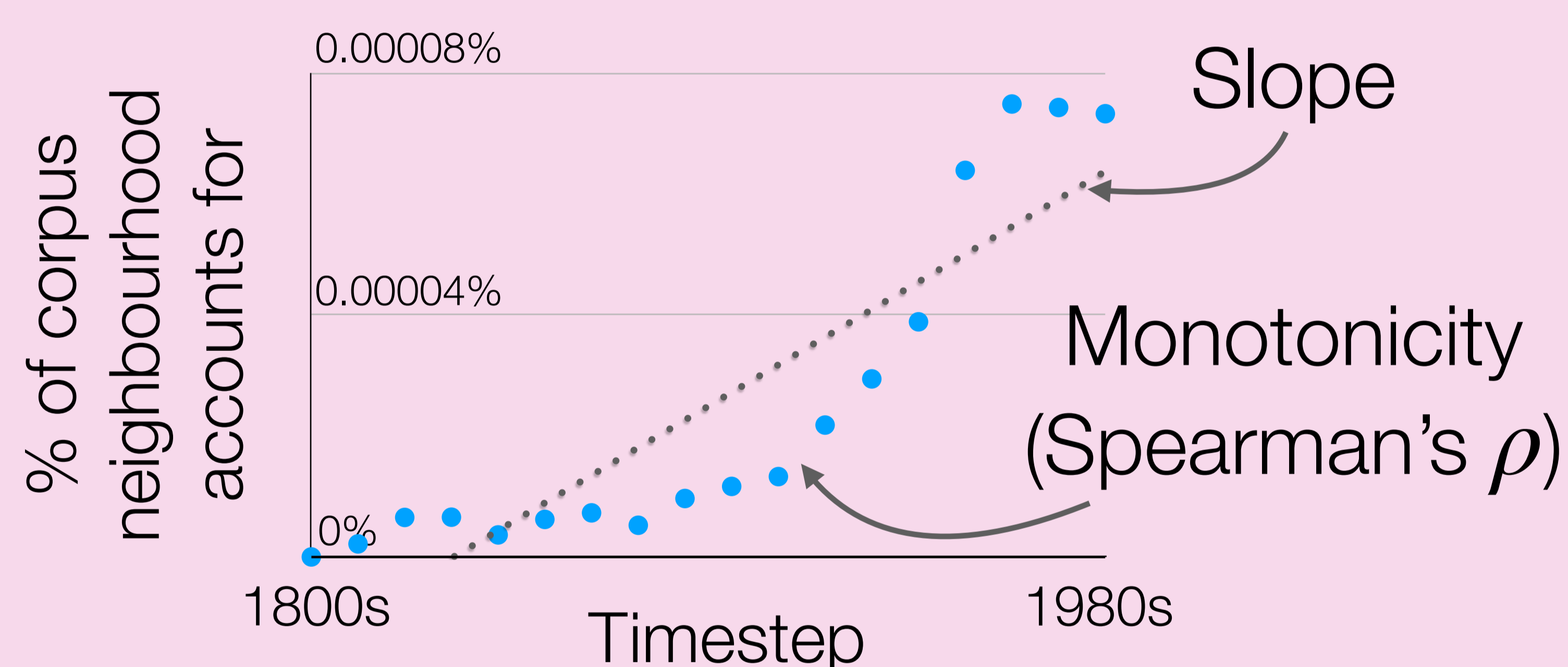
Method:



Target word's neighbourhood (Word2Vec/RoBERTa space)

Density: $\log(\#\text{neighbours} + 1)$

Frequency growth:



Result: less evidence for demand hypothesis for , possibly due to creative tendencies in neologism formation

