



Data and methods

Findings

```
graph TD; A[Monthly Reddit comments  
2013/06 - 2016/05] -- "Filter  
Only English  
Top 100,000 words  
Remove spam, robots" --> B[1.6 billion posts]; B --> C[57 billion tokens  
190 million trigrams]; B --> D[14 million users  
100 million threads  
330,000 subreddits]; C -- "Acronyms  
Blends  
Clippings  
Derivations  
Exclamations  
Onomatopoeia  
Respellings" --> E[1451 successful innovations]; C --> F[600 failed innovations];
```

Monthly Reddit comments
2013/06 - 2016/05

Filter
Only English
Top 100,000 words
Remove spam, robots

1.6 billion posts

57 billion tokens
190 million trigrams

Acronyms
Blends
Clippings
Derivations
Exclamations
Onomatopoeia
Respellings

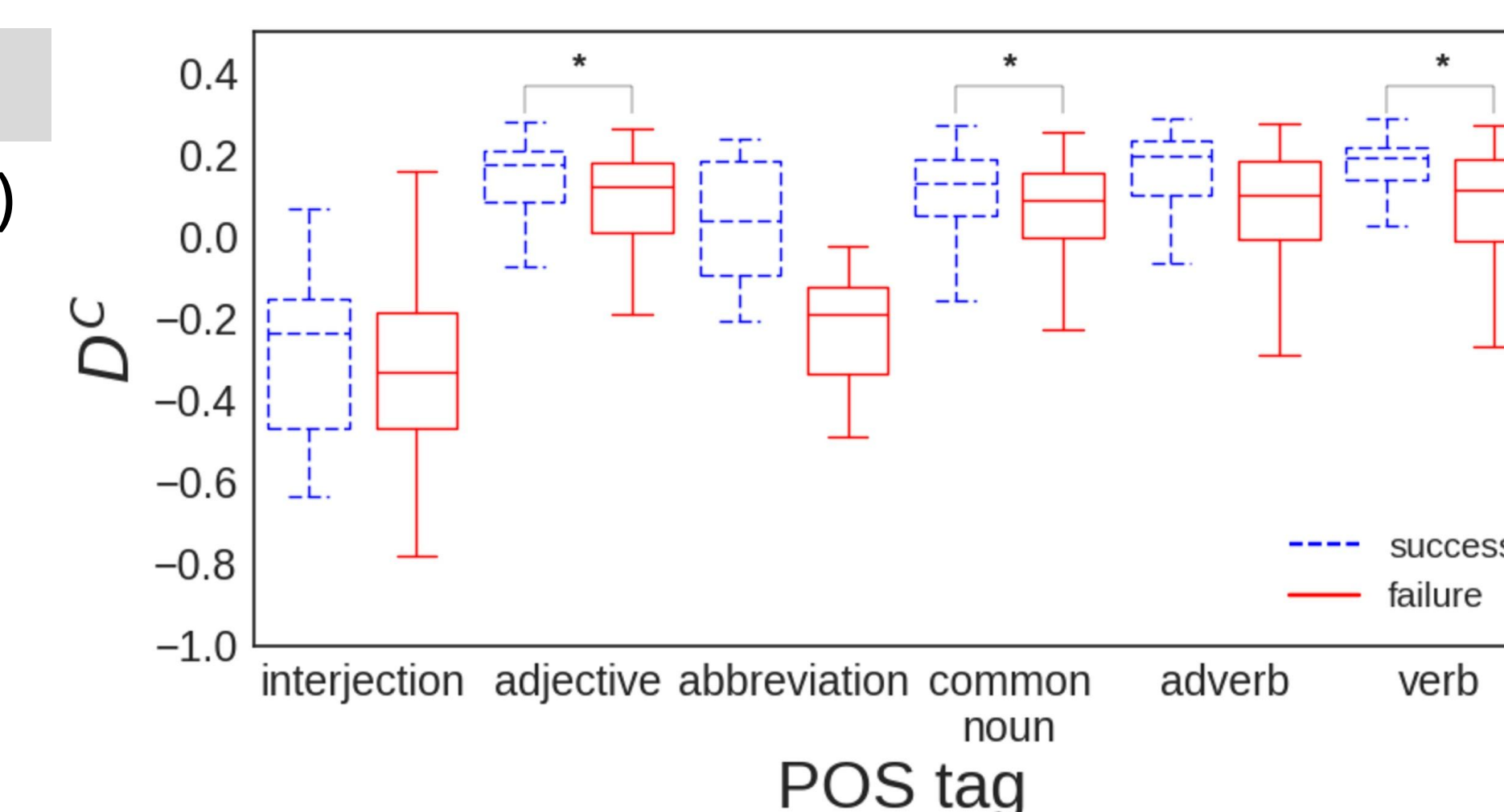
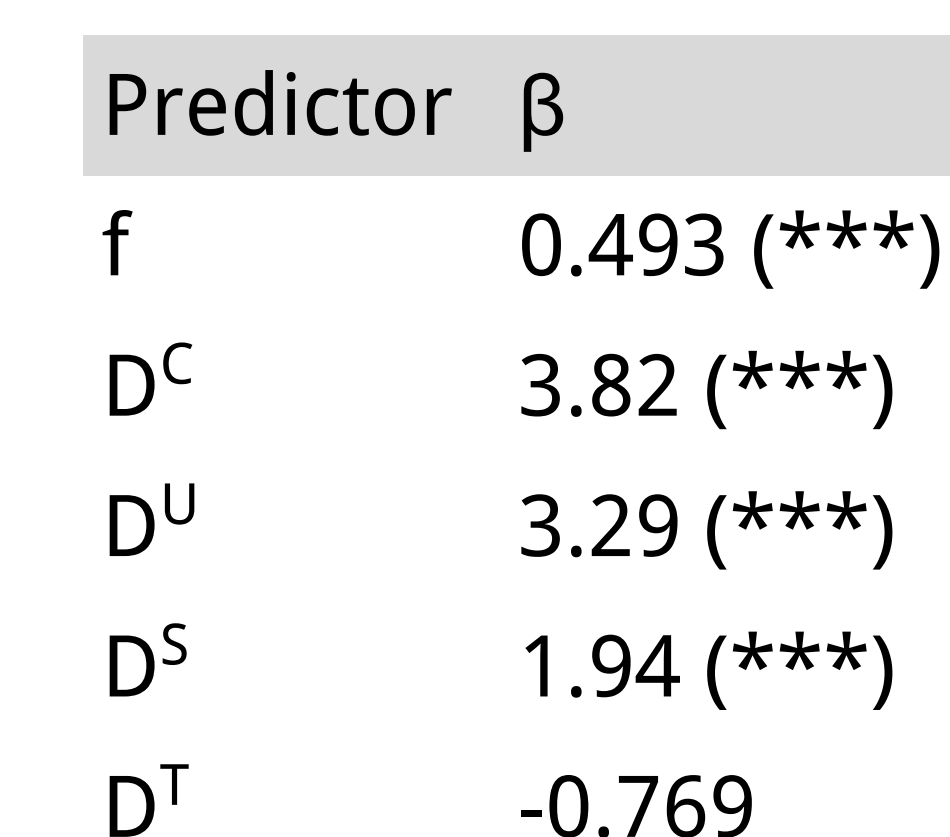
14 million users
100 million threads
330,000 subreddits

1451 successful innovations

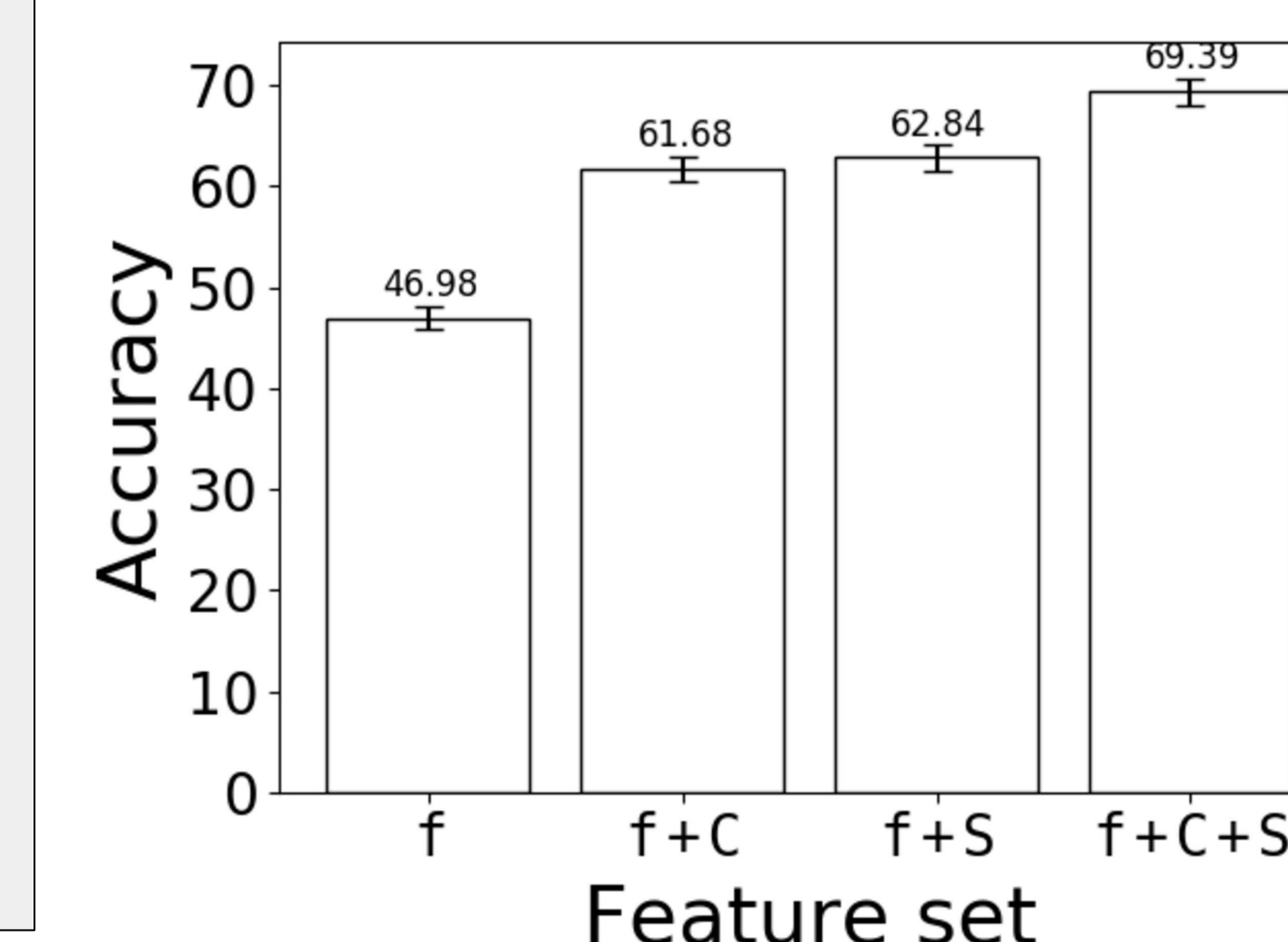
600 failed innovations

RQ1: Higher context dissemination predicts innovation success. Consistent across part-of-speech categories (generated with POS tagger trained on Twitter [2]).

RQ2: Higher user and subreddit dissemination predicts innovation success; thread dissemination is insignificant.

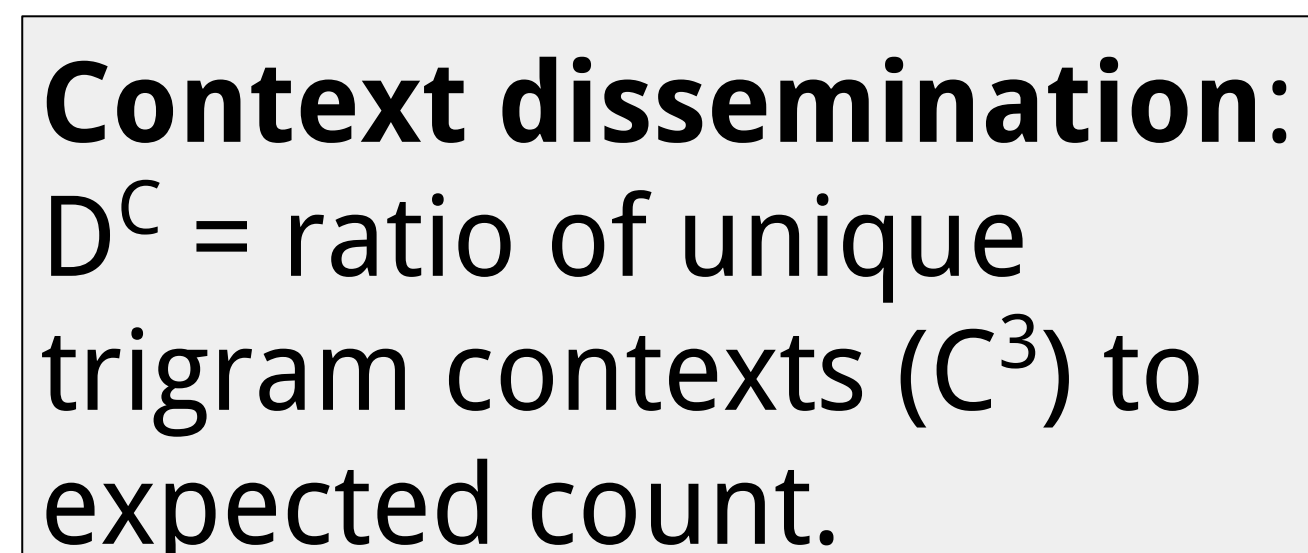


Context+social dissemination (f+C+S) contributes more to predicting success than context or social alone: social and linguistic factors contribute differently to success.

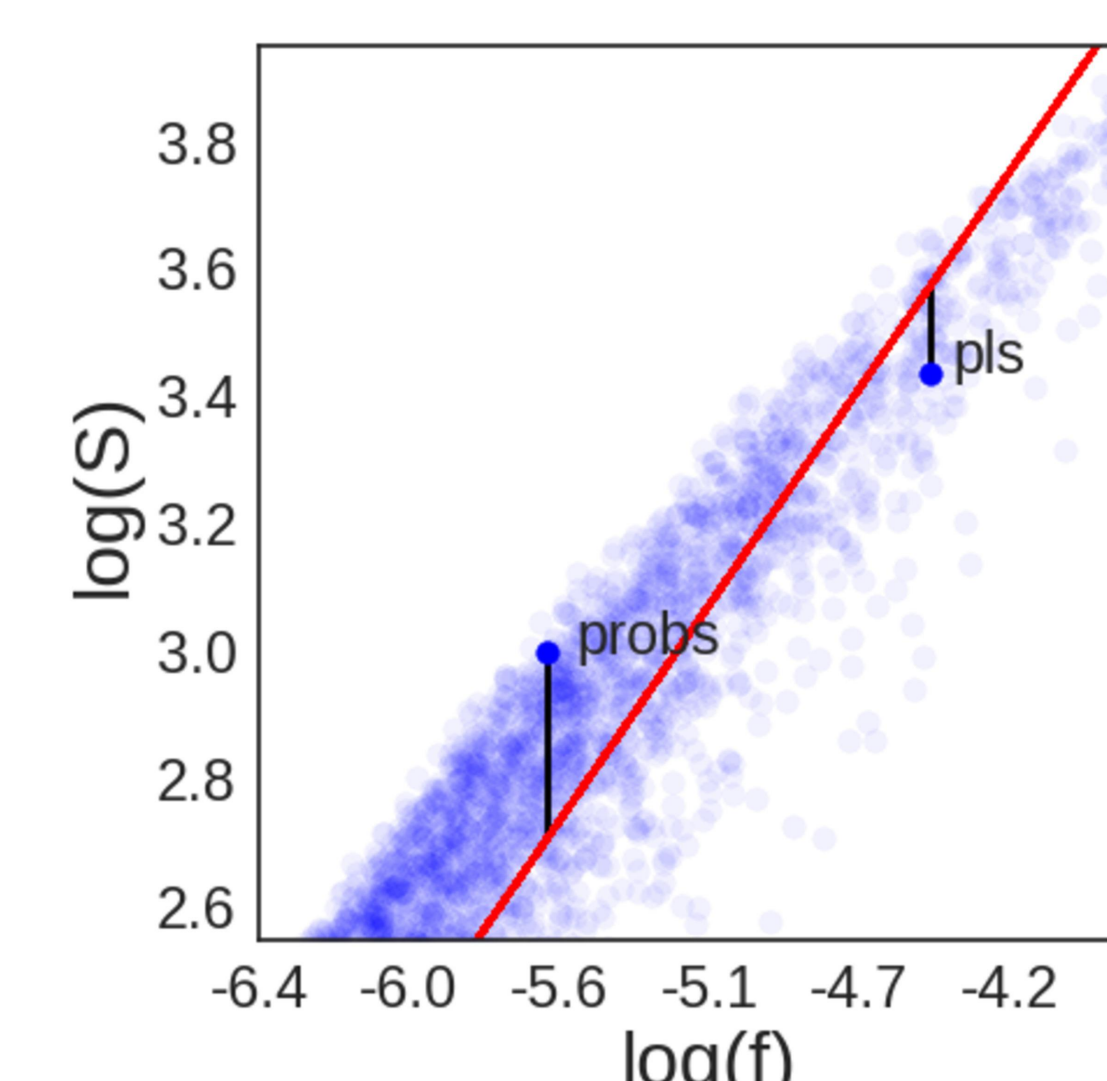
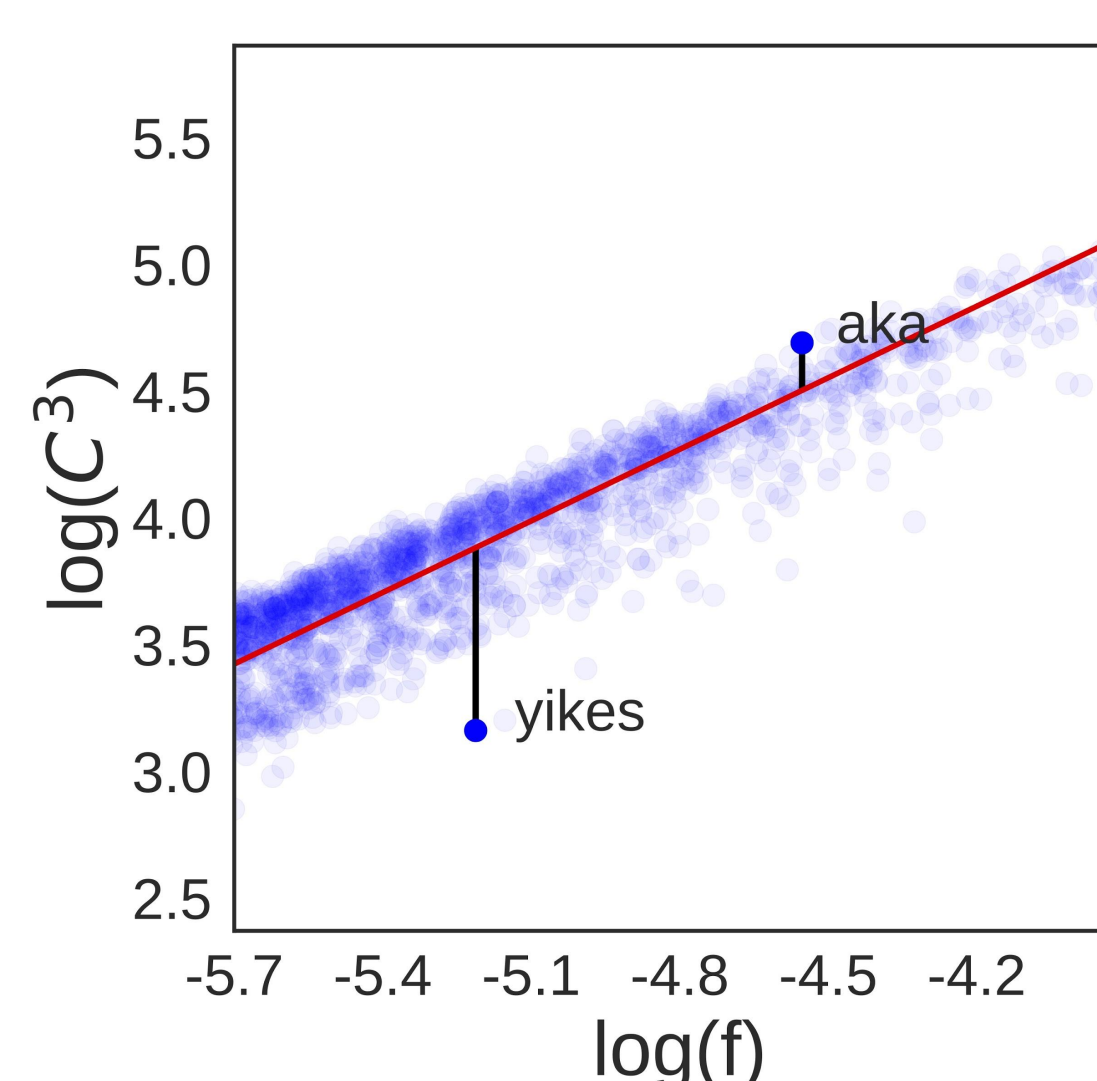


Our work validates the **Diversity** criterion of the FUDGE model for predicting successful innovations [4]. It also supports the need for comparison tests when studying language change, namely comparing variables at different stages of change. Future work will investigate more nuanced versions of context dissemination (syntactic, semantic contexts).

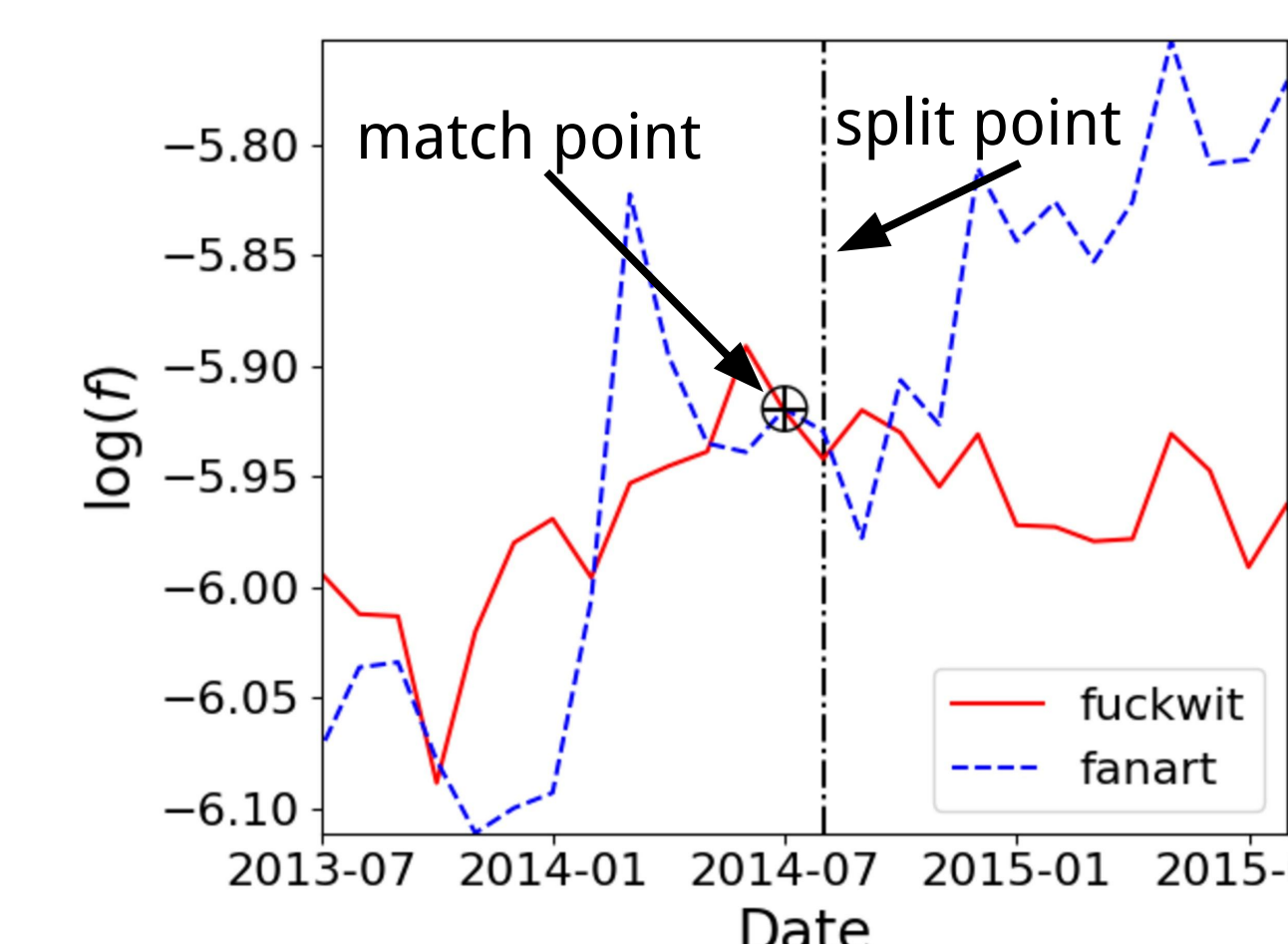
Full paper (in submission):
<https://arxiv.org/abs/1709.00345>



Social dissemination
among users, subreddits,
threads [2]:
 D^U, D^S, D^T = ratio of
actual to expected count.

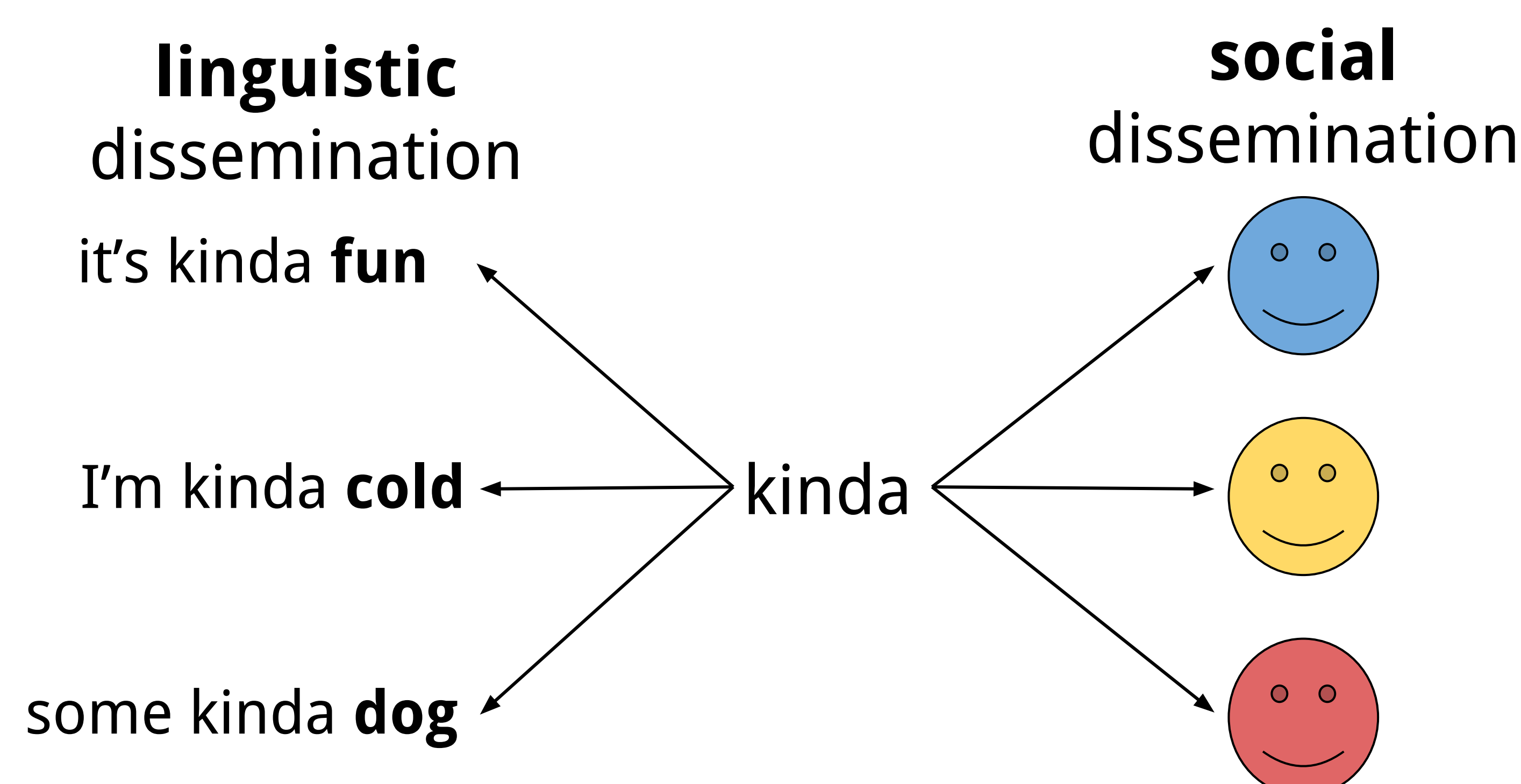
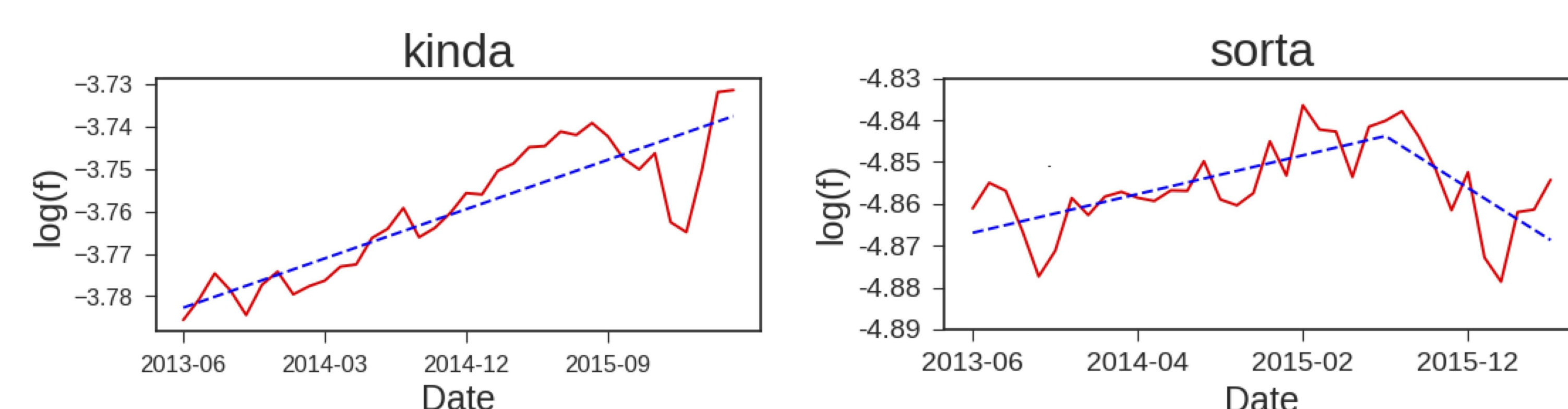


We match successful and failed innovations on frequency before the split point and learn a classifier to predict innovation success.



References

- [1] Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLoS ONE*, 6(5), 1–12.
- [2] K. Gimpel et al. (2011). Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proceedings of ACL*, 42–47.
- [3] Grieve, J., Nini, A., & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, 20(2), 1–29.
- [4] Metcalf, A. (2004). *Predicting new words: The secrets of their success* (Vol. 18). New York: Houghton Mifflin.
- [5] Partington, A. (1993). Corpus evidence of language change: The case of the intensifier. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 177–192). Philadelphia: John Benjamins Publishing.



RQ1: Do successful lexical innovations exhibit higher dissemination across linguistic contexts than failed innovations?

RQ2: Do successful lexical innovations tend to have higher dissemination across social contexts than failed innovations?