Marginality and orthographic variation in (lol)

Joshua McNeill

Seminar in Language Variation

3 December 2020

Code available at https://osf.io/mgdpu/

Phenomenon

Spelling varies

Examples of (lol) from Twitter:

1. @markeasterbrook @MacFinlay @rarchbarch **LOL**. Never seen that.
2. @QueenofCobden **lolololol** I think I was the first. _Ã™Â□Ã‰
3. @anceldelambert @Liberal_Lunacy @JeffHollandaise **lolcow** milk, fresh winge salt, all churned into the memeyest kek butter.
4. @Wario64 except for the volcano **lol**

Previous Work

Second Dialect Acquisition

e.g., the work of Jennifer Nycz

- Inherently focuses on geography and acquired that to which one does not already have access

Twitter and Orthographic Variation

- Large-scale lexical variation (e.g., Bamman et al., 2014)
- Representation of spoken language in writing (e.g., Tatman, 2016)

Social Network Analysis

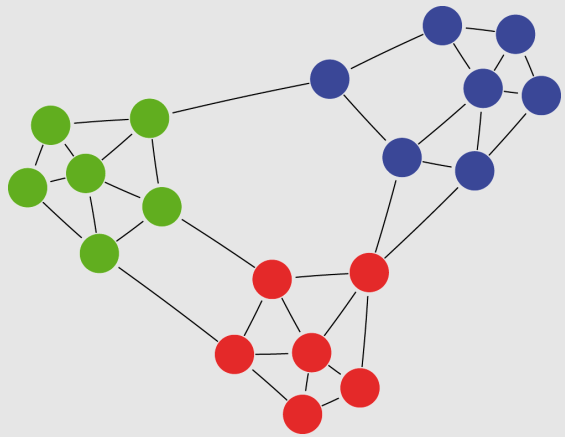
Milroy (1987)

Constructed a simple index representing integration into a given neighborhood

Modern Techniques

Community Detection

- Newman and Girvan (2004) aimed to maximize modularity



Social Network Analysis

Milroy (1987)

Constructed a simple index representing integration into a given neighborhood

Modern Techniques

Centrality measures

- Brin and Page (1998) created PageRank to sort search engine results

$$PR(A) = (1 - d) + d\left(\frac{PR(T1)}{C(T1)} + \ldots + \frac{PR(Tn)}{C(Tn)}\right)$$

Research Question

Do individuals use more or fewer orthographic variants of (lol) as their centrality in a given community goes up?

Data

Collection done for McNeill (2018)

1. Mined Twitter between January and February of 2017
2. Used Gephi to detect communities and calculate centralities
3. Reduced to 1,139 users in 13 communities, 3,938 tokens, 83 spelling variants

Coding

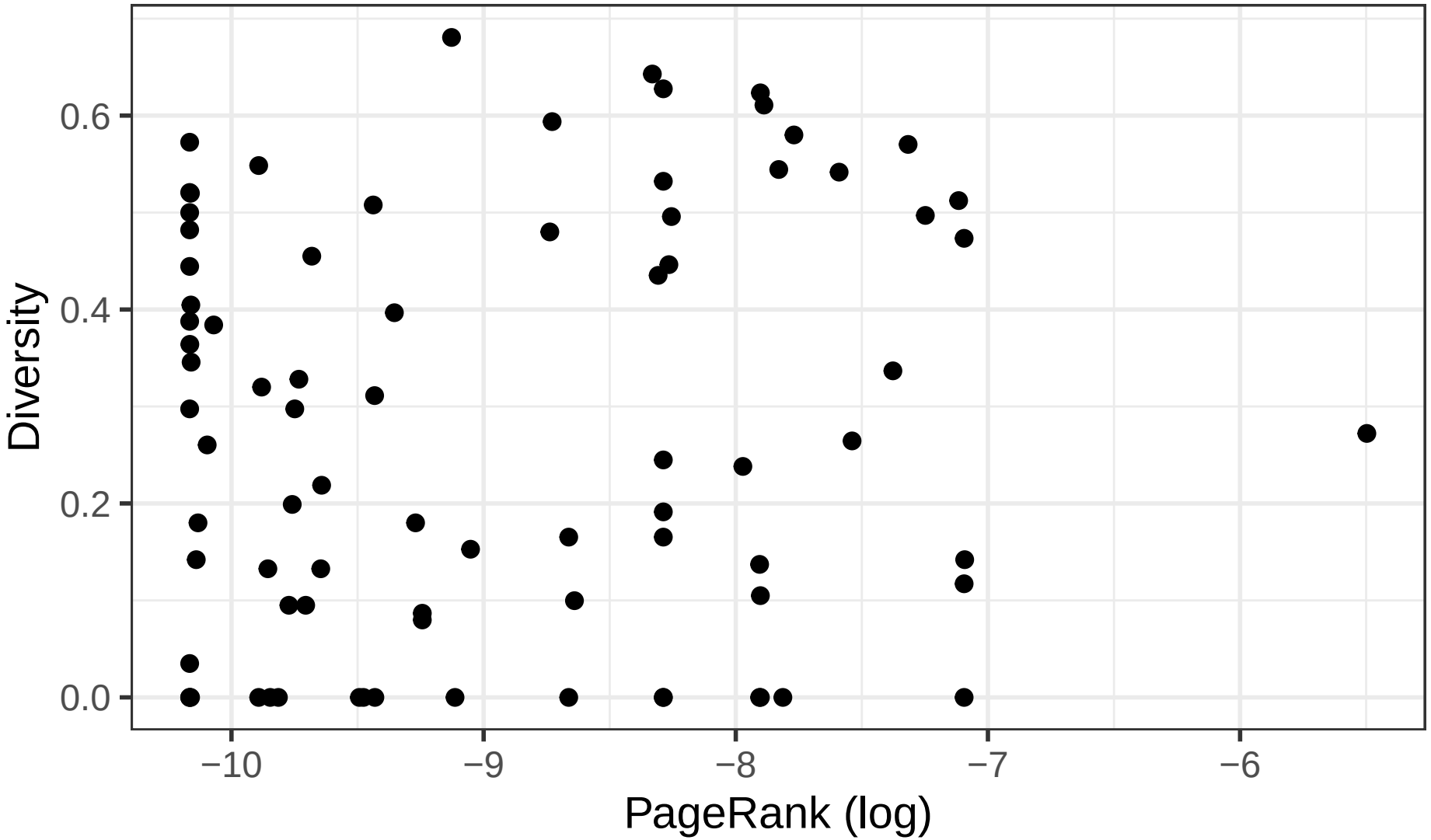
Community #, PageRank, and Simpson's diversity index:

$$D = 1 - \sum_{i=1}^{R} p_i^2$$

Summary of Communities

Community	Mode	Diversity	Members who used (lol)
173	lol	0.45	46
302	lol	0.42	316
572	lol	0.45	90
756	lol	0.41	25
799	lol	0	1
1032	lol	0.61	358
1097	lol	0.53	57
1227	lol	0.57	74
1291	lol	0.49	15
1917	lol	0.44	138
2067	lol	0	2
2265	LOL	0.71	8
6817	lol	0.52	10

Diversity by PageRank, ≥10 tokens



Conclusion

There is no relationship between diversity and PageRank for the spelling of (lol)

- Variation is the result of something else, perhaps pragmatics
- All communities have the same norm $\rightarrow$ Nothing to adapt to

Further Analysis

Discourse analysis of those with the highest and lowest PageRanks may be revealing

Data and code avaiable at https://osf.io/mgdpu/

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160. https://doi.org/10.1111/josl.12080

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

McNeill, J. (2018). *LOL sur Twitter: Une approche du contact de langues et de la variation par l'analyse des réseaux sociaux* (Master's thesis). Université du Québec à Montréal. Montreal, QC.

Milroy, L. (1987). *Language and Social Networks* (2nd) [original-date: 1980]. Blackwell.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks [arXiv: cond-mat/0308217]. *Physical Review E*, *69*(2), 1–16. https://doi.org/10.1103/PhysRevE.69.026113

Tatman, R. (2016). “I’m a spawts guay”: Comparing the Use of Sociophonetic Variables in Speech and Twitter. *University of Pennsylvania Working Papers in Linguistics*, *22*(2).