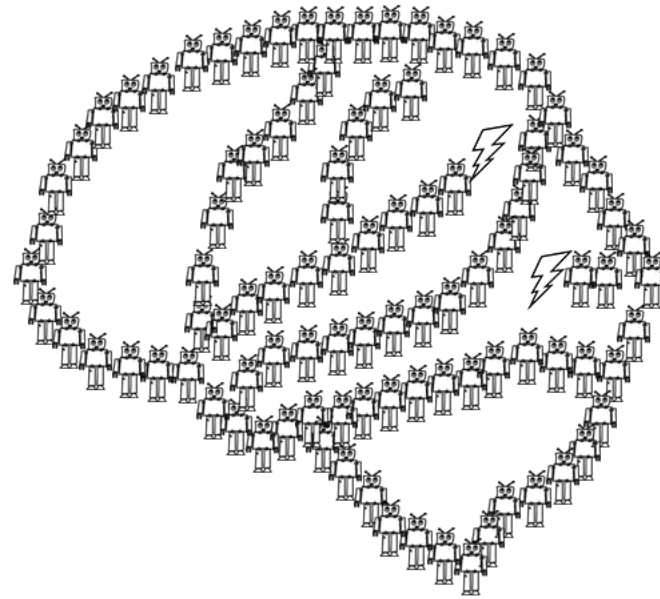


Virtual MathPsych/ICCM 2020

virtual.mathpsych.org

July 2020



How Does Modern Life Affect Memory Retrieval: Analyzing News Headlines

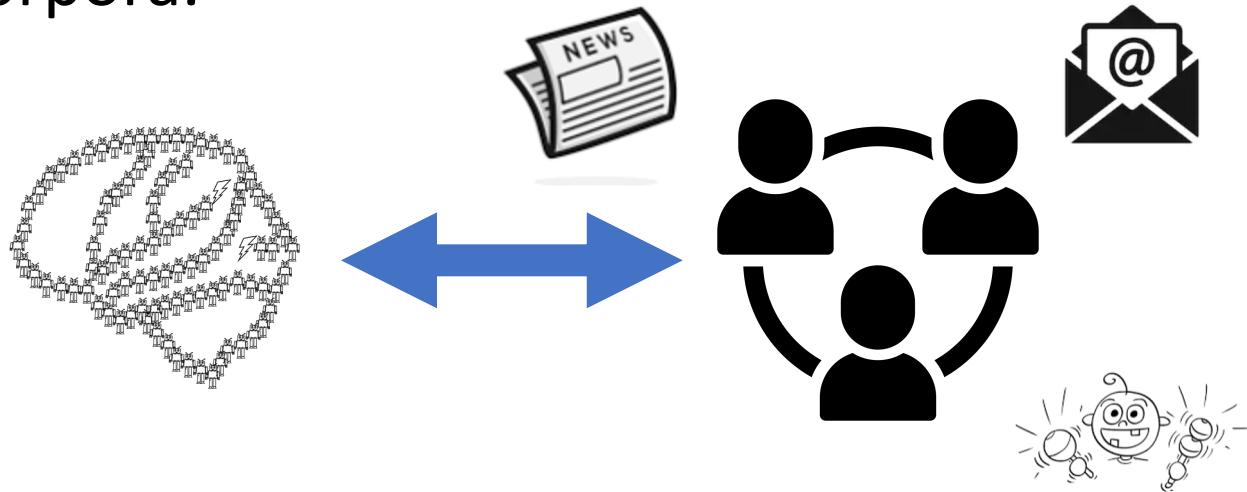
Cao, Kesong; Afrasiabi, Mohsen; Austerweil, Joseph L.

University of Wisconsin-Madison, United States of America



Introduction

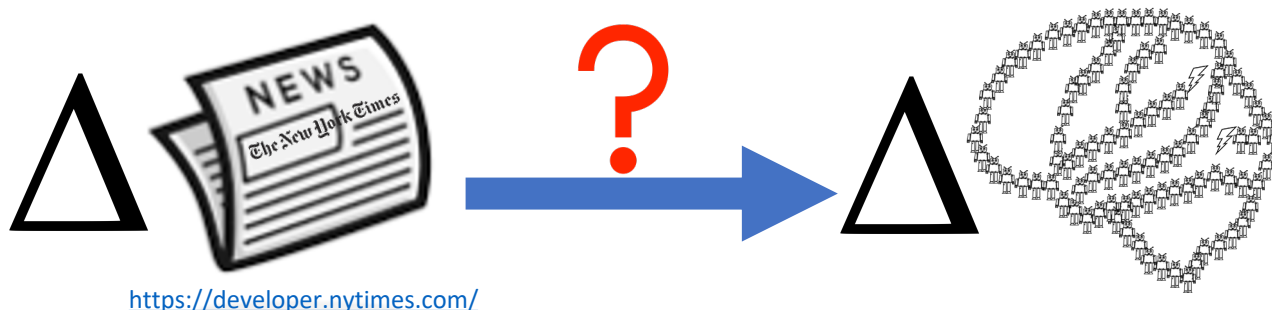
- Human memory adapts to the environment.
- Environmental sources reflect memory demands.
- Anderson & Schooler (1991) illustrated this using three corpora.



Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96(4), 703–719.
Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, 2(6), 396-408.

Introduction

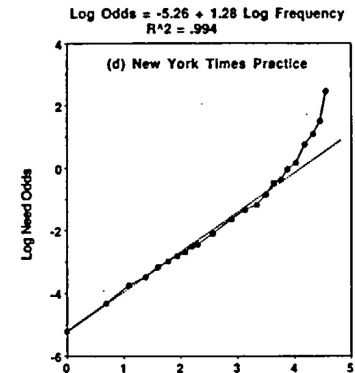
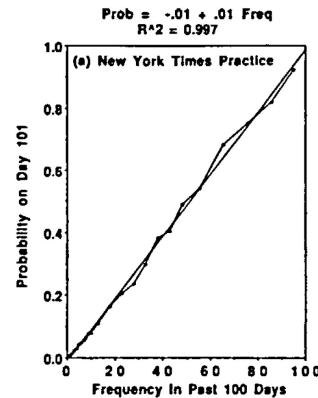
- **Changes** of environmental sources should reflect **changes** in memory demands.
- Modern technological advances are speeding up our pace of life.
- Are there changes in *New York Time* headlines?
- If yes, do those changes reflect on memory?



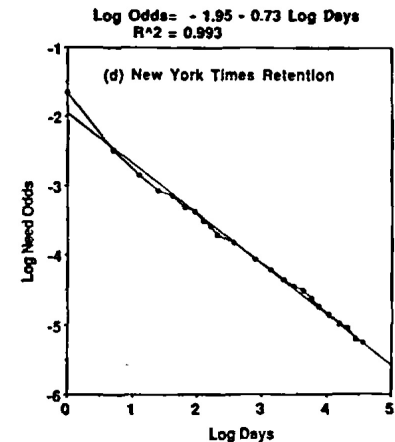
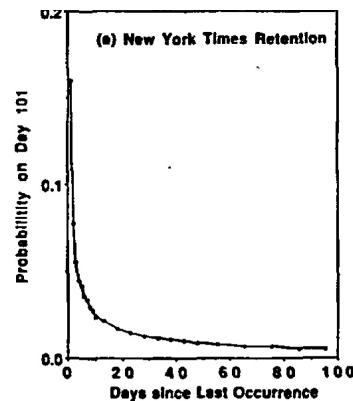
Previous work on *New York Times*

Anderson & Schooler (1991) looked at frequency, recency, and spacing of **words** in *NYT* headlines from 1986 to 1987.

Frequency: higher freq ->
higher prob recurring

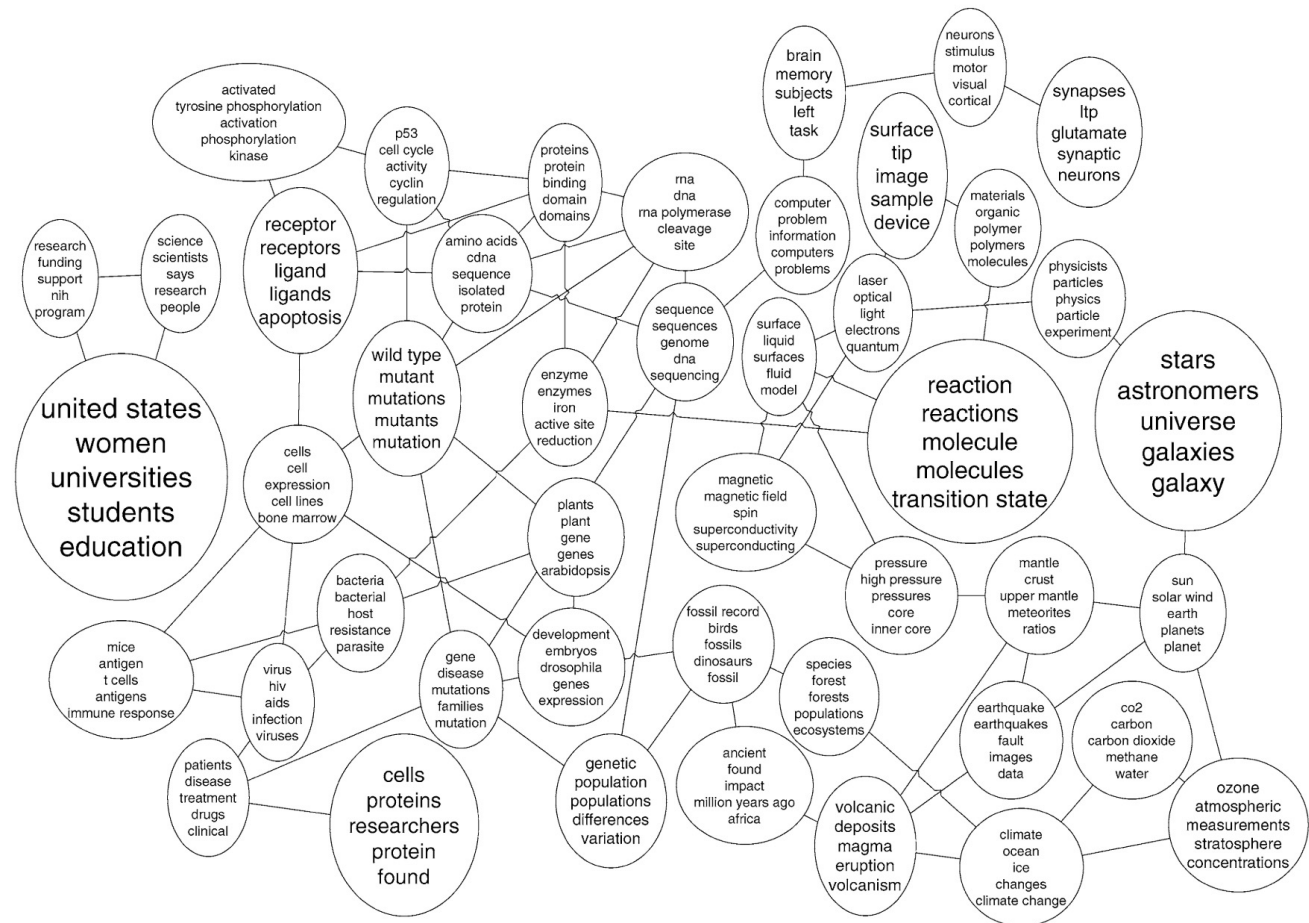


Recency: More recent ->
higher prob recurring



Work on time series analyses of text

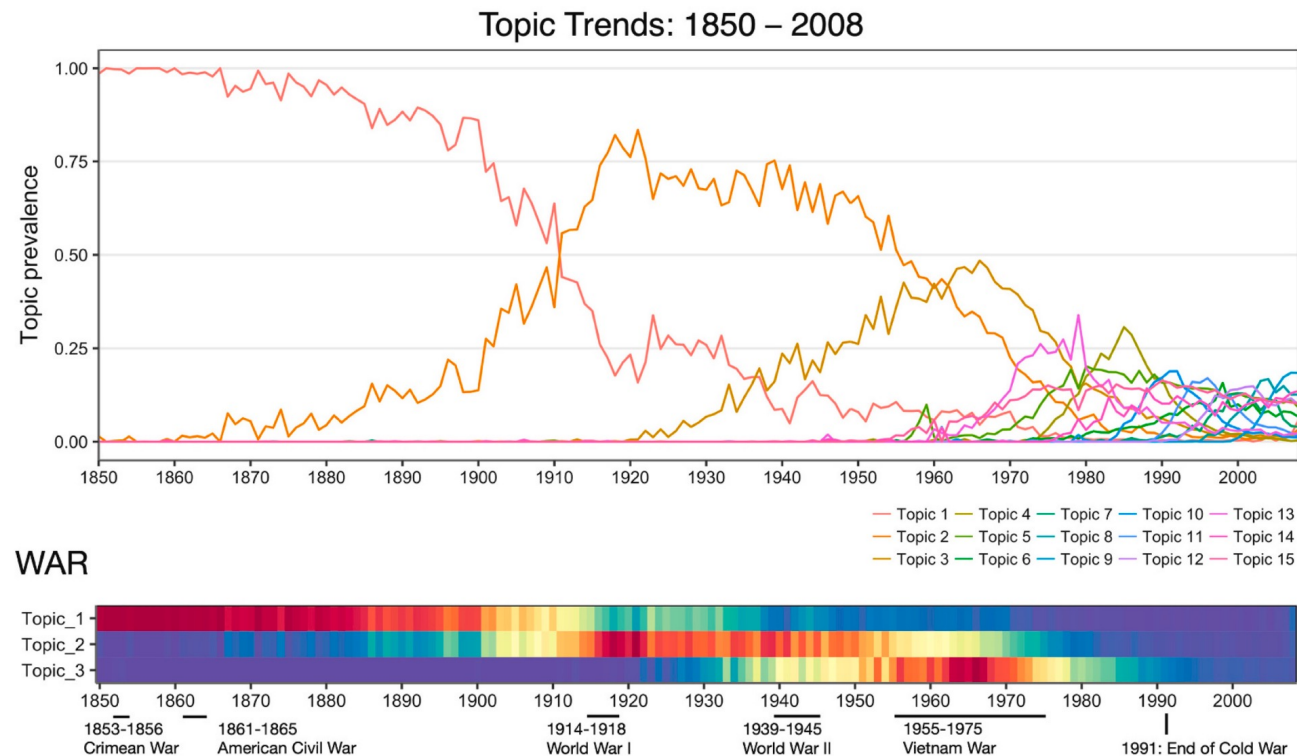
- Blei & Lafferty (2007) applied CTM topic modeling on *Science* articles from 1990 to 1999.



Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17-35.

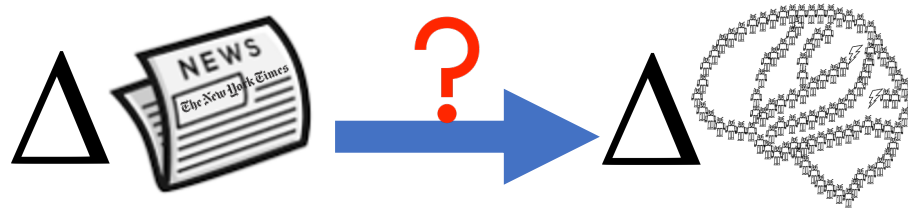
Work on time series analyses of text

- Li et al. (2020) decomposed the word “risk” using LDA topic modeling and performed content analyses using *NYT* and Google Books Ngram corpora.



Li, Y., Hills, T., & Hertwig, R. (2020). A brief history of risk. *Cognition*, 203(October 2019), 104344.

Question



Given:

- (Hypothesized) bi-directional relationship between human memory and the environment
- Accelerating speed of news (e.g. social media, notifications on electronic devices)
- *New York Times* headlines as a good proxy of event occurrences in the environment

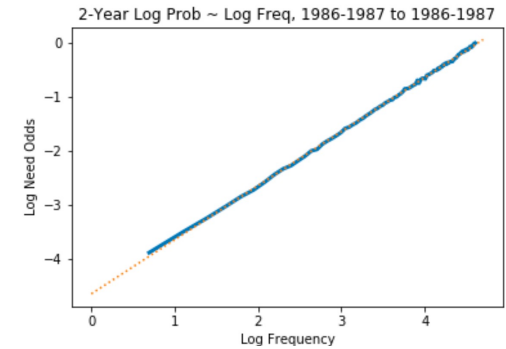
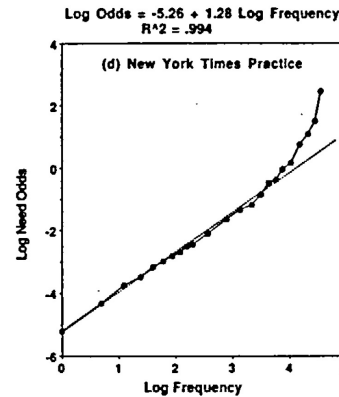
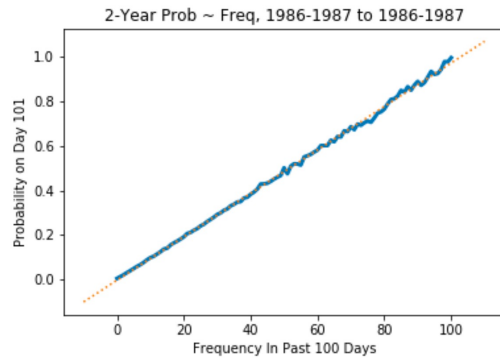
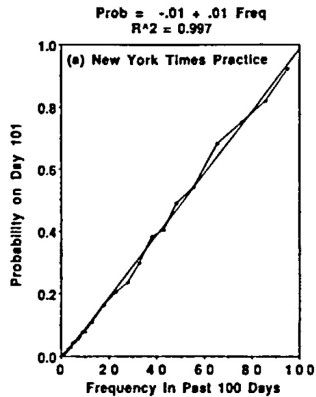
What are the dynamics of events as measured in *New York Times* headlines?

Methods

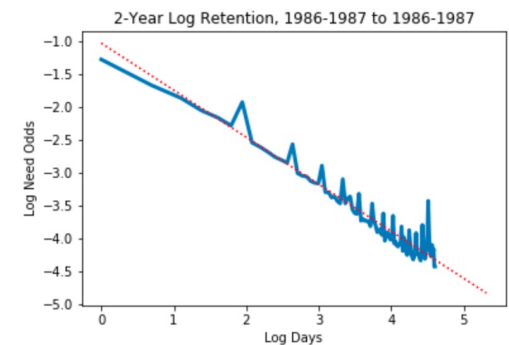
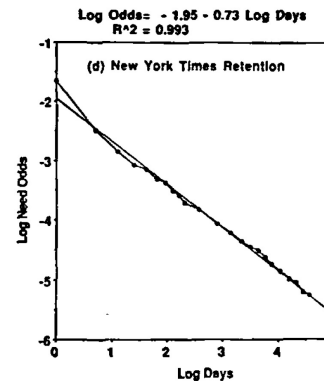
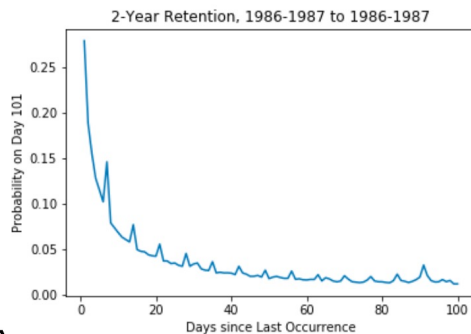
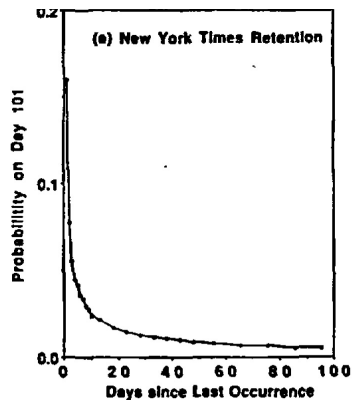
- Data source:
<https://developer.nytimes.com/docs/archive-product/1/overview>
- Processing steps
 - Download monthly archives of NYT from 1919 to 2019
 - Extract headlines with date information
 - Strip headlines into individual words (tokenization)
 - All lowercase, alphanumeric only, no punctuations
 - eXaMpLE123*^&! -> example123
 - y'all''''?.\$%@ -> y'all (keeping apostrophes)
 - Remove stopwords, as headlines are relatively short and therefore sensitive to high frequency stopwords
 - Ready for our computations! But first: verification

Previous work on *New York Times*

Frequency: higher freq -> higher prob recurring



Recency: More recent -> higher prob recurring



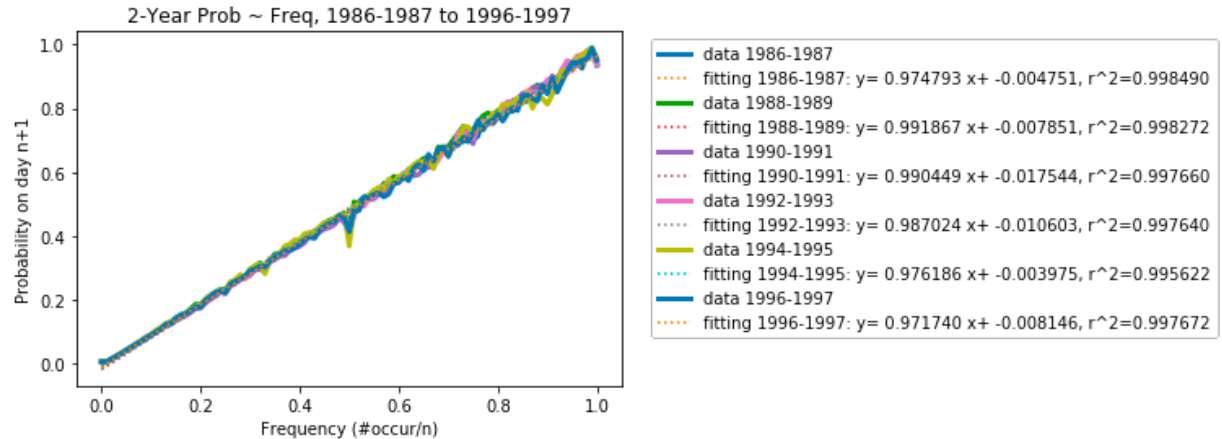
Methods Cont'd

- Frequency and recency analysis:
 - Repeat for each year from 1919 to 2019
 - Calculate prob of occurrences \sim freq / recency
 - Fit data into a linear regression
 - Store the parameters
 - Plot linear regression parameters
- Top words:
 - Sort words by number of occurrences
 - Store top 100 words per year for content analysis

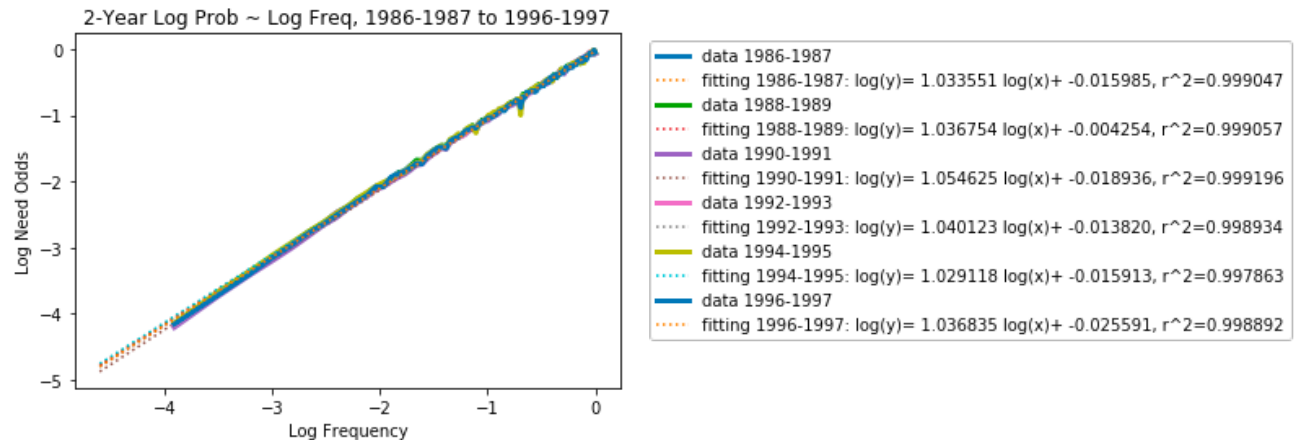
Results: frequency

Neighboring years seems very similar at first glance

- normal scale

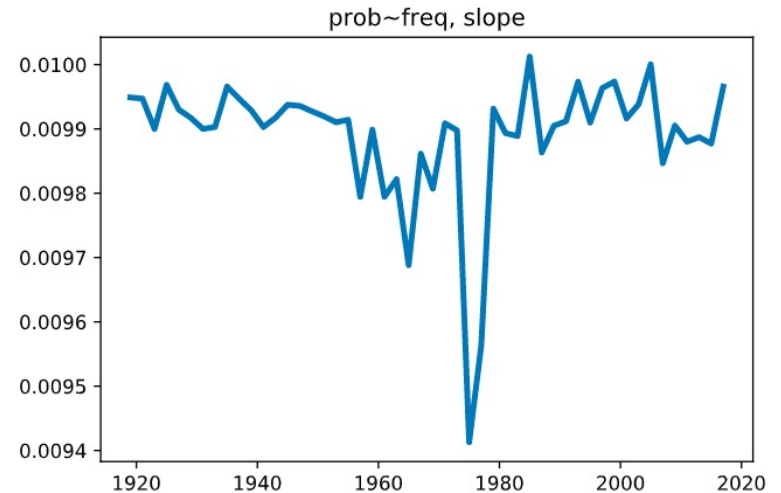
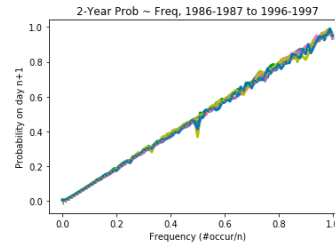


- log-log scale

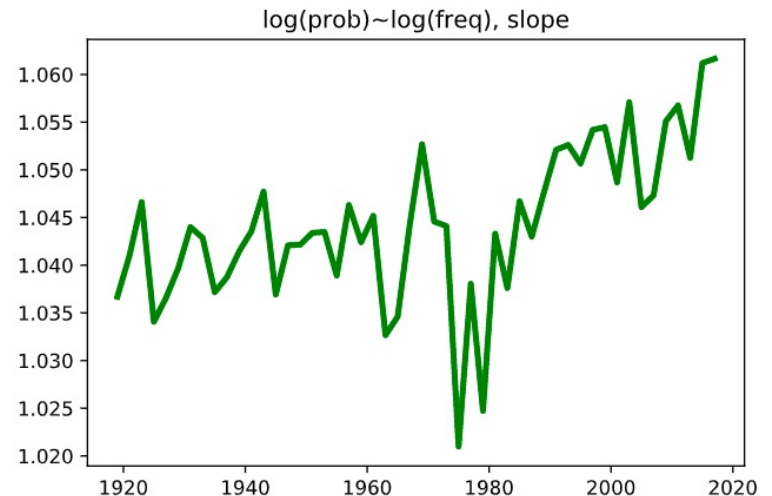
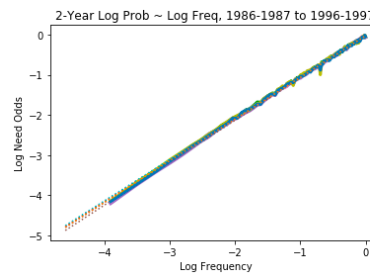


Results: frequency

For the normal scale, its slope remains fairly constant.

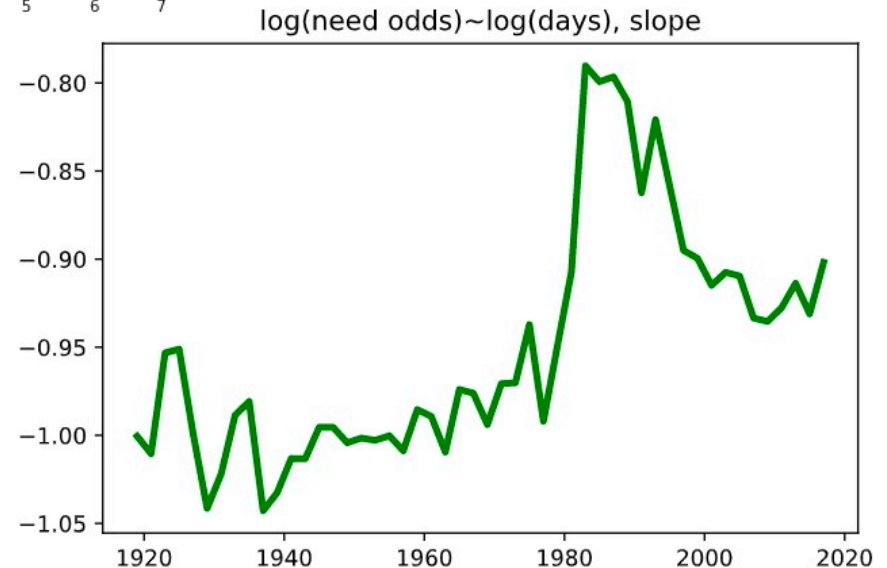
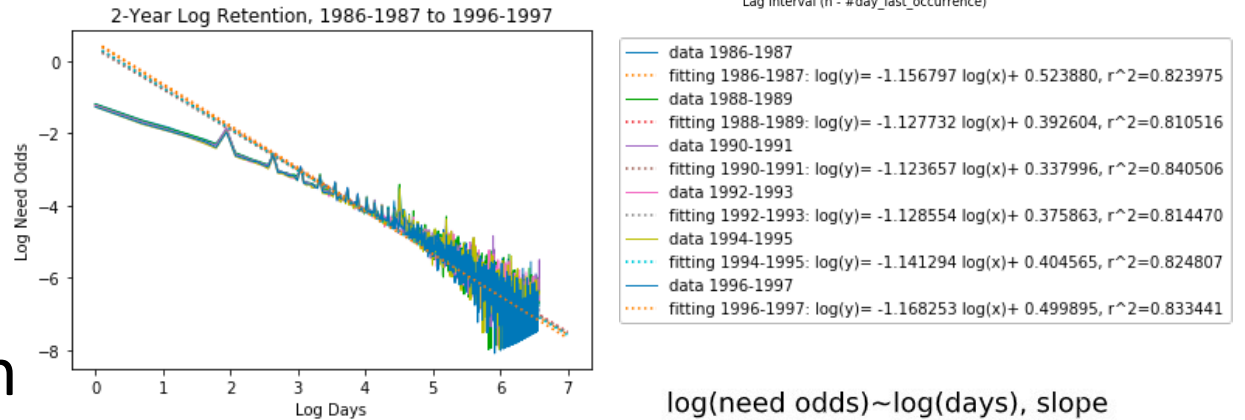
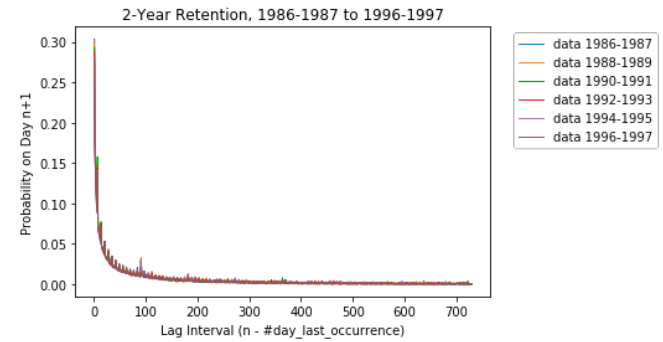


But for the log-log scale, its slope is slowly increasing over the years!



Results: recency

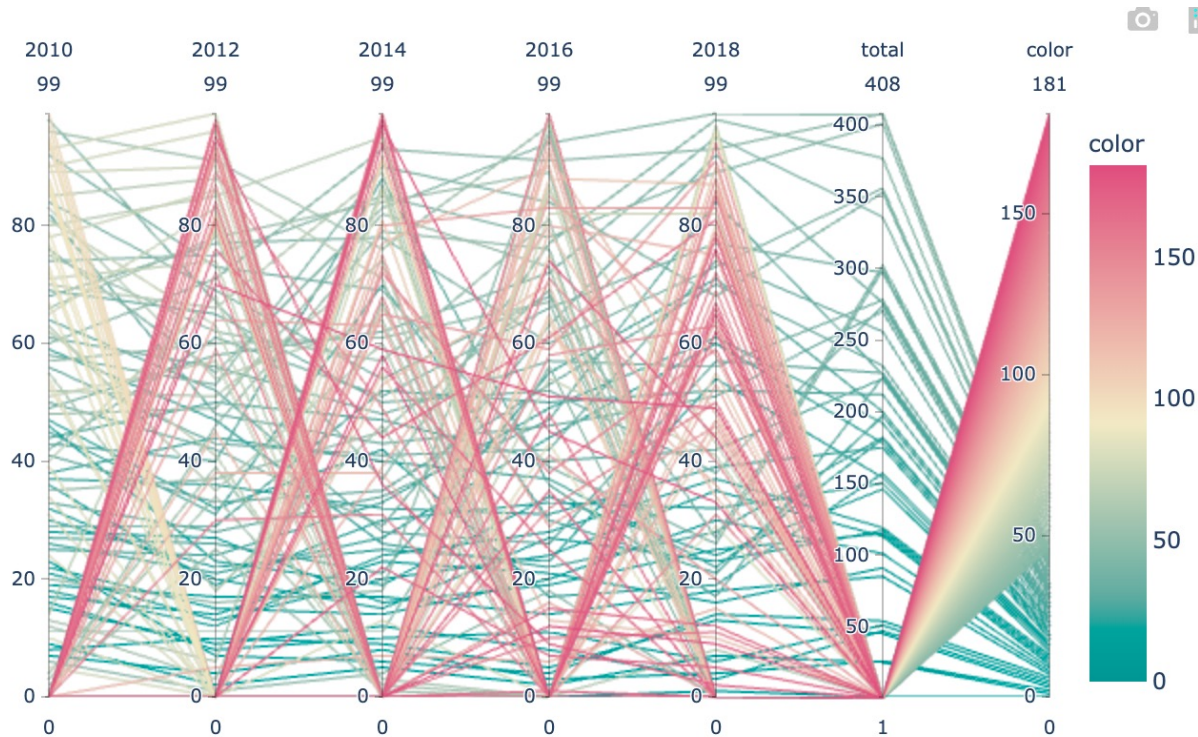
- News altogether is not faster.
- But faster news is much faster.
- The increasing slope in log-log recency plot also confirms this.



Results: top words

- Example: top words in 2018

'trump' 'new' 'what's' 'will' 'us'
'tv' 'says' 'dies' 'day' 'the'
'briefing' 'york' 'one' 'now' 'may'
'week' 'it's' 'get' 'house' 'world'
'china' 'first' 'say' 'war' 'police'
'dead' 'city' 'listen' 'time' 'women'
'best' 'back' 'california' 'big'
'american' 'life' 'like' 'make' 'man'
'just' 'white' '2019' 'evening'
'democrats' 'black' 'years' 'north'
'court' 'still' 'home' '5' 'million'
'art' 'don't' 'president' 'case' 'two'
'trade' 'love' 'dealbook' 'weekend'
'right' 'see' 'gets' 'know' '2' 'deal'
'take' 'health' 'de' 'want' 'takes'
'shows' 'people' '2018' 'south' 'go'
'fight' 'news' 'power' 'here's'
'behind' 'homes' 'help' 'america'
'john' 'state' 'way' 'judge' 'plan'
'good' 'another' 'death' 'end'



Future Work

- Memory modeling on the spacing effect
 - Will follow Walsh et al. (2018) for model comparison
 - We are now fitting the models including a Hawkes process as a theoretically motivated possible unification.
- Topic modeling
 - Will start with LDA as in Li et al. (2020), see slide 6
- More data: GitHub, Twitter, citations...

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., Krusmark, M., Myung, J. I., Pitt, M. A., & Zhou, R. (2018). Mechanisms underlying the spacing effect in learning: A comparison of three computational models. *Journal of Experimental Psychology: General*, 147(9), 1325.

Thank you!

- Recap: faster news are now much faster.
- Play with a subset of NYT data here:
<https://bit.ly/demo-nyt>
- Questions and suggestions are greatly appreciated!

This project is funded by the University of Wisconsin-Madison Letters & Science Honors Program through a Summer Senior Thesis Research Grant awarded by the L&S Honors Program.



L&S HONORS PROGRAM
College of Letters & Science



The Austerweil Lab thanks its previous and current funders.

