

# Measuring and Explaining Political Sophistication through Textual Complexity

Kenneth Benoit   Kevin Munger   Arthur Spirling

American Journal of Political Science, 2019

June 10, 2025

# Overview

- 1 Motivation
- 2 Limitations of Traditional Indices
- 3 New Approach
- 4 Validation
- 5 Application: State of the Union
- 6 Contributions
- 7 Implications
- 8 LLMs?
- 9 Discussion

# Why Measure Textual Sophistication?

- Political language appears to have grown simpler over two centuries, yet evidence is mixed.
- Common readability indices (e.g., Flesch Reading Ease) were built for school texts, not political speech.
- Goal: provide a domain-specific, statistically grounded measure of complexity with interpretable uncertainty.

# Problems with Flesch and Friends

- Rely almost exclusively on sentence length and syllable counts.
- Ignore parts-of-speech distribution, rhetorical devices, and political jargon.
- Scores lack probabilistic meaning and fixed coefficients date to the 1940s.

# Crowdsourced Pairwise Comparisons

- 20k+ Mechanical Turk comparisons of 25-word State of the Union (SOTU) snippets.
- Task: identify which snippet is easier to understand.
- Outcomes scaled with a Bradley–Terry model to infer latent “ease” parameters.

# Feature Engineering

- Parts of speech ratios (nouns, verbs, adjectives, etc.).
- Syntactic depth indicators (avg. dependency tree height).
- Dynamic word rarity from Google Books (1800–2000).
- Classic length syllable metrics retained for comparability.

# Model Estimation

- Logistic regression links features to pairwise outcomes.
- $\theta = X\beta$  estimated with bias-reduction; bootstrapped uncertainty.
- Rescaled 0–100 for interpretability (100 = 5th-grade baseline).

## Correlation with FRE (Figure 1)

- New measure correlates  $r \approx 0.70$  with Flesch Reading Ease but diverges at the extremes.
- Provides sharper discrimination of political texts.



## Complexity over Time (Figure 2)

- Overall downward trend since 1800, steepest pre-1910.
- Postwar speeches remain easier than 19th-century addresses but seldom beat the 5th-grade baseline.
- Written vs. spoken paired addresses (1945–1980): spoken versions are 10–14% easier.

# Key Contributions

- 1 Domain-specific, crowd-informed sophistication scale.
- 2 Open-source software and dynamic rarity tables for replication.
- 3 Uncertainty estimates allow formal hypothesis testing.
- 4 Framework adaptable to other corpora and languages.

# Why It Matters

- Reassesses claims of “dumbing down” with richer evidence.
- Enables covariate analysis by author, era, or delivery medium.
- Offers methodological template for inductive definition of concepts in law, media, and comparative politics.

# LLMs

- Don't use them to replace complicated tasks
- Make tasks simple enough for a crowdfworker
- Double-code a percentage of tasks – validate LLMs on this task

Questions?