

Silent prosody and the Bolinger effect

Saahil Sundaresan and Arto Anttila

Stanford University

Annual Meeting on Phonology 2026

University of Maryland

October 9, 2026

Words in English sentences show varying degrees of prominence

WE HAVE A PLACE, ALL OF US
IN A LONG STORY, A STORY WE CONTINUE
BUT WHOSE END WE WILL NOT SEE

✗ robotic

We HAVE a PLACE, **ALL** of us,
IN a **LONG** STORY,
a STORY we **CONTINUE**
but WHOSE END we WILL NOT SEE

✓ natural

– *George W. Bush, first inaugural address (2001)*

What explains English sentential prominence?

- 1. **Mechanical view:** Prominence is stress assigned by syntax.

(Chomsky et al. 1956; Chomsky and Halle 1968; Liberman and Prince 1977; Cinque 1993)

- ▶ Phrasal stress can be computed algorithmically on syntactically parsed text.
- ▶ Assumes lexical stress: content words are stressed; function words are not.

- 2. **Meaningful view:** Prominence highlights information. (Bolinger 1972)

I have a **POINT** to make

I have a **POINT** to **EMPHASIZE**

Unpredictable (informative) words are **accented**.

Predictable (uninformative) words are **unaccented**.

- For empirical evaluation, we need to quantify these notions.

Computing mechanical stress

- Nuclear Stress Rule (= NSR) and Compound Stress Rule (= CSR) (Chomsky and Halle 1968, 18) apply cyclically to syntactic trees from the innermost constituent outward.
- In SPE, 1 = primary stress, 2 = secondary stress, etc. A smaller number means 'more prominent' ($1 > 2 > \dots$).

[[John's _D]	[[[black _A	[board _N]] _N	[eraser _N]] _{NP}	[was	stolen] _{VP}	
1	1	1	1	1	1	init
	[1	2]				CSR
	[1	3	2]			CSR
[2	1	4	3]			NSR
[3	2	5	4		1]	NSR
John's	black	board	eraser	was	stolen	

Quantifying information: Shannon information theory

“Information is the resolution of uncertainty”

- **Shannon (1948):** Information is a probabilistic measure (a number) that can be calculated from the text. It tells you how much a word reduces your uncertainty (entropy) about the world; unexpected messages convey more information.

I am going to the ...

OFFICE.

MUNICIPAL ADMINISTRATION BUILDING.

Quantifying information (Shannon 1948)

- **Frequency:** how often a word occurs, overall
 - ▶ $P(\text{States})$ overall in English text is not very high
- **Predictability:** how likely a word is *in a given context*
 - ▶ $P(\text{States} \mid \text{We the People of the United})$ is very high
- **Surprisal:** Negative log-transform of predictability; $-\log P(w|c)$
 - ▶ property of a word **token**: a particular instance in a particular context
- **Informativity:** average surprisal of w over all contexts; $\mathbb{E}_{c \sim P(c|w)} [-\log P(w|c)]$
 - ▶ property of a word **type**: one value per word, held in “cold storage”
 - ▶ $\approx$ “how much entropy does this word tend to reduce, on average”
 - ▶ commonly bigram informativity (e.g., Cohen Priva 2015, Jo et al. in progress)

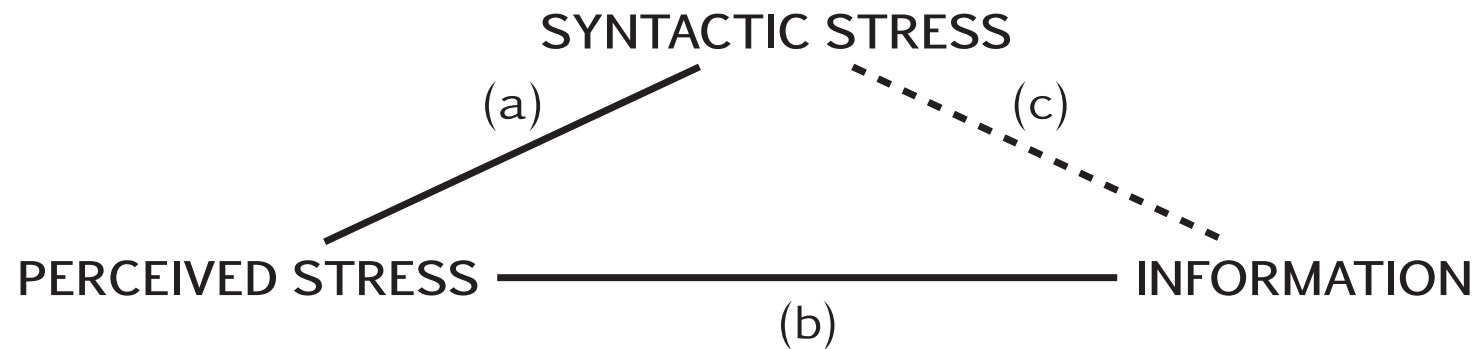
Both syntax and information contribute to perceived stress

- **The syntactic hypothesis** predicts a positive correlation between mechanical (syntactic) stress and perceived prominence (Chomsky and Halle 1968).
- **The information hypothesis** predicts a positive correlation between informativity and perceived prominence (Bolinger 1972).
- Empirical evidence suggests that **both** independently contribute to perceived stress (Anttila et al. 2020; Clapp and Anttila 2021; Shapiro and Anttila 2021; Jo et al. in progress).
 - ▶ Human-annotated prominence patterns in presidential inaugural addresses.
 - ▶ Contributes to a view of language as optimizing for communicative efficiency.

Research Question: Do stress and information interact in writing?

- **Bolinger:** Clearly, perceived (spoken) stress depends on information. This is what orators pay attention to when they read a passage aloud.
- **Chomsky and Halle:** Clearly, perceived (spoken) stress is also influenced by the syntactic structure of the sentence.
- But beyond speech production and perception, people also have freedom in how they *write*. Does mechanical (syntactic) stress matter in writing?

Decomposing sentential prominence



- (a) ✓ Syntactic stress predicts perception (Chomsky and Halle 1968)
- (b) ✓ Information predicts perception (Bolinger 1972)
- (c) ? Is syntactic stress related to information?

We will focus on (c) today. Bolinger (1957) proposed an interesting hypothesis.

Bolinger on good writing

“The writer should ... maneuver the semantic heavy stress into the position of the mechanical loud stress.”

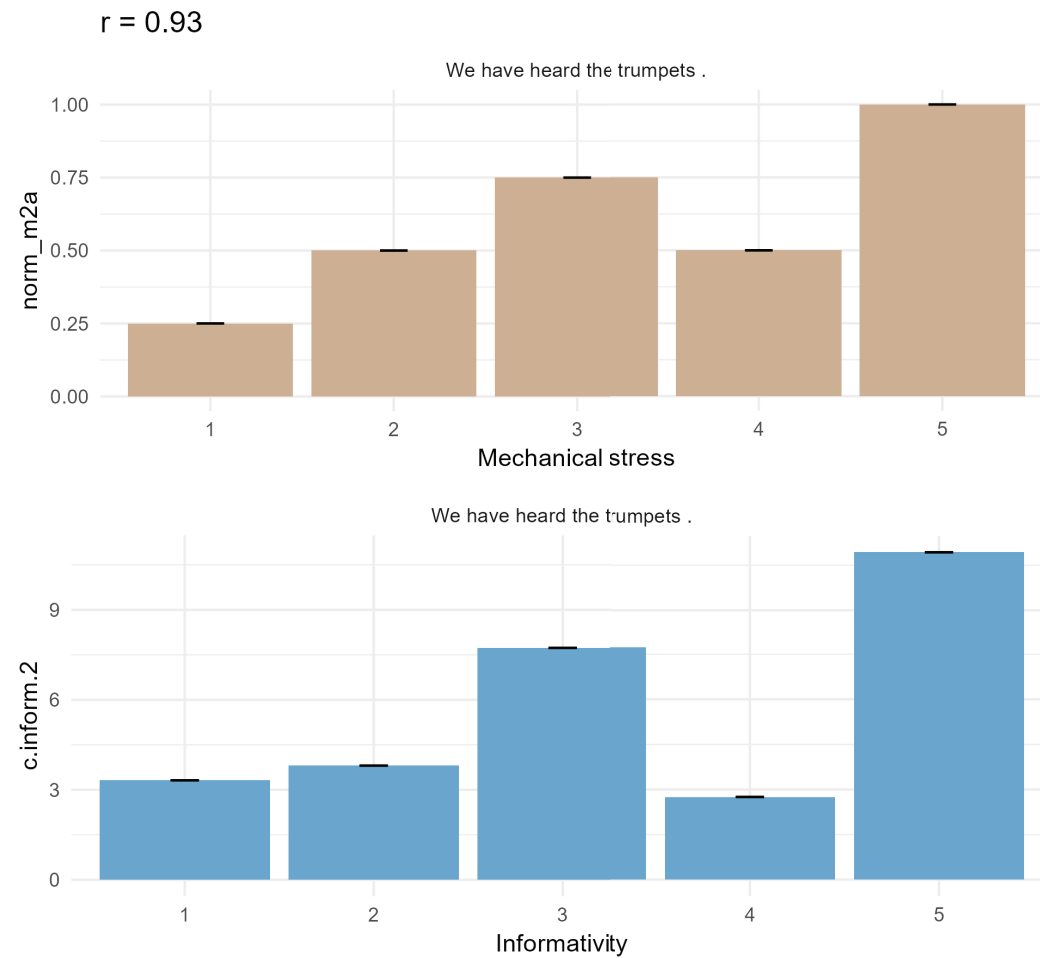
– Bolinger 1957: 235

- Good writers
 - ▶ are aware of an implicit stress contour
 - ▶ place informative words into syntactically stressed positions
- Cf. **silent** or **covert prosody** (Fodor 2002; Breen 2014; Frazier and Gibson 2015, a.o.), which we approximate using a version of the SPE algorithm.
- **Hypothesis:** In good prose, mechanical stress and information go together.
 - ▶ This is reasonable if you take an information-optimizing view of language.

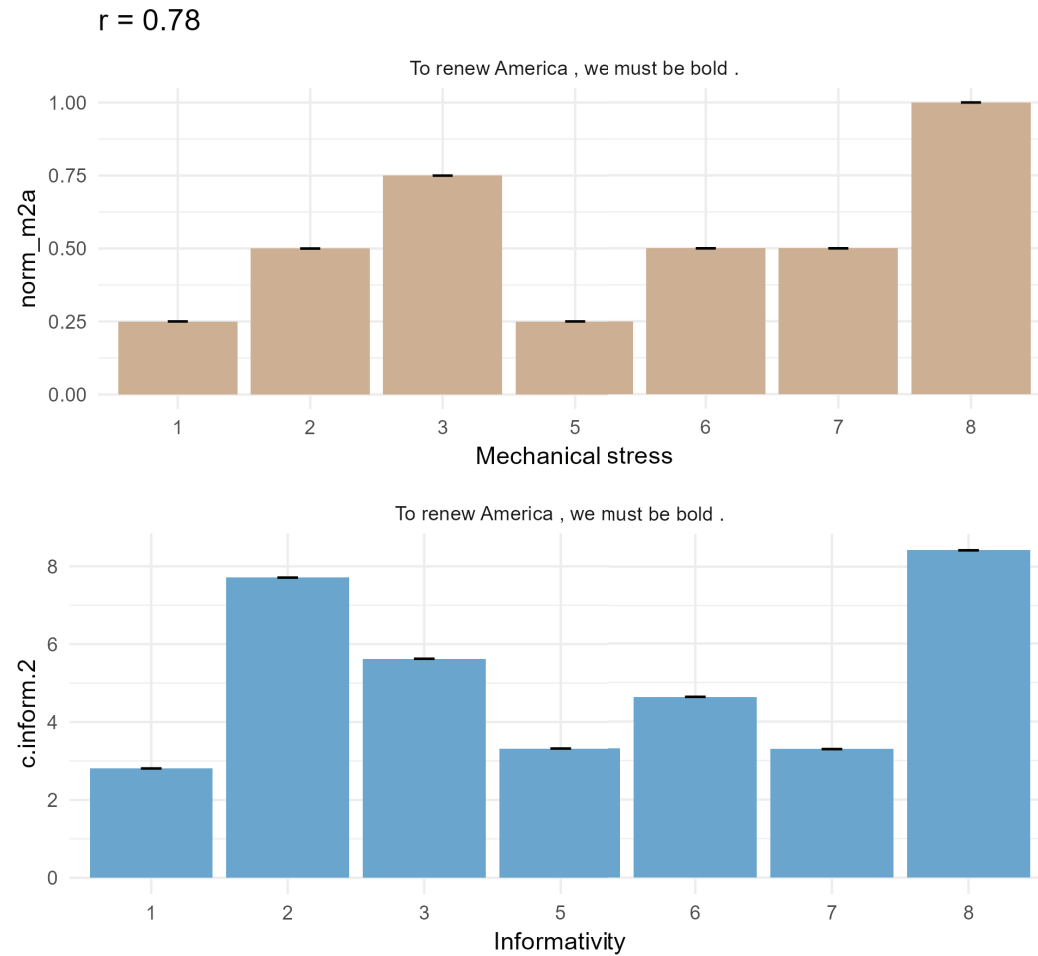
Illustrating stress-information harmony

- The following three sentences are from Clinton's (1993) inaugural.
 - ▶ "We have heard the trumpets" (virtually perfect alignment)
 - ▶ "To renew America, we must be bold" (somewhat worse)
 - ▶ "Technology is almost magical" (worst)
- Note that our "normalized" stress (norm_m2a) reverses the SPE scale.

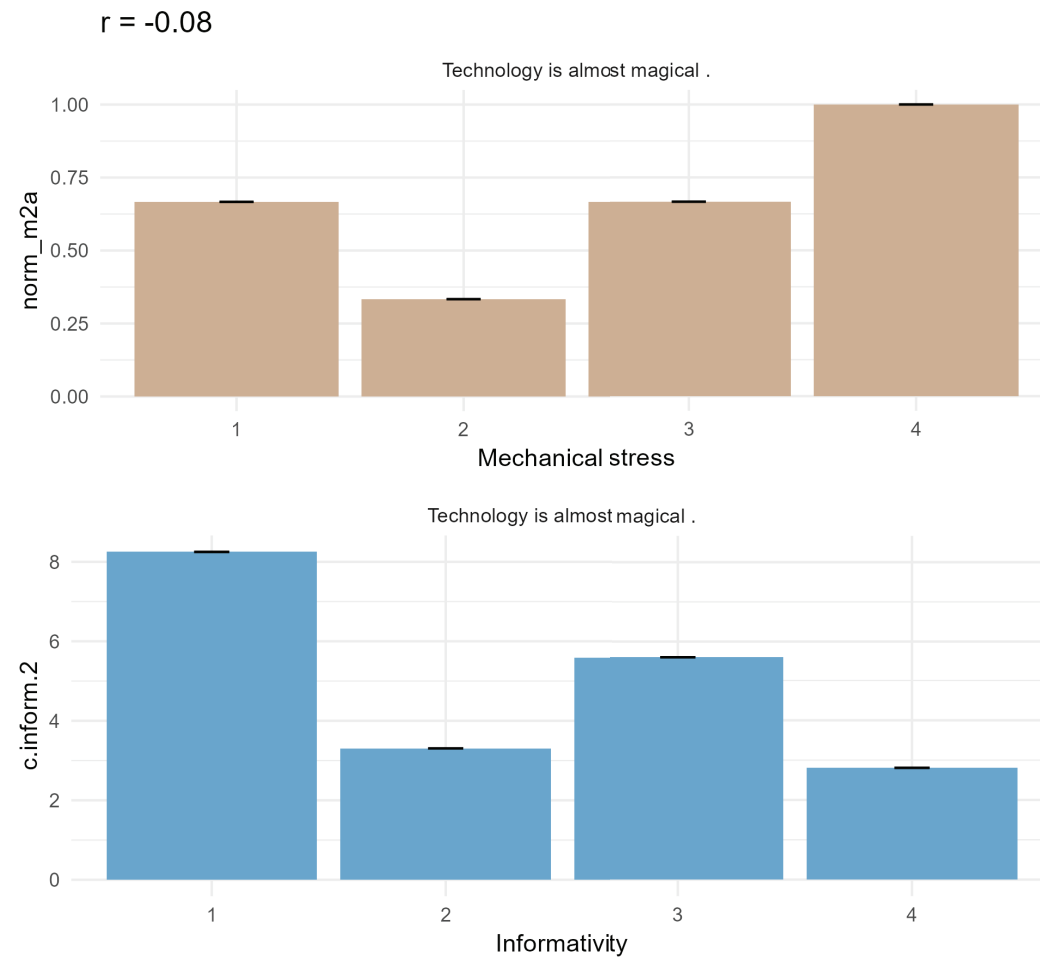
“We have heard the trumpets” (almost perfect alignment)



“To renew America, we must be bold” (imperfect alignment)



“Technology is almost magical” (poor alignment)



Our study

□ Corpora:

- ▶ Oratorical: Inaugurals of six recent U.S. presidents (Woolley and Peters 1999, curated by Anttila et al. 2020, $N = 15,531$ words)
- ▶ Non-oratorical: California Driver's Handbook (California DMV 2024, $N = 12,756$ words).

□ Syntactic stress:

- ▶ METRICALTREE (<https://metricaltree.stanford.edu/>) implements a version of the SPE-style Nuclear and Compound Stress Rules on the Stanford Parser.

□ Information content:

- ▶ Offline: Bigram informativity (e.g., Piantadosi et al. 2011, Cohen Priva 2015) as a context-invariant measure of information in cold storage.
- ▶ Online: Next-token probabilities from LLMs, as an on-the-fly, context-aware estimate of predictability.

Regression modeling

- Mixed effects linear regression using lme4; all predictors normalized (z-scored)

```
informativity ~ mechanical stress (mech) +  
               word position (wpos) +  
               number of segments (nseg) +  
               number of syllables (nsyll) +  
               part of speech (pos) +  
               mech stress (mech): POS +  
               (1|president) (only in the Presidents Corpus)
```

Results by part of speech

informativity $\sim$ mech + wpos + nseg + nsyll + pos + mech:pos
(+ (1|president) in Presidents)

Effect of mechanical stress on bigram informativity, b (SE)

	Presidents	DMV
	$N = 15,531$	$N = 12,756$
mech (adjectives, baseline)	0.15 (0.069)	0.30 (0.073)
<i>Mech $\times$ POS: difference from the adjective slope</i>		
Verbs	1.08*** (0.086)	1.03*** (0.085)
Nouns	0.10 (0.081)	0.11 (0.082)
Function words	-0.46*** (0.075)	0.09 (0.077)

- Only verbs have a significantly steeper slope than adjectives, in both corpora.
- Verbs are also significantly steeper than nouns (Presidents: +0.98; DMV: +0.92; both $p < 10^{-40}$).

Predictors z-scored. * $p < .05$, *** $p < .001$.

F-tests for nested models

For each part of speech separately, we compare:

full: informativity $\sim$ mech + wpos + nseg + nsyll

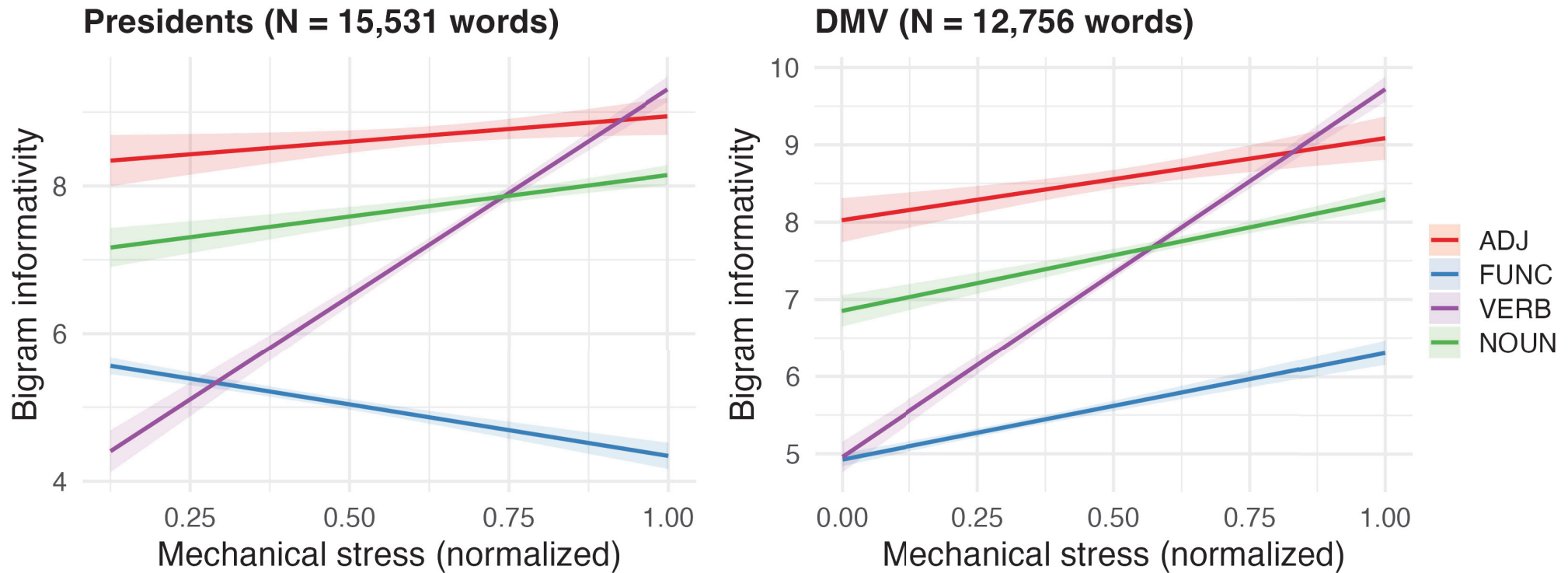
reduced: informativity $\sim$ wpos + nseg + nsyll

% of residual variance explained by adding mechanical stress

	Presidents		DMV	
	% var	F	% var	F
Verbs	9.21%	235.2	16.72%	488.6
Nouns	0.16%	5.2	0.88%	26.4
Adjectives	0.03%	0.3	1.00%	7.6
Function words	3.12%	201.8	0.24%	11.4

- Mechanical stress explains about 3× (Presidents) and 16× (DMV) more variance in verbs than in any other part of speech.
- The same results emerge when we look at LLM contextual predictabilities.

Model predictions



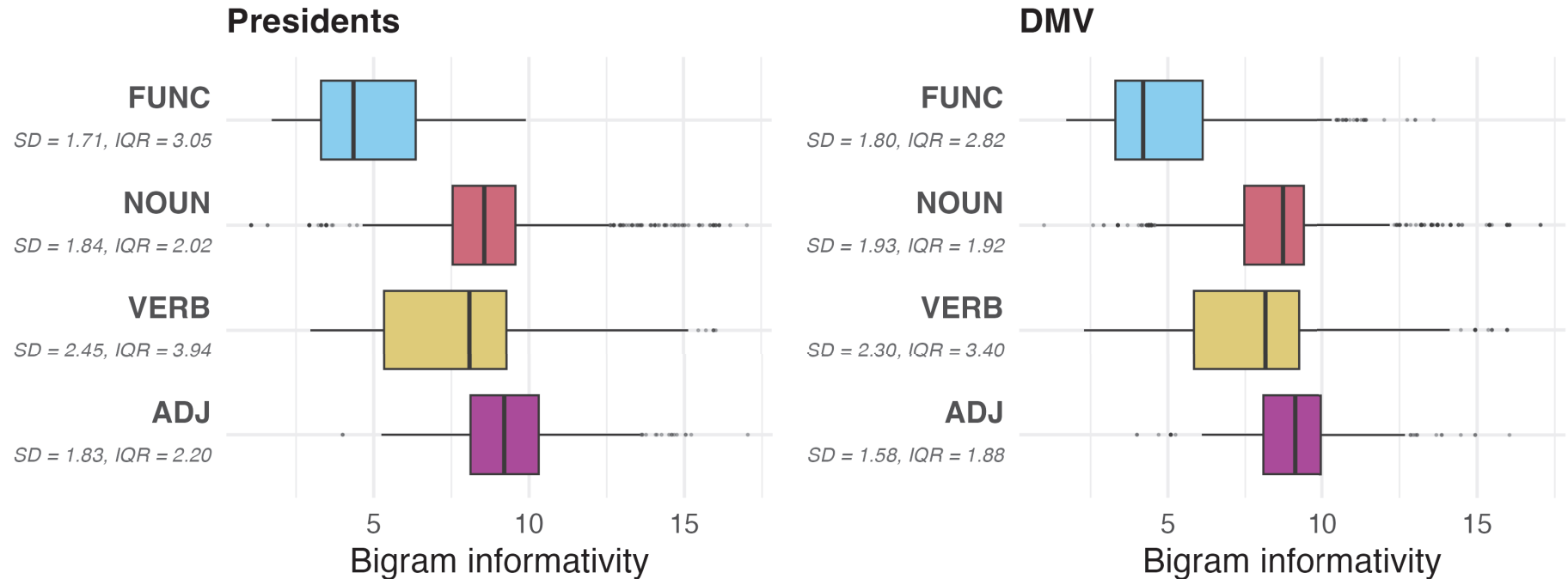
- Predicted bigram informativity $\sim$ mechanical stress, by part of speech (other predictors at their means).

The key results

- Two corpora: Presidents, DMV.
- Three different informativity measures:
 - ▶ bigram informativity from a global corpus (Wikipedia documents)
 - ▶ bigram informativity from a local corpus (just the presidents data)
 - ▶ LLM contextual predictability
- Across all five models, only two predictors were consistently and reliably positive with significance:
 - ▶ number of segments in the word (as expected)
 - ▶ **norm_mech:pos_catVERB** (our main finding)

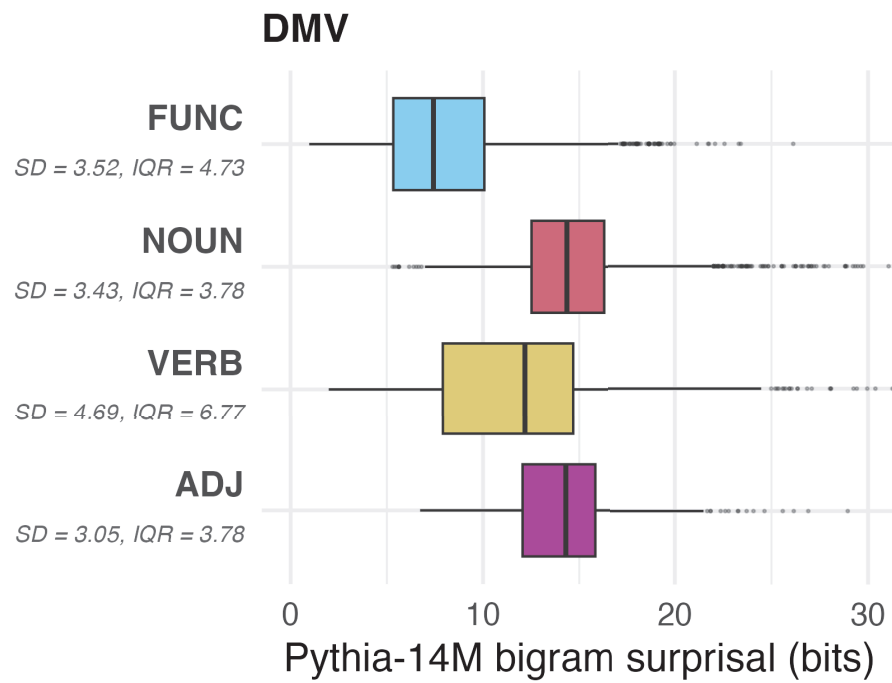
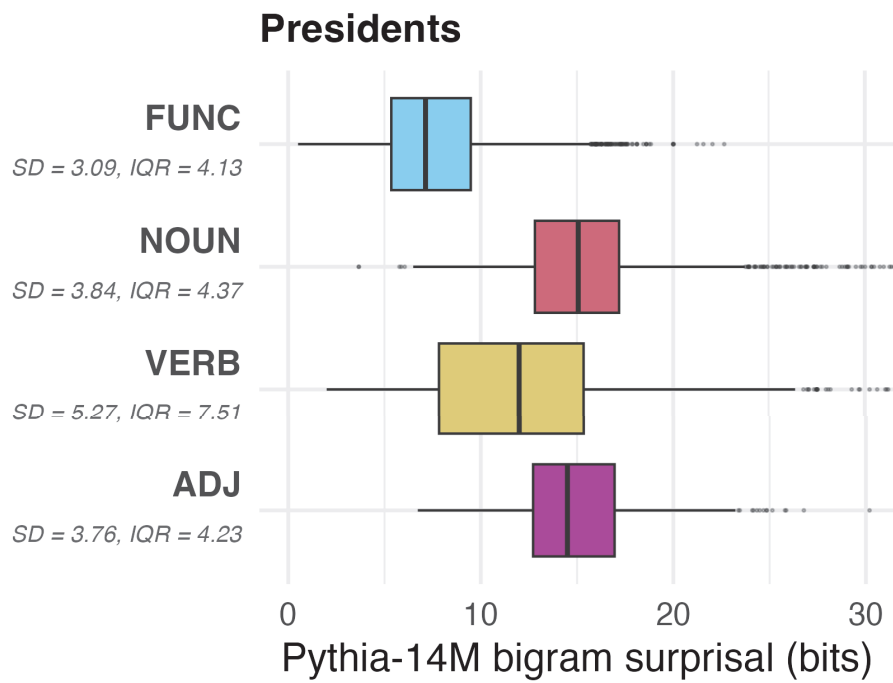
What makes verbs special?

Informativity distributions



- Verbs vary the most. Function words are mostly uninformative; nouns and adjectives are mostly informative.

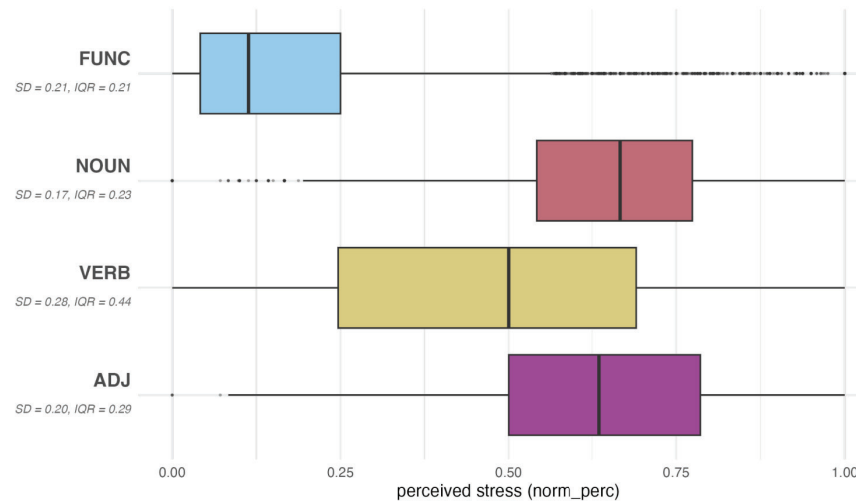
Contextual (LLM) surprisal distributions



The ceiling/floor effect

□ Here's an intuition:

- ▶ **Nominals:** High degree of perceived stress drowns informativity.
- ▶ **Function words:** Little perceived stress, no room for information.
- ▶ **Verbs:** Middle ground, hence the Bolinger effect can emerge.



The Prague School Hypothesis

- Functional Sentence Perspective (e.g., Firbas 1992): Given < New.
- **Prediction:** Information should be positively correlated with the word's linear position in the sentence.
- This turns out to be false once syntactic stress is included as a predictor in the model ($b = -0.067, SE = 0.018, p = 10^{-4}$).
- **Tentative conclusion:** Information is distributed in the sentence, not by linear position, but by phrasal stress.

Main takeaways

- **WRITING:** Good writers are
 - aware of implicit stress (Chomsky and Halle 1968)
 - place informative words into implicitly stressed positions
 - but the effect only emerges in verbs.
- **SPEECH:** Informativity
 - predicts perceived stress (Bolinger 1972)
 - but only in verbs and function words.
- **DISTRIBUTION OF INFORMATION:**
 - is sensitive to phrasal stress
 - not to linear position (not Given < New)

Work in progress

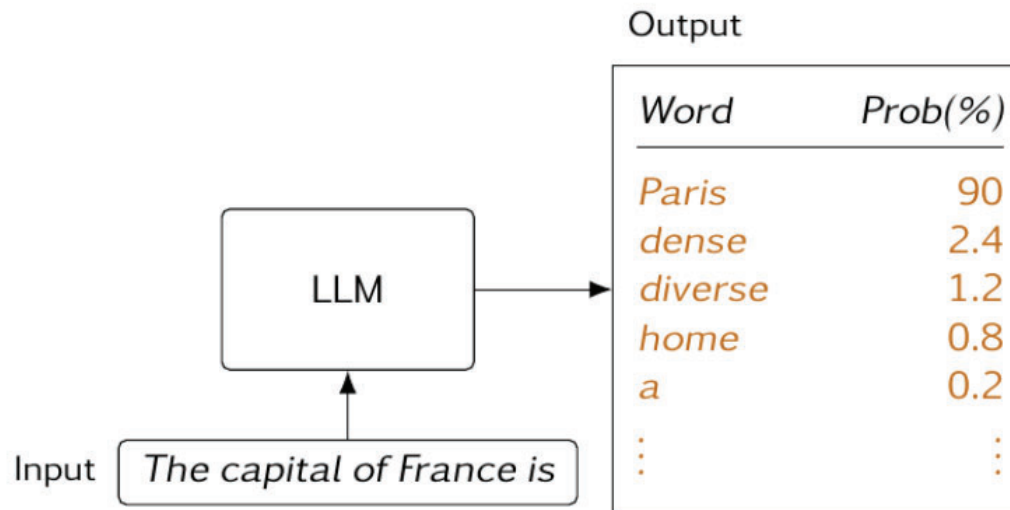
- Including two interactions:
 - ▶ mechanical stress : part of speech (already included)
 - ▶ mechanical stress : linear position of word in sentence (new)
- Predicting mechanical stress from informativity (turning the model upside down)
 - ▶ Preliminary models suggest the uniquely privileged position of verbs still holds in this condition.
- Adding more random effects (word, sentence)
- Separating auxiliaries from verbs (contra Penn Treebank)
 - ▶ Our results hold even if we exclude auxiliaries from the verb category.

Acknowledgements

- We thank Vivienne Fong, and the participants of LINGUIST 215 in Spring 2025 (Emiyare Ikwut-Ukwa, Oliver Lee, Jonathan Merchan, and Joseph Sarmenta) for helpful input. All errors are ours.
- Statistical analyses and visualization were done in R (R Core Team 2026; Bates et al. 2015), `ggplot2` (Wickham 2016), and `sjPlot` (Lüdecke 2025).

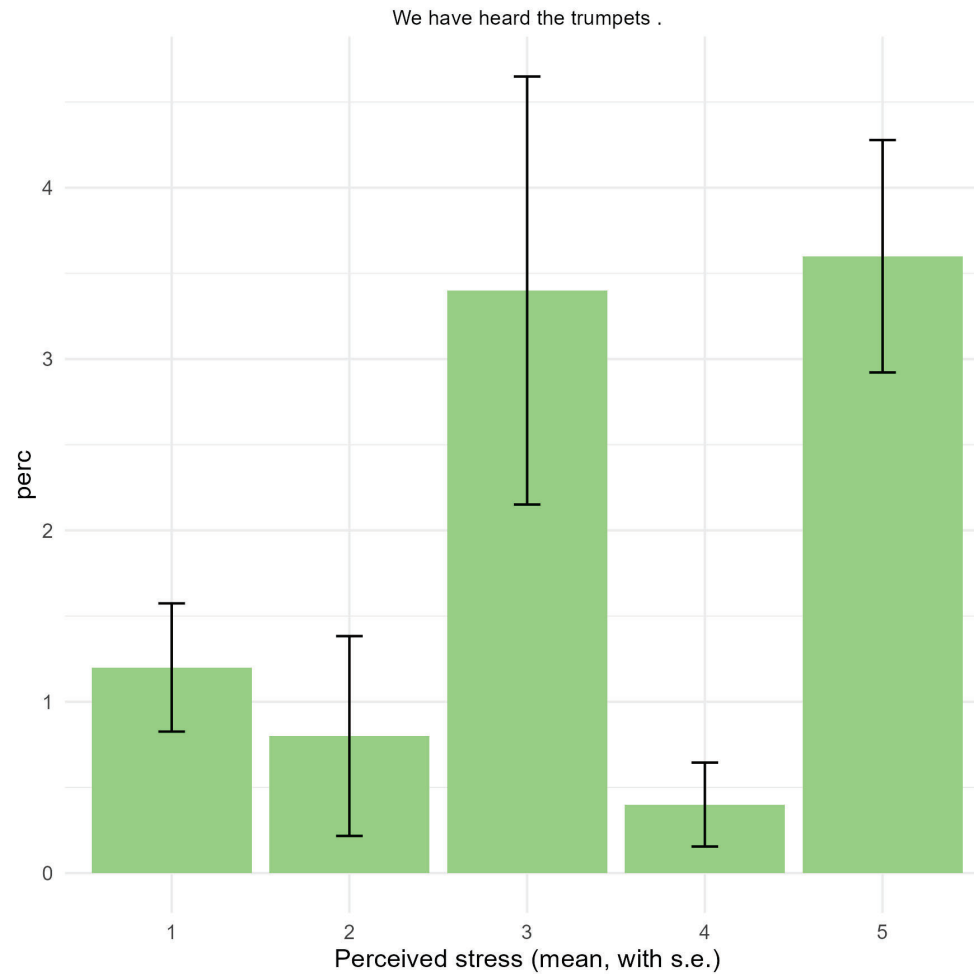
Thank you!

LMs output a probability distribution over next tokens

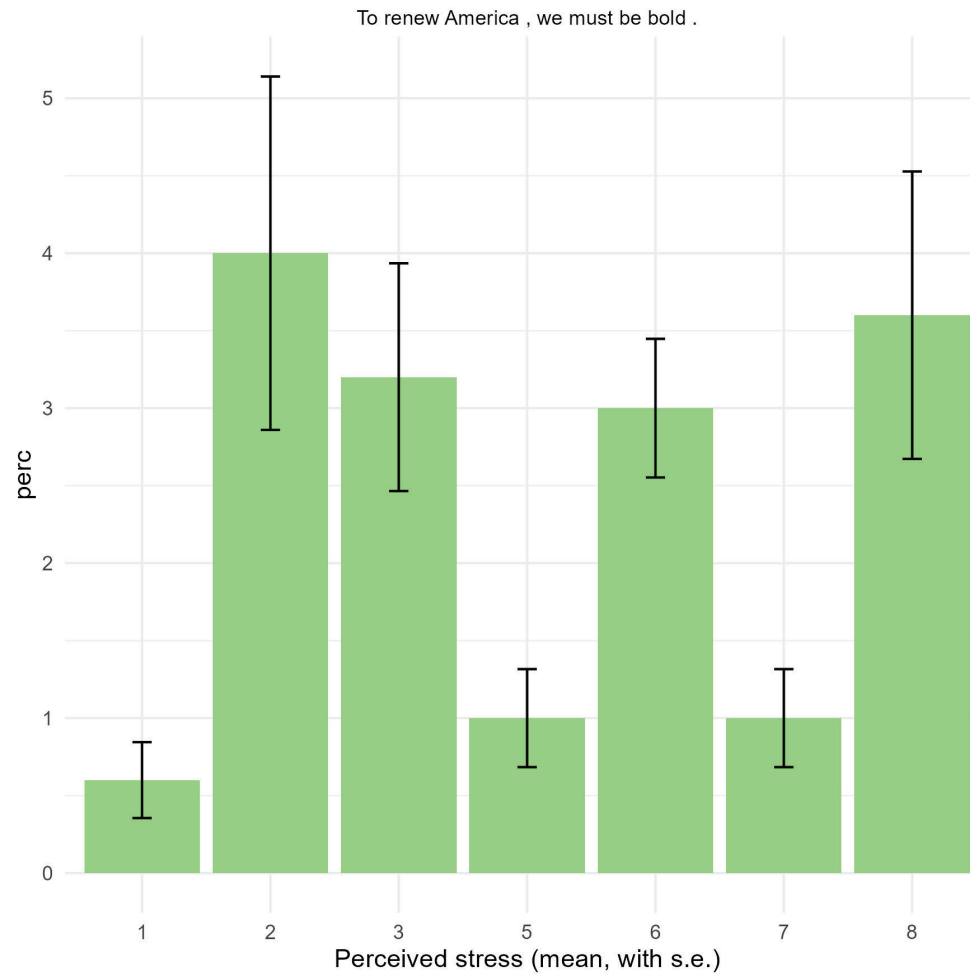


- We know the real next word for any n -gram context in our corpus. The probability a LM assigns to that true next word is, by definition, its contextual predictability.

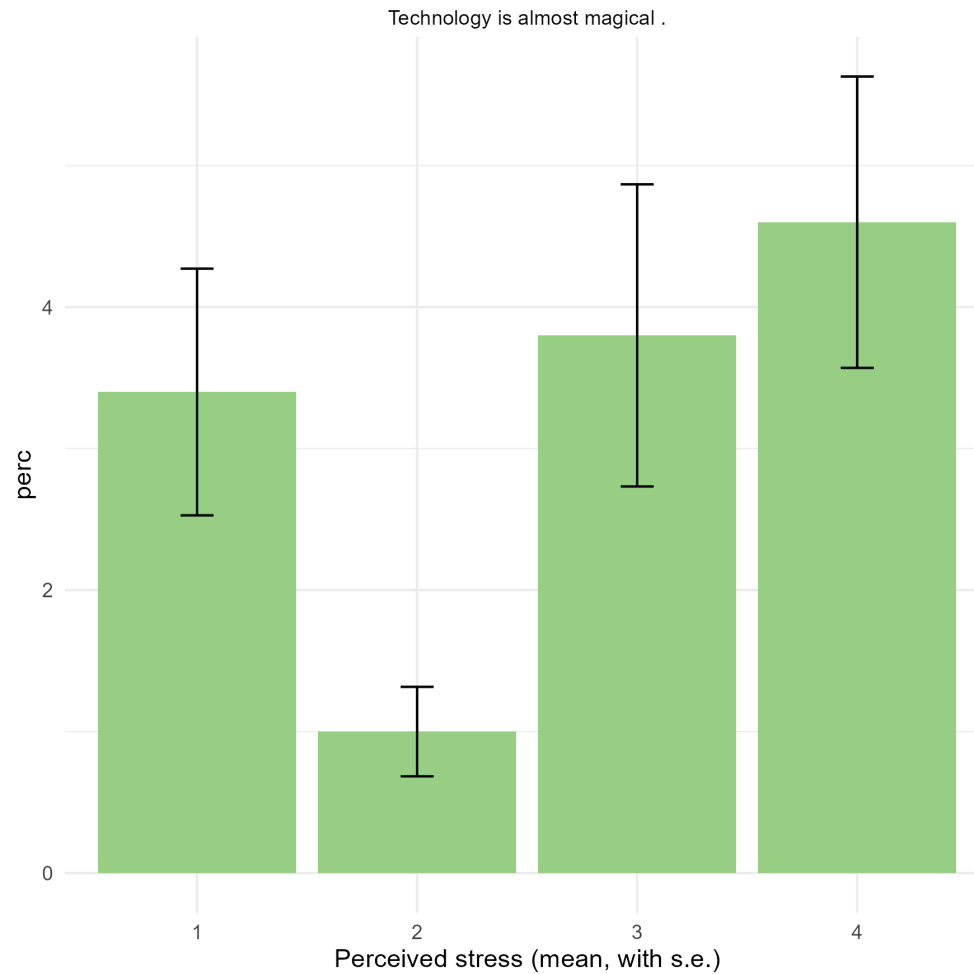
Perceived stress (N = 5): "We have heard the trumpets"



Perceived stress (N = 5): "To renew America, we must be bold"



Perceived stress (N = 5): "Technology is almost magical"



References I

- Anttila, Arto, Timothy Dozat, Daniel Galbraith, and Naomi Shapiro. 2020. Sentence stress in presidential speeches. In *Prosody in Syntactic Encoding*, ed. Gerrit Kentner and Joost Kremers, 17–50. Berlin/Boston: Walter De Gruyter. Pre-print available on lingbuzz/004303.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.
- Bolinger, Dwight L. 1957. Maneuvering for stress and intonation. *College Composition and Communication* 8:234–238.
- Bolinger, Dwight L. 1972. Accent is predictable (if you are a mind reader). *Language* 48:633–644.
- Breen, Mara. 2014. Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass* 8:37–50.
- California DMV. 2024. California driver's handbook. URL <https://www.dmv.ca.gov/portal/handbook/california-driver-handbook/>, sacramento, California.
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam, Morris Halle, and Fred Lukoff. 1956. On accent and juncture in English. In *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*, 65–80. The Hague: Mouton Co.
- Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24:239–298.
- Clapp, William, and Arto Anttila. 2021. To predict or to memorize: Prominence in inaugural addresses. In *Supplemental Proceedings of the 2020 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America.
- Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory phonology* 6:243–278.

References II

- Firbas, Jan. 1992. *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.
- Fodor, Janet Dean. 2002. Psycholinguistics cannot escape prosody. In *Proc. SpeechProsody 2002*, 83–90. DOI: 10.21437/SpeechProsody.2002-12.
- Frazier, Lyn, and Edward Gibson. 2015. *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor*. Springer.
- Jo, Jinyoung, Sean Choi, and Arto Anttila. in progress. English phrasal stress. Ms., Stanford University and Santa Clara University.
- Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249–336.
- Lüdecke, Daniel. 2025. *sjplot: Data visualization for statistics in social science*. URL <https://CRAN.R-project.org/package=sjPlot>, r package version 2.9.0.
- Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108:3526–3529.
- R Core Team. 2026. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell system technical journal* 27:379–423.
- Shapiro, Naomi Tachikawa, and Arto Anttila. 2021. On the phonology and semantics of deaccentuation. In *Proceedings of the 2020 Annual Meeting on Phonology*, ed. Ryan Bennett, Richard Bibbs, Mykel L. Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel, and Maya Wax Cavallaro. Washington, DC: Linguistic Society of America.

References III

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. URL <https://ggplot2.tidyverse.org>.

Woolley, John T., and Gerhard Peters. 1999. The American Presidency Project. *Santa Barbara, CA* Available from World Wide Web: <http://www.presidency.ucsb.edu/ws>.