Stress-Information Alignment

Arto Anttila
Stanford University

1 Preliminaries

(1) Words in an English sentence show 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 (George W. Bush, first inaugural)

(2) Based on the inaugurals of six U.S. presidents, we present new evidence for the following:
- (a) Hierarchical phrasal prominence is real.
- (b) Phrasal prominence is a matter of stress, not merely pitch accent.
- (c) Amount of information and degree of phrasal stress are aligned.
- (d) This is true of cyclic stress. Postcyclic stress shows split behavior.

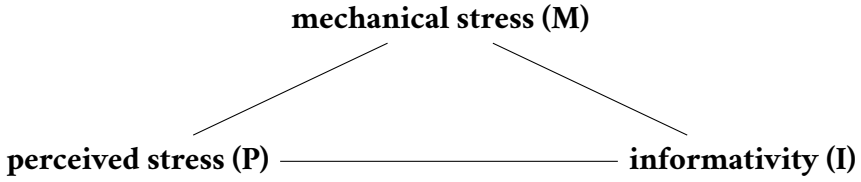
(3) Phrasal stress has two main functions (Jespersen 1920):

Mechanical stress reflects syntax and falls on the rightmost content word of a phrase in a cyclic fashion (Chomsky and Halle 1968, Liberman and Prince 1977, Cinque 1993).

Meaningful stress highlights constituents with the greatest freight of information (Bolinger 1972).

(4) Mechanical and meaningful stress can conflict. Here mechanical stress wins:
A: How much did they pay you for participating in the experiment?
B: Five FRANCS. (Ladd, 1996, 166)

Figure 1: Decomposing sentential prominence



(5) Hypotheses:
- (a) Mechanical stress predicts perceived stress (M→ P). (Chomsky and Halle 1968)
- (b) Informativity predicts perceived stress (I→ P). (Bolinger 1972)
- (c) Mechanical stress predicts informativity. (M → I) (Bolinger 1957)

Hypotheses (a) and (b) are well known; (c) is the topic of this talk.

(6) We now have the tools to tease apart the components in **Figure 1**.

Mechanical stress: MetricalTree (Anttila et al. 2020) written by Timothy Dozat implements an updated version of the SPE stress algorithm building on the output of the Stanford Parser (Klein and Manning 2003; Chen and Manning 2014; Manning et al. 2014). Check out `https://metricaltree.stanford.edu/` for a web app developed by Juan Solis.

Perceived stress: MetricGold (Shapiro 2019). A word's perceived stress level heard by native speaker annotators in the Presidents Project (2016–). Presidents: Obama, George W. Bush, Clinton, George H. W. Bush, Reagan, Carter. The current corpus has 10 inaugurals, 11 annotators, about 78,000 data points (word-annotation pairs) and counting.

Informativity: The weighted average of the negative log probability of a word given the previous word (Piantadosi et al. 2011; Cohen Priva 2012, 2015) calculated for each word in the inaugural corpus. The informativity and accentuation variables were developed by Naomi Shapiro for the project reported on in Shapiro and Anttila 2021.

(7) Here's the sentence *The American dream endures* (Carter 1977, Annotator 3):

	The	American	dream	endures
MECHANICAL STRESS:	4	3	2	1
PERCEIVED STRESS:	1	4	6	8
INFORMATIVITY:	2.75	5.65	5.70	4.60
WORD NUMBER:	1	2	3	4

(8) By representing each component as a vector of numbers we can test interesting hypotheses.

(9) **The Bolinger Hypothesis**: Mechanical stress predicts informativity in good prose.

(10) "[T]he writer should make [mechanical and meaningful stress] coincide as nearly as he can by maneuvering the semantic heavy stress into the position of the mechanical loud stress; that is, toward the end." (Bolinger, 1957, 235).

(11) Examples of good and bad writing (Bolinger, 1957, 235-256)

(a) The "f" turned to "h" which became silent in its turn. (bad)
(b) The "f" turned to "h" which in turn became silent. (good)

[H]e defines Canada as the place where

(c) Canadian bacon was invented (bad)
(d) they invented Canadian bacon (good)

(12) **The Prague School Hypothesis:** Linear order predicts informativity: old information comes before new information ("linear modification" in Firbas 1992).

(13) The components can be closely correlated. For a striking example, see **Figure 2**. Visualization and statistical analyses were done in `R` (R Core Team 2024) and `ggplot2` (Wickham 2016).

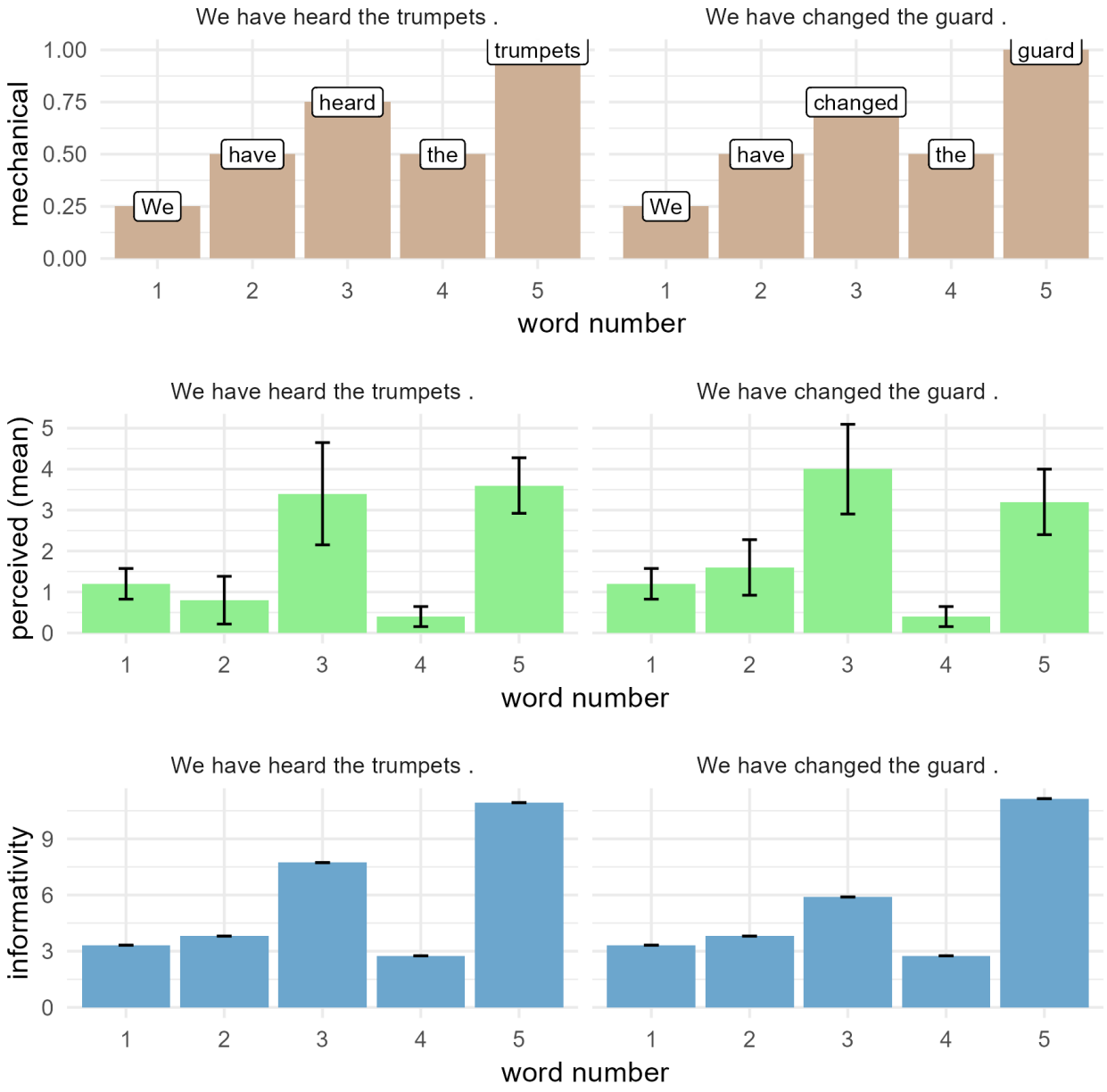


Figure 2: Sentences 107 and 108 from Clinton's 1993 inaugural

2 Where do informative words go?

(14) Which predicts a word's informativity better: its mechanical stress level (Bolinger 1957) or its linear position in the sentence? (cf. Firbas 1992)?

(15) The current inaugural corpus contains $78,604$ data points (word-annotation pairs). Here we only included nouns, verbs, adjectives, and function words ($N = 65,748$), excluding adverbs, numerals, interjections, punctuation, contractions, and `NA`s where mechanical or perceived stress was missing ($N = 12,856$), which constitutes 16.4% of all data.

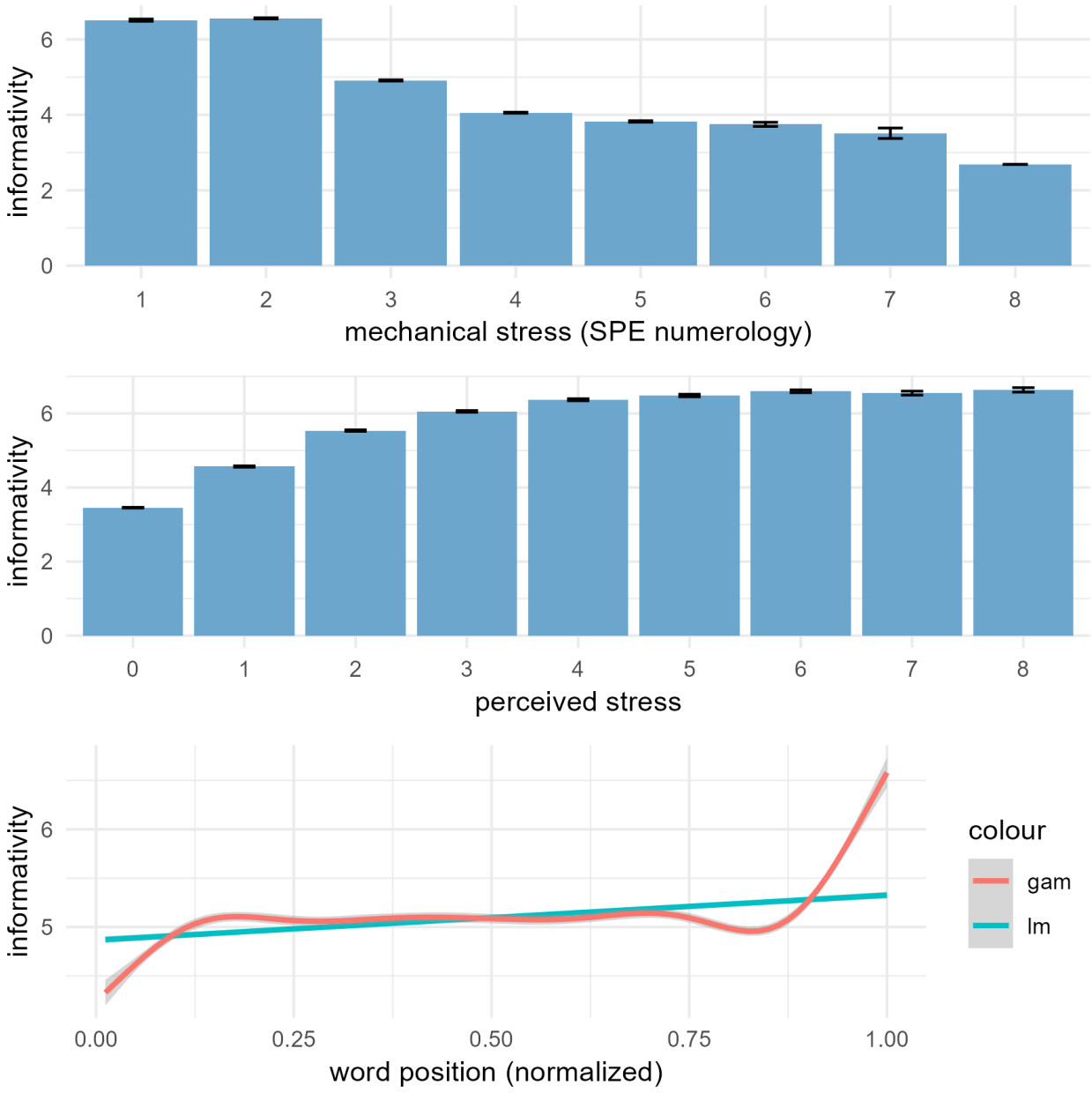


Figure 3: Mean informativity by perceived stress, mechanical stress, and linear position.

(16) **Figure 3** shows that mechanical and perceived stress are both correlated with informativity (in opposite directions: a higher SPE number means less stress). Word position is positively correlated with informativity, but the effect is tiny, as confirmed by simple linear regression. R^2 = variance described by the model, `info` = informativity, `perc` = perceived stress, `mech` = mechanical stress, `wpos` = word position

Model	R^2	Content of predictor
`lm(info ~ perc)`	0.219	perceived = mechanical + meaningful + word position
`lm(info ~ mech)`	0.193	mechanical
`lm(info ~ wpos)`	0.003	word position

(17) Here's a model summary (predictors centered and standardized). Informativity is
  (a) negatively associated with mechanical stress: $b = -0.99, SE = 0.01, p < 0.001$
  (b) positively associated with perceived stress: $b = 1.05, SE = 0.01, p < 0.001$
  (c) positively associated with word position: $b = 0.13, SE = 0.01, p < 0.001$

(18) If mechanical stress and word position are both included in the same model, the latter reverses sign and becomes **negative**, suggesting a confound. Once mechanical stress is explicitly present word number is no longer positively correlated with informativity.

(19) Here's a model with both predictors:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.097160	0.007774	655.66	<2e-16 ***
`mechanical stress`	-1.104886	0.008448	-130.79	<2e-16 ***
`word position`	**-0.305426**	0.008448	-36.16	<2e-16 ***

(20) **Summary: Information tracks hierarchical stress, not word order (in good prose).**

(21) This is prose. Hierarchical phrasal stress may also explain aspects of POETIC CLOSURE in verse (Smith 1968; Blumenfeld 2015; Smith and Anttila 2024).

(22) An alternative hypothesis not addressed here: Uniform Information Density (Jaeger and Levy 2006, Meister et al. 2021).

3 How about meaningful stress?

(23) (a) Mechanical stress: a vector of numbers computed by an updated SPE algorithm.
  (b) Perceived stress: a vector of numbers provided by native speaker annotators.
  (c) Meaningful stress: the difference between the two (cf. Shapiro and Anttila 2021).

(24) **Accentuation Index (AI)** = perceived stress − mechanical stress

(25) $AI = 0$ Mechanical (= cyclic) stress corresponds exactly to native speaker perceptions.
  $AI > 0$ Accentuation: The word has more stress than expected.
  $AI < 0$ Deaccentuation: The word has less stress than expected.

(26) Empirically, Accentuation Index is related to at least
  focus (Rooth 1992; Wagner 2021)
  deaccentuation (Ladd, 1980, 50-69)
  eurhythmy (Hayes 1984)
  paralanguage (Ladd, 1996, 33-41)

(27) Deaccentuation is illustrated in **Figure 4** and **Figure 5**. It is attributed to the following constraint by Shapiro and Anttila (2021) inspired by Wagner 2012, 2021:

Anti-Epistrophe Constraint: Avoid nuclear stress on a word that has a segmentally identical copy (reasonably near) on its left.

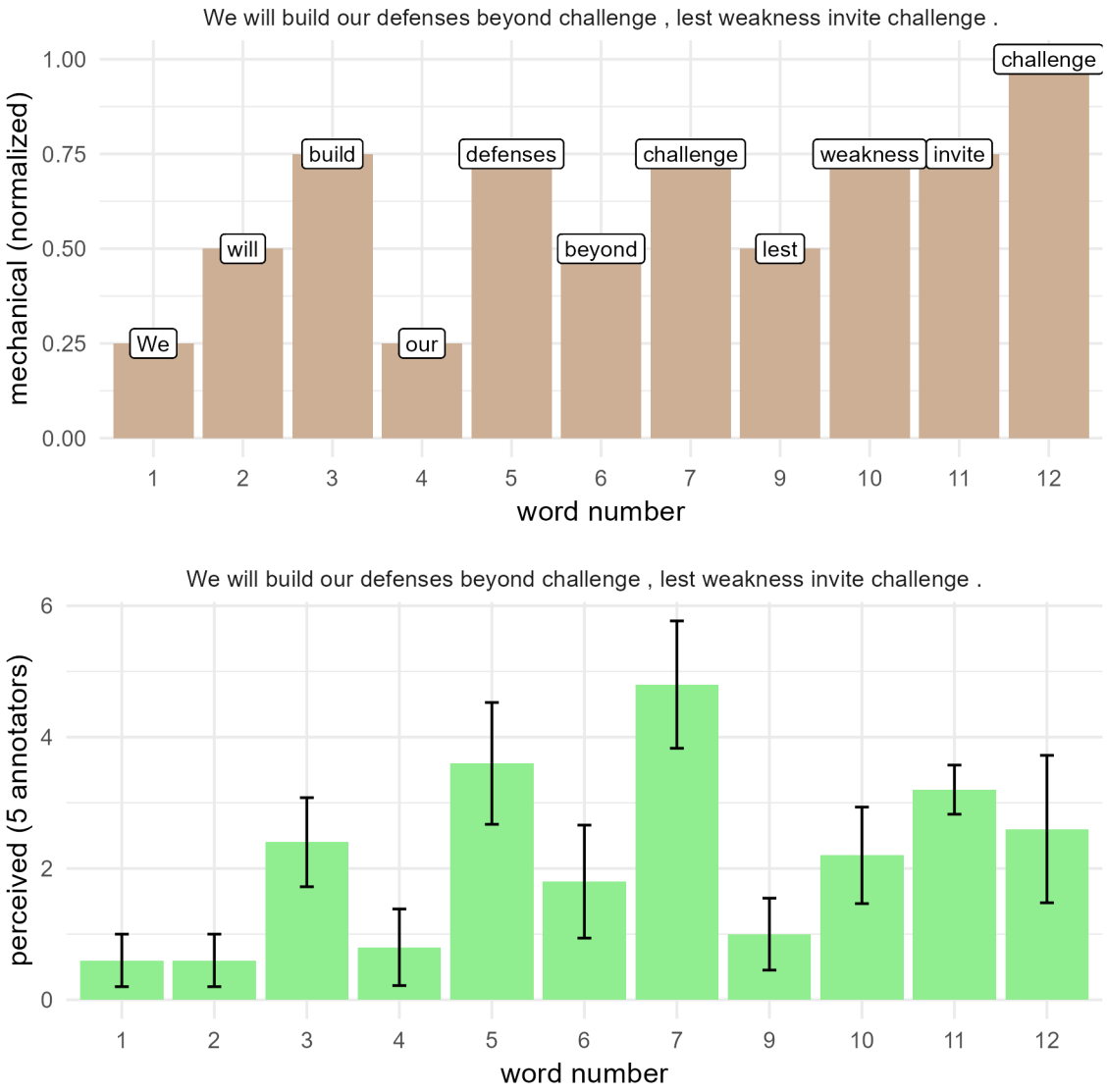


Figure 4: Stress retraction from *challenge* (12) to *invite* (11) (George W. Bush 2001)

(28) Deaccentuation is a postcyclic stress retraction rule of (some dialects of) English that moves the stress that violates the Anti-Epistrophe Constraint to the closest stressable word on the left (Shapiro and Anttila 2021).

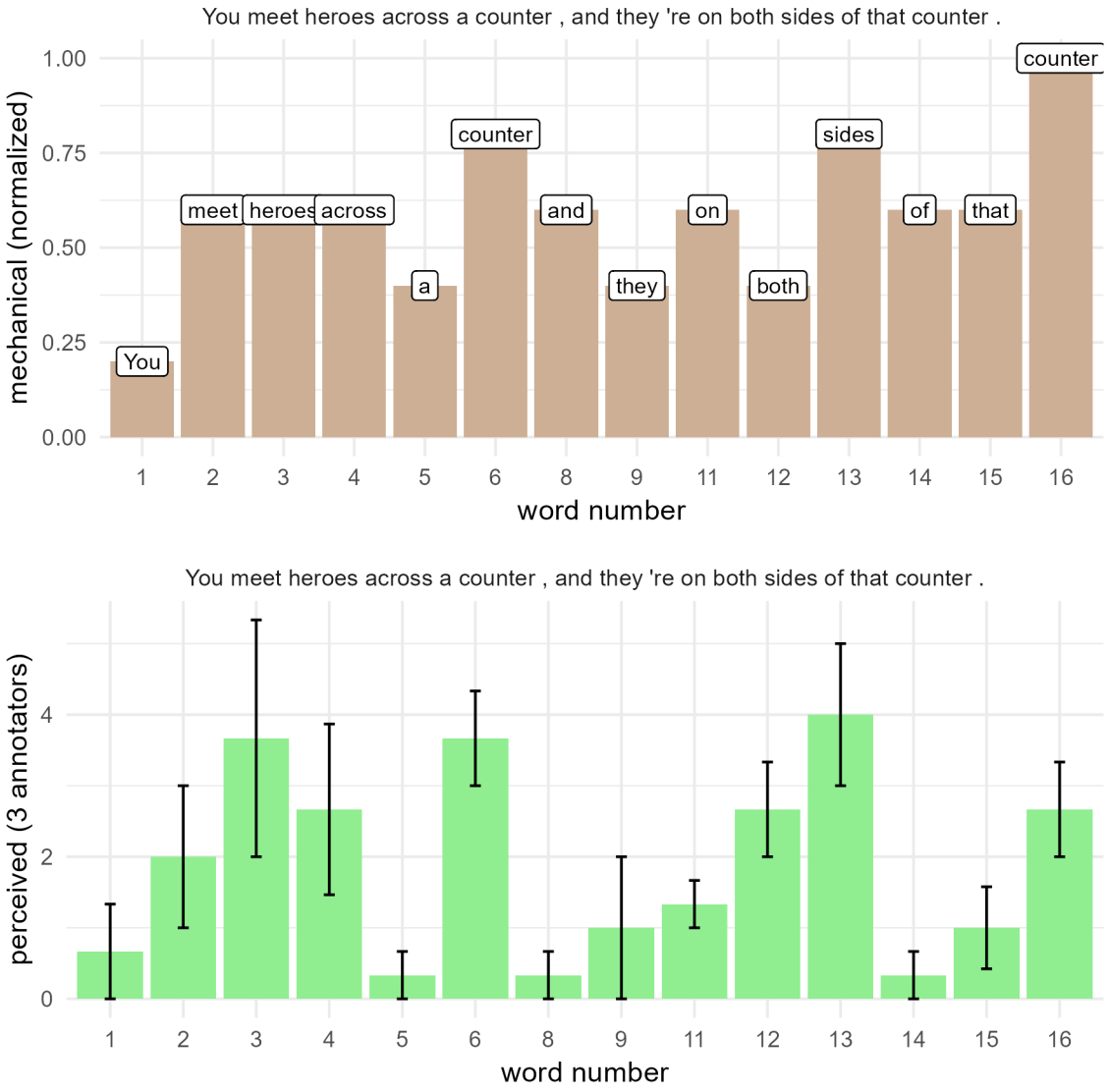


Figure 5: Stress retraction from *counter* (16) to *sides* (13) across two function words (Reagan 1981)

(29) Is Accentuation Index related to informativity? **Figure 6** suggests the following conjecture:

Accentuation is about information: Words with $AI > 0$ tend to be more informative.

Deaccentuation is not about information: Words with $AI < 0$ show no clear pattern.

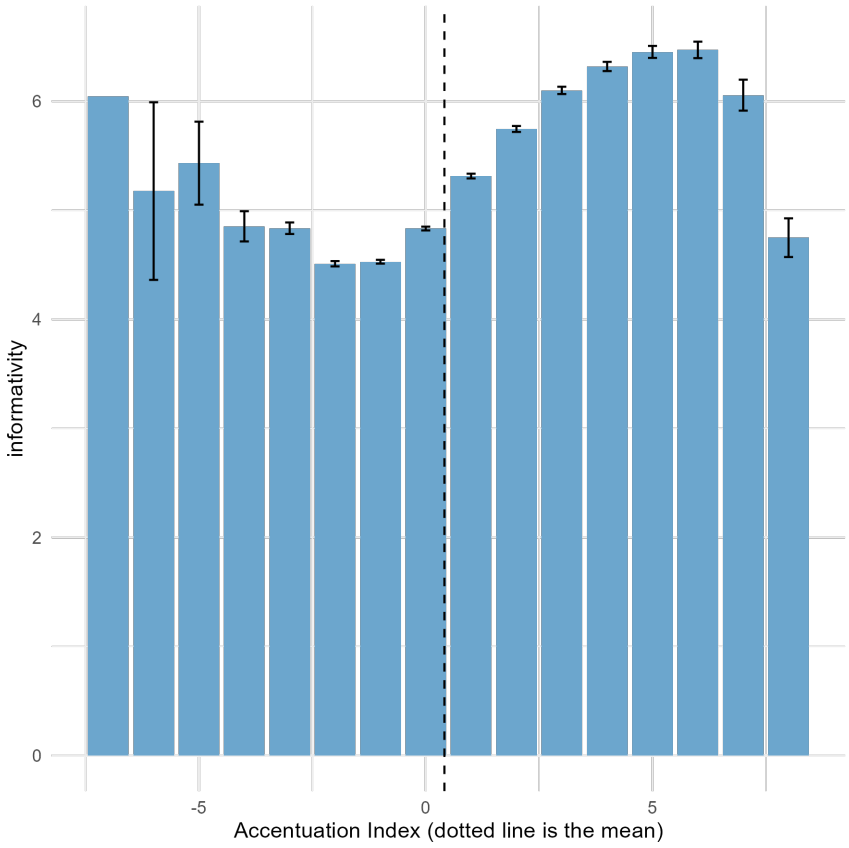


Figure 6: Informativity by Accentuation Index

4 Suggested exercise for those interested

(30) Let's define Bolinger Index as R^2 of `lm(info ~ mechanical)`. I suspect that R^2 is

high in oratorical prose prepared by professional speechwriters.

low in speech (where intonation is available), especially under time pressure.

(31) You'll need MetricalTree (`https://metricaltree.stanford.edu/`) and a corpus with informativity scores to test the above conjecture. You won't need perceived stress.

5 Conclusions

(32) Hierarchical stress above the word is real (*pace* Gussenhoven 2011, 2015, 2022). Stress is not binary (accented vs. not) but a matter of degree (see also Clapp and Anttila 2021).

(33) Stress and informativity are aligned in a way that depends on the domain:

(a) clearly in the case of cyclic stress

(b) clearly in the case of postcyclic accentuation

(c) but not in the case of postcyclic deaccentuation

Appendix A: Practice data (Shapiro and Anttila 2021)

Naomi Shapiro's GitHub repository `https://github.com/tsnaomi/AMP-2020-Tutorial` contains the data frame `inaugurals.csv` with the inaugural addresses of George W. Bush (2001) and Barack Obama (2009) plus the R script for our tutorial:

`https://web.stanford.edu/~anttila/research/AMP-2020-Tutorial-Handout.pdf`

Appendix B: The SPE stress algorithm

Cyclicity and Stress Subordination: Stress is assigned cyclically from the inside out, assigning [1 stress] to a designated word and reducing stress elsewhere by one (stress subordination).

The Nuclear Stress Rule (NSR): Assign [1 stress] to the rightmost vowel bearing the feature [1 stress]. Applies to phrases (NP, VP, AP, S).

The Compound Stress Rule (CSR): Skip over the rightmost word and assign [1 stress] to the rightmost remaining [1 stress] vowel; if there is no [1 stress] to the left of the rightmost word, then try again without skipping the word. Applies to words (N, A, V).

(A convenient summary can be found in Baart 1987.)

Acknowledgements

Funding: This research was funded by the Roberta Bowman Denning Initiative in the Digital Humanities as part of the project *Prose Rhythm and Linguistic Theory*, the Vice-Provost for Undergraduate Education at Stanford University, and Stanford Introductory Seminars Plus.

Collaborators: The individuals who made this research possible include Frankie Conover, Timothy Dozat, Elena Felix, Daniel Galbraith, Julia Mendelsohn, Shina Penaranda, Naomi Shapiro, Liam Smith, Madeline Snigaroff, Juan Solis, Connor Toups, Alexander Wade, Amy Wang, and Annalisa Welinder. I thank them all.

Data: The inaugurals data were obtained from Peters and Woolley (1999–) *The American Presidency Project*, `http://www.presidency.ucsb.edu/index.php`.

References

Anttila, Arto, Timothy Dozat, Daniel Galbraith, and Naomi Shapiro. 2020. Sentence stress in presidential speeches. In *Prosody in Syntactic Encoding*, eds. Gerrit Kentner and Joost Kremers, 17–50. Berlin/Boston: Walter De Gruyter. Preprint available on lingbuzz/004303.

Baart, Joan. 1987. Focus, Syntax, and Accent Placement. PhD diss, University of Leiden.

Blumenfeld, Lev. 2015. Meter as faithfulness. *Natural Language & Linguistic Theory* 33: 79–125.

Bolinger, Dwight L. 1957. Maneuvering for stress and intonation. *College Composition and Communication* 8 (4): 234–238.

Bolinger, Dwight L. 1972. Accent is predictable (if you are a mind reader). *Language* 48: 633–644.

Chen, Danqi, and Christopher D. Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 740–750. https://aclanthology.org/D14-1082.pdf.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.

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. 2012. Sign and signal: Deriving linguistic generalizations from information utility. PhD diss, Stanford University.

Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory phonology* 6 (2): 243–278.

Firbas, Jan. 1992. *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2011. Sentential prominence in English. In *The Blackwell Companion to Phonology*, eds. Marc Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2778–2806. Malden, MA: Wiley-Blackwell.

Gussenhoven, Carlos. 2015. Does phonological prominence exist? *Lingue e linguaggio* 14 (1): 7–24.

Gussenhoven, Carlos. 2022. Just how metrical is the Autosegmental-Metrical model? Evidence from pitch accents in Nubi, Persian, and English. In *Prosody and prosodic interfaces*. Oxford University Press.

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15 (1): 33–74.

Jaeger, T, and Roger Levy. 2006. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19.

Jespersen, Otto. 1920. *Lehrbuch der Phonetik: Mit 2 Tafeln*. Leipzig und Berlin: B. G. Teubner.

Klein, Dan, and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 423–430. https://aclanthology.org/P03-1054.pdf.

Ladd, D. Robert. 1980. *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press. Cambridge Studies in Linguistics 79.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60. https://aclanthology.org/P14-5010.pdf.

Meister, Clara, Tiago Pimentel, Patrick Haller, Lena Jäger, Ryan Cotterell, and Roger Levy. 2021. Revisiting the Uniform Information Density Hypothesis. https://arxiv.org/abs/2109.11635.

Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108 (9): 3526–3529.

R Core Team. 2024. *R: A language and environment for statistical computing*. Vienna, Austria. R Foundation for Statistical Computing. https://www.R-project.org/.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1 (1): 75–116.

Shapiro, Naomi Tachikawa. 2019. MetricGold. A software package for the metrical annotation of English prosody, https://github.com/tsnaomi/metric-gold.

Shapiro, Naomi Tachikawa, and Arto Anttila. 2021. On the phonology and semantics of deaccentuation. In *Proceedings of the 2020 Annual Meeting on Phonology*, eds. 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.

Smith, Barbara Herrnstein. 1968. *Poetic Closure: A Study of How Poems End*. Chicago and London: The University of Chicago Press.

Smith, Liam, and Arto Anttila. 2024. Metrical tension and prose cadence. A talk to be presented at AMP 2024, Rutgers University.

Wagner, Michael. 2012. A givenness illusion. *Language and Cognitive Processes* 27 (10): 1433–1458.

Wagner, Michael. 2021. Prosodic focus. In *The Wiley Blackwell Companion to Semantics*, eds. Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann. John Wiley & Sons, Inc. First Edition.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer. https://ggplot2.tidyverse.org.