

Information and copula variation in African American Language

Michael Senko, Lorena Martín Rodríguez,
Amir Zur, and Dan Jurafsky

NWAV 54
October 23, 2026

Copula variation in AAL

Variable realizations of the copula:

1. You are in trouble. ← full surface copula
2. You're in trouble. ← contracted surface copula
3. You $\emptyset$ in trouble. ← zero copula

Grammatical factors

subject



You **[COP]** in trouble



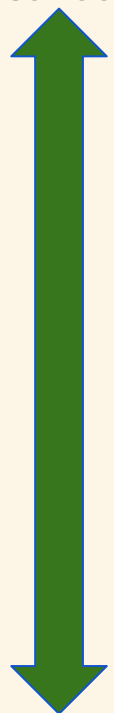
copula slot



predicate

Copula constraint hierarchy

+surface



+zero

NP: My mother **[COP]** the vice president.

Loc: Like you **[COP]** on a farm or something.

Adj: He **[COP]** successful but he **[COP]** dirty.

V+ing: Everybody **[COP]** looking at your new shirt.

Gonna: You **[COP]** gonna respect us.

The constraint hierarchy is consistent

- Why does the constraint hierarchy reoccur?

“More serious is the absence of convincing explanations for certain recurrent effects... why the following grammatical constraint hierarchy should pattern as it does” (Rickford 1998: 758)

- Particularly important given the stylistic diversity and local specificity of AAL varieties across the US

Information and syntactic reduction

Corpus analysis methodology

- 222 sociolinguistic interviews from the Corpus of Regional African American Language (CORAAL)
- Information derived from across all of CORAAL (N = 1.2m bigrams)
- Copula variation: 4 interviews from Washington, DC, ~2015
 - Hand annotated for syntactically licit copula instances (N = 418)



Language and information

- The more probable a linguistic unit is in context, the less information the unit carries in that context

He (is | 's | $\emptyset$) old

He (is | 's | $\emptyset$) petulant

Why reduce?

- Speakers reduce uninformative units and enhance informative ones

He (is | 's | $\emptyset$) petulant

Speaker-internal: hard to retrieve

Listener-oriented: hard to predict

Measuring information in CORAAL

- Possible that the copula doesn't surface!
 - Throw out all copula and auxiliary forms
 - Count every unigram (1.2m) and bigram (1.2m-1)
 - Create context bigram

You ? gonna respect us



→ 4 unigrams, 3 bigrams

Measuring you ? gonna

- We want to know how strongly the predicate constrains the copula context
- How to represent the amount of information contained in a bigram?

Informativeness: $P(\text{you} \mid \text{gonna})$

$$= C(\text{you}, \text{gonna}) / C(\text{gonna})$$

A worked example

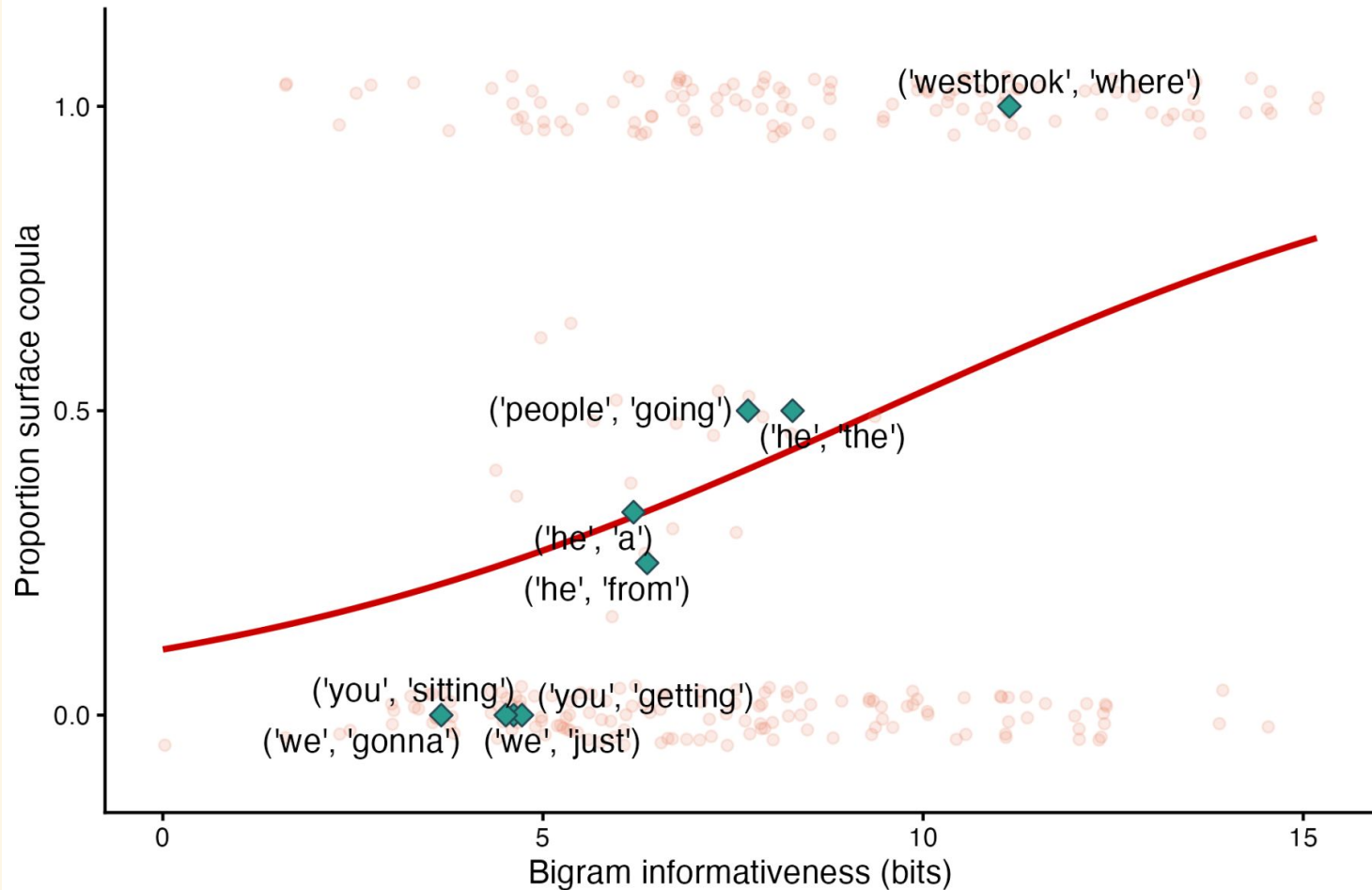
- Negative log2 compresses extreme values into bits
 - 1 bit = choice between two equally likely possibilities

they ? peeling
they ? peeling

we ? peeling
we ? peeling

$$-\log_2 \left(\frac{\text{count}(\text{they}, \text{peeling})}{\text{count}(\text{peeling})} \right) = 1$$

Informativeness of selected bigrams



More info → more material in copula slot

- The less probable a copula context, the more information it contains
- Speakers are more likely to realize surface copula when the context is more informative
 - Constrained by the predicate (word after copula slot)

Logistic regression model: formula

$$\begin{aligned} \text{surface copula} \sim & \beta_0 + \beta_1 \text{ speech rate} + \beta_2 \text{ bigram informativeness} \\ & + \beta_3 \text{ prosodic stress} + \beta_4 \text{ predicate type} \\ & + \beta_5 \text{ preceding segment} + \beta_6 \text{ following segment} \\ & + (1 \mid \text{speaker}) + (1 \mid \text{context bigram}) \end{aligned}$$

Logistic regression model: results

Term	Estimate	Std. Error	z value	p value
<i>Intercept</i>	4.285	1.094	3.919	0
Speech rate	-0.261	0.059	-4.397	0
Bigram informativeness	0.203	0.076	2.688	0.007
Prosodic stress	-0.473	0.183	-2.584	0.01
Predicate: gonna	-2.716	0.768	-3.535	0
Predicate: locative	-0.874	0.514	-1.698	0.089
Predicate: noun phrase	-0.169	0.447	-0.379	0.705
Predicate: verb + <i>ing</i>	-0.869	0.464	-1.871	0.061
Preceding segment: vowel	0.015	0.374	0.041	0.967
Following segment: vowel	-0.953	0.432	-2.206	0.027

Can information explain the hierarchy?

- Across the full corpus...
 - Measure informativeness for each predicate type
 - (*gonna*, *V+ing*, *Adj*, *Loc*, *NP*)

pronoun ? predicate



Revisiting the constraint hierarchy

+surface



+zero

NP: My mother **[COP]** the vice president.

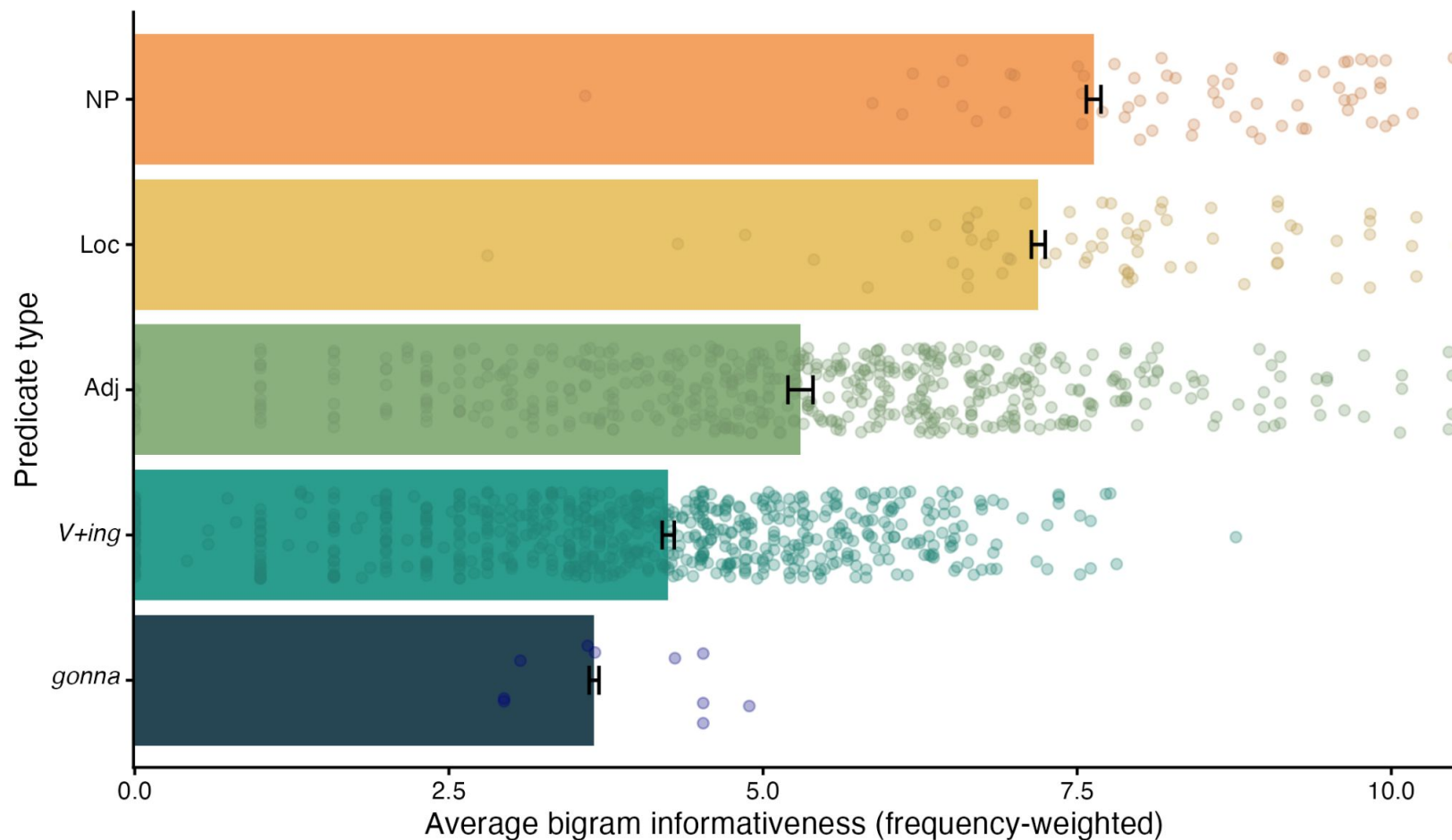
Loc: Like you **[COP]** on a farm or something.

Adj: He **[COP]** successful but he **[COP]** dirty.

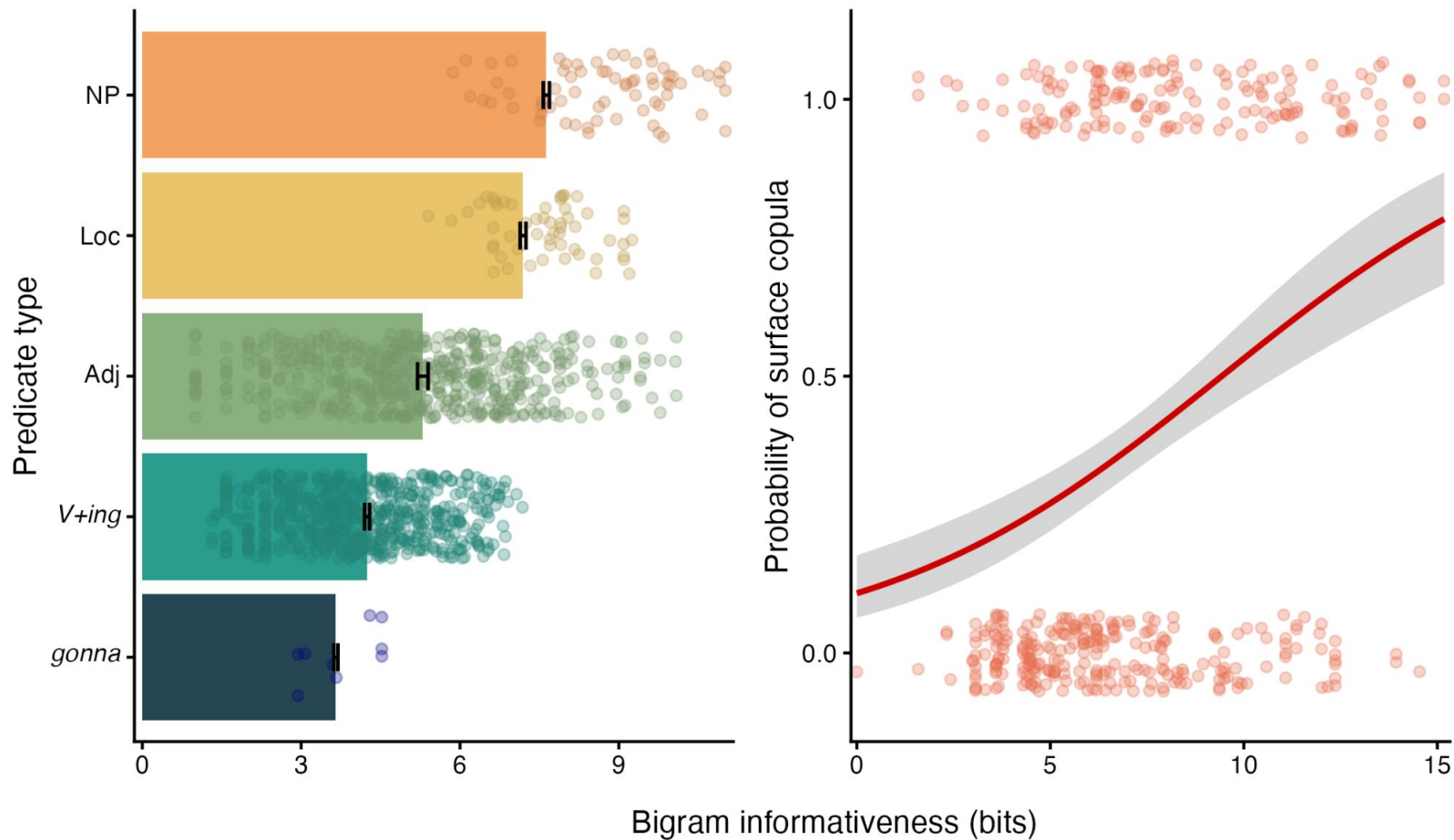
V+ing: Everybody **[COP]** looking at your new shirt.

Gonna: You **[COP]** gonna respect us.

Average information content by predicate type



Informativeness unifies the copula constraint hierarchy



The persistence of the copula constraint hierarchy reflects pressure to reduce predictable elements

Predicate types are ordered by average predictability in context

References

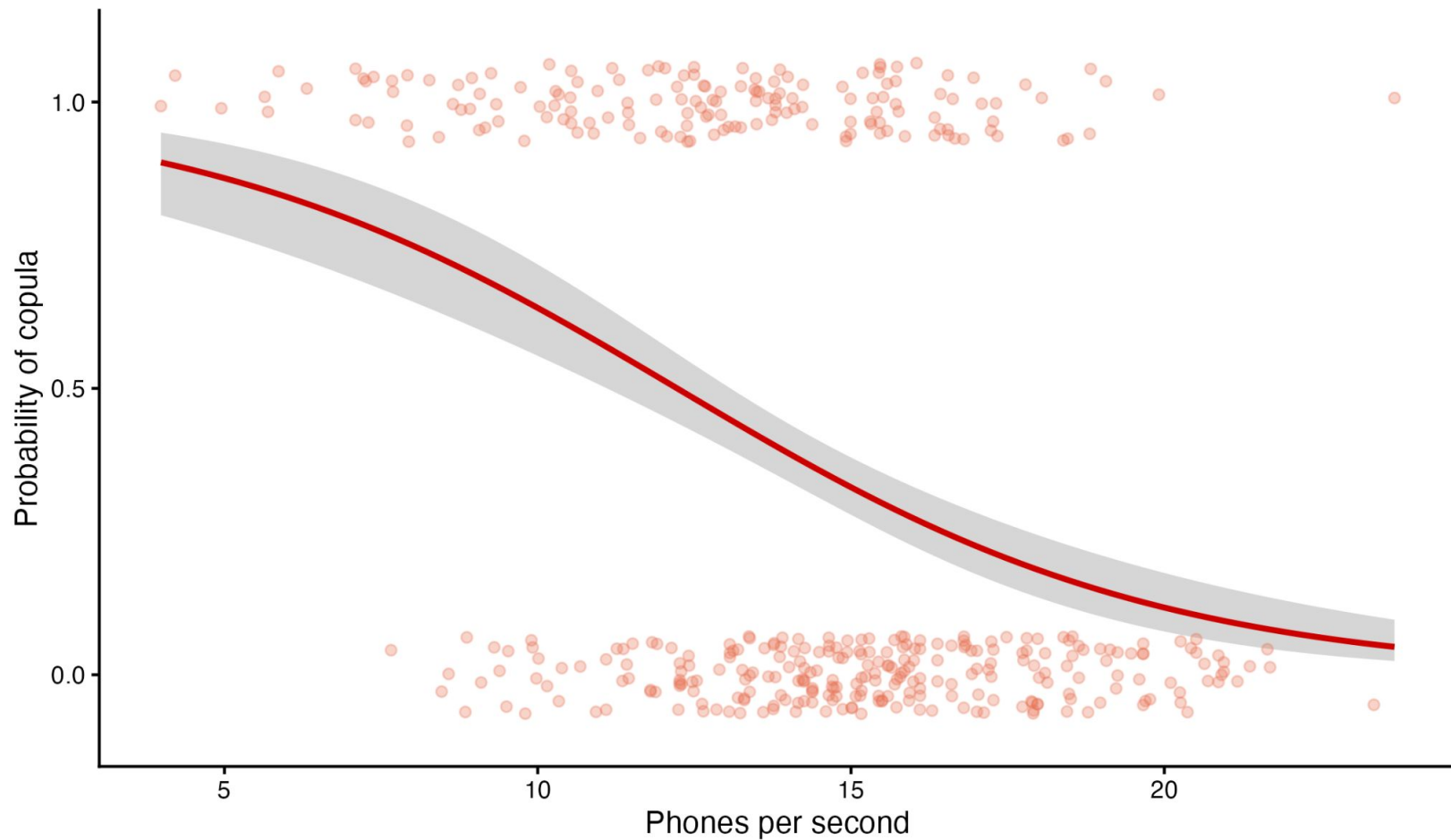
- Blake, R. (1997). Defining the envelope of linguistic variation: The case of “don’t count” forms in the copula analysis of African American Vernacular English. *Language Variation and Change*, 9(1), 57–79. <https://doi.org/10.1017/S0954394500001794>
- Bresnan, J. (2021). Formal Grammar, Usage Probabilities, and Auxiliary Contraction. *Language*, 97(1), 108–150. <https://doi.org/10.1353/lan.2021.0003>
- Bybee, J. (2001). *Phonology and Language Use* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511612886>
- Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 30(30).
- Kendall, T., & Farrington, C. (2023). *The Corpus of Regional African American Language* (Version 2023.06) [Data set]. The Online Resources for African American Language Project. <https://doi.org/10.7264/1ad5-6t35>
- King, S. (2020). From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans’ Speech. *Annual Review of Linguistics*, 6(1), 285–300. <https://doi.org/10.1146/annurev-linguistics-011619-030556>
- Poplack, S. (Ed.). (2000). *The English history of African American English*. Blackwell Publishers.
- Rickford, J. R. (1998). The Creole Origins of African American Vernacular English: Evidence from Copula Absence. In S. S. Mufwene, J. R. Rickford, G. Bailey, & J. Baugh (Eds.), *African American English*. Routledge.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sharma, D., & Rickford, J. R. (2009). AAVE/creole copula absence: A critique of the imperfect learning hypothesis. *Journal of Pidgin and Creole Languages*, 24(1), 53–90. <https://doi.org/10.1075/jpcl.24.1.03sha>

Thank you!

Questions?

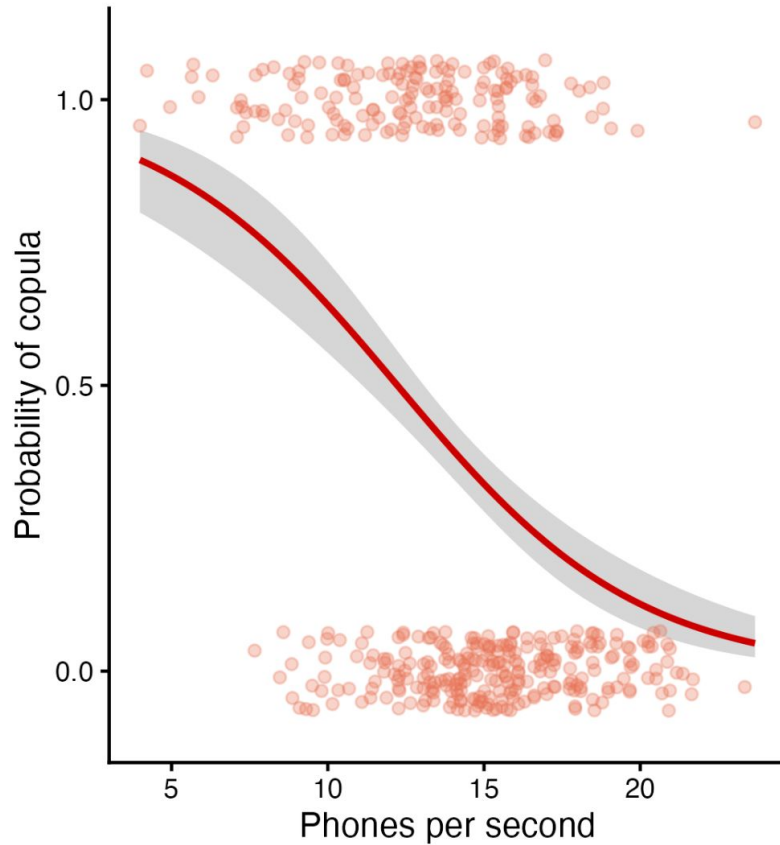
Speech rate and other puzzles

Copula variation by speech rate

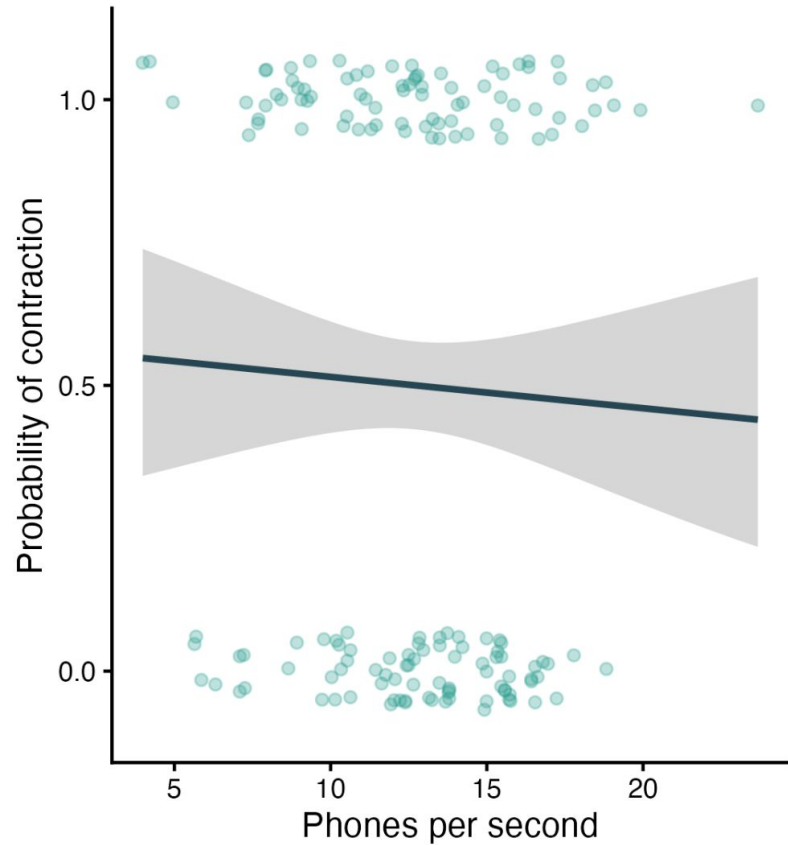


Effect of phoneme rate on...

Copula maintenance

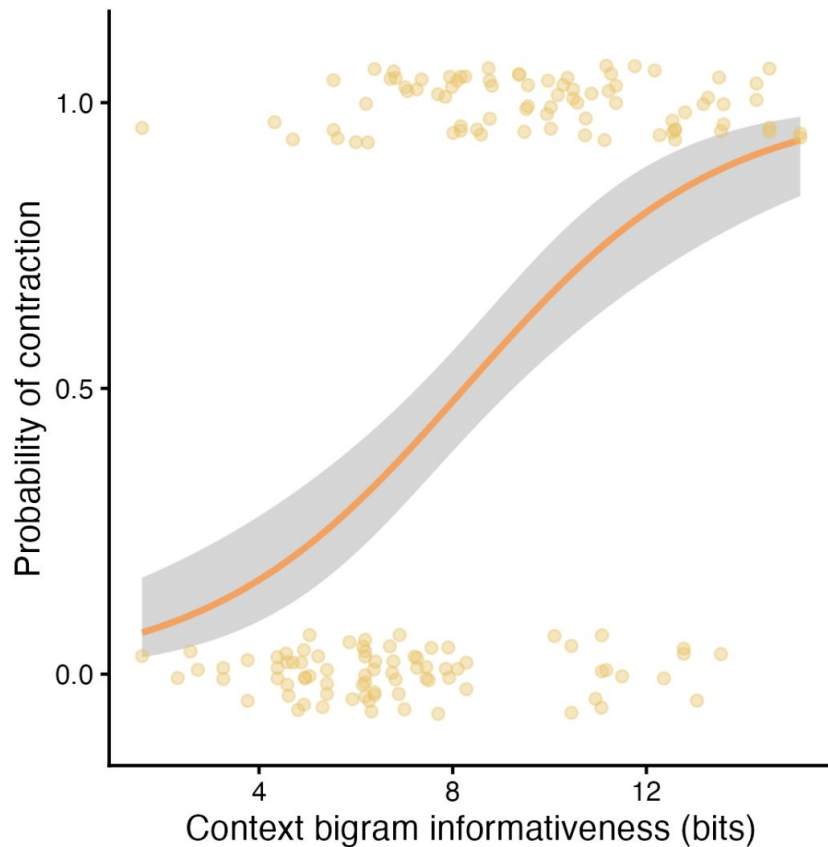


Copula contraction

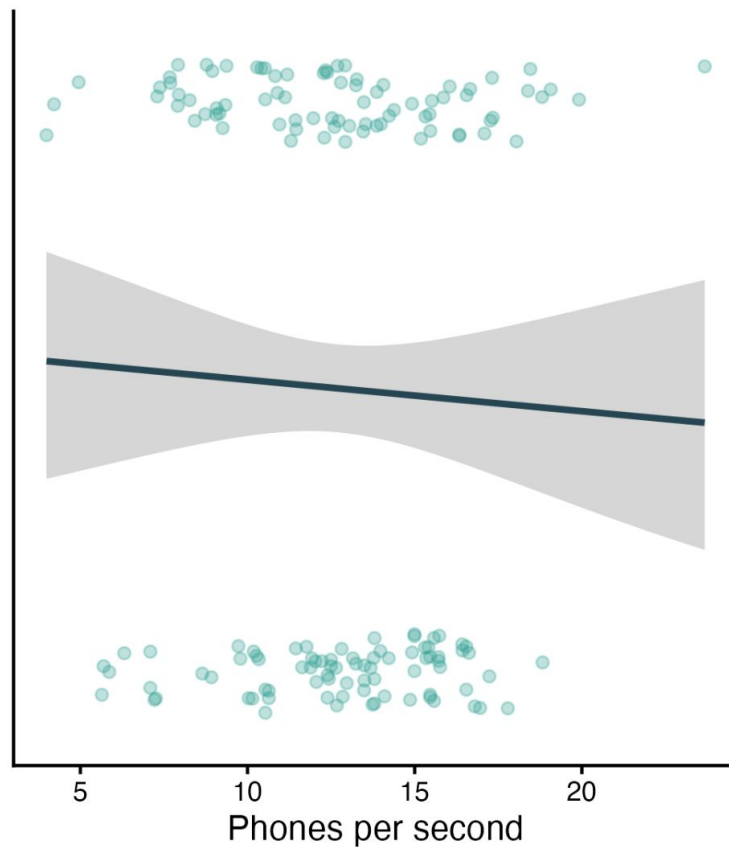


Probability of copula contraction by..

Bigram informativeness



Phoneme rate



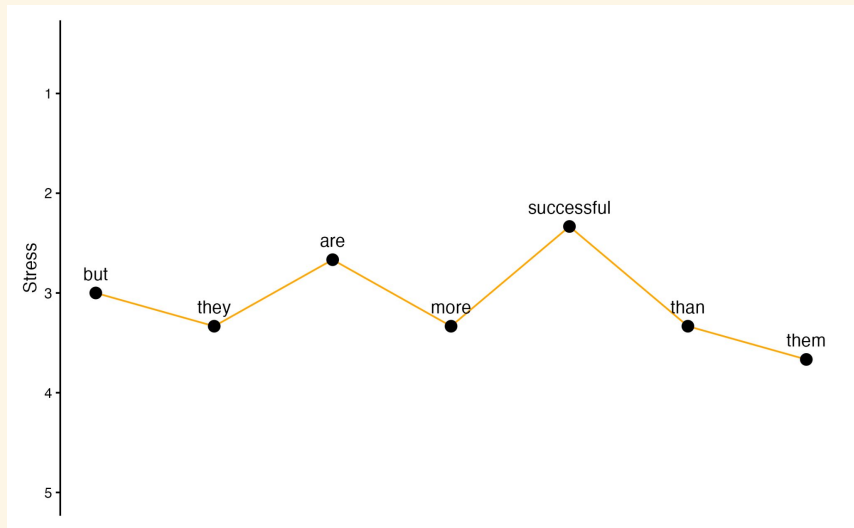
Prosody and morphosyntactic reduction

- Metrical structure governs the distribution of prominence in the whole phrase
- English auxiliary contraction depends on the distribution of phrasal and lexical stress
 - I (will | 'll) do it. → / and [AUX] both unstressed
 - Yes, I will. → [AUX] receives stress, no contraction
- Prosody may be a deciding factor when there is more than one syntactic parse available

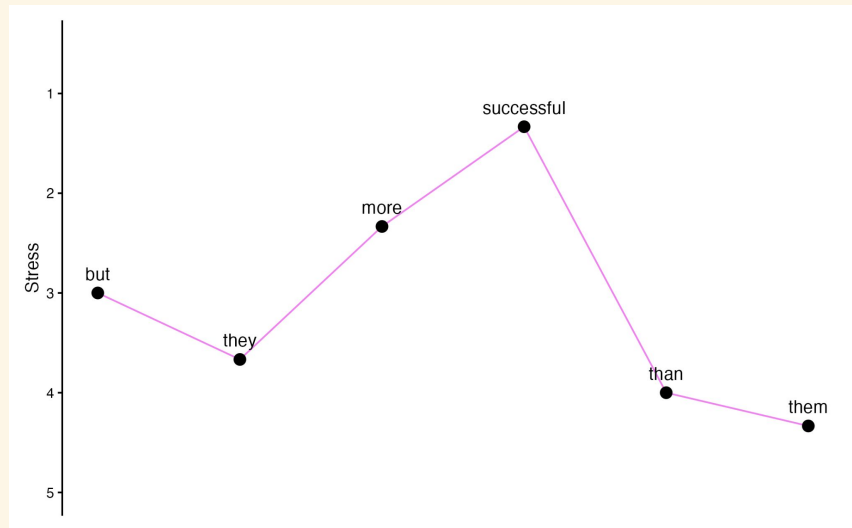
Assigning stress to AAL

- Syntax predicts variation by admitting alternative choices among constituents
- Phonology acts as a filter
 - Assigns high probability to well-formed sequences
 - Lower probability to less well-formed sequences
- MetricalTree implements SPE rules to assign stress
 - Hierarchical stress contour → one word carries primary
 - 1 = primary stress, 2 = secondary stress, etc. up to 10

“but they [COP] more successful than them”

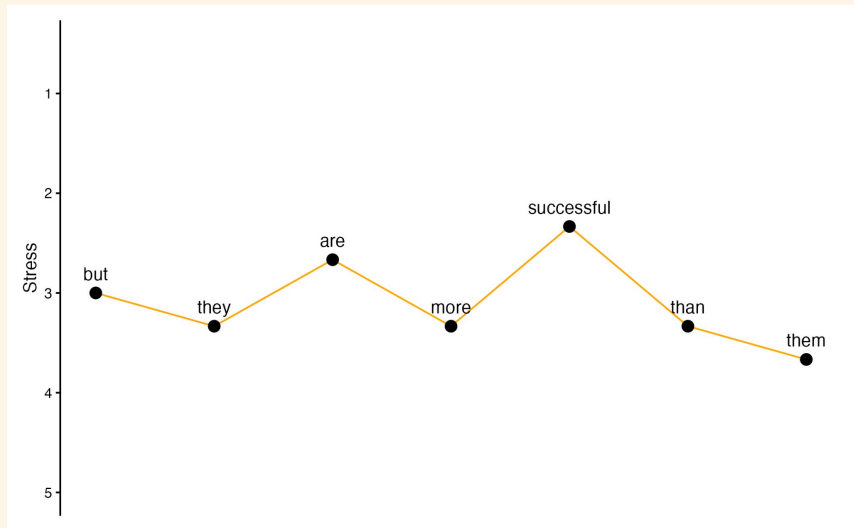


surface copula

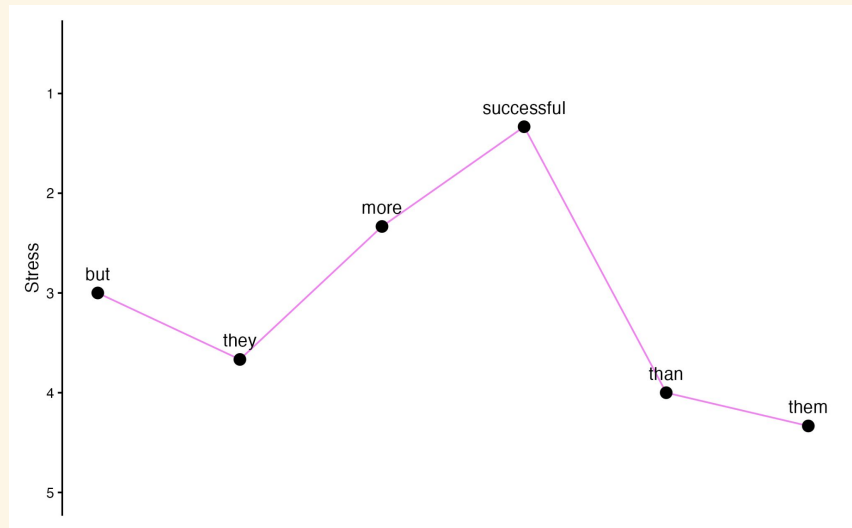


zero copula

“but they [COP] more successful than them”



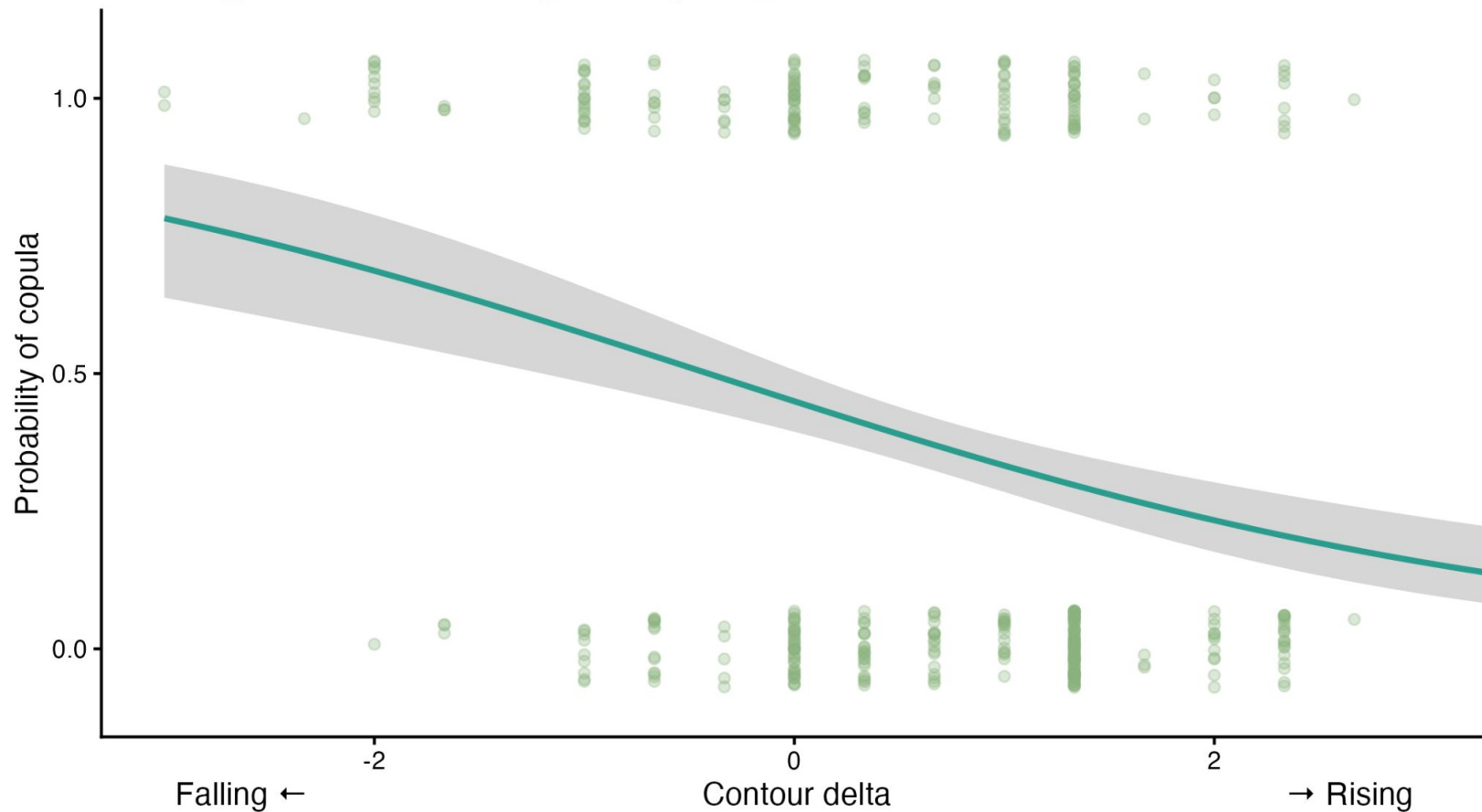
$3.33 - 3.33 = 0$
flat: no preference



$3.66 - 2.33 = 1.33$
rising: prefer zero

Probability of copula by context bigram stress contour

Preceding stress minus following stress: primary=1



By what mechanism? Stress × prominence!

- Words with more information receive greater acoustic prominence; orthogonal to metrical stress

+surface



NP: My mother **[COP]** *the* **vice president.**

Loc: Like you **[COP]** *on a* **farm** or something.

Adj: He **[COP]** **successful** but he **[COP]** **dirty.**

V+ing: Everybody **[COP]** **looking** at your new shirt.

Gonna: You **[COP]** **gonna** respect us.

+zero