Language Production Part 2

Thomas Hikaru Clark

9.19/9.190 Guest Lecture
November 29th, 2023

Warm-up #1

The wool socks were made by hand in Peru

The runner tagged by the catcher was called out

The cotton sweaters are made of is grown in India

Warm-up #2

Warm-up #2

List as many examples of linguistic alternations as you can (we've seen several in class)

Bonus points if you can describe one in another language

Goals for Today

- Assess the influence of two potential pressures on language production:
  - **Ease of production**
  - **Robust communication**
- Evaluate the design of language production experiments
- Use computational models to explain human behavior in language production

Availability-Based Production

What the theory claims: Speakers choose between alternatives based on what is most available; “easy-first” production

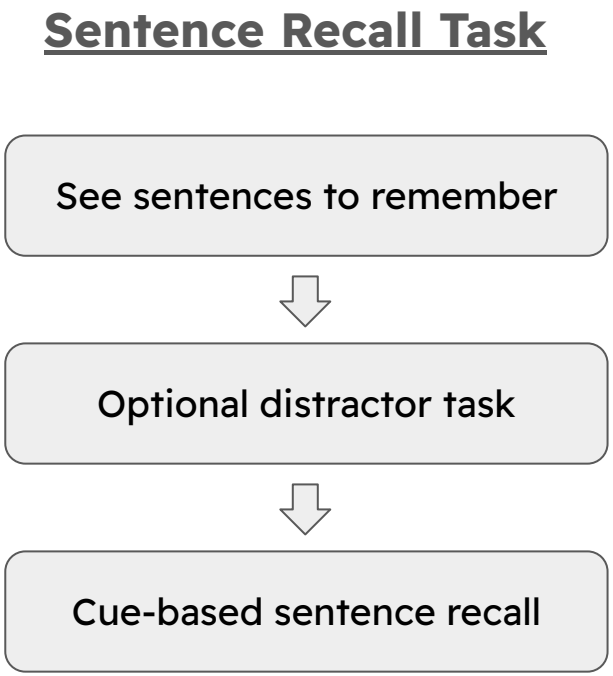
- Short
- Frequent
- Discourse-given
- . . .

Availability-Based Production: Bock (1987), Ferreira & Dell (2000), Zhan & Levy (2018)

Availability-Based Production

The evidence:
Mainly comes from psycholinguistic behavioral experiments.

Question – how do you *experimentally* measure speaker's choices in language production?



Ferreira & Dell (2000)

Experiment Demo

- You will see 3 sentences on the screen sequentially
- Then you will see cues consisting of 2 words
- For each cue, recall the sentence that contained those words

The rich actress donated a million dollars to the college

The football coach knew the goalkeeper skipped practice

A life preserver fell into the sea from the ship’s deck

actress million

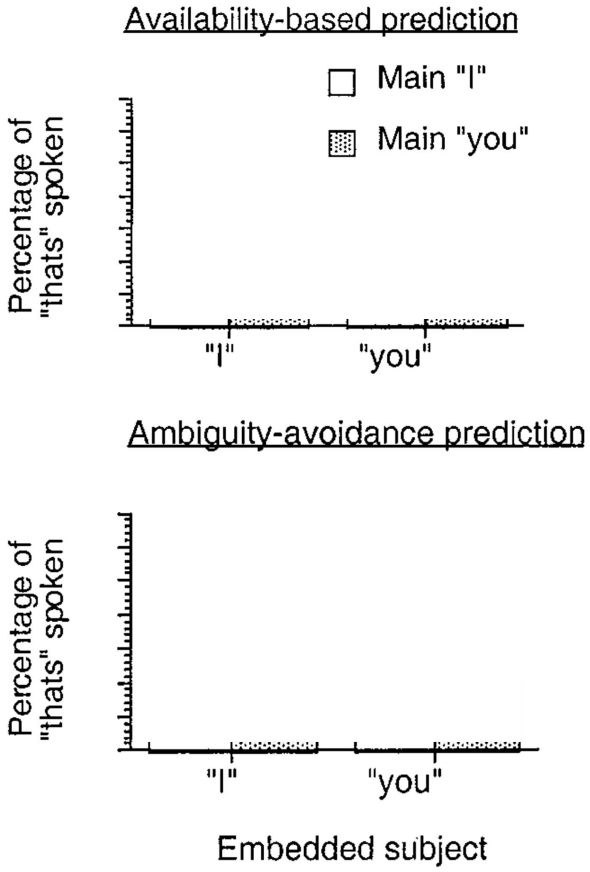
sea ship

coach practice

Availability-Based Production

The evidence:
Speaker choices in sentences like

a) I knew (that) I missed practice
b) You knew (that) I missed practice
c) I knew (that) you missed practice
d) You knew (that) you missed practice

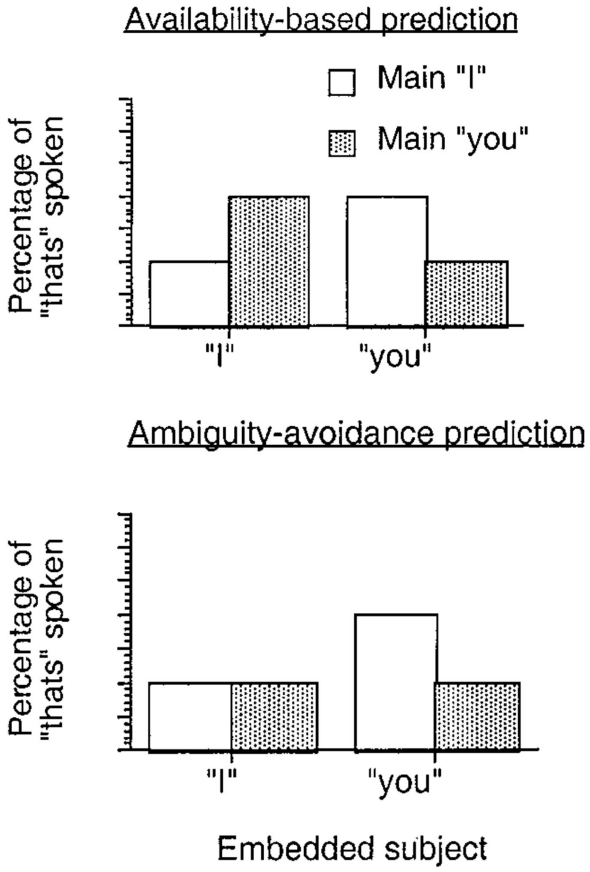


Ferreira & Dell (2000)

Availability-Based Production

The evidence:
Speaker choices in sentences like

a) I knew (that) I missed practice
b) You knew (that) I missed practice
c) I knew (that) you missed practice
d) You knew (that) you missed practice



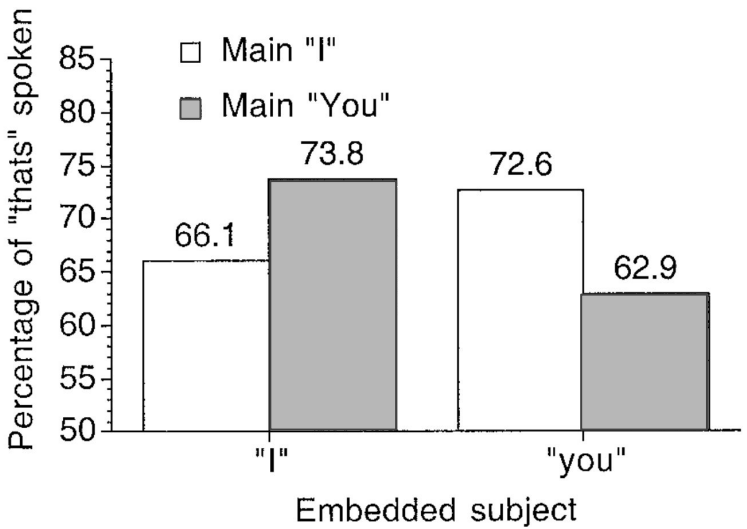
Ferreira & Dell (2000)

Availability-Based Production

The evidence:

Speaker choices in sentences like

a) I knew (that) I missed practice
b) You knew (that) I missed practice
c) I knew (that) you missed practice
d) You knew (that) you missed practice



Ferreira & Dell (2000)

Check for Understanding

English has many phrasal verbs composed of a verb + preposition (though the meaning is often non-compositional!)

take out, pick up, freak out

In transitive phrasal verbs, the object can usually come either before or after the preposition.

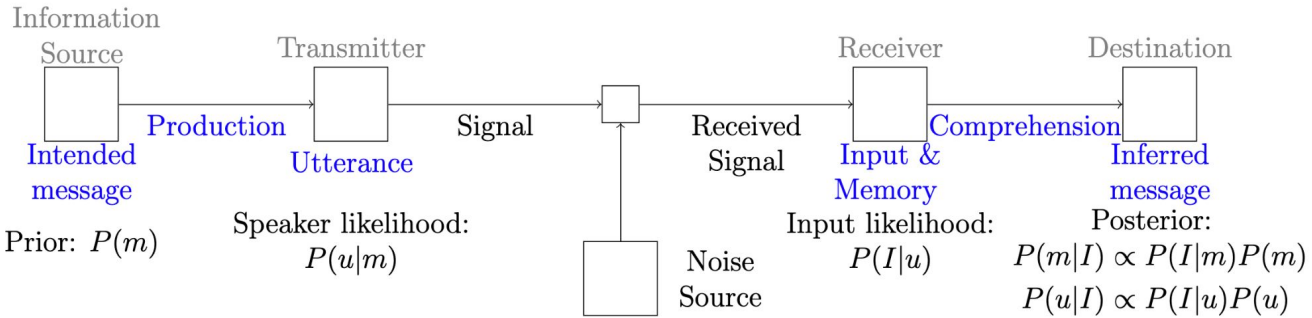
What would availability-based production predict about speaker choices when the object of the verb is LONG vs. SHORT?

pick up the box vs. pick the box up

pick up the heavy box of used books vs. pick the heavy box of used books up

Uniform Information Density

What the theory claims: Speakers distribute information uniformly throughout an utterance; avoid spikes/troughs in surprisal

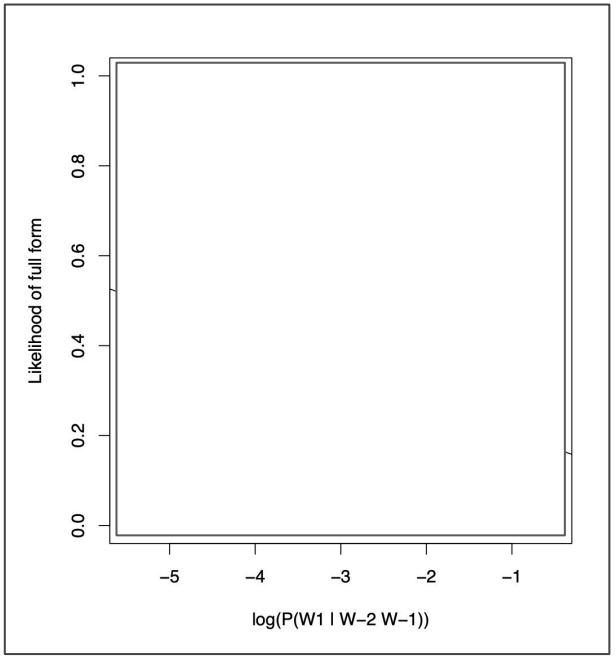


Uniform Information Density: Aylett & Turk (2004), Jaeger (2006), Levy & Jaeger (2007)

Uniform Information Density

The evidence: Optional “that” omission in non-subject relative clauses

a) Did you read the book I wrote?
b) Did you read the book **that** I wrote?
c) Did you read the book inmates wrote?
d) Did you read the book **that** inmates wrote?

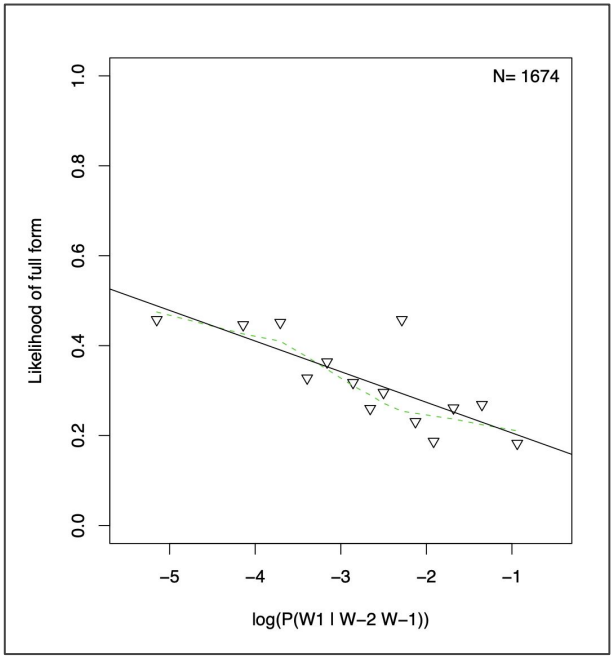


Uniform Information Density: Aylett & Turk (2004), Jaeger (2006), **Levy & Jaeger (2007)**

Uniform Information Density

The evidence: Optional “that” omission in non-subject relative clauses

a) Did you read the book I wrote?
b) Did you read the book **that** I wrote?
c) Did you read the book inmates wrote?
d) Did you read the book **that** inmates wrote?



Uniform Information Density: Aylett & Turk (2004), Jaeger (2006), **Levy & Jaeger (2007)**

Check for Understanding

Imagine you look at a corpus of conversational speech and compute the **surprisal** and **duration** of each word.

What would the UID theory predict about the relationship between surprisal and duration of a word?

Check for Understanding

Factor	ΔR^2 (%)	$F(1,\sim1365)$	Significance (p)
Main factors			
Content vs function (lexical class)	.96	49.0	< .00005
Frequency	.25	12.9	< .0005
Frequency^2	.22	11.1	< .001
Following conditional probability	.65	33.3	< .00005
Previous conditional probability	.11	5.7	< .02
Interactions			
Lexical class × previous conditional probability	.24	12.3	.0005
Lexical class × log rate	.05	2.5	.11

DUR1: Raw syllabic duration

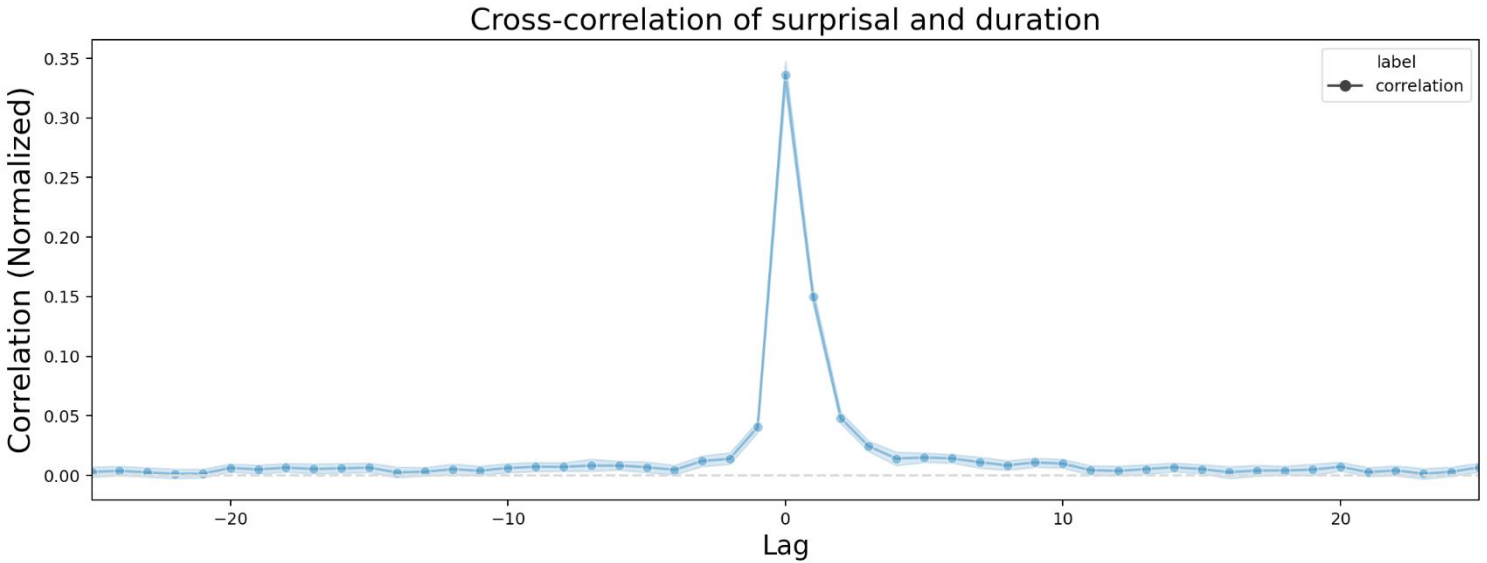
	Regression Results	*r = .6081*	*r² = 0.3698*
Redundancy Factor	Unique Contrib. to r^2	F(1,89531)	*p* value
wf	10.11%	14361.29	.001
trigram	01.93%	2736.84	.001

Table 1
Fixed effects summary for model of Buckeye word durations.

	β	SE	t	$p(\chi^2)$
Intercept	0.0257	0.0057	4.48	–
Baseline duration	0.5879	0.0150	39.32	<0.0001
Syllable count	0.0592	0.0104	5.71	<0.0001
Speech rate	−0.3406	0.0077	−43.97	<0.0001
Bigram prob. given previous	−0.0102	0.0007	−15.00	<0.0001
Bigram prob. given following	−0.0205	0.0007	−30.55	<0.0001
Orthographic length	0.0437	0.0167	2.62	0.0089
Part of Speech = Adjective	0.0033	0.0032	1.04	(<0.0001)
Part of Speech = Adverb	−0.0172	0.0042	−4.09	–
Part of Speech = Verb	−0.0275	0.0022	−12.54	–
Informativity given following	0.0244	0.0023	10.77	<0.0001

Aylett & Turk (2004), Bell et al. (2009), Seyfarth et al. (2014)

Check for Understanding



Case Study 1
Mandarin Classifiers

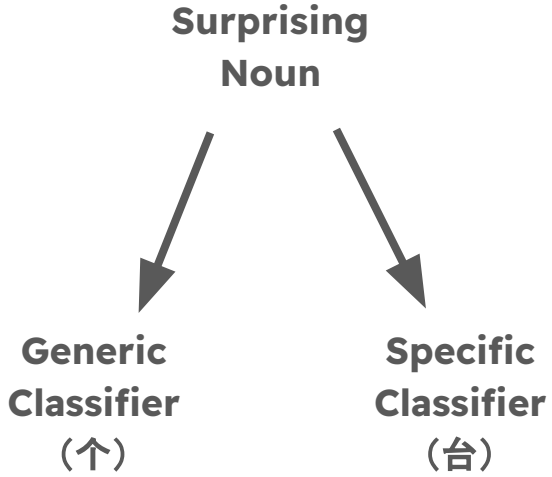
Zhan & Levy (2018)

Mandarin Classifiers

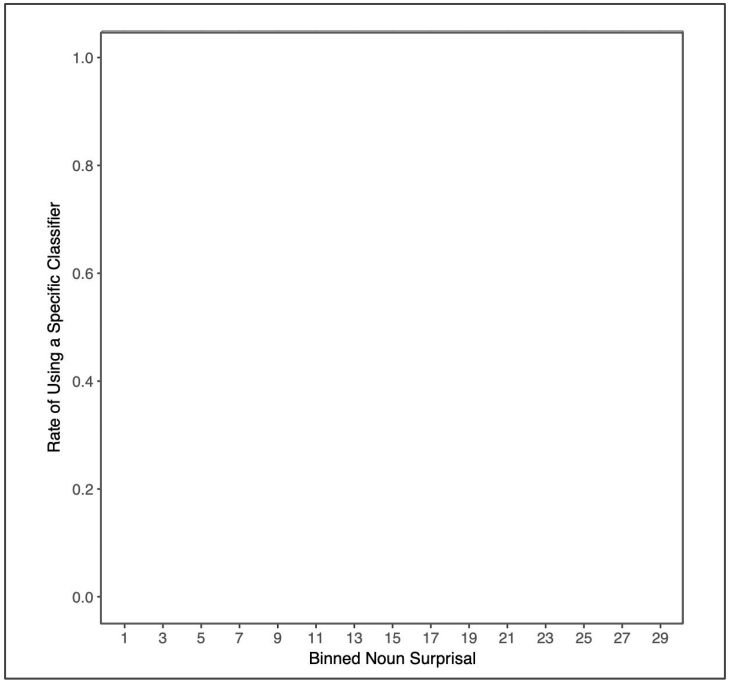
Availability vs. UID

(1) 我 卖了 三 台 电脑
wo mai-le san **tai** diannao
I sold three CL.machinery computer (“I sold three comput-ers”)

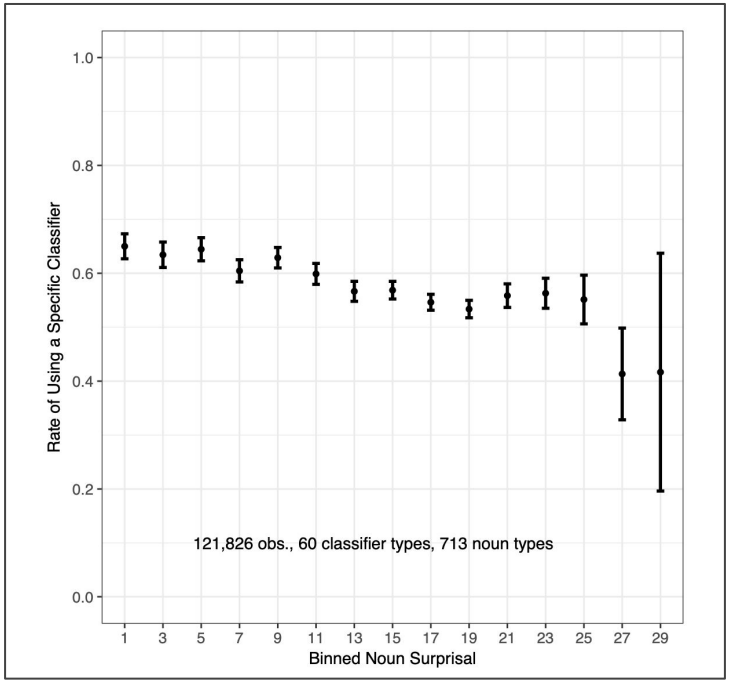
(2) 我 卖了 三 个 电脑
wo mai-le san **ge** diannao
I sold three CL.general computer (“I sold three computers”)



Mandarin Classifiers



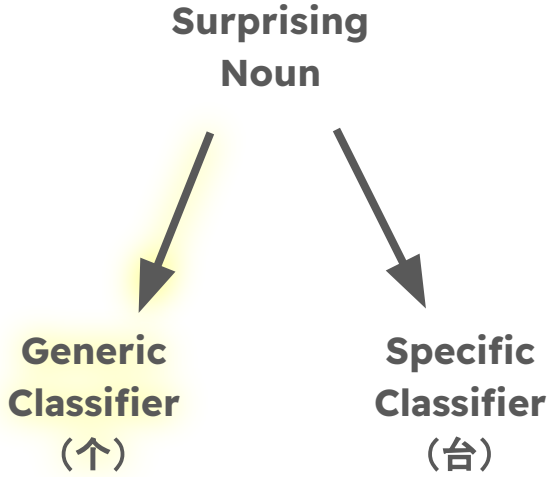
Mandarin Classifiers



Mandarin Classifiers

(1) 我 卖了 三 台 电脑
wo mai-le san **tai** diannao
I sold three CL.machinery computer (“I sold three comput-ers”)

(2) 我 卖了 三 个 电脑
wo mai-le san **ge** diannao
I sold three CL.general computer (“I sold three computers”)



How do you interpret this result?

Case Study 2
Russian Comparatives

Clark, Wilcox, Gibson & Levy (2022)

Russian Comparatives

Availability vs. UID

‘This is more important than your work’

Explicit 🐢 Construction

это	важнее	чем	твоя	работа
eto	*vazhnyeye*	*chem*	*tvoya*	*rabota*
this.NOM.SG	important.COMPARATIVE	than	your.NOM.SG	work.NOM.SG

Russian Comparatives

Availability vs. UID

‘This is more important than your work’

Explicit 🐢 Construction

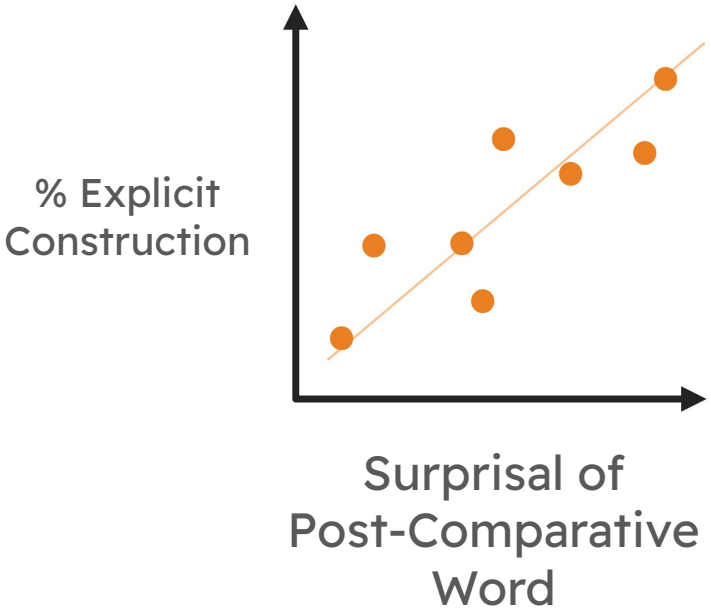
это	важнее	чем	твоя	работа
eto	*vazhnyeye*	*chem*	*tvoya*	*rabota*
this.NOM.SG	important.COMPARATIVE	than	your.NOM.SG	work.NOM.SG

Short ⚡ Construction

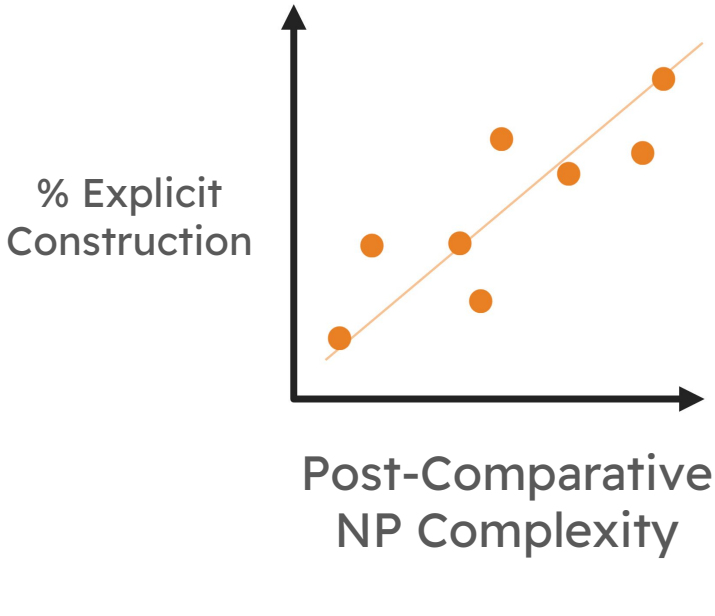
это	важнее	твоей	работы
eto	*vazhnyeye*	*tvoyej*	*raboty*
this.NOM.SG	important.COMPARATIVE	your.GEN.SG	work.GEN.SG

Russian Comparatives

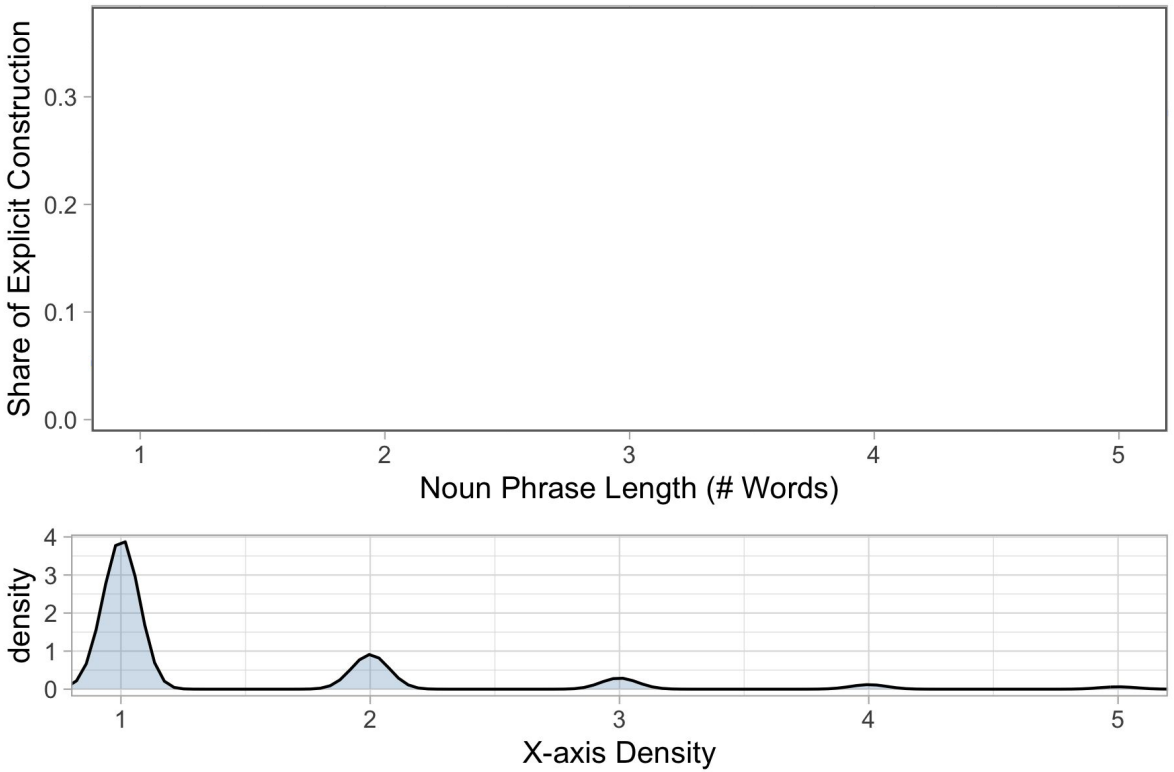
Prediction of UID



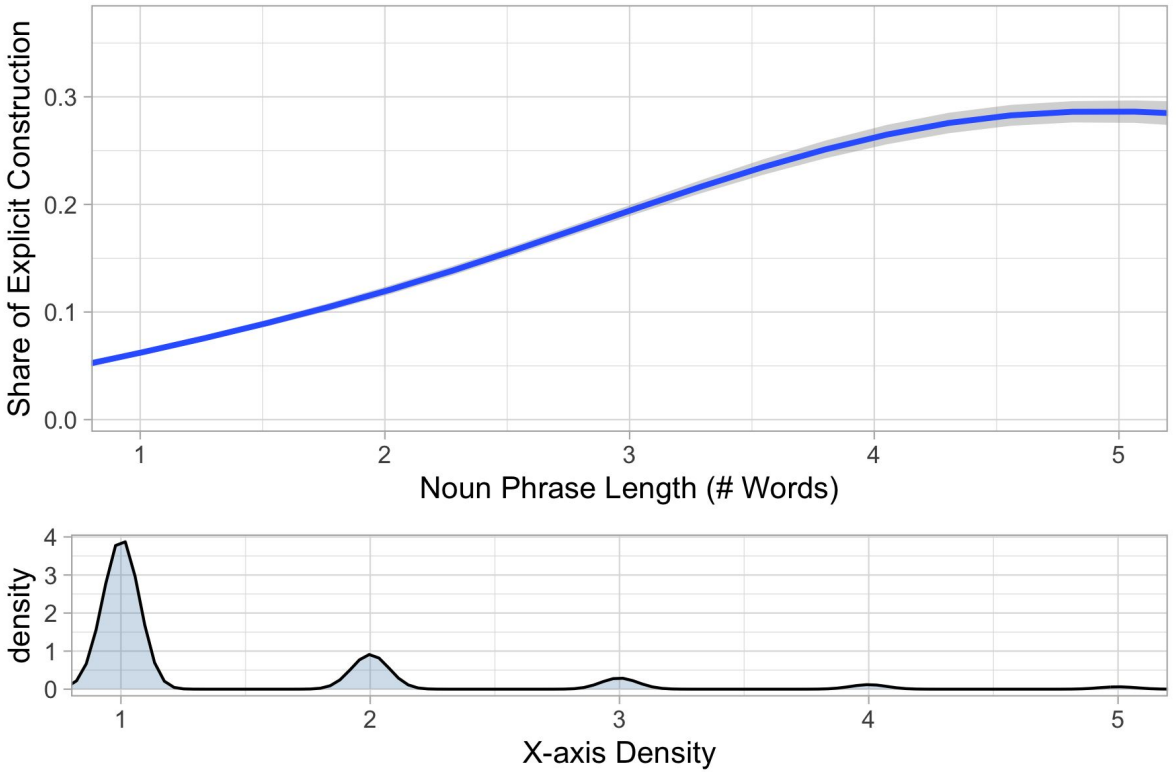
Prediction of Availability



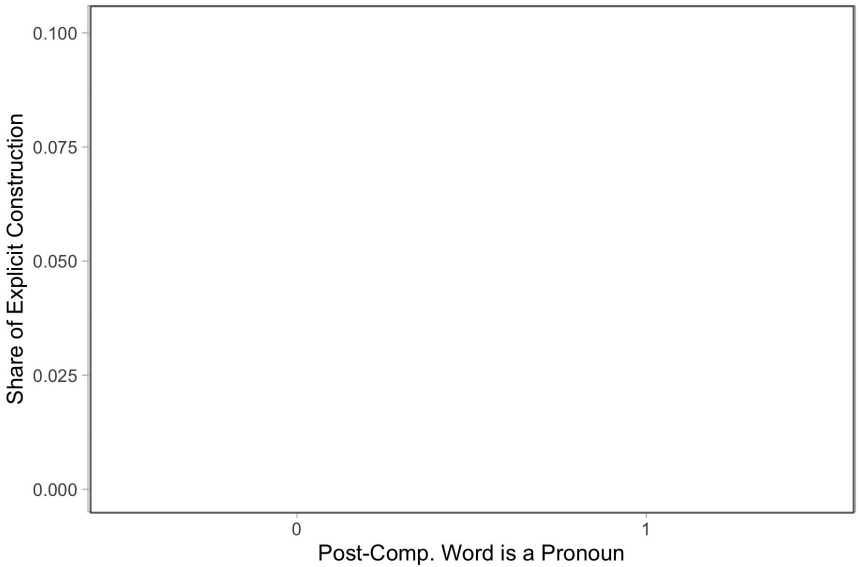
Noun Phrase Complexity

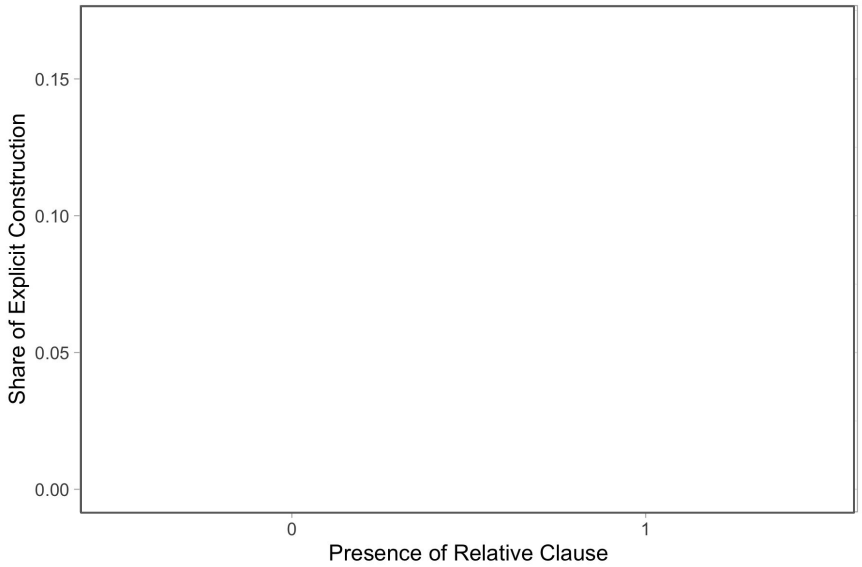


Noun Phrase Complexity

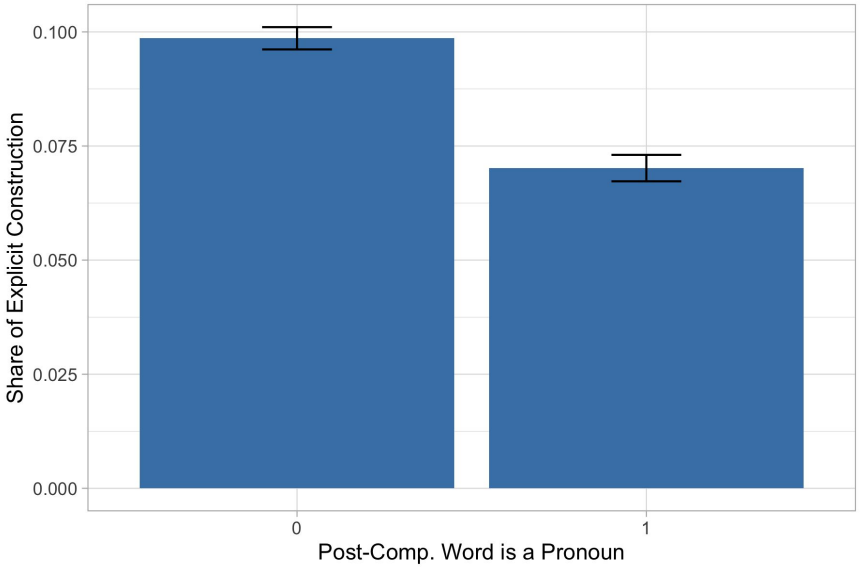


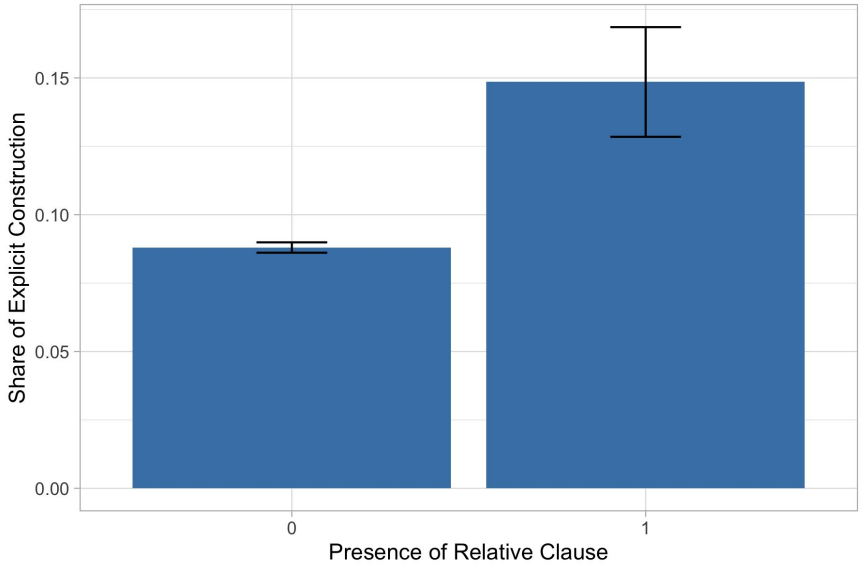
Noun Phrase Complexity



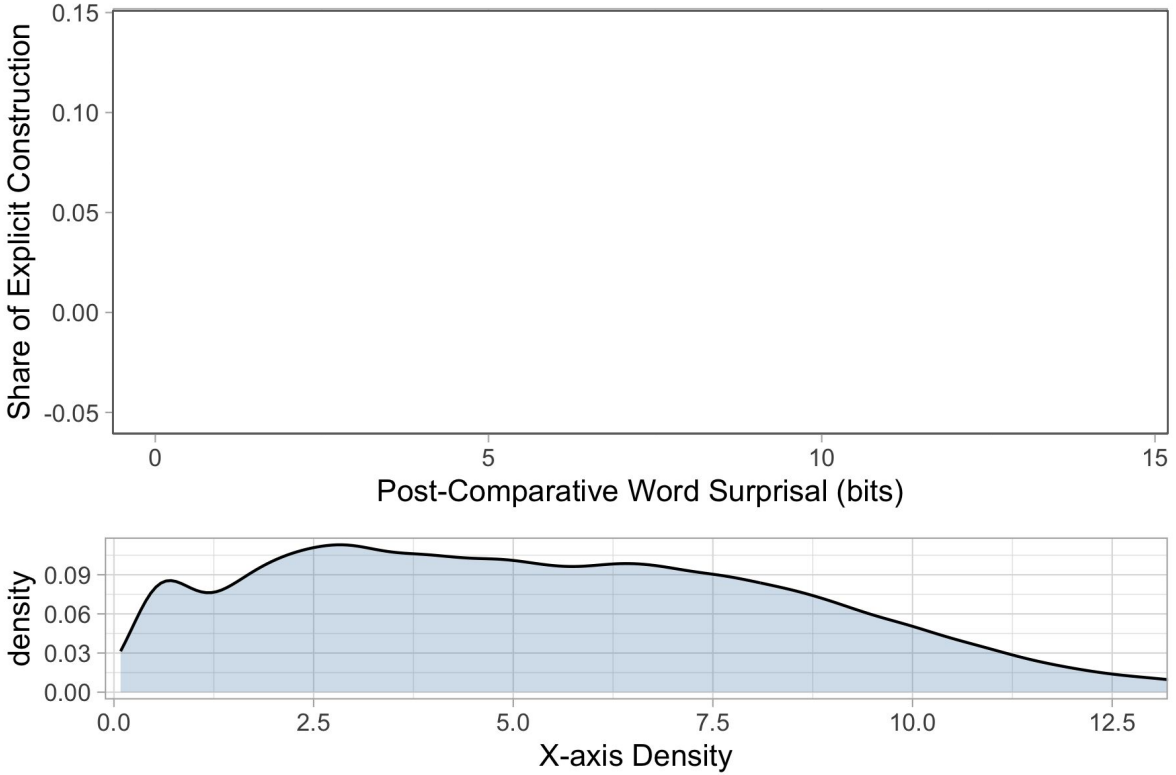


Noun Phrase Complexity

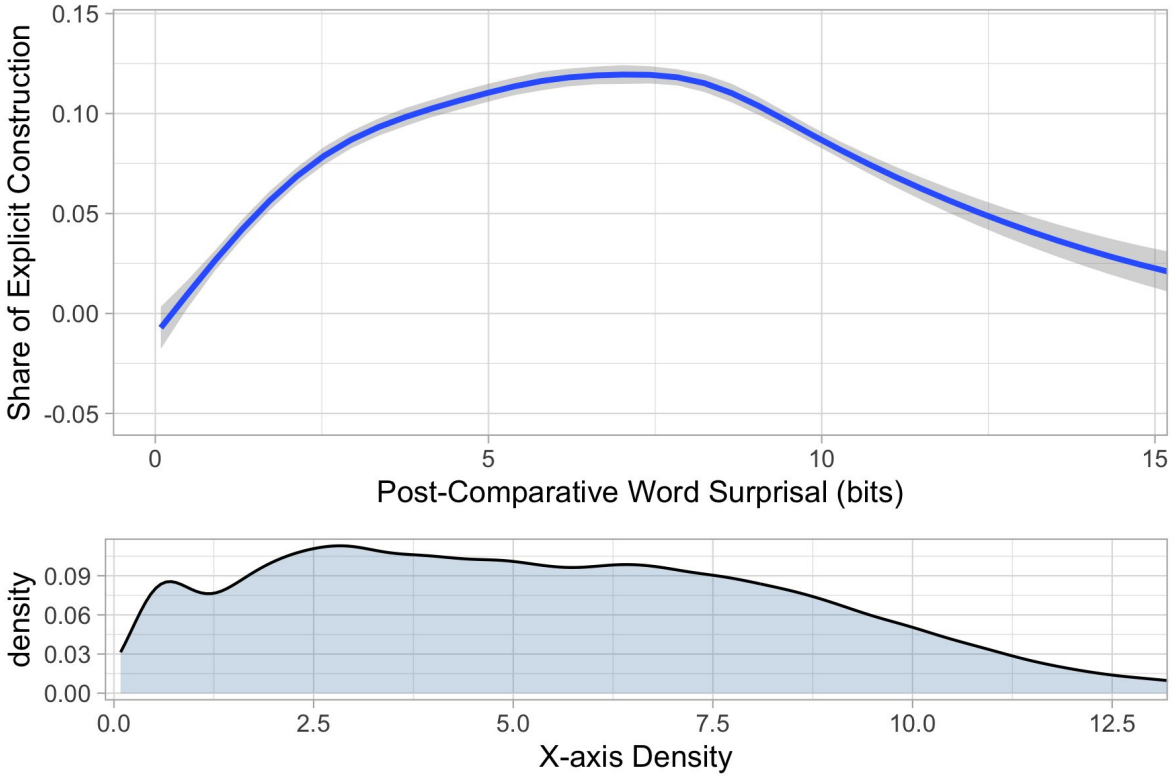




Surprisal


Share of Explicit Construction
0.15
0.10
0.05
0.00
-0.05
0
5
10
15
Post-Comparative Word Surprisal (bits)
density
0.09
0.06
0.03
0.00
0.0
2.5
5.0
7.5
10.0
12.5
X-axis Density

Surprisal

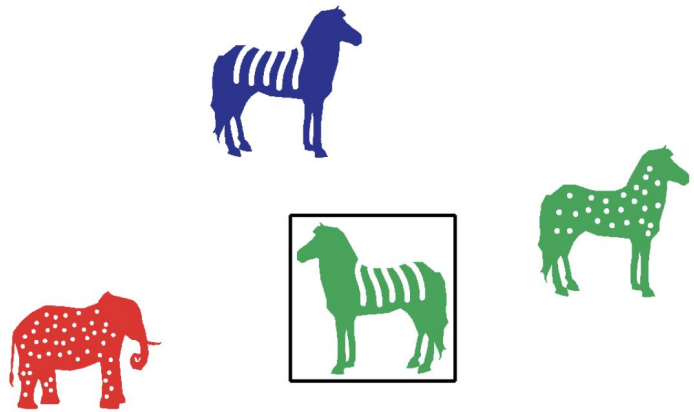

Share of Explicit Construction
0.15
0.10
0.05
0.00
-0.05
0
5
10
15
Post-Comparative Word Surprisal (bits)
density
0.09
0.06
0.03
0.00
0.0
2.5
5.0
7.5
10.0
12.5
X-axis Density

Production Experiment

Brainstorming: how might we measure people’s usage of the explicit vs. short construction?

Production Experiment

- Russian native speakers (N=100) recruited via Prolific
- Visual stimulus paired with sentence completion task
- Scenes are manipulated to elicit more complex noun phrases (due to disambiguation)

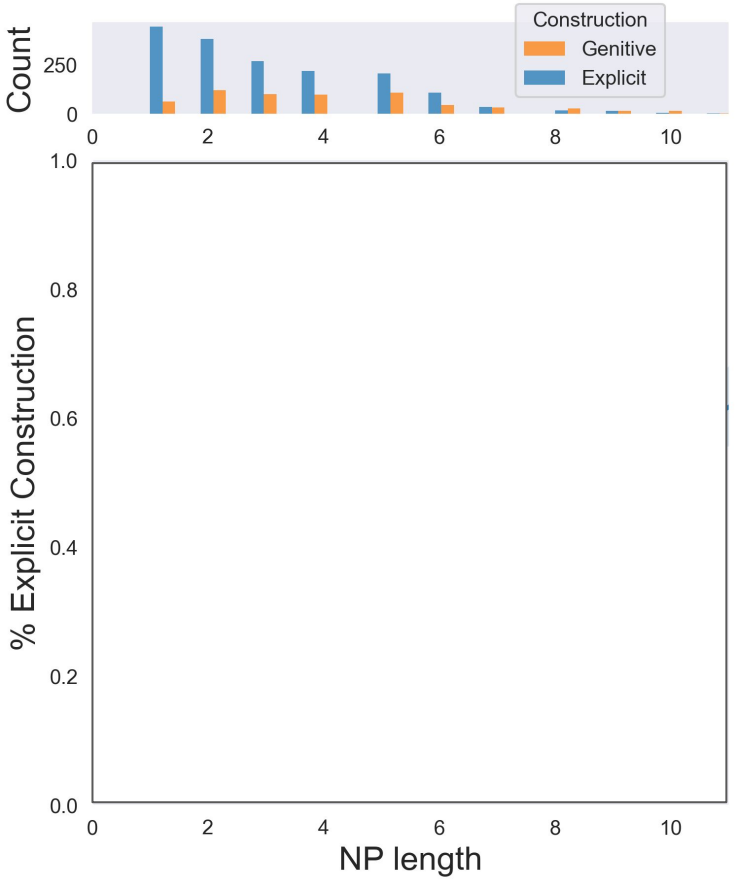


Саша быстрее ...

Завершите предложение:

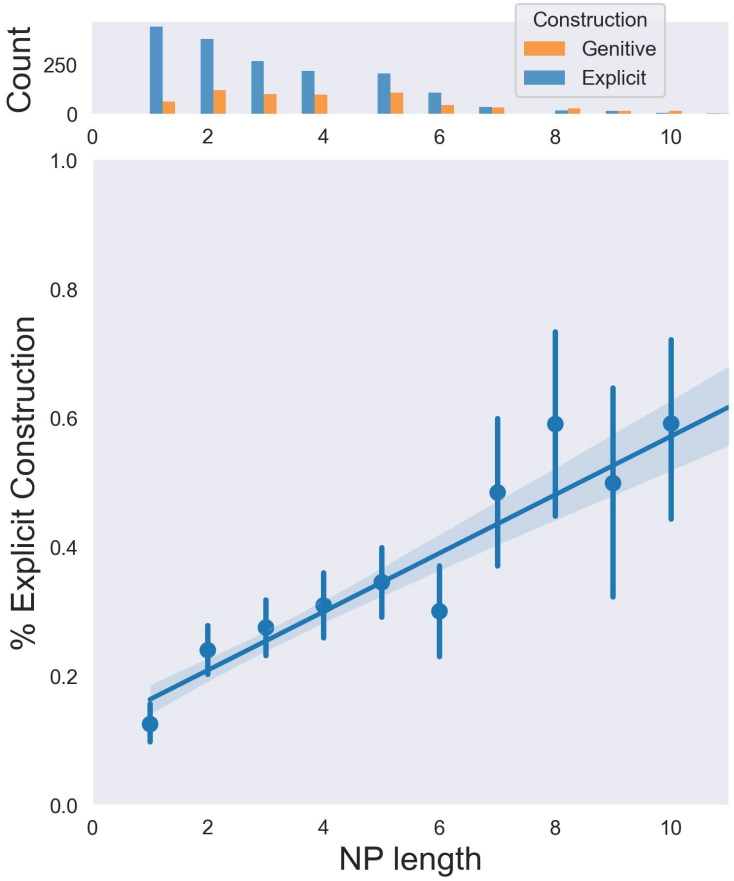
→ Следующая страница

Production Experiment



Production Experiment

- Noun phrase length was a strong predictor of using the explicit construction



Summary so far

- We have discussed two hypothesized pressures on language production: availability and uniform information density
- We have assessed evidence for these two pressures using data from behavioral language production experiments
- There seems to be evidence supporting both pressures, and they are not mutually exclusive

Towards a Unified Theory

Theory Desiderata

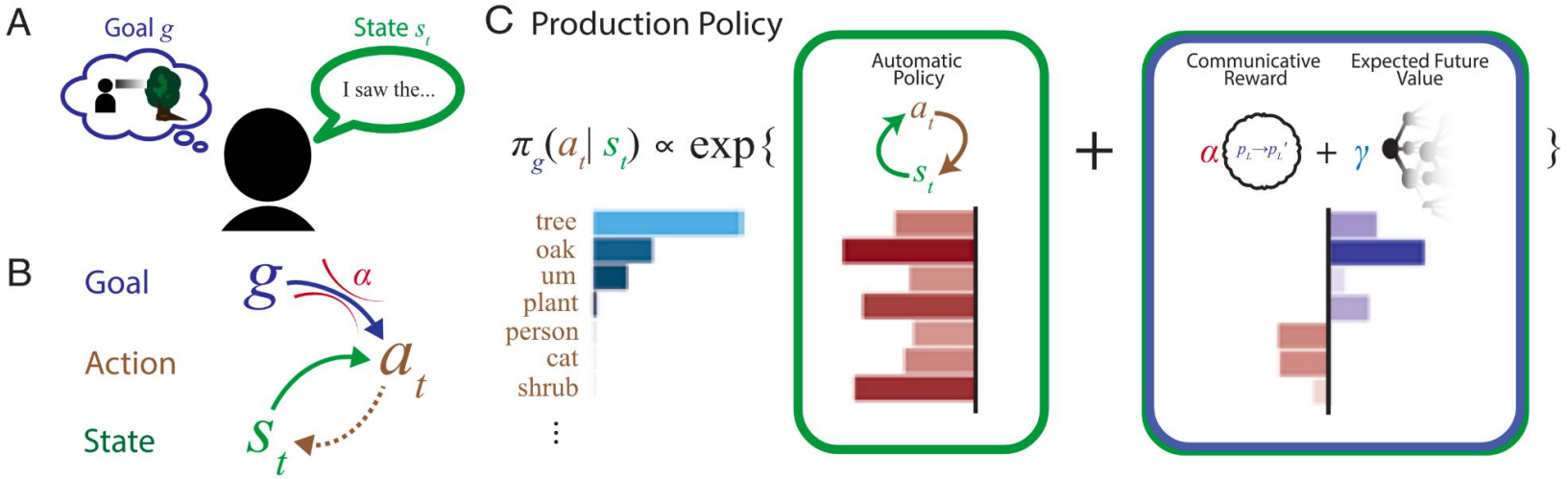
- Account for speaker choices across a range of behavioral phenomena
- Reconcile existence of both availability-based and information-theoretic effects
- Be computationally implemented and testable

Rate-Distortion Theory of Control (RDC)

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Information-theoretic principles in incremental language production

Richard Futrell[a,1]



Futrell (2023)

Rate-Distortion Theory of Control (RDC)

Value function trades off communicative reward and control information

$$V_g^{\pi}(s) = \mathop{\mathbb{E}}_{a \sim \pi_g(\cdot|s)} \Big[\underbrace{\alpha R_g(a \mid s)}_{\text{Reward}} + \underbrace{\gamma V_g^{\pi}(s')}_{\text{Future value}} \Big] - \underbrace{\mathrm{I}_{\pi}[A : g \mid s]}_{\text{Control information}}$$

RDC Example 1: Dative Alternation

Consider the following two constructions:

Send ***Adele*** ***a card***

Send ***a card*** ***to*** ***Adele***

RDC Example 1: Dative Alternation

Choice between construction xy and yx

Meaning is equivalent given goal: $R_g(xy) = R_g(yx)$

$$\pi_g(x) = \sigma\left(\underbrace{\ln \frac{\pi_0(x)}{\pi_0(y)}}_{\text{Predictability}} + \underbrace{\alpha\,(1-\gamma)\,\Delta R_g}_{\text{Reward differential}} + \underbrace{\gamma \ln \frac{\pi_0(y \mid x)}{\pi_0(x \mid y)}}_{\text{Planning}}\right)$$

$$\pi_0(a \mid s) = \sum_g p(g \mid s)\pi_g(a \mid s)$$

"Is action x more frequent than action y in context across communicative goals?"

GPT-3

RDC Example 1: Dative Alternation

Choice between construction xy and yx

Meaning is equivalent given goal: $R_g(xy) = R_g(yx)$

$$\pi_g(x) = \sigma\left(\underbrace{\ln \frac{\pi_0(x)}{\pi_0(y)}}_{\text{Predictability}} + \underbrace{\alpha\,(1-\gamma)\,\Delta R_g}_{\text{Reward differential}} + \underbrace{\gamma \ln \frac{\pi_0(y \mid x)}{\pi_0(x \mid y)}}_{\text{Planning}}\right)$$

$$\pi_0(a \mid s) = \sum_g p(g \mid s)\pi_g(a \mid s)$$

"Does action x make action y more predictable as the second element?"

GPT-3

RDC Example 1: Dative Alternation

Table 1. Fixed-effect coefficients of a Bayesian logistic regression (75) predicting double object vs. prepositional object form for the dative alternation

Predictor	Coefficient	95% Posterior CrI
(Intercept)	−0.68	[−1.64, 0.23]
Verb Semantics	−0.36	[−1.10, 0.38]
Length	−0.14	[−0.25, −0.05]
Definiteness	0.89	[0.31, 1.61]
Animacy	1.24	[0.35, 2.24]
Predictability	0.95	[0.75, 1.25]
Planning	0.87	[0.68, 1.15]

Verb Semantics indicates that the meaning annotation is "Communication" rather than "Abstract" or "Transfer." Length, Definiteness, and Animacy predictors are values for the recipient NP minus the theme NP. Definiteness is coded as indefinite = 0, definite = 1, pronominal = 2. Length is length in characters of each phrase. Predictability and Planning are as in Eq. **10**, using GPT-3 for the automatic policy π_0.

RDC Example 2: Filled Pauses

Filled pauses like “uh” and “um”

Convey no intrinsic reward, but can still provide value according to RDC

$$\frac{\pi_g(e)}{\pi_g(a_g)} = \exp\left\{\underbrace{\ln \frac{\pi_0(e)}{\pi_0(a_g)}}_{\text{Predictability}} - \underbrace{\alpha\,(1-\gamma)\,R_g(a_g)}_{\text{Target reward}} + \underbrace{\gamma \ln \pi_0(a_g \mid e)}_{\text{Planning}}\right\}$$

This behavioral matches known predictors of disfluencies:

Schachter et al. (1991), Hartsuiker & Notebaert (2010), Harmon & Kapatsinski (2015)

RDC Example 2: Filled Pauses

Interactive equation:

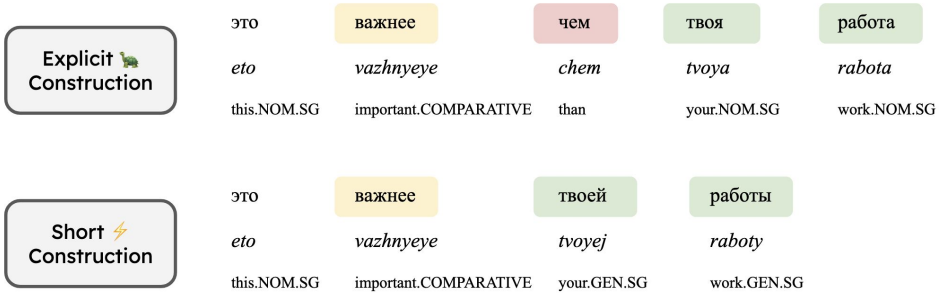
https://disfluency.streamlit.app

Discussion

How might the RDC Theory explain the behavior described earlier in the Mandarin Classifiers or Russian Comparatives case studies?

(1) 我 卖了 三 台 电脑
wo mai-le san **tai** diannao
I sold three CL.machinery computer (“I sold three comput-ers”)

(2) 我 卖了 三 个 电脑
wo mai-le san **ge** diannao
I sold three CL.general computer (“I sold three computers”)



Goals for Today, Revisited

- Assess the influence of two potential pressures on language production:
  - **Ease of production**
  - **Robust communication**
- Evaluate the design of language production experiments
- Use computational models to explain human behavior in language production

Thanks!

References:

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Bock, K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, *26*(2), 119–137. https://doi.org/10.1016/0749-596X(87)90120-3

Clark, T., Wilcox, E. G., Gibson, E., & Levy, R. (2022). Evidence for Availability Effects on Speaker Choice in the Russian Comparative Alternation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/1q19f8vt

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730

Florian Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Futrell, R. (2023). Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences*, *120*(39), e2220593120. https://doi.org/10.1073/pnas.2220593120

Levy, R., & Jaeger, T. F. (2006). *Speakers optimize information density through syntactic reduction*. *19*, 849–856.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Zhan, M., & Levy, R. (2019). Availability-Based Production Predicts Speakers' Real-time Choices of Mandarin Classifiers. *arXiv: Computation and Language*.