



# Syntactic Planning and the Information Threshold

---

Joel C. Wallenberg<sup>1</sup>

(with collaborators: Rachael Bailes<sup>2</sup>, Anton K. Ingason<sup>3</sup> and Christine Cuskley<sup>2</sup>)

<sup>1</sup> University of York, UK

<sup>2</sup> Newcastle University, UK

<sup>3</sup> University of Iceland, Iceland

# “Constraints on the Adaptiveness of Information in Language” (CAIL)

- <https://cail-project.github.io/>
- Collaboration with Christine Cuskley and Rachael Bailes
- ESRC Secondary Data Analysis Initiative (SDAI), grant #ES/T005955/1

# Today's Talk

Information Uniformity and Language

What It's For: Noise Resistance

What It Does: Syntactic Planning

Diachronic Study 1: Information Theoretic Constant Rate Effects:  
OV to VO

Diachronic Study 2: Adding Adjunct Fronting with V2

Theoretical Implications and Future Work

# Information Uniformity and Language

What It's For: Noise Resistance

What It Does: Syntactic Planning

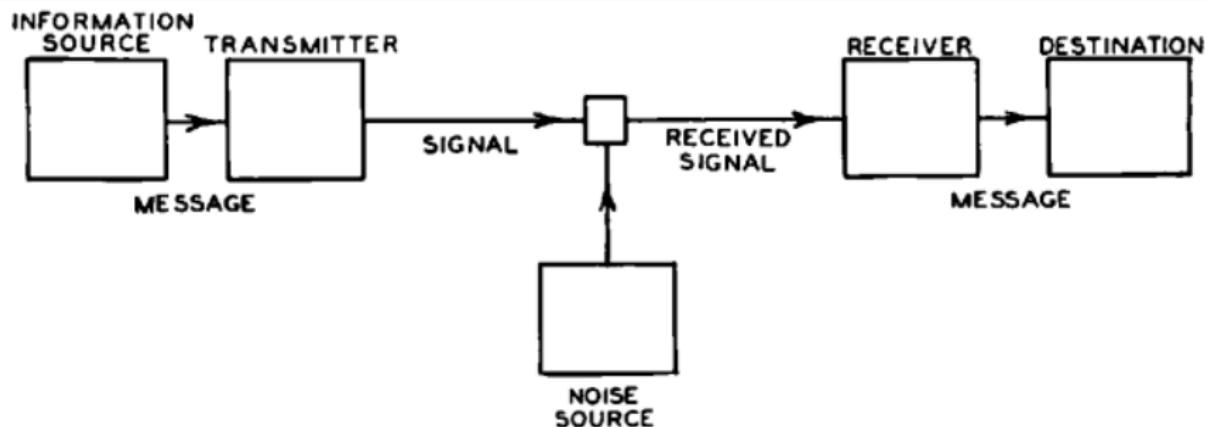
Diachronic Study 1: Information Theoretic Constant Rate

Effects: OV to VO

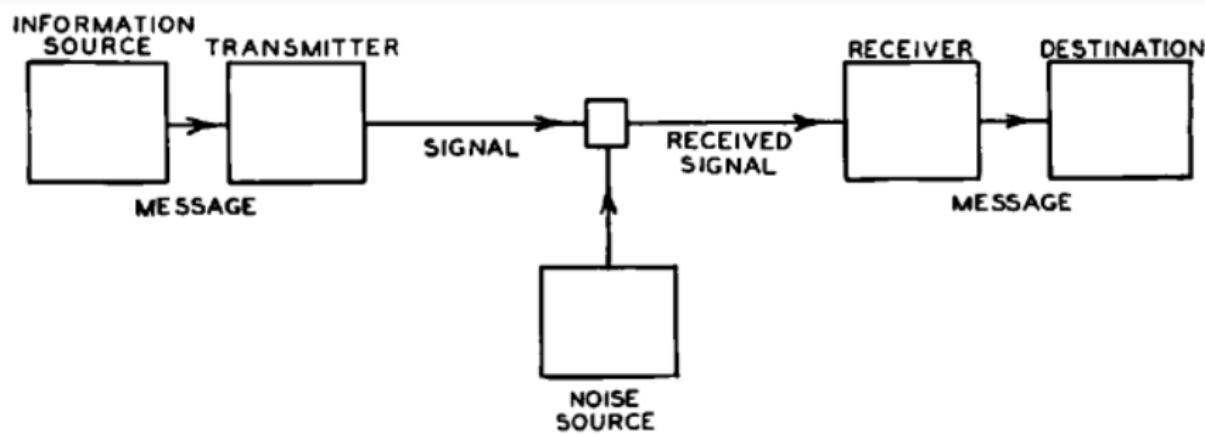
Diachronic Study 2: Adding Adjunct Fronting with V2

Theoretical Implications and Future Work

# Information theory and language

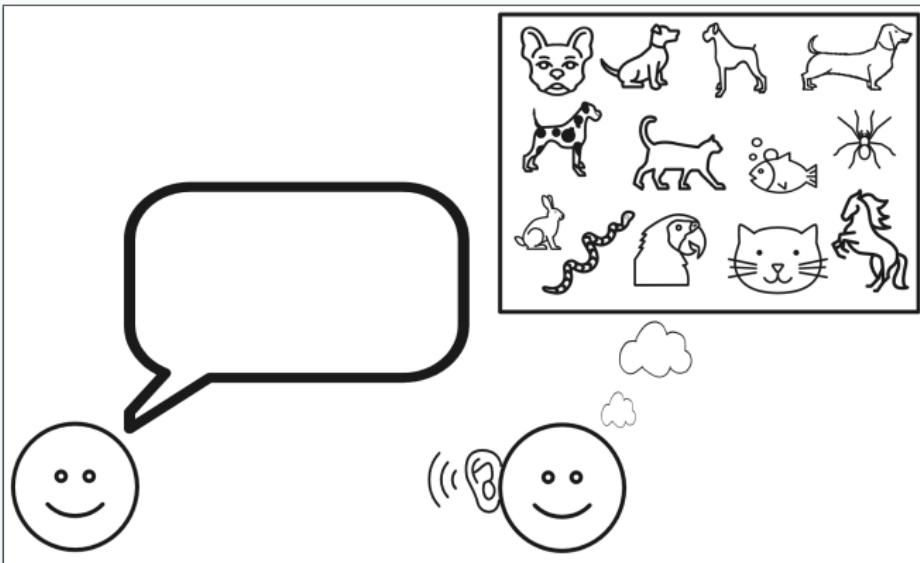


# Information theory and language



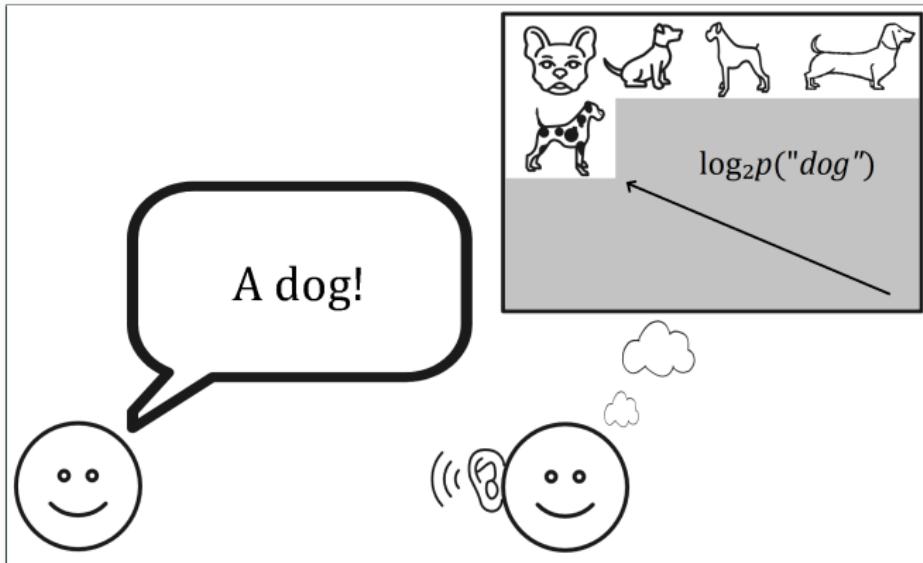
- **Key Insight:** The amount of information a sender can theoretically communicate about an event is the uncertainty ("entropy") the receiver has about the event beforehand, which may be reduced by a signal (Hartley, 1928; Shannon, 1948).

# Information content of words



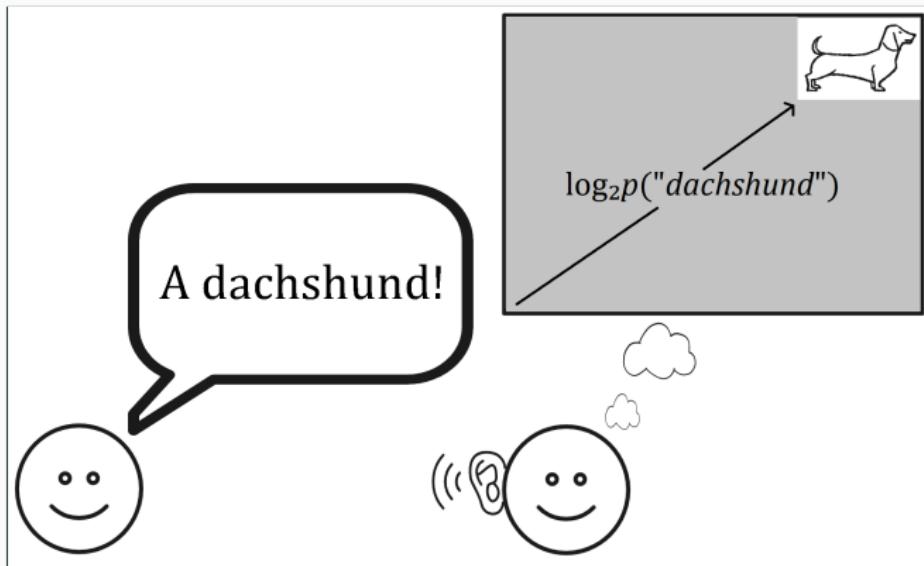
Receiver begins with a **set of expected outcomes** that could result from the message they receive. This set of expected outcomes is the amount of **uncertainty** they have about the message.

# Information content of words



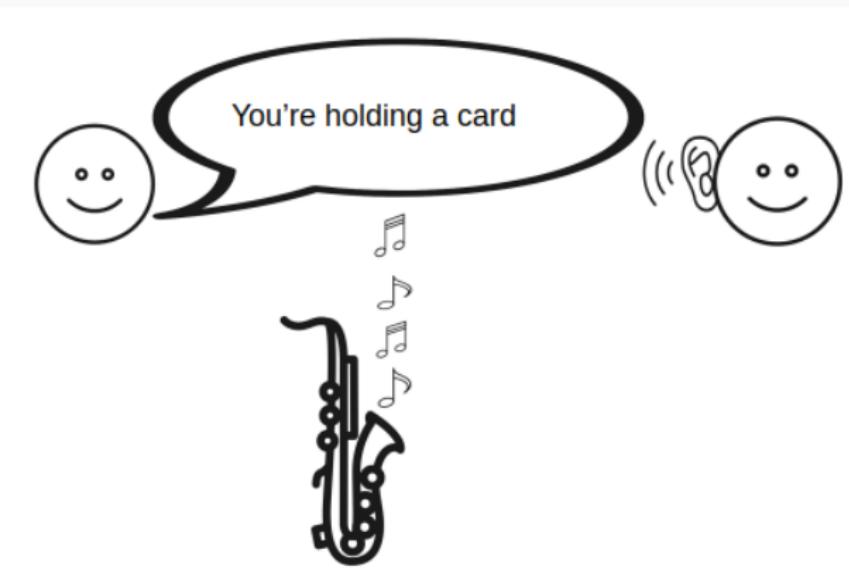
Sender uses a word that **reduces the receiver's uncertainty** by some amount proportional to the word's frequency. Here, the word is relatively frequent.

# Information content of words

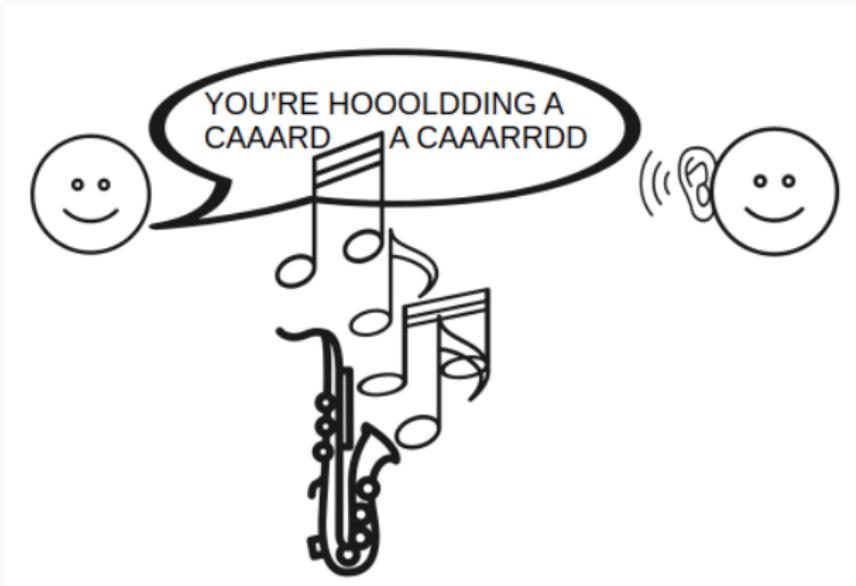


The more infrequent the word, the more uncertainty is reduced.

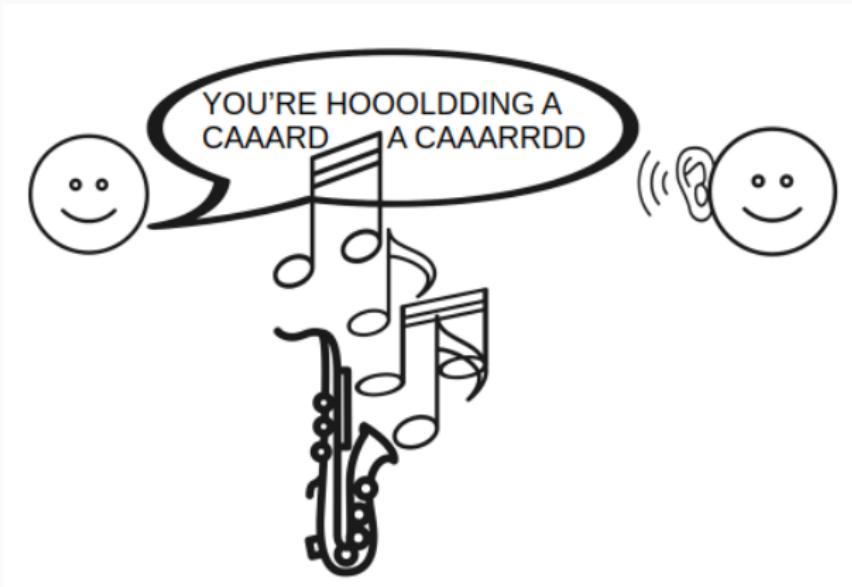
## Noise and redundancy



## Preference for redundancy (and low surprisal)



## Preference for redundancy (and low surprisal)



(Shannon, 1948; Fenk & Fenk, 1980; Fenk-Oczlon, 2001; Aylett & Turk, 2004; Levy & Jaeger, 2007; Frank & Jaeger, 2008; Jaeger, 2010; Turk, 2010; Chingacham et al., 2023)

## Information Uniformity and Language

**What It's For: Noise Resistance**

What It Does: Syntactic Planning

Diachronic Study 1: Information Theoretic Constant Rate  
Effects: OV to VO

Diachronic Study 2: Adding Adjunct Fronting with V2

Theoretical Implications and Future Work

## Reordering for uniformity: functional for noise resistance?

- Signalling error reduced arbitrarily given sufficient redundancy (Shannon, 1948)
- Does reordering confer noise resistance?

# Reordering for uniformity: functional for noise resistance?

- Signalling error reduced arbitrarily given sufficient redundancy (Shannon, 1948)
- Does reordering confer noise resistance?



Cognition

Volume 214, September 2021, 104754



## Noise resistance in communication: Quantifying uniformity and optimality

Christine Cuskley, Rachael Bailes, Joel Wallenberg

Show more ▾

+ Add to Mendeley   Share   Cite

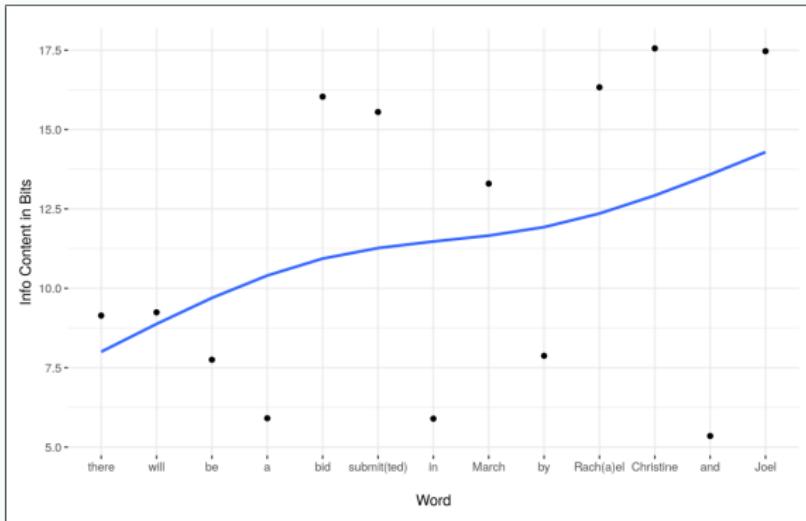
---

<https://doi.org/10.1016/j.cognition.2021.104754>

Get rights and content

# Information distributions of sentences

Deriving the information content of each word means we can derive a **distribution of information content values** for a given sentence.



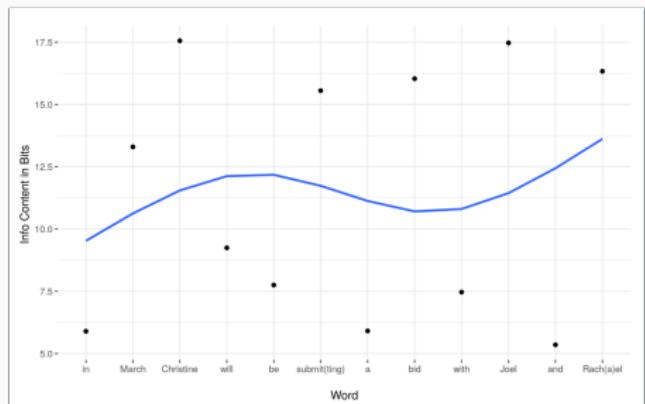
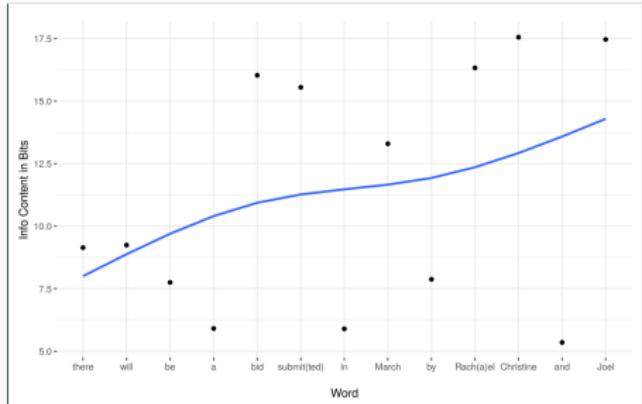
## Uniformity of information distributions

---

These distributions can vary in terms of how the information is spread or clustered. The **order of elements** in a sentence derives **more uniform** or **more asymmetric** distributions of information (Cuskley, Bailes, & Wallenberg, 2021).

# Uniformity of information distributions

These distributions can vary in terms of how the information is spread or clustered. The **order of elements** in a sentence derives **more uniform** or **more asymmetric** distributions of information (Cuskley et al., 2021).



We can measure how uniform an information distribution is with Deviation of the Rolling Mean (DORM) (Cuskley et al., 2021).

## DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	6.15	10.4

## DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
<b>6.79</b>	<b>6.15</b>	10.1	9.25	6.15	10.4
	<b>6.47</b>				

# DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	<b>6.15</b>	<b>10.1</b>	9.25	6.15	10.4
6.47		<b>8.12</b>			

# DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	<b>10.1</b>	<b>9.25</b>	6.15	10.4
6.47	8.12		<b>9.67</b>		

## DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	<b>9.25</b>	<b>6.15</b>	10.4
6.47	8.12	9.67		<b>7.70</b>	

# DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	<b>6.15</b>	<b>10.4</b>
6.47	8.12	9.67	7.70		<b>8.29</b>

## DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	6.15	10.4
6.47	8.12	9.67	7.70	8.29	

Sample variance of rolling means = 1.33 bits

**low DORM → more uniform**

**high DORM → more lopsided**

## UIDO: Uniform information density optimisation

- An algorithm that finds the most uniform/dispersed/smooth distribution of a given set of values (Cuskley et al., 2021).
- Not absolute lowest DORM possible; but robust enough and computationally tractable.
- Useful **calibration** for utterance DORMs & establishing a baseline.

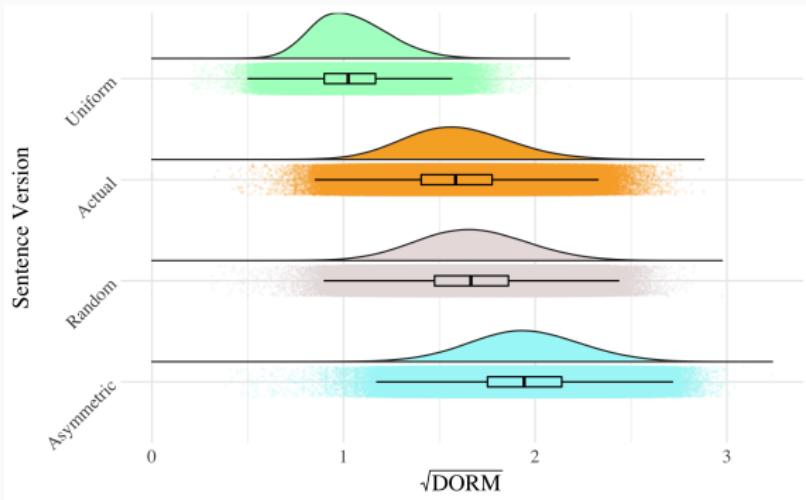
## UIDO: Uniform information density optimization

- For a given array of values (e.g. information values of words in a sentence):
  - the array of information content values is ordered greatest to least
  - starting with the second and penultimate value in the array and moving inward, every other number is swapped, mixing up the large and small values
  - DORM is calculated for the resulting array
  - if this is lower than the original DORM, the array is kept
  - otherwise, the algorithm proceeds with the original array.
- Useful calibration for interpreting utterance DORMs & establishing a baseline, as well as current study use

## Corpus-Based Simulation

- The Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE; Ecay, 2015)
- 628,083 sentences exactly 10 words in length
- DORMs for 4 versions: maximally uniform/dispersed, original, random, maximally asymmetric

# Uniformity of distributions



**Figure 1:** Distributions of the square root of DORM values for 628,083 sentences prior to noise simulation.

## Single-unit and clustered noise

---

- Single-unit noise: three random single units “knocked out” per distribution per trial.

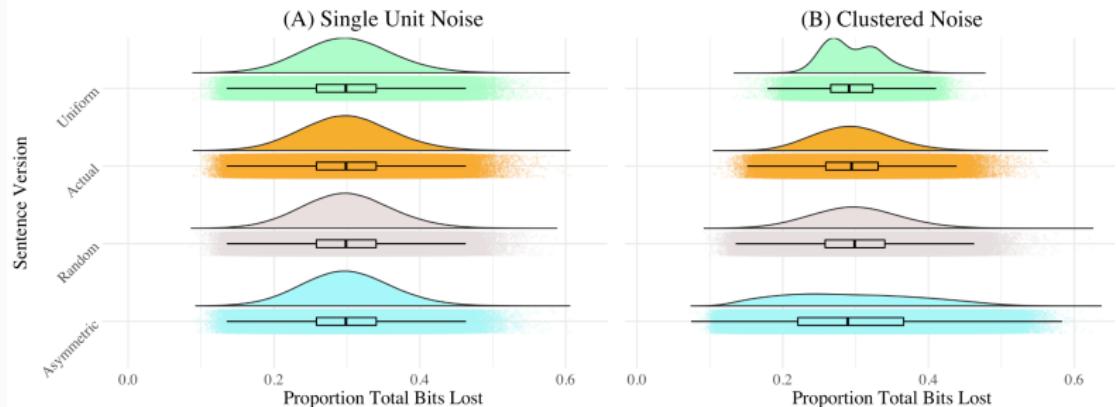
## Single-unit and clustered noise

- Single-unit noise: three random single units “knocked out” per distribution per trial.
- Clustered noise: three sequential units “knocked out” per distribution per trial.
  - More similar to naturally occurring noise events that span multiple linguistic units.

## Single-unit and clustered noise

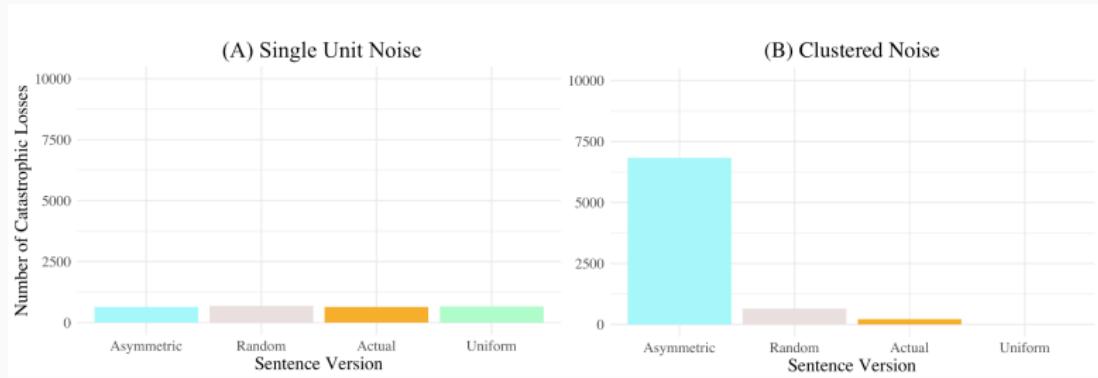
- Single-unit noise: three random single units “knocked out” per distribution per trial.
- Clustered noise: three sequential units “knocked out” per distribution per trial.
  - More similar to naturally occurring noise events that span multiple linguistic units.
- Knocked out item positions matched across all 4 versions of all sentences in each trial.

# Information loss in conditions of noise



**Figure 2:** Distributions of the proportion of information lost in 628,083 trials in each condition under 3 single-word noises (A) and clustered noise (B)

# Catastrophic Failures



**Figure 3:** The number of sentences with “catastrophic” information failure in each condition (a noise event knocking out  $\geq 50\%$  of the total information content) under 3 single-word noises (a) and clustered noise (b)

## Interim Conclusions: What It's For

---

- Hyperdispersing information prevents catastrophic losses,

## Interim Conclusions: What It's For

---

- Hyperdispersing information prevents catastrophic losses,
- ...particularly where continuous noise spans multiple categorical linguistic units.

## Interim Conclusions: What It's For

- Hyperdispersing information prevents catastrophic losses,
- ...particularly where continuous noise spans multiple categorical linguistic units.
- More uniform orders may decrease overall info loss, probably because of Zipfian vocabularies.

## Interim Speculations: Language and “Ruin”

- Thresholds: metaphor of insurance reserving and underwriting.
- Signalling happens between every linguistic level, as does noise (Aylett & Turk, 2004).

**acoustic** → **segmental/allophonic** → **phonemic** → **morphemic**  
→ **morphosyntactic** → **propositional/utterance-function**

- Humans make use of inference at every level, so do not need perfect **mutual information**, but do need some.

## Information Uniformity and Language

What It's For: Noise Resistance

What It Does: Syntactic Planning

Diachronic Study 1: Information Theoretic Constant Rate  
Effects: OV to VO

Diachronic Study 2: Adding Adjunct Fronting with V2

Theoretical Implications and Future Work

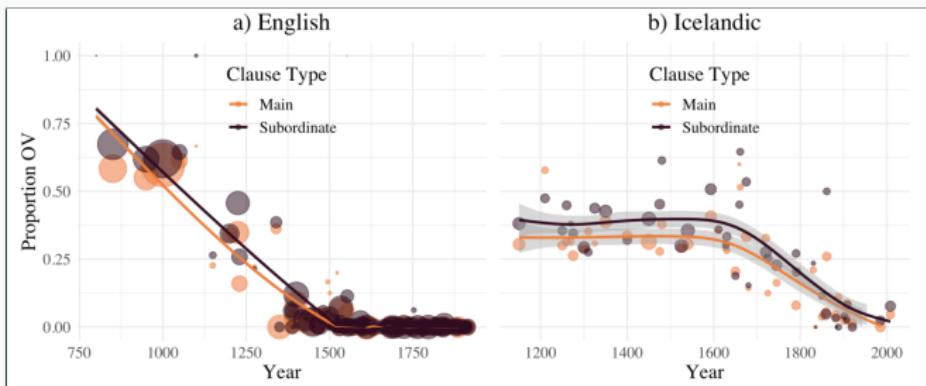
## Diachronic Study 1: Constant Rate Effects

---

- When a change is in progress for a given linguistic variable, some contexts favour one of its variants over another **without affecting the change** (Kroch, 1989).

# Diachronic Study 1: Constant Rate Effects

- When a change is in progress for a given linguistic variable, some contexts favour one of its variants over another **without affecting the change** (Kroch, 1989).



e.g. In the change from OV to VO in English & Icelandic, OV is consistently favoured in subordinate clauses throughout the steady decrease in its overall proportion (see also Pintzuk & Taylor 2006).

# CRE in English and Icelandic OV to VO



*languages*



*Article*

## Smooth Signals and Syntactic Change

Joel C. Wallenberg <sup>1,\*</sup>, Rachael Bailes <sup>1</sup>, Christine Cuskley <sup>1</sup> and Anton Karl Ingason <sup>2,\*</sup>

- Based on this account of information uniformity, (Wallenberg, Bailes, Cuskley, & Ingason, 2021) predicted a previously undetected argument-type effect in the English and Icelandic OV to VO changes.

# CRE in English and Icelandic OV to VO



Article

## Smooth Signals and Syntactic Change

Joel C. Wallenberg <sup>1,\*</sup>, Rachael Bailes <sup>1</sup>, Christine Cuskley <sup>1</sup> and Anton Karl Ingason <sup>2,\*</sup>

- Based on this account of information uniformity, (Wallenberg et al., 2021) predicted a previously undetected argument-type effect in the English and Icelandic OV to VO changes.
- During the change, speakers had access to both constituent orders, making this an ideal case for testing whether language users choose more informationally uniform constructions.

# OV-to-VO in English

## Middle English:

- (1) Mi feader & Mi moder for-þi þt ich nule      be  
My father and my mother because that I not+would you  
forsaken; habbe forsake me.

forsake have forsaken me

“Because I would not forsake you, my father and mother have  
forsaken me”

(*St. Juliana*, northern Herefordshire/southern Shropshire, date: c1225;  
ID CMJULIA-M1,106.172 from the *Penn Parsed Corpus of Middle  
English* 2 PPCME2, 2000)

# OV-to-VO in Icelandic

## Historical Icelandic:

- (2) a. ... og sannleikurinn mun yður frelsa  
and the truth will you free  
“... and the truth will free you.”

(*Oddur Gottskálksson's New Testament*, date: 1540; ID 1540.NTJOHN.REL-BIB, 204.662 from *Icelandic Parsed Historical Corpus* (*IcePaHC*, 2009))

- b. ... en eg skal sjá yður aftur.  
but I shall see you-pl again  
“... but I shall see you again”

(*Oddur Gottskálksson's New Testament*, date: 1540; ID 1540.NTJOHN.REL-BIB, 223.1305 from *IcePaHC*)

## English and Icelandic OV to VO

- Syntactic constituents (and their lexical content) occur at different frequencies, which means they have different information content values.

# English and Icelandic OV to VO

- Syntactic constituents (and their lexical content) occur at different frequencies, which means they have different information content values.

Constituent Type	Average Information Content (PPCMBE; Kroch et al., 2016)
Pronominal DP	low ( $\approx 11.7$ bits)
Nominal DP	high ( $> 13.7$ bits)
Lexical Verb	mid ( $\approx 13.5$ bits)

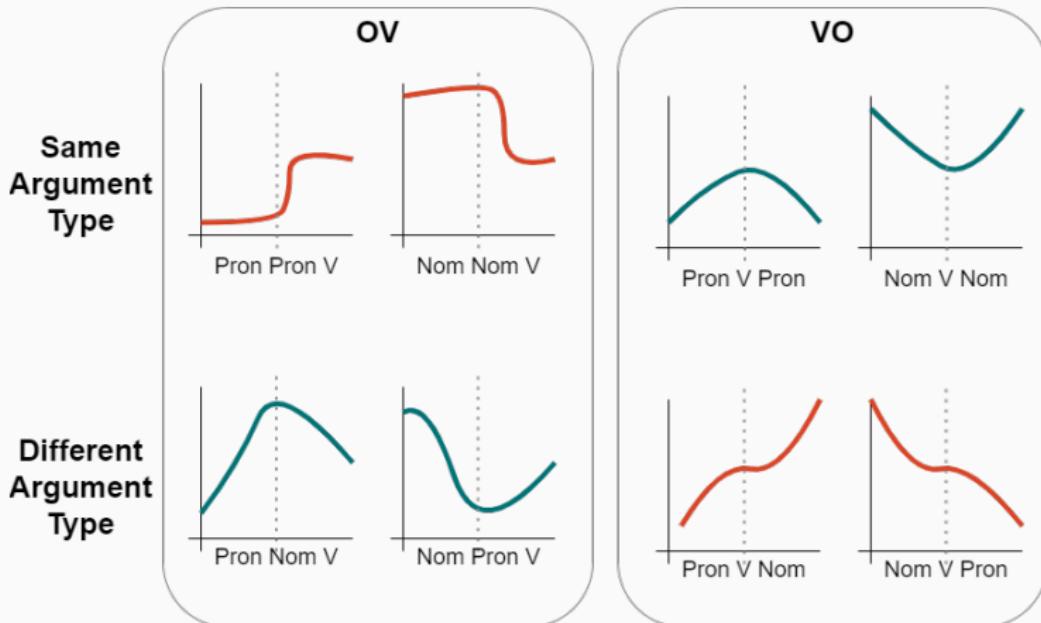
# English and Icelandic OV to VO

- Syntactic constituents (and their lexical content) occur at different frequencies, which means they have different information content values.

Constituent Type	Average Information Content (PPCMBE; Kroch et al., 2016)
Pronominal DP	low ( $\approx 11.7$ bits)
Nominal DP	high ( $> 13.7$ bits)
Lexical Verb	mid ( $\approx 13.5$ bits)

- We can therefore make predictions about the ordering of elements that speakers will prefer, when they have a choice - i.e., constituent orders that yield more uniform information distributions.

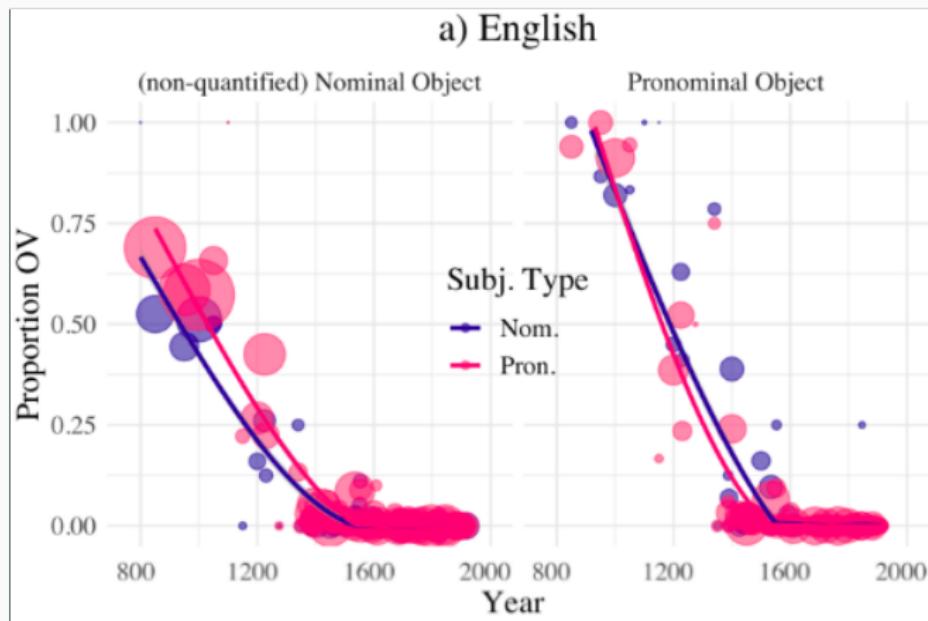
# Predictions



- OV is disfavoured when Sbj and Obj are the **same type**
- OV is favoured when Sbj and Obj are **different types**

## Results:

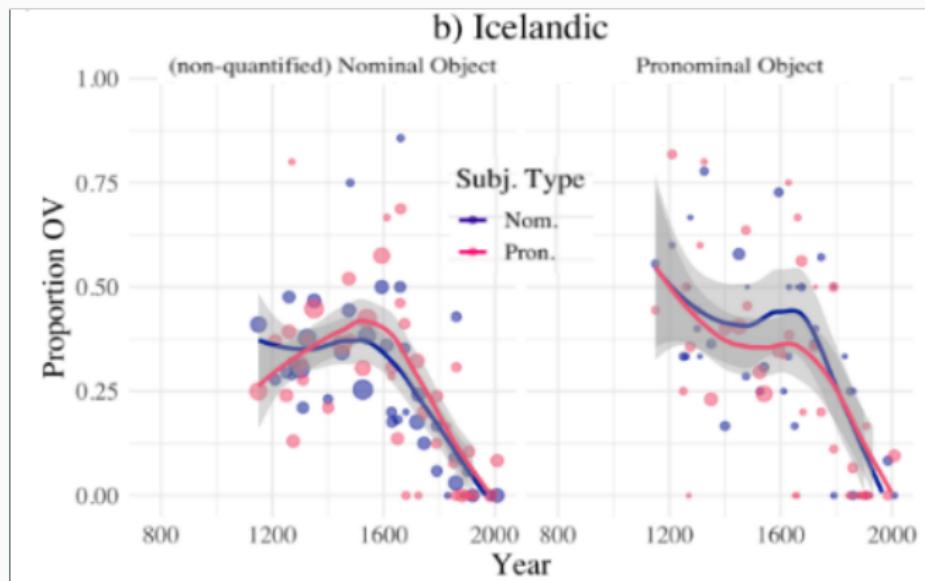
OV is favoured when Subject and Object are different types



*YCOE and Penn Parsed Corpora of Historical English* (Taylor, Warner, Pintzuk, & Beths, 2003; Kroch & Taylor, 2000; Kroch, Santorini, & Delfs, 2004; Kroch et al., 2016)

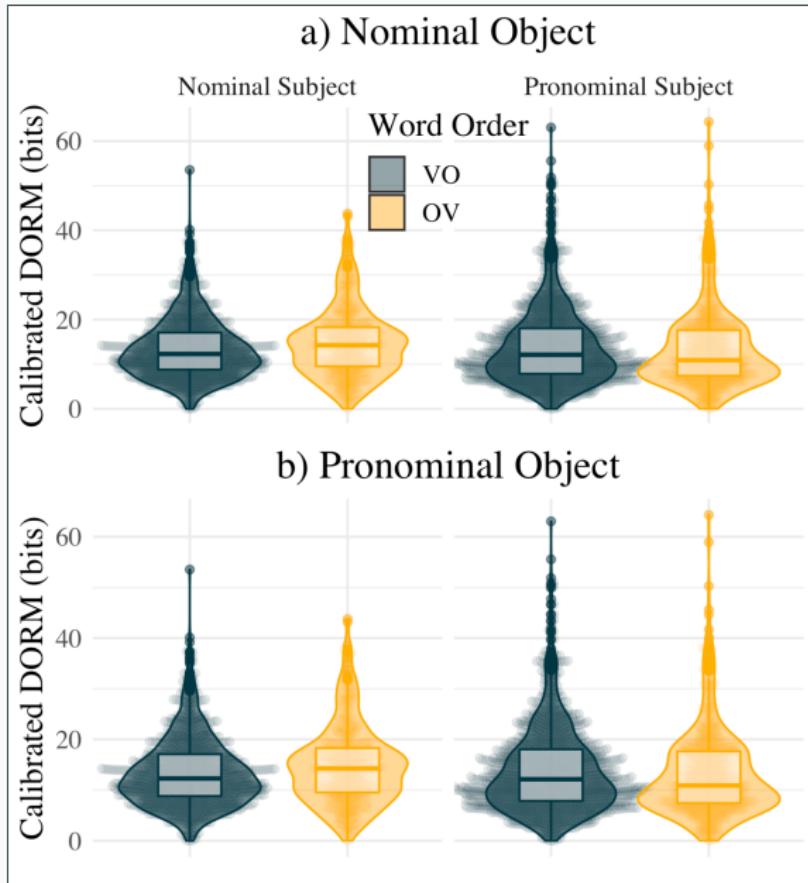
## Results:

OV is favoured when Subject and Object are different types



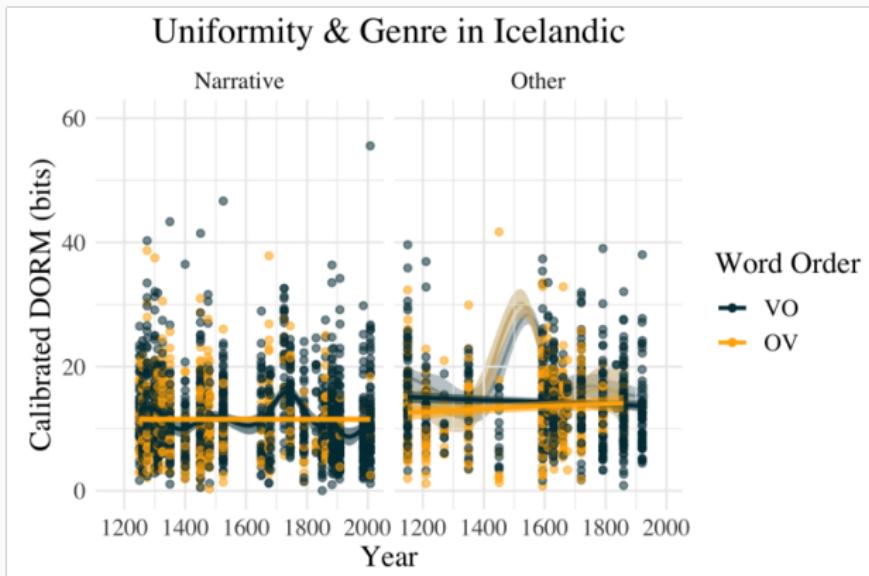
*IcePaHC (Wallenberg, Ingason, Sigurðsson, & Rögnvaldsson, 2011)*

# Results: DORMs by Obj, Sbj Type



## Results: A threshold?

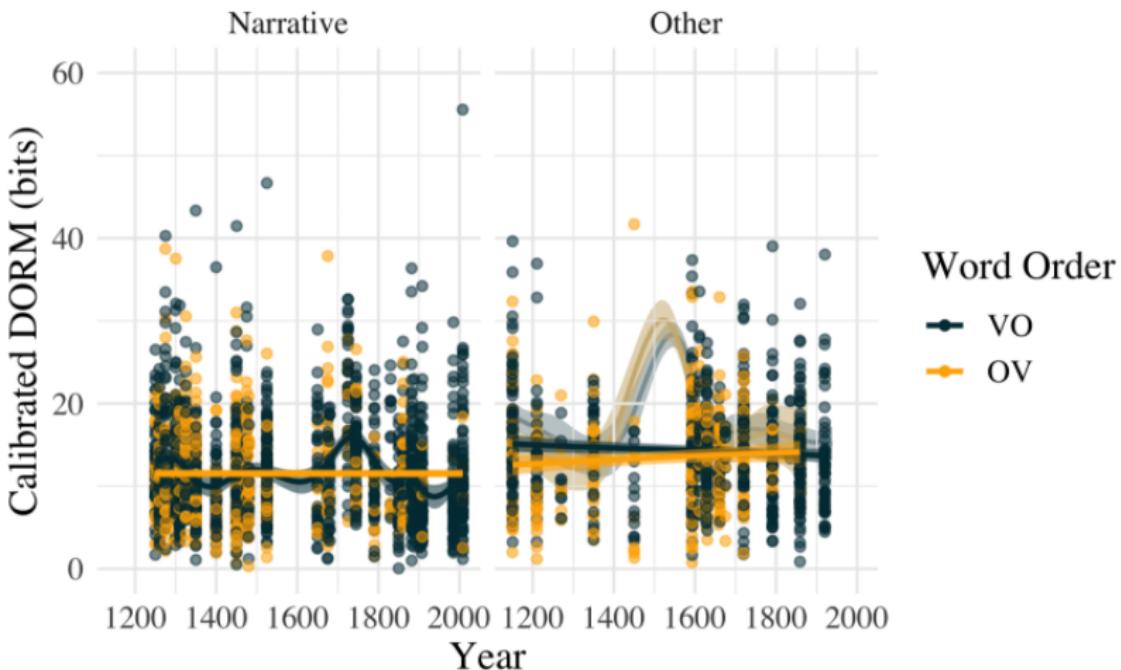
Information density remains constant



- The average uniformity of sentences is constant across the history of Icelandic.

...but the outlier indicates a complex planning problem

## Uniformity & Genre in Icelandic



## Diachronic Study 2: Adding V2

- Wallenberg et al. (2021) showed that a pressure for information uniformity creates contextual effects in the OV-to-VO change.
- OV-to-VO progresses across the argument-type contexts at the same rate (“Constant Rate Effect” (Kroch, 1989) and subs).
- But we did not account for how Subject-Aux inversion under V2 interacts informationally with OV/VO.
- We now consider Icelandic main clauses with adjunct XPs...

# Examples: adjunct fronting under V2

## VO – no fronting

- (3) Jón hefur keypt bók  
Jón has bought a.book  
í dag.  
in today  
'Jón has bought a book  
today.'

## OV – no fronting

- (4) Pessi sami riddari vildi  
This same knight wanted  
eigi gaum gefa  
not attention give  
'This same knight didn't  
want to pay attention  
[to...].'  
(1475.AEVINTYRI.NAR-REL.,933)

## VO – adjunct fronting

- (5) Í dag hefur Jón  
in today has Jón  
keypt bók.  
bought a.book  
'Today, Jón has bought a  
book.'

## OV – adjunct fronting

- (6) Aldrei hafði kóngsson  
never has a prince  
slíkan grip séð.  
such a.thing seen  
'Never has a prince seen  
such a thing'  
(1450.VILHJALMUR.NAR-  
SAG,21.327)

## Diachronic Study 2: Adding V2

---

- From earlier study, constituent order affects distribution uniformity.

## Diachronic Study 2: Adding V2

- From earlier study, constituent order affects distribution uniformity.
- Remember: syntactic constituents occur at different frequencies, which means they have different information content values.

Constituent Type	Average Information Content (PPCMBE; Kroch et al., 2016)
Pronominal DP	low ( $\approx 11.7$ bits)
Aux	low-ish ( $\approx 12.4$ bits)
Lexical Verb	mid ( $\approx 13.5$ bits)
Nominal DP	high ( $> 13.7$ bits)

Finally, Adjunct XPs are comparable to (not least because in many cases they include) Nominal DPs, and are therefore treated here as **high**

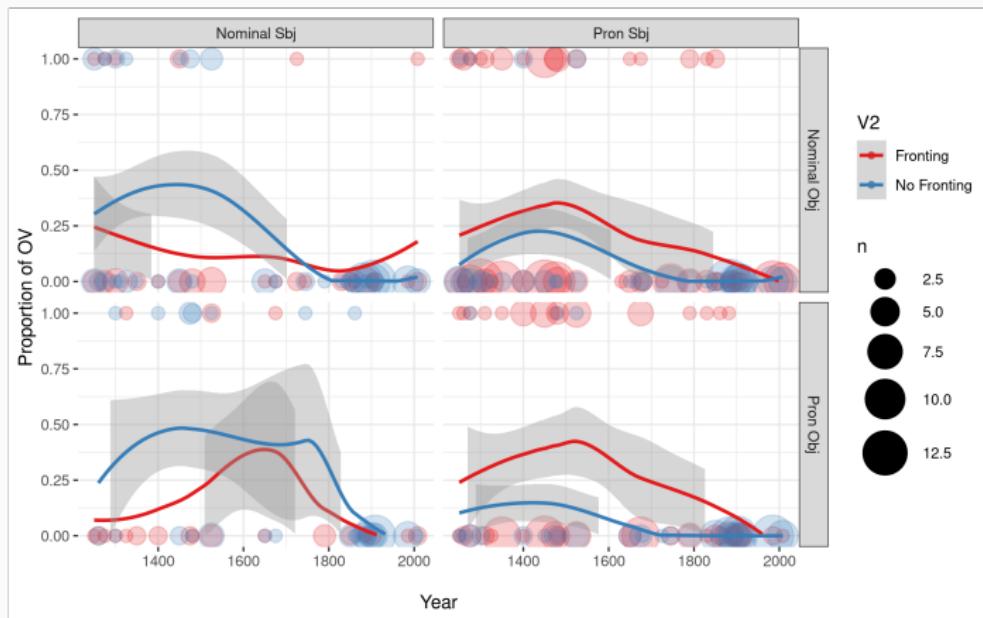
## Study 2: Adding V2

---

- We might also expect that V2 will also give rise to information theoretic effects that interact with OV/VO, as argument type did in previous study.
- So that's what we looked for...

## Preliminary:

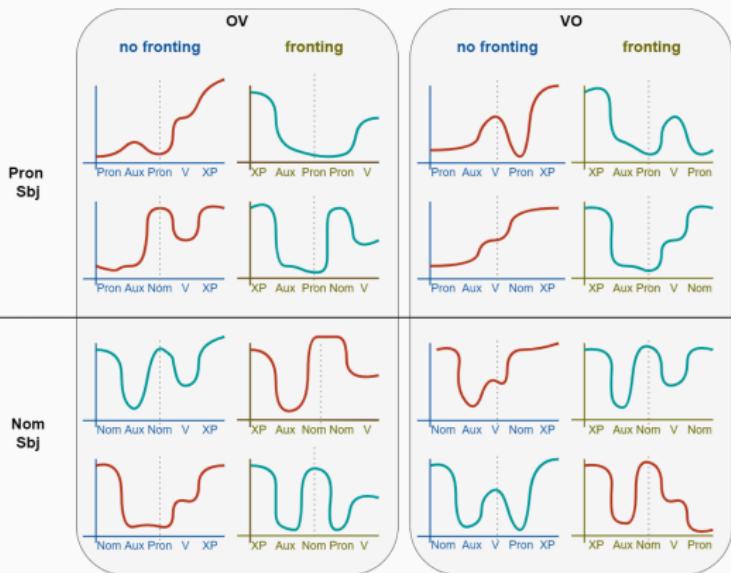
Is V2 a “context” for OV/VO (in the CRE sense)?



- Predictions not borne out if V2-fronting conditions choice of OV/VO.

# Study 2 Predictions

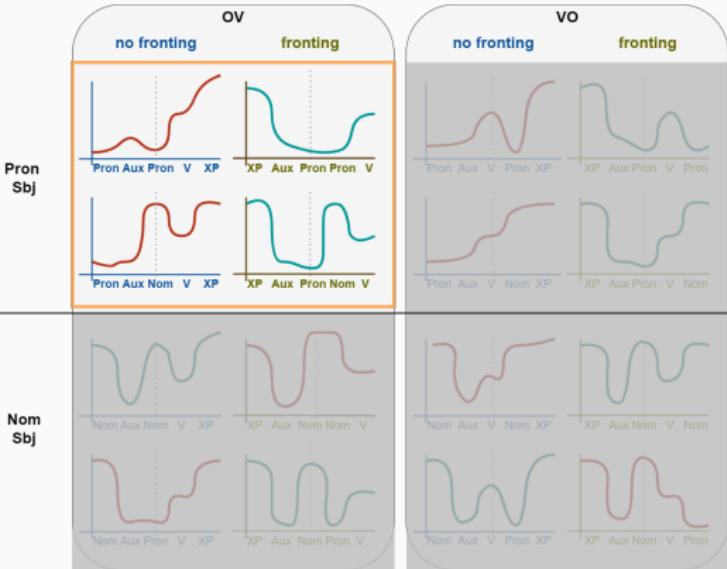
Speakers deploy fronting when it yields more uniformity



If speakers *deployed* adjunct-fronting to maximise information uniformity (given other parameters), they would conform to this pattern

# Study 2 Predictions

## Given OV and Pron Sbj:



+ Pron Obj:

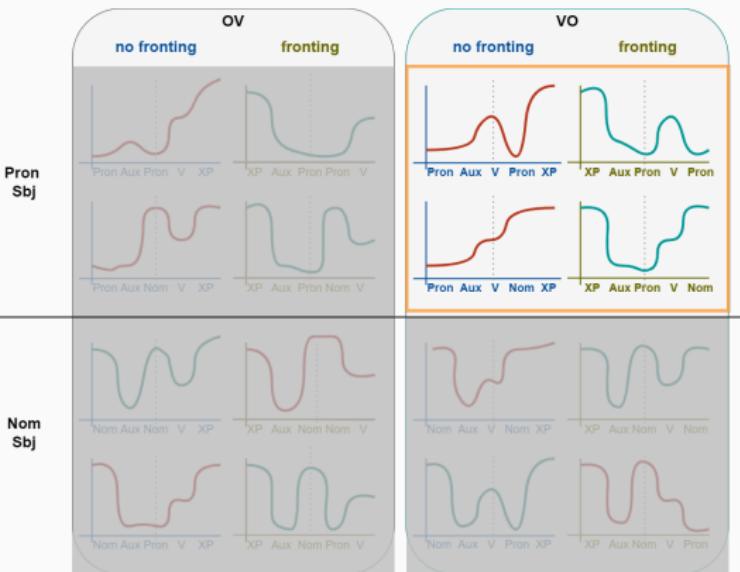
- Fronting helps balance out the unavoidable 3-unit informational troughs (Pron-Aux-Pron & Aux-Pron-Pron)

+ Nom Obj:

- Fronting means avoidance of clustered information peak (Nom-V-XP)

# Study 2 Predictions

## Given VO and Pron Sbj:



+ Pron Obj:

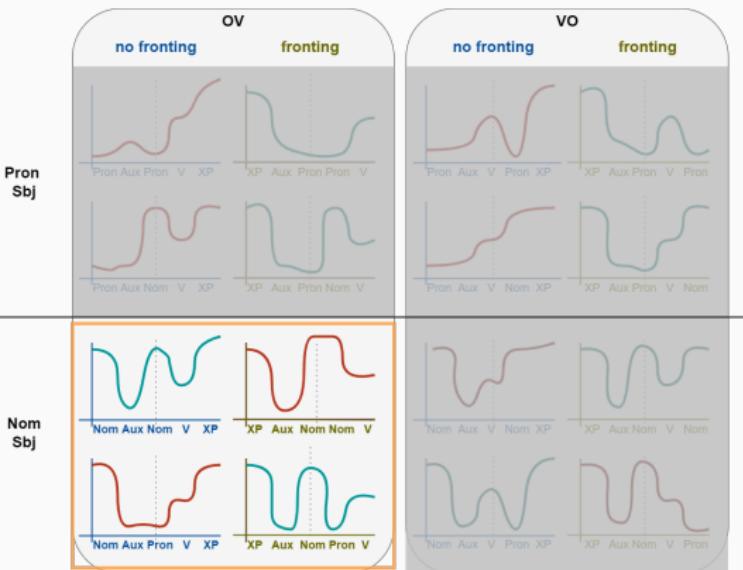
- Fronting may yield a bit more uniformity

+ Nom Obj

- Pressure from 2-unit informational troughs and peaks (Pron-Aux & Nom-XP)
- Maximally asymmetric distribution can be avoided by fronting

# Study 2 Predictions

## Given OV and Nom Sbj:



Distribution symmetries aren't so obviously different, so pressures probably aren't strong. But:

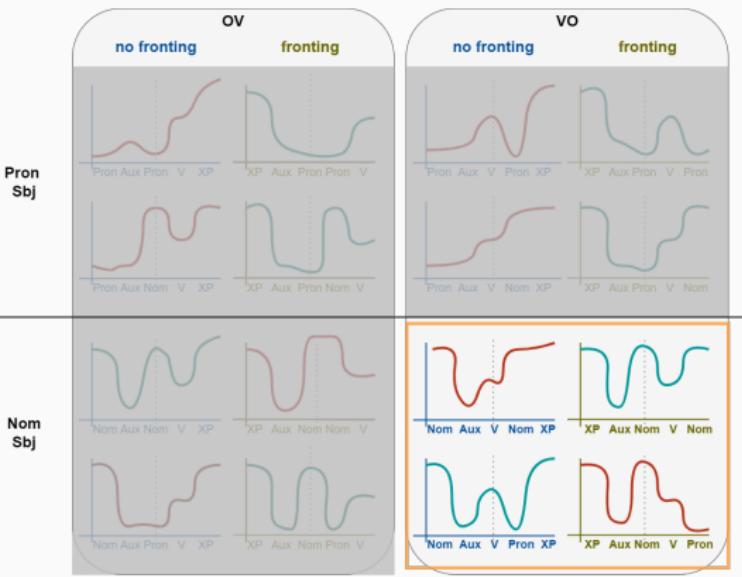
+ Nom Obj:

- Fronting yields 2-unit informational peak (Nom-Nom), so *may* be disfavoured

+ Pron Obj:

- Fronting *might* be more uniform (if not more symmetrical)

# Study 2 Predictions Given VO and Nom Sbj:



Distribution symmetries not obviously different, so pressures probably aren't strong. But:

+ Nom Obj:

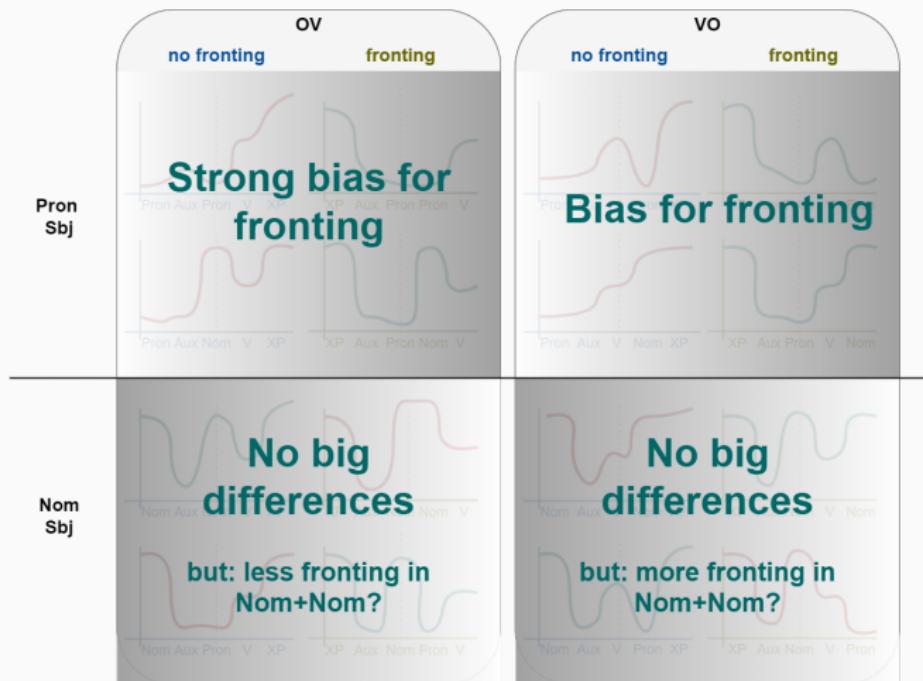
- Fronting *may* yield slightly more uniform distribution (by avoiding clustered peak of Nom+XP)

+ Pron Obj:

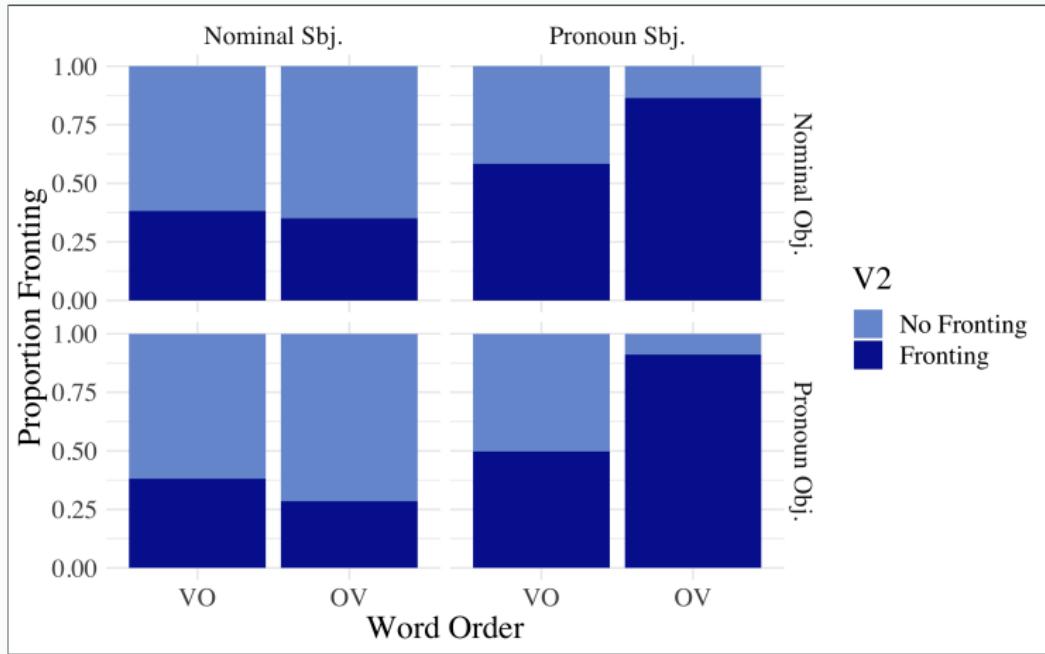
- Non-fronting might be more symmetrical

# Study 2 Predictions

Speakers deploy fronting when it yields more uniformity



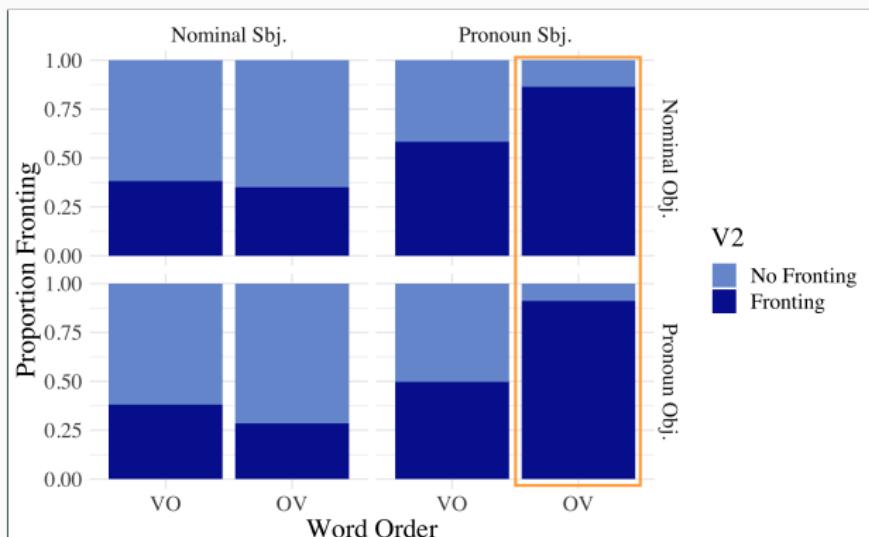
## Study 2 Results



- Overall, fronting appears where we'd expect if speakers are trying to maximise the uniformity of information distributions with the order of constituents.

# Study 2 Results: Given OV and Pron Sbj

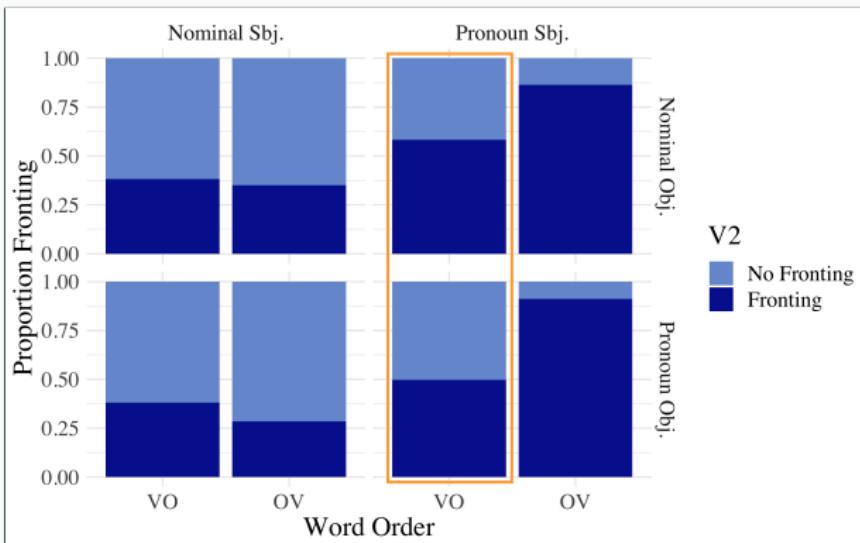
Two pressures for fronting



- Pron Sbj + Nom Obj:  
Avoid long peak  
(Nom-V-XP)
- Pron Sbj + Pron Obj:  
Avoid maximal  
asymmetry

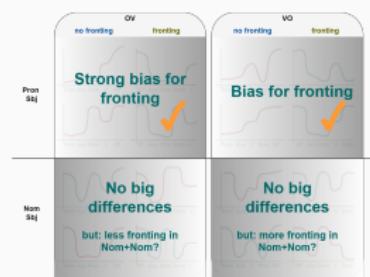


# Study 2 Results: Given VO and Pron Sbj

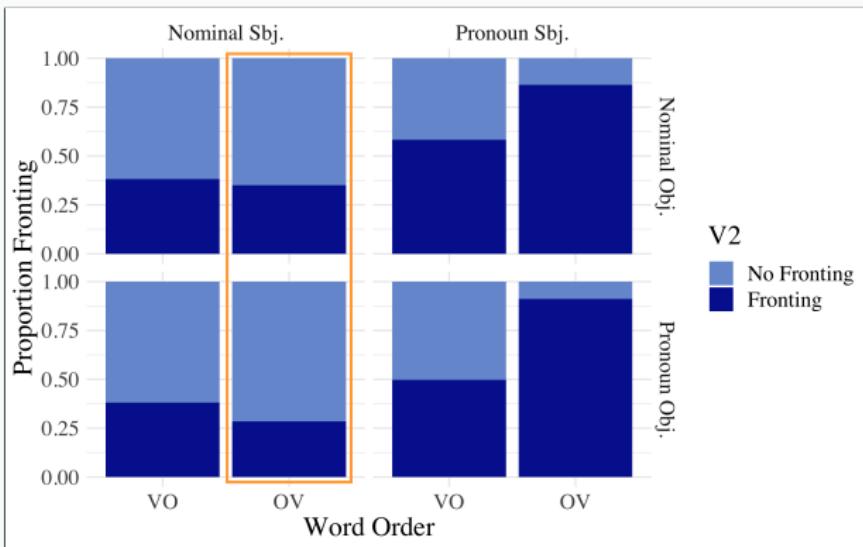


## Bias for fronting

- Stronger (>50%) for Pron Sbj + Nom Obj
- Avoidance of maximal asymmetry



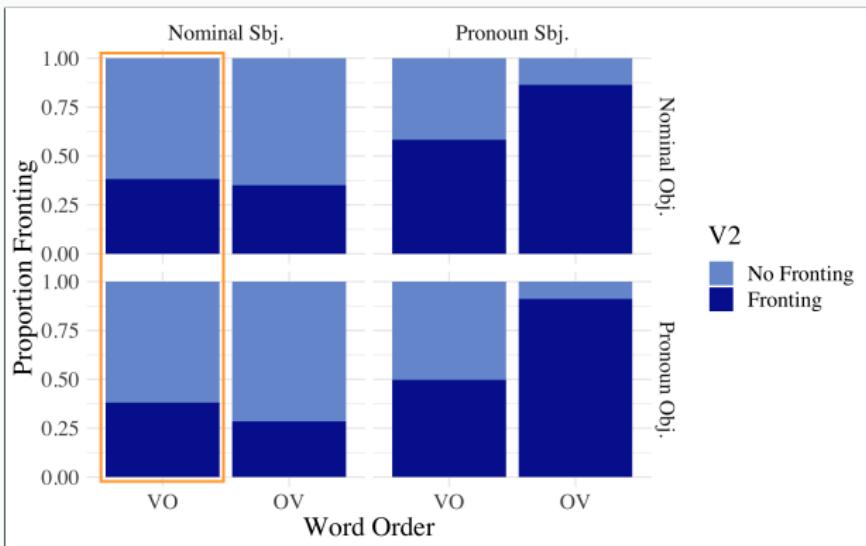
# Study 2 Results: Given OV and Nom Sbj



- Less fronting than in VO overall
- No big difference between Pron/Nom Obs, but not the direction we'd expected (less fronting to avoid Nom+Nom)



# Study 2 Results: Given VO and Nom Sbj



- No big difference between Pron/Nom Obs



## Information Uniformity and Language

What It's For: Noise Resistance

What It Does: Syntactic Planning

Diachronic Study 1: Information Theoretic Constant Rate

Effects: OV to VO

Diachronic Study 2: Adding Adjunct Fronting with V2

## Theoretical Implications and Future Work

## Theoretical Implications

---

- The diachronic studies show speakers use the syntactic resources available to optimize for information uniformity in planning.

# Theoretical Implications

---

- The diachronic studies show speakers use the syntactic resources available to optimize for information uniformity in planning.
- The way they do so suggests:
  1. Speakers select lexical array first (incl. Obj, Sbj types).
  2. Speakers then select right- or left-headed vP/VP.
  3. Then, movement
    - (i.e. fronting modulates information uniformity under the constraint of OV or VO, rather than vice versa).
    - (cf. Speyer, 2010 where Sbj type conditions topicalization in English)
- Time Separation Theorem (Kauhanen & Walkden, 2018): given this order of operations, once the basic lexical content is decided, maximum/minimum uniformity is strictly bounded.

## Further Work: Non-V2 Object Topicalization

---

- Fits well with work showing topicalization is sensitive to accent clash (Speyer, 2008, 2010).

## Further Work: Non-V2 Object Topicalization

- Fits well with work showing topicalization is sensitive to accent clash (Speyer, 2008, 2010).

(11) Joel she likes (but Bill she doesn't).

(12) Joel the cat likes.

(13) Joel Viola likes.

(14) Joel, Viola likes.

From PPCMBE, nominal objects ( $\chi^2 = 260$ ,  $p < 2 \times 10^{-16}$ ):

	Fronted	In Situ	Prop. Fronted
Pron Sbj	631	20,071	0.031
Nom Sbj	119	16,808	0.0071

## Further Work: Non-V2 Object Topicalization

---

- Informational uniformity and degree of accent clash could both be continuous variables, predicting slow change over time:

## Further Work: Non-V2 Object Topicalization

---

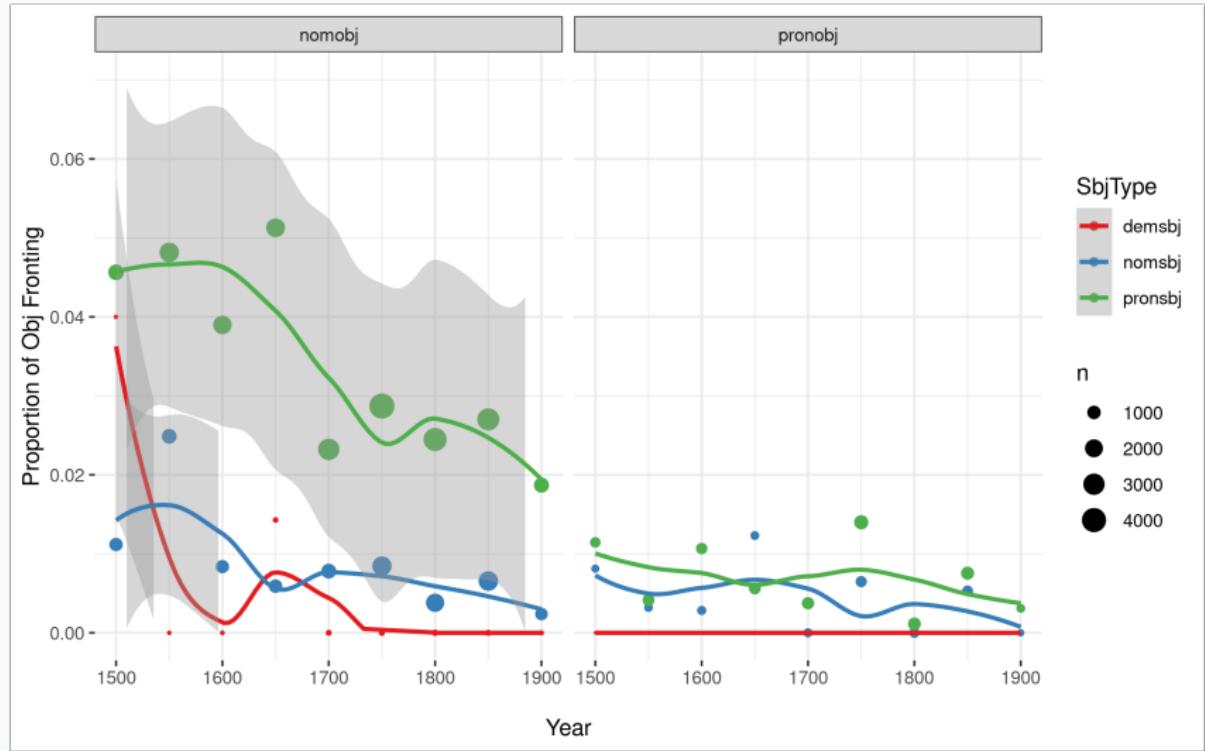
- Informational uniformity and degree of accent clash could both be continuous variables, predicting slow change over time:
  - The two syntactic forms overlap in function (Prince, 1998), and therefore compete in use (Kroch, 1994).

## Further Work: Non-V2 Object Topicalization

---

- Informational uniformity and degree of accent clash could both be continuous variables, predicting slow change over time:
  - The two syntactic forms overlap in function (Prince, 1998), and therefore compete in use (Kroch, 1994).
  - The competition is mitigated by specialization, but along a continuous dimension, and so total specialization is impossible and slow change results (Wallenberg, 2016).

## Further Work: Non-V2 Object Topicalization



Note: significant effects of **Year**, **Obj**, **Sbj**, sig **Obj:Sbj** interaction, but

## Further Work: Non-V2 Object Topicalization

---

- Both dimensions are likely in operation, but can they be distinguished?
  - (15) Joel, Bob likes.
  - (16) Joel, men like.
- Calibrated DORM and accent clash are both bounded dimensions.

## Conclusions: uniformity and thresholds

---

- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.

## Conclusions: uniformity and thresholds

---

- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.
- **What It Does:** causes speakers to use their syntactic resources to disperse information as uniformly as possible across a sentence.

## Conclusions: uniformity and thresholds

---

- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.
- **What It Does:** causes speakers to use their syntactic resources to disperse information as uniformly as possible across a sentence.
- Presumably there is a threshold for allowable losses, and so a threshold for sentence DORM.

## Conclusions: uniformity and thresholds

---

- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.
- **What It Does:** causes speakers to use their syntactic resources to disperse information as uniformly as possible across a sentence.
- Presumably there is a threshold for allowable losses, and so a threshold for sentence DORM.
- **Other Implications:** derivational ordering effects, where sbj/obj type and perhaps headedness precede movement.

## Conclusions: uniformity and thresholds

---

- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.
- **What It Does:** causes speakers to use their syntactic resources to disperse information as uniformly as possible across a sentence.
- Presumably there is a threshold for allowable losses, and so a threshold for sentence DORM.
- **Other Implications:** derivational ordering effects, where sbj/obj type and perhaps headedness precede movement.
- Minimizing DORM is a tough planning problem, making it hard to do with other constraints.

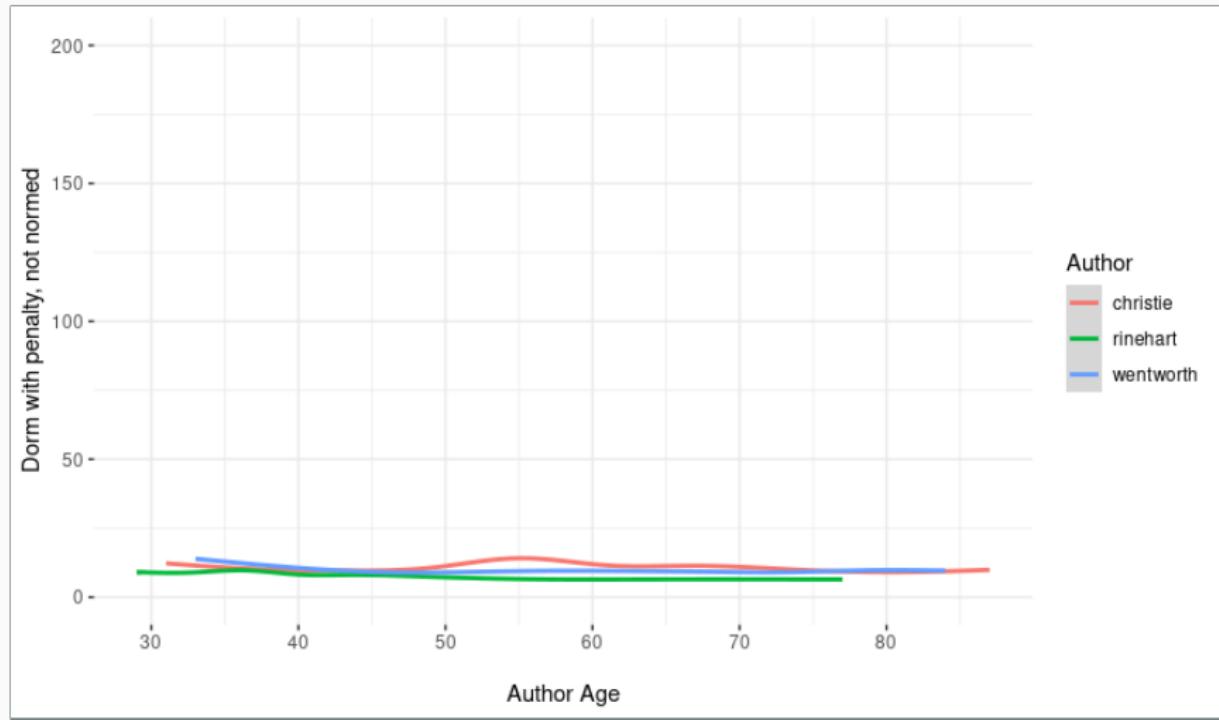
## Conclusions: uniformity and thresholds

---

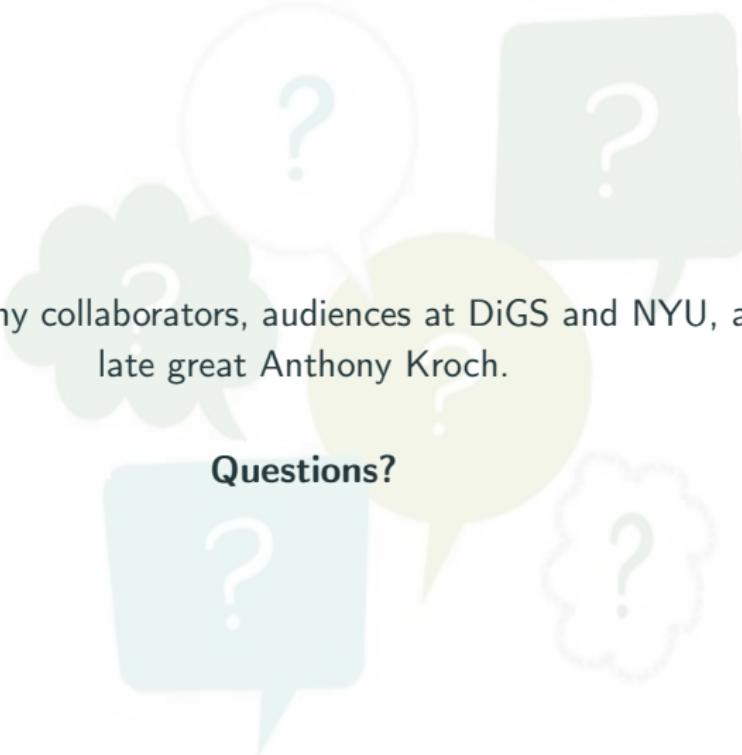
- **What It's For:** prevents catastrophic losses, which could lead to loss of proposition/utterance-level meaning.
- **What It Does:** causes speakers to use their syntactic resources to disperse information as uniformly as possible across a sentence.
- Presumably there is a threshold for allowable losses, and so a threshold for sentence DORM.
- **Other Implications:** derivational ordering effects, where sbj/obj type and perhaps headedness precede movement.
- Minimizing DORM is a tough planning problem, making it hard to do with other constraints.
- **Exploratory Work:** information uniformity can be a continuous dimension of specialization, slowing change in non-V2 topicalization.

## More Exploratory Work: Anything to See Here?

3 Writers; 200,000-500,000 sentences each



# Acknowledgements, and thanks for listening!



Thank you to my collaborators, audiences at DiGS and NYU, and the late great Anthony Kroch.

**Questions?**

## References i

---

- 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.
- Chingacham, A., Demberg, V., & Klakow, D. (2023). A data-driven investigation of noise-adaptive utterance generation with linguistic modification. In *2022 ieee spoken language technology workshop (slt)* (pp. 353–360).
- Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, 214, 104754.
- Ecay, A. (2015). *The Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP)*. Public release 1. <https://github.com/uoy-linguistics/pyccle>.

## References ii

---

- Fenk, A., & Fenk, G. (1980). Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, 27, 402.
- Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, 45, 431–448.
- Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).
- Hartley, R. V. (1928). Transmission of information 1. *Bell System technical journal*, 7(3), 535–563.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1), 23–62.
- Kauhanen, H., & Walkden, G. (2018). Deriving the constant rate effect. *Natural Language & Linguistic Theory*, 36(2), 483–521.

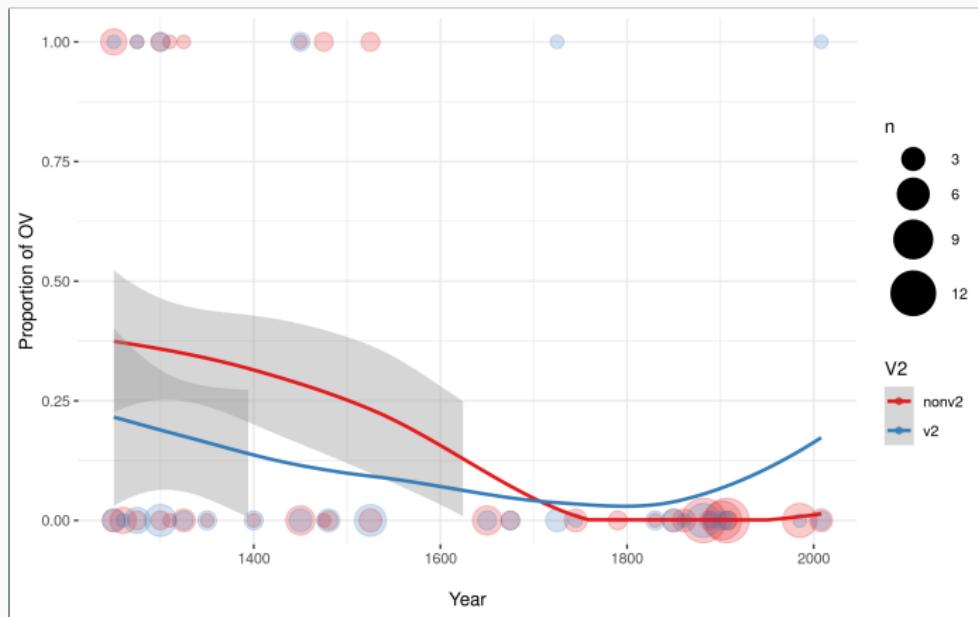
- Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, 199–244.
- Kroch, A. S. (1994). Morphosyntactic variation. In K. B. et al (Ed.), *Papers from the 30th regional meeting of the Chicago Linguistics Society: Parasession on variation and linguistic theory*.
- Kroch, A. S., Santorini, B., & Delfs, L. (2004). *Penn-Helsinki Parsed Corpus of Early Modern English, release 3*. (Size 1.8 Million Words)
- Kroch, A. S., Santorini, B., & Diertani, A. (2016). *Penn Parsed Corpus of Modern British English 2nd edition, release 1*. (Size ~ 2.8 million words.)
- Kroch, A. S., & Taylor, A. (2000). *Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition, release 4*. (Size: 1.3 million words.)

- Levy, R. P., & Jaeger, F. T. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).
- Pintzuk, S., & Taylor, A. (2006). The loss of OV order in the history of English. In A. van Kemenade & B. Los (Eds.), *Blackwell handbook of the history of English* (pp. 247–278). Blackwell.
- Prince, E. (1998). On the limits of syntax, with reference to left-dislocation and topicalization. *Syntax and semantics*, 281–302.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
- Speyer, A. (2008). *Topicalization and clash avoidance: On the interaction of prosody and syntax in the history of English with a few glimpses at German* (Unpublished doctoral dissertation). University of Pennsylvania.

- Speyer, A. (2010). *Topicalization and stress clash avoidance in the history of english*. Berlin/New York: Mouton de Gruyter.
- Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*.  
<http://www-users.york.ac.uk/~lang22/YcoeHome1.htm>.
- Turk, A. (2010). Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis. *Laboratory phonology*, 1(2), 227–262.
- Wallenberg, J. C. (2016). Extraposition is disappearing. *Language*, 92(4), e237–e256.
- Wallenberg, J. C., Bailes, R., Cuskley, C., & Ingason, A. K. (2021). Smooth signals and syntactic change. *Languages*, 6(2), 60.

- Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F., & Rögnvaldsson, E. (2011). *Icelandic Parsed Historical Corpus (IcePaHC)*. (Version 0.9. Size: 1 million words. URL  
[http://www.linguist.is/icelandic\\_treebank](http://www.linguist.is/icelandic_treebank))

## Study 2 Results: OV disfavoured by adjunct-fronting through OV-to-VO



## Study 2 Results: Count data for V2 in Icelandic

