



# Information Theoretic Constant Rate Effects

---

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

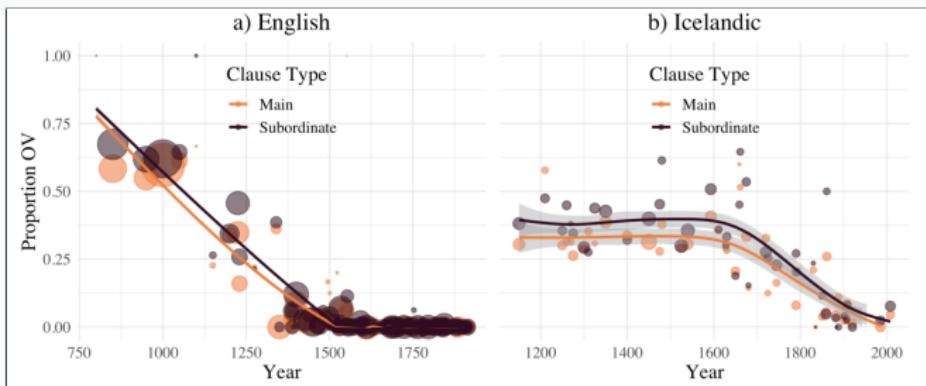
<sup>3</sup> University of Iceland, Iceland

## 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).

# 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).

# Information Theoretic Constant Rate Effects

Today's talk:

- summarises CRE of subject/object type in OV to VO
  - predicted and explained by our information theoretic account of language use
- expands on earlier findings on Icelandic, adding fronting under V2
- explains an apparent exception to the CRE
- discusses some theoretical implications for syntax and the CRE.

## Information Uniformity and Language (a crash course)

Study 1: Information theoretic CRE in OV to VO

Study 2: Adding V2

Discussion

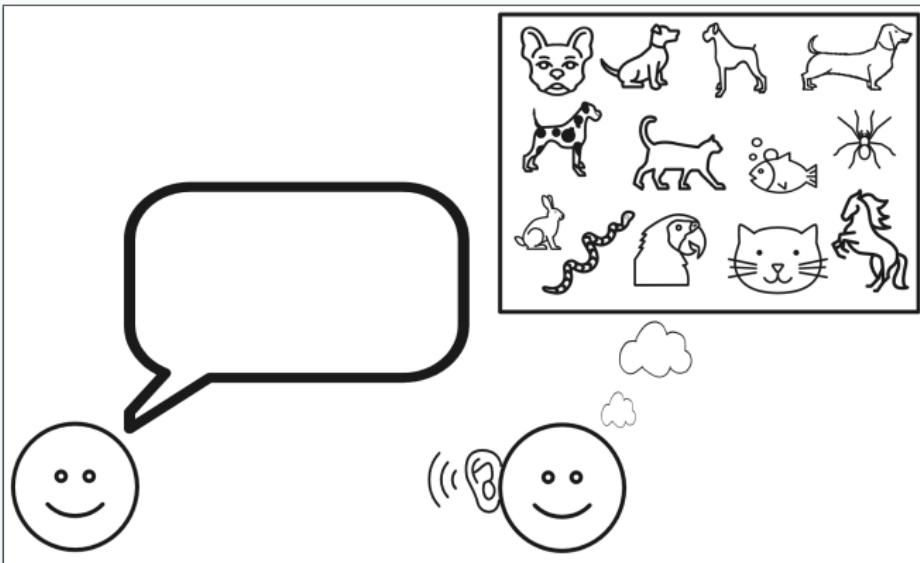
# Information content of words

Information Theory quantifies information and describes how communication systems work (Shannon, 1948).

**Key point for today:**

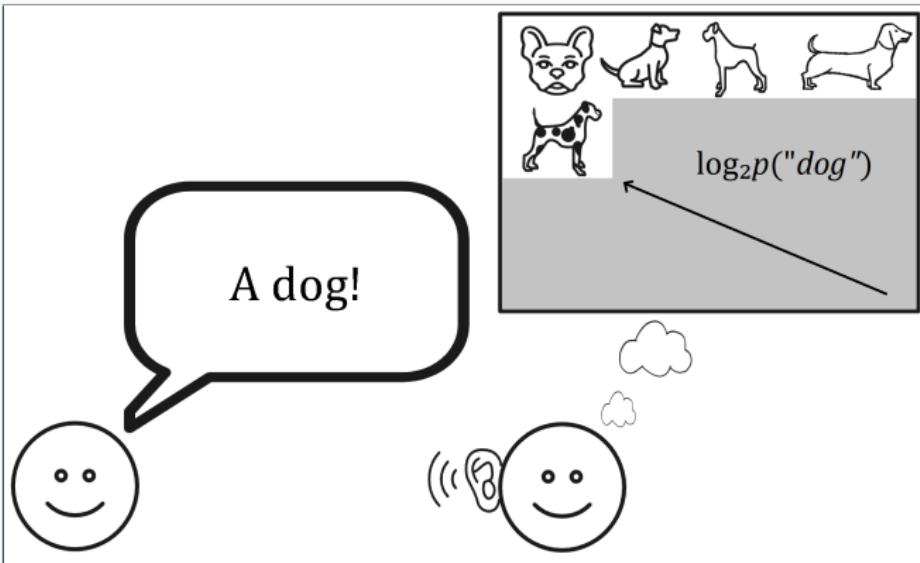
- ⇒ Low frequency words are high information,  
high frequency words are low information.

# Information content of words



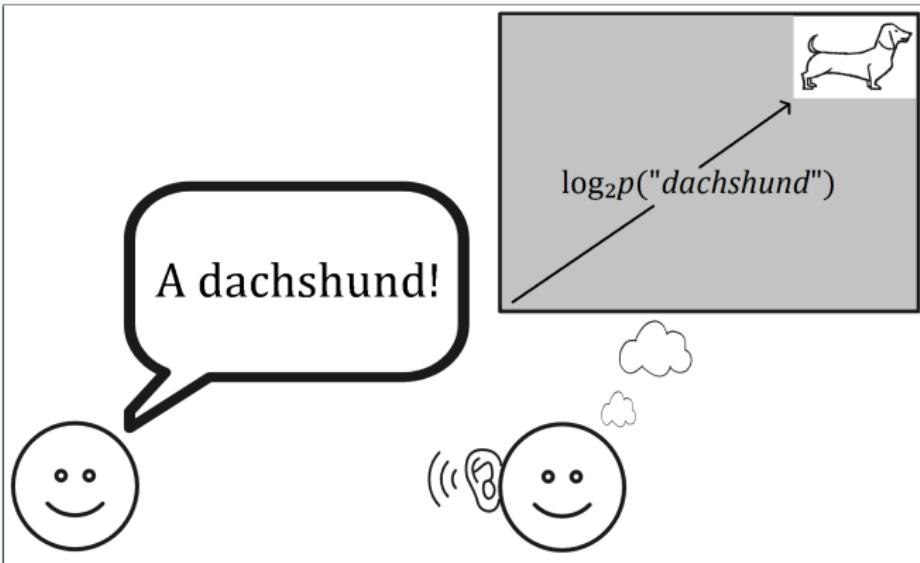
Receiver begins with a **set of expected outcomes** that could result from the message they receive. This set of expected outcomes is the amount of **uncertainty** they have about the message.

# 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.

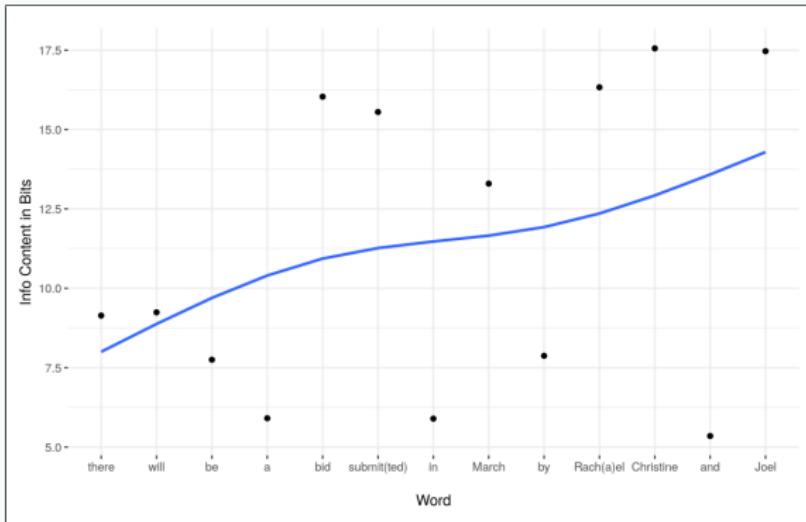
# Information content of words



The more infrequent the word, the more uncertainty is reduced. This is the same thing as saying **the more infrequent the word, the more information it has.**

# 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.



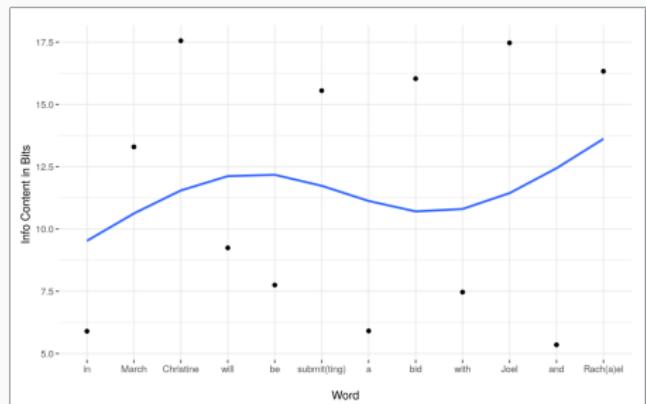
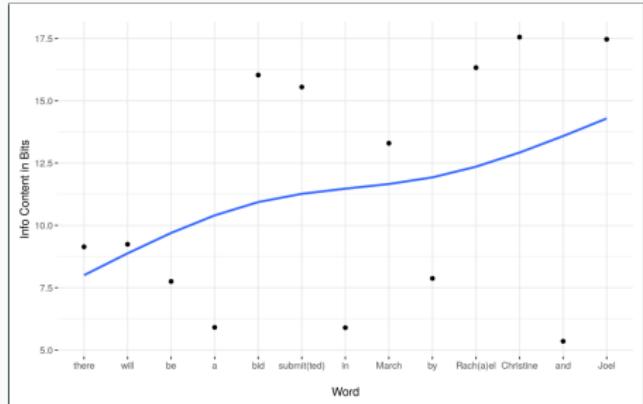
## Uniformity of information distributions

---

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 (Cuskley et al., 2021) - details available in the Q&A upon request!

## Information Uniformity and Language (a crash course)

**Study 1: Information theoretic CRE in OV to VO**

Study 2: Adding V2

Discussion

# 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 CRE 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 CRE 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*)

## CRE in English and Icelandic OV to VO

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

# CRE in 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)

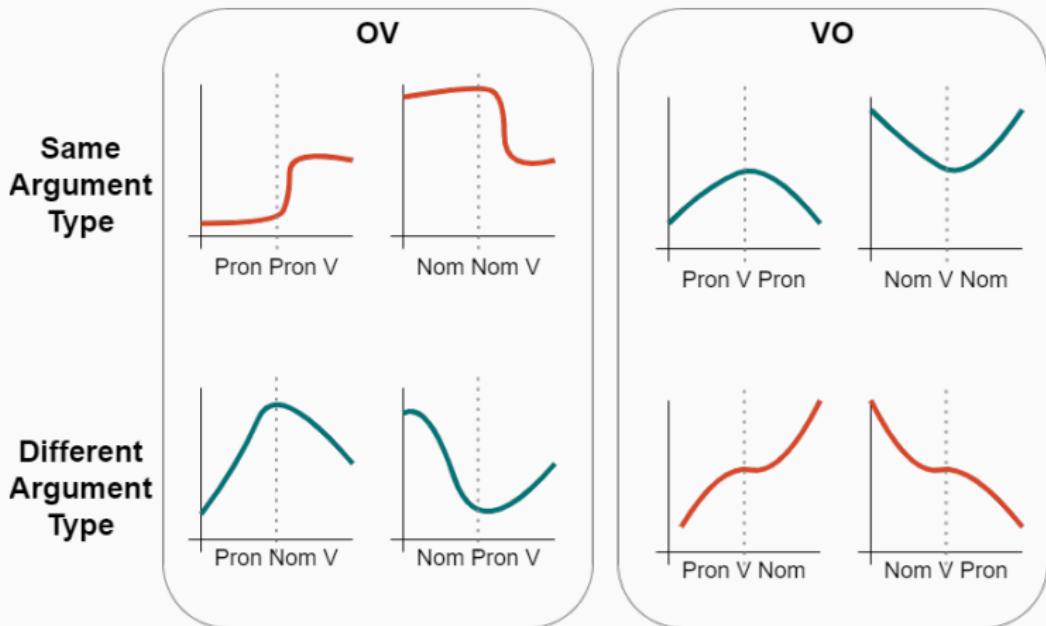
# CRE in English and Icelandic OV to VO

- 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.

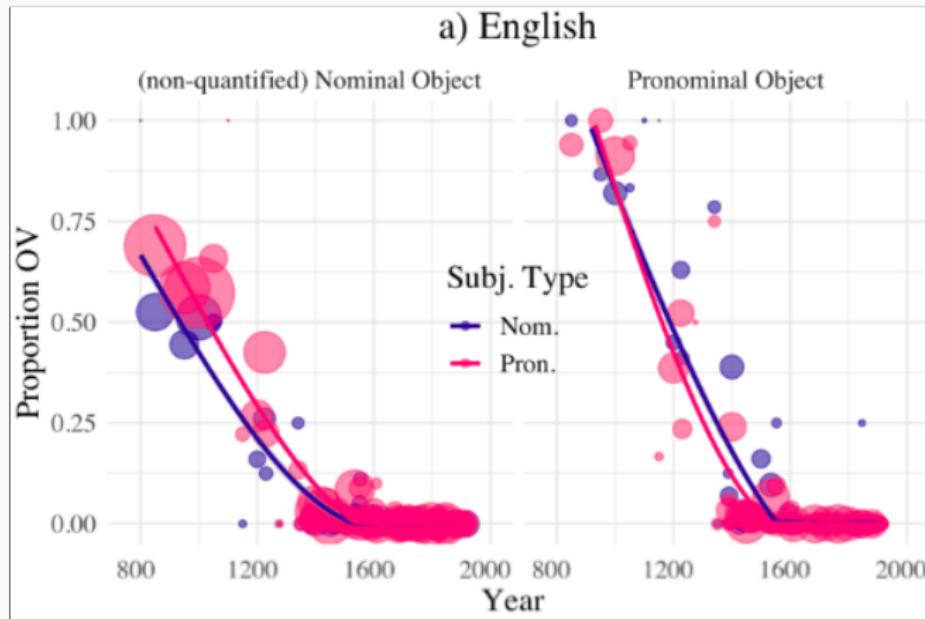
# Study 1 Predictions



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

## Study 1 Results:

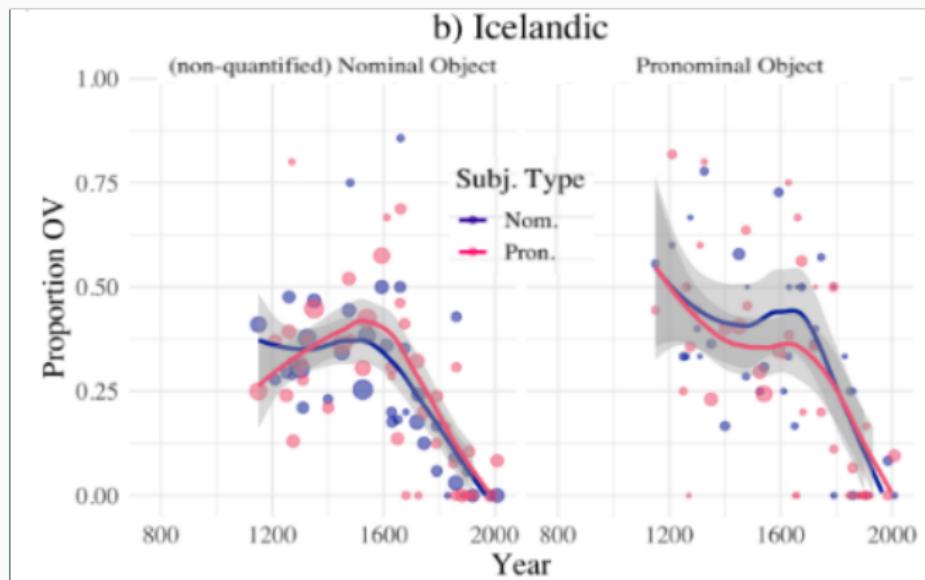
OV is favoured when Subject and Object are different types



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

## Study 1 Results:

OV is favoured when Subject and Object are different types



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

## Information Uniformity and Language (a crash course)

Study 1: Information theoretic CRE in OV to VO

Study 2: Adding V2

Discussion

## 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 (CRE).
- 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)

## Study 2: Adding V2

---

- From Study 1, constituent order affects distribution uniformity.

## Study 2: Adding V2

- From Study 1, 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**

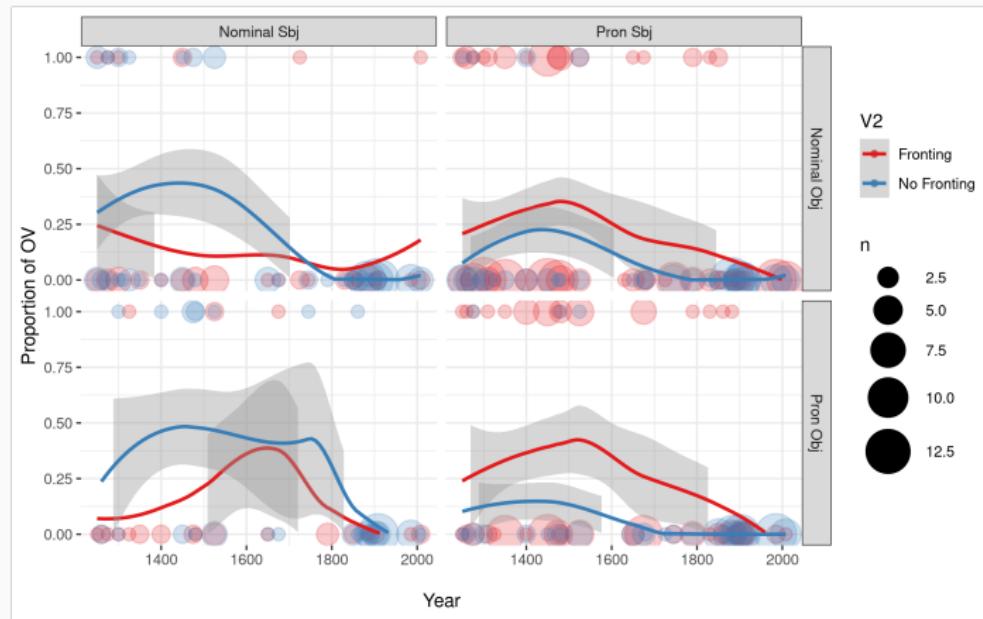
## Study 2: Adding V2

---

- We might also expect that V2 will also give rise to an information theoretic CRE in OV/VO, as argument type did in Study 1.
- So that's what we looked for...

## Preliminary:

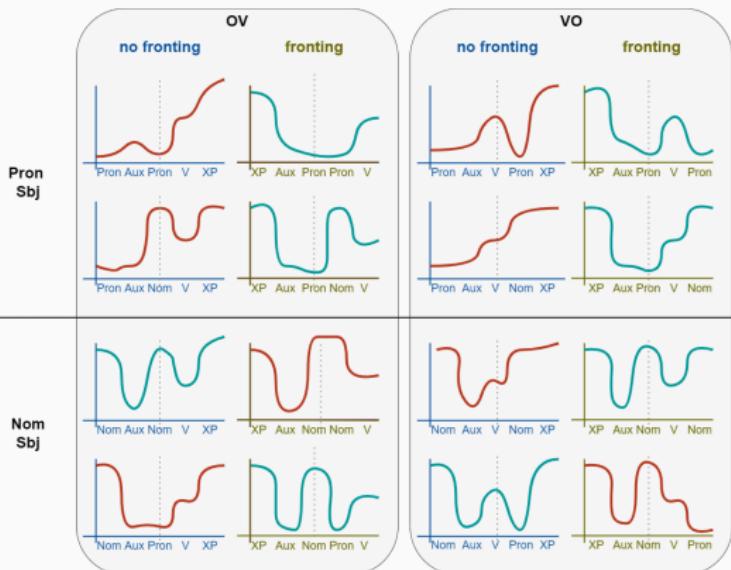
# Is V2 a “context” for OV/VO (in the CRE sense)?



- If V2 were a conditioning context for OV/VO, we'd expect CRE.
- This *may* apply to Pron Sbj + Nom Obj, but not in other cases.
- Predictions not borne out if V2-fronting conditions choice of OV/VO.

# Study 2 Predictions

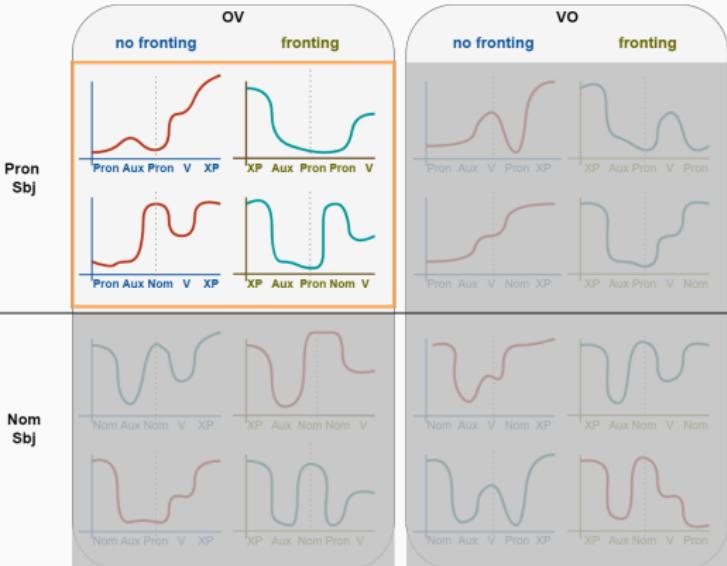
Speakers deploy fronting when it yields more uniformity



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:

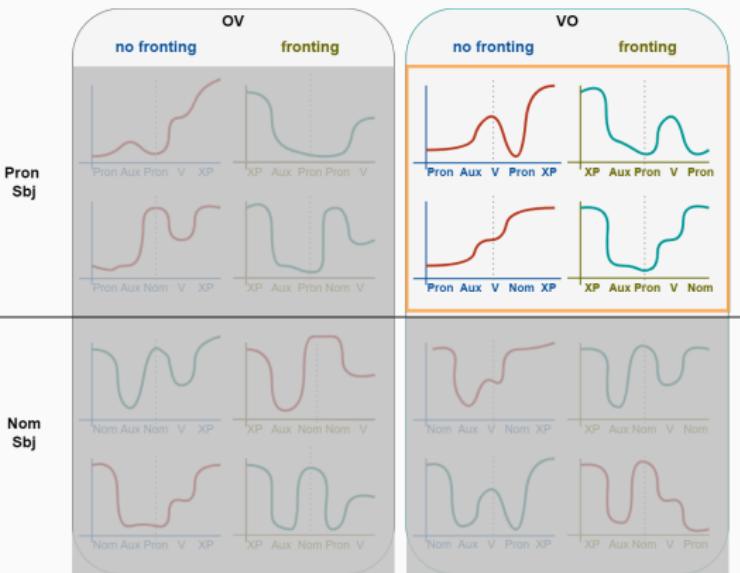
- Fronting helps balance out the unavoidable 3-unit informational troughs (Pron-Aux-Pron & Aux-Pron-Pron)

+ Nom Obj:

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

# Study 2 Predictions

## Given VO and Pron Sbj:



+ Pron Obj:

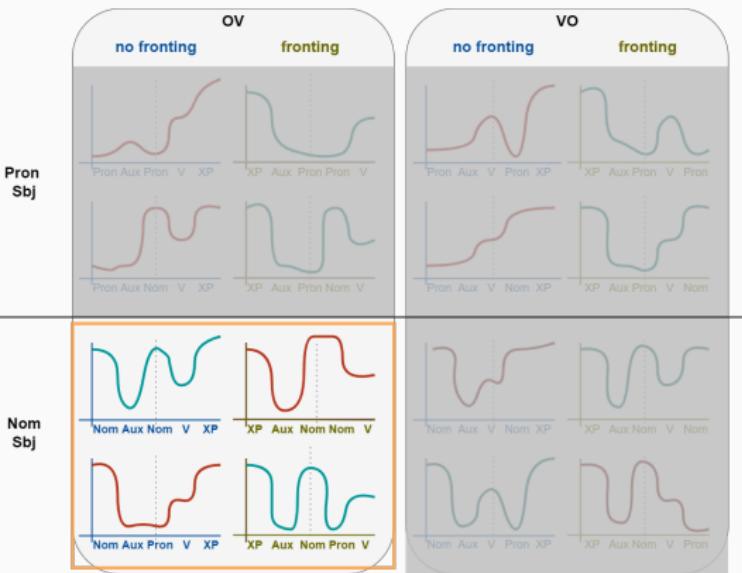
- Fronting may yield a bit more uniformity

+ 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:

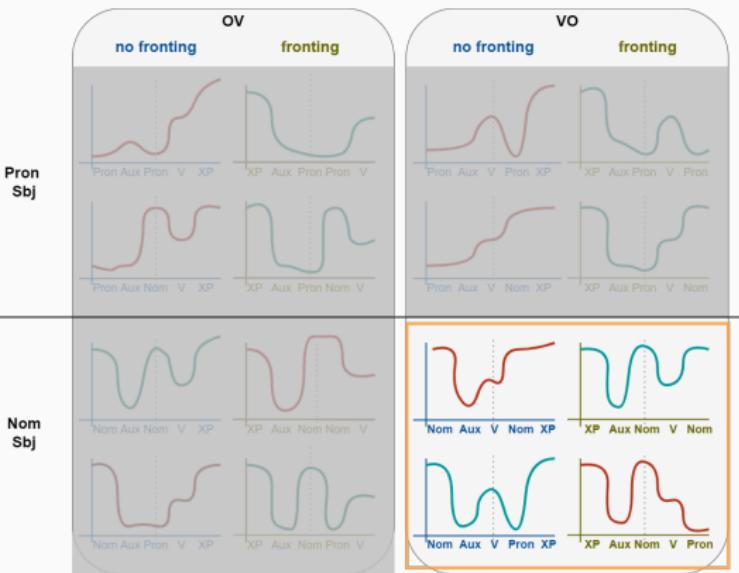
+ Nom Obj:

- 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:

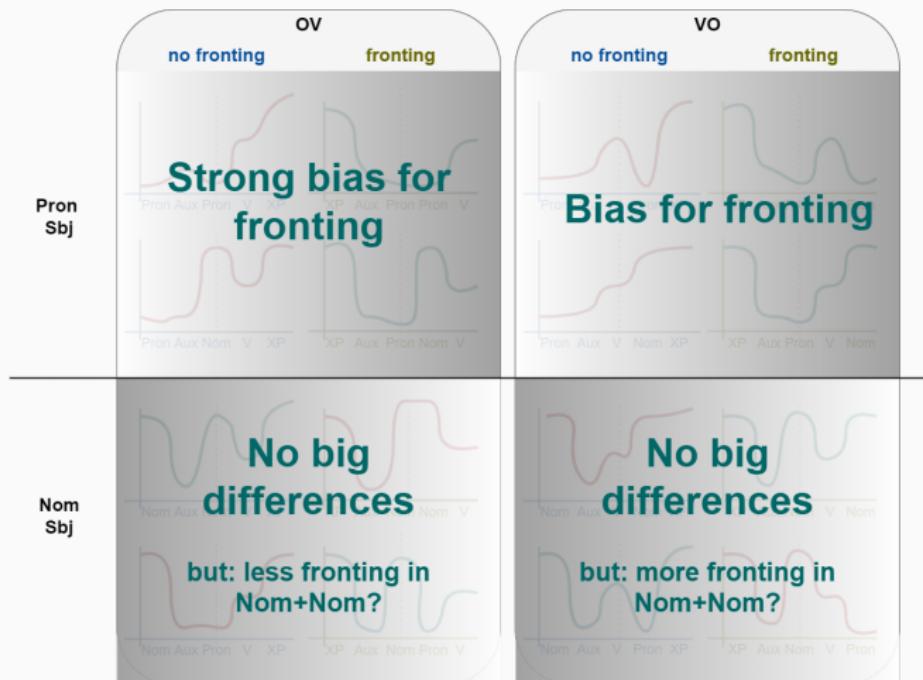
- Fronting *may* yield slightly more uniform distribution (by avoiding clustered peak of Nom+XP)

+ Pron Obj:

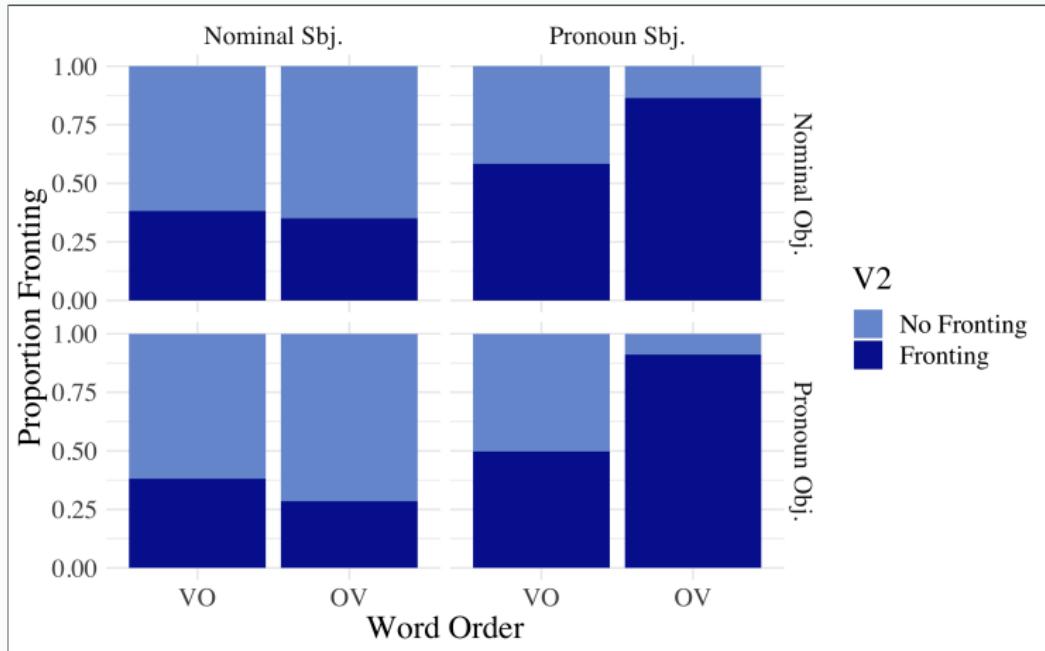
- Non-fronting might be more symmetrical

# Study 2 Predictions

Speakers deploy fronting when it yields more uniformity



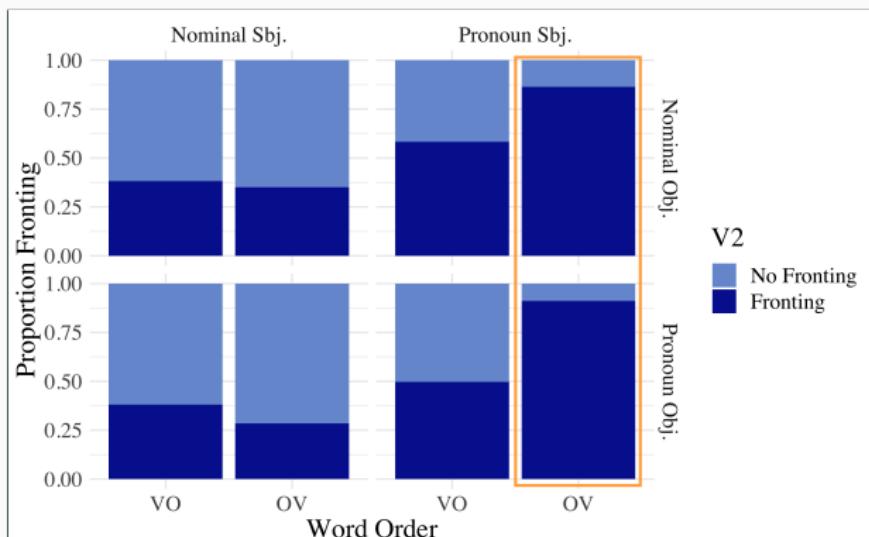
## Study 2 Results



- 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

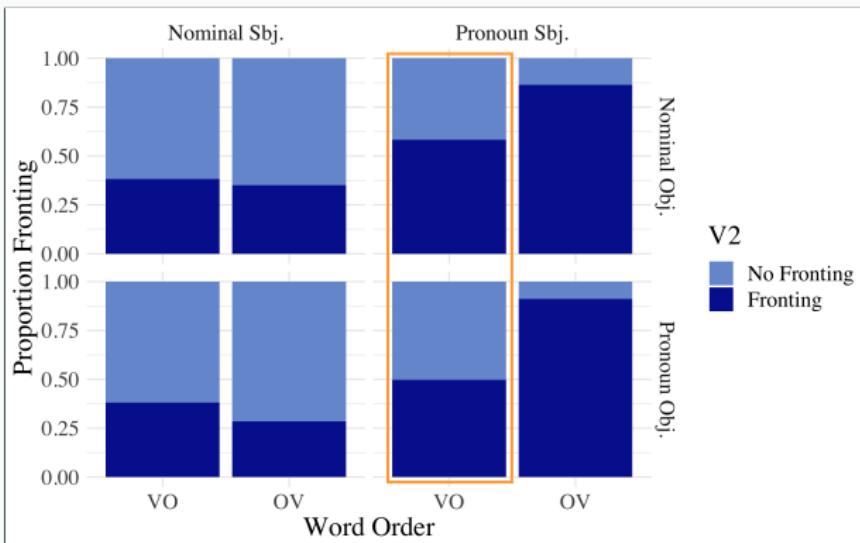
Two pressures for fronting



- 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

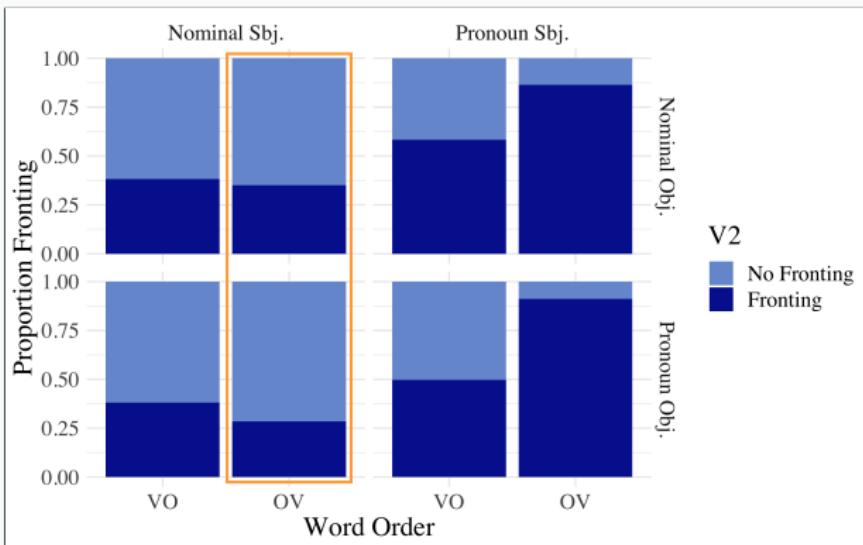


## Bias for fronting

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



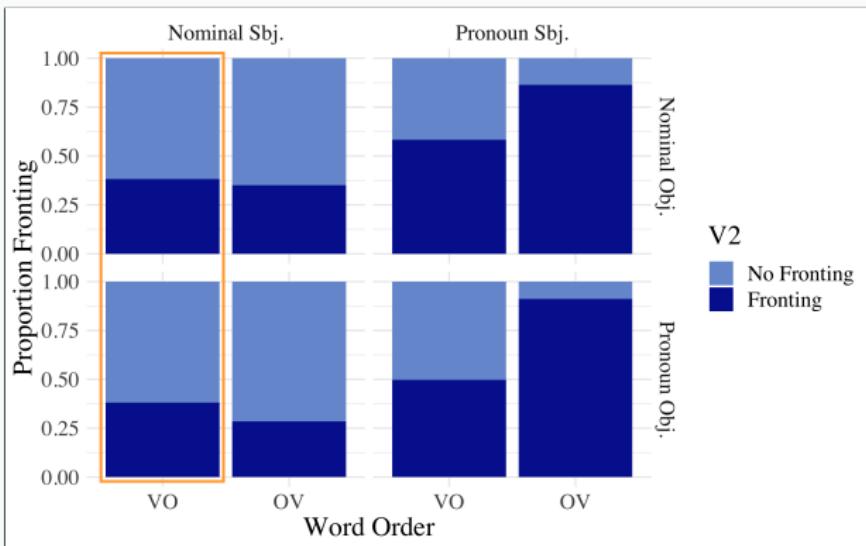
# Study 2 Results: Given OV and Nom Sbj



- 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 (a crash course)

Study 1: Information theoretic CRE in OV to VO

Study 2: Adding V2

### Discussion

## Theoretical Implications

---

- As in Study 1, Study 2 shows speakers use the syntactic resources available to optimize for information uniformity in planning.

## Theoretical Implications

---

- As in Study 1, Study 2 shows 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)

## Theoretical Implications

---

- As in Study 1, Study 2 shows 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)
- Explains lack of CRE for OV within V2 “contexts”.

## Theoretical Implications

---

- As in Study 1, Study 2 shows 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)
- Explains lack of CRE for OV within V2 “contexts”.
- Time Separation Theorem (Kauhanen & Walkden, 2018): given this order of operations, once the basic lexical content is decided, maximum/minimum uniformity is strictly bounded.

Thanks for listening!

Questions?



- Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, 214, 104754.
- 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., 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.)

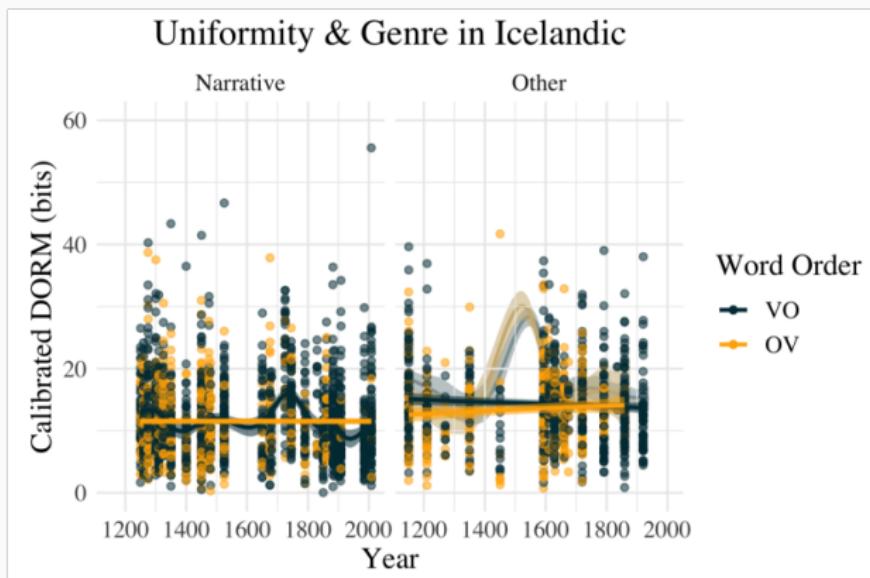
## References ii

---

- Kroch, A. S., & Taylor, A. (2000). *Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition, release 4.* (Size: 1.3 million words.)
- 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.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
- 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>.
- Wallenberg, J. C., Bailes, R., Cuskley, C., & Ingason, A. K. (2021). Smooth signals and syntactic change. *Languages*, 6(2), 60.

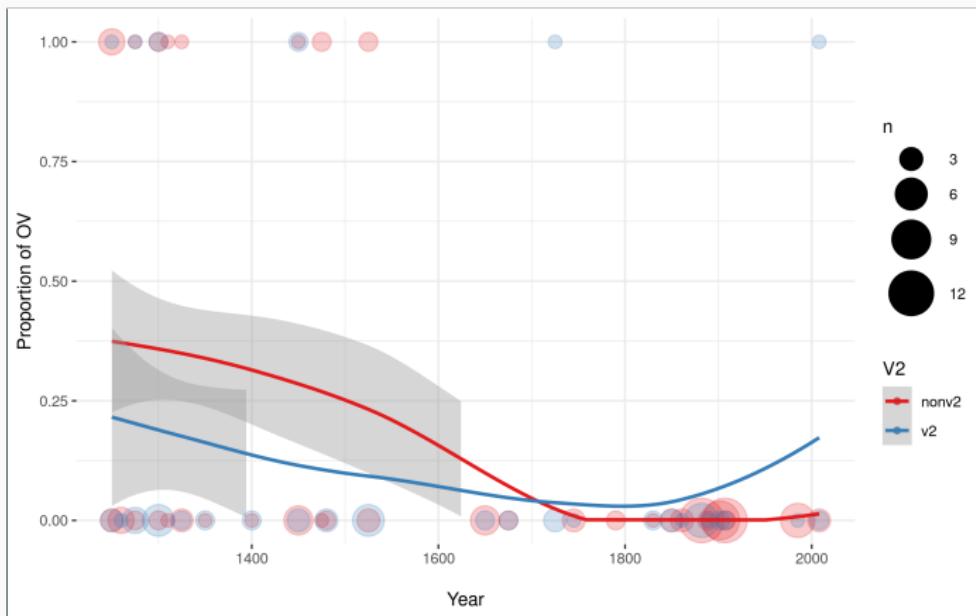
Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F., & Rögnvaldsson, E. (2011). *Icelandic Parsed Historical Corpus (IcePaHC)*. (Version 0.9. Size: 1 million words. URL  
[http://www.linguist.is/icelandic\\_treebank](http://www.linguist.is/icelandic_treebank))

# Study 1 Results: Information density remains constant through OV-to-VO

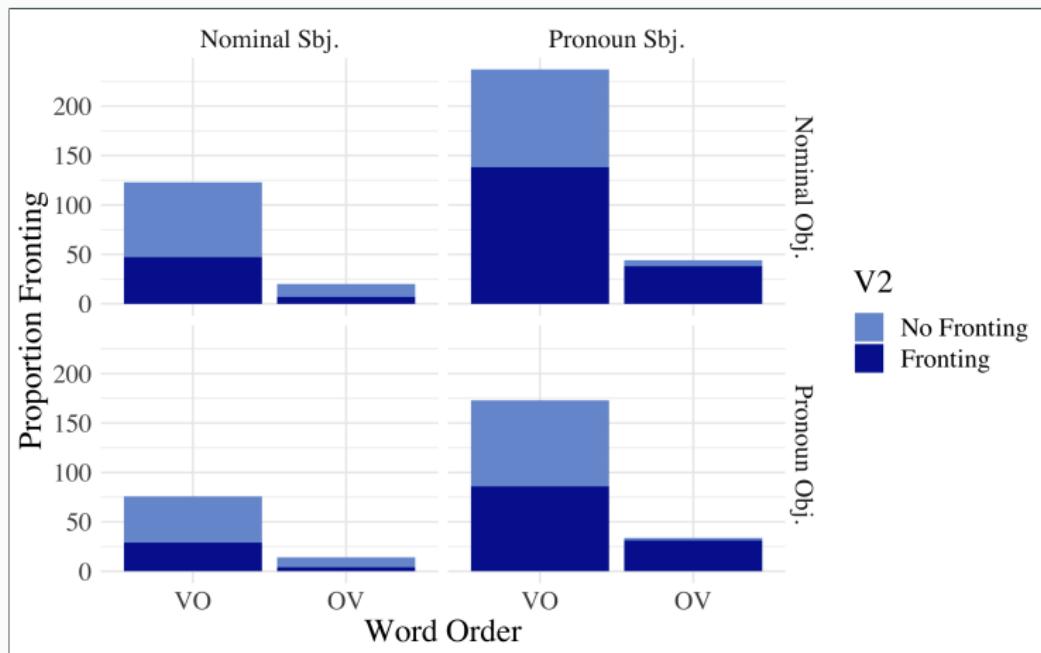


- The average uniformity of sentences is constant across the history of Icelandic.

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



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



# 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

- Algorithm that finds the most uniform distribution of a given set of values (Cuskley et al., 2021)
- Not absolute lowest DORM possible; but robust enough and computationally tractable.
- 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