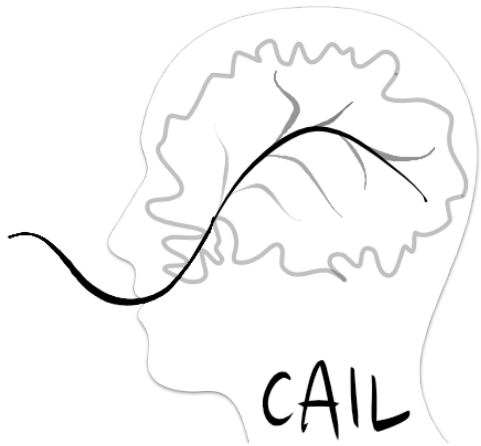
An information theoretic approach to language change

Joel C. Wallenberg
joel.wallenberg@ncl.ac.uk



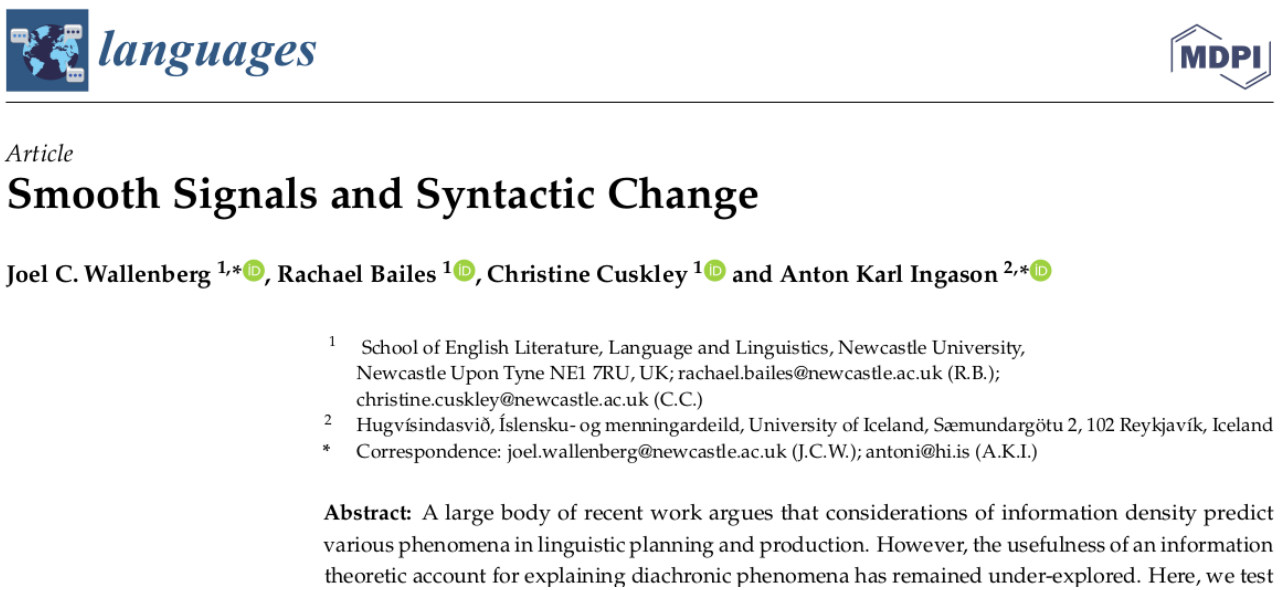
April 12, 2021

A Mystery in Language Change

- A **Constant Rate Effect** (first described in Kroch 1989): when a change in some linguistic variable is in progress, some linguistic contexts favour one variant over another **without changing the change** (i.e. stopping, slowing, accelerating, etc.).
- Using information theory, we can not only explain a CRE, but predict the existence of one.
- This CRE shows speakers unconsciously solving a complex planning problem to achieve **information uniformity**.

“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



languages

MDPI

Article

Smooth Signals and Syntactic Change

Joel C. Wallenberg [1,*], Rachael Bailes [1], Christine Cuskley [1] and Anton Karl Ingason [2,*]

[1] School of English Literature, Language and Linguistics, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK; rachael.bailes@newcastle.ac.uk (R.B.); christine.cuskley@newcastle.ac.uk (C.C.)
[2] Hugvísindasvið, Íslensku- og menningardeild, University of Iceland, Sæmundargötu 2, 102 Reykjavík, Iceland
* Correspondence: joel.wallenberg@newcastle.ac.uk (J.C.W.); antoni@hi.is (A.K.I.)

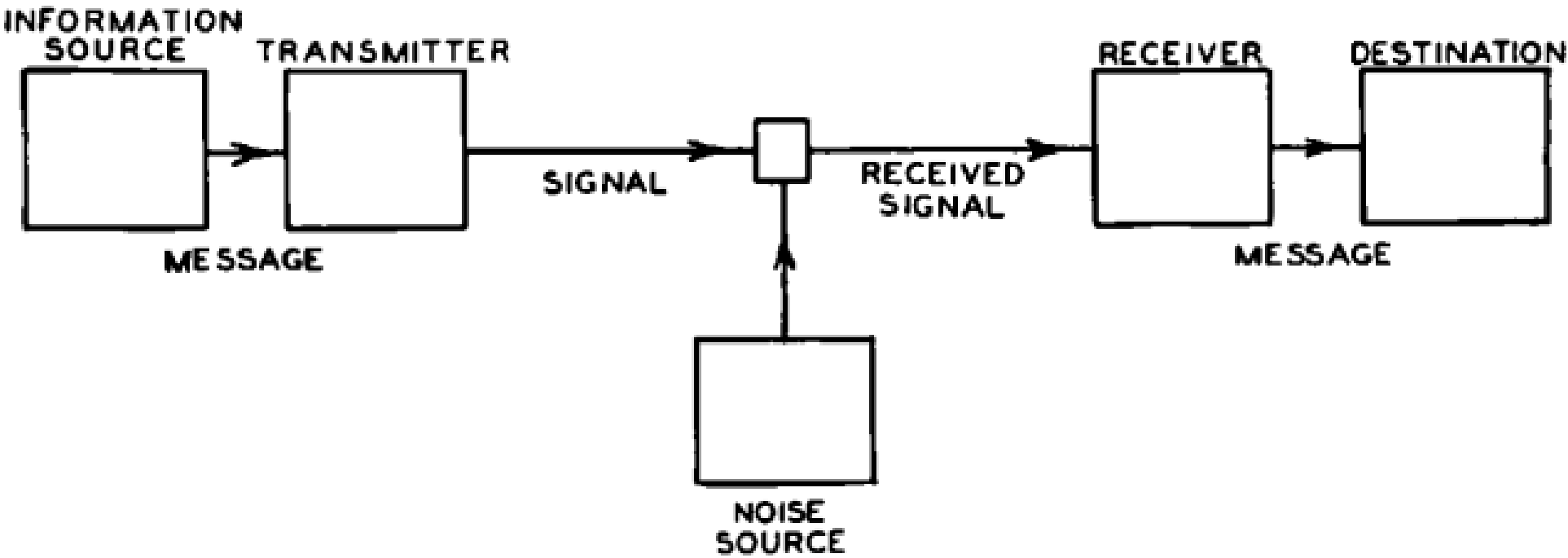
Abstract: A large body of recent work argues that considerations of information density predict various phenomena in linguistic planning and production. However, the usefulness of an information theoretic account for explaining diachronic phenomena has remained under-explored. Here, we test

Wallenberg et al. (2021)

Outline

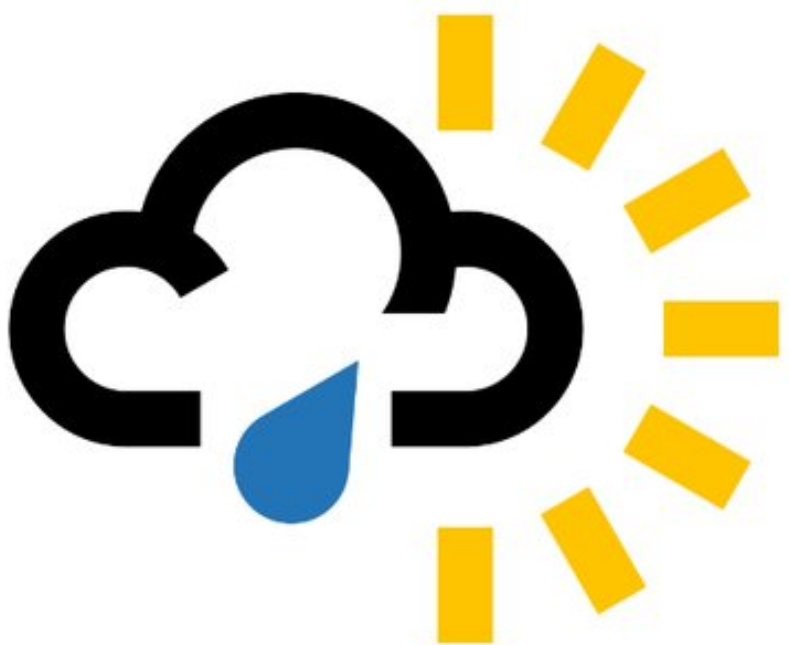
Crash course: Information theory and language

- **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).



Crash course: Information theory and language

- **Before the receiver gets any signal:** for a more uncertain event, more information could be communicated.
- **If the receiver gets a signal:** a low probability signal has given the receiver more information than a high probability one, regardless of how uncertain the event was.



Crash course: Information theory and language

- Shannon (1948)'s formula for information in an event with n discrete outcomes with probabilities $\mathrm{p}_1 ... \mathrm{p}_n$:

$$\sum_{1}^{n} p_i log_2 \frac{1}{p_i}$$

- The $log_2 \frac{1}{p_i}$ part is the *information content* of an outcome.
- Lower probability signals provide more information when received, though they show up less often.
- The unit of information is a **"bit"**!

"Information Uniformity" in Sentences

- Suppose morphemes, words, phrases are signals to the overall interpretation/function of an utterance.

  low probability → high information content

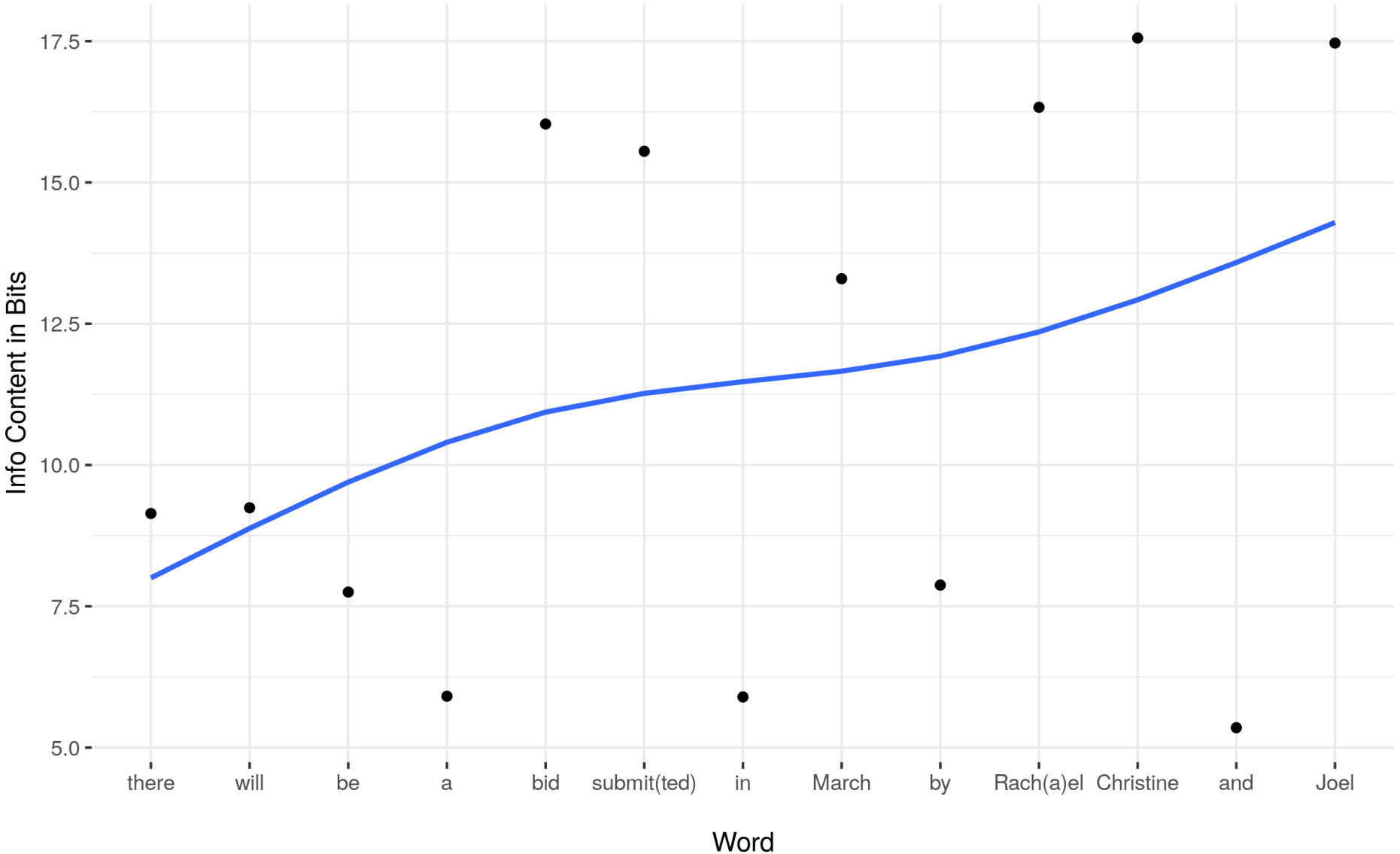
- Speakers tend to spread information across utterances as uniformly as possible, perhaps to mitigate effects of "noise":

  (Fenk and Fenk 1980; Aylett and Turk 2004; Levy and Jaeger 2007; Cuskley, Bailes & Wallenberg, *Forthcoming*)

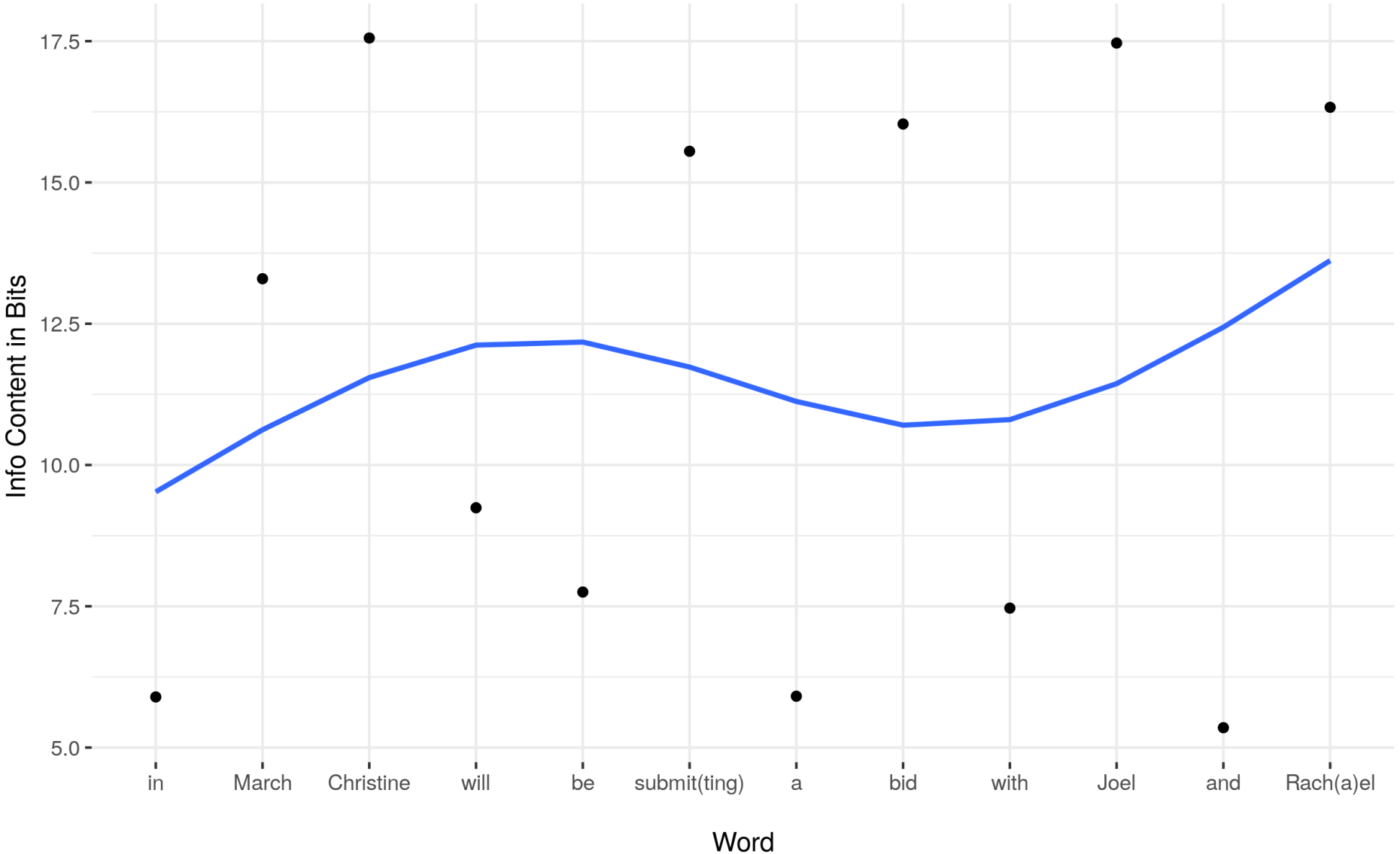
(1) How big is the family [(that) you cook for]?

- If *that* is deleted, more information is carried by *you*, creating a denser pocket of information.

"Information Uniformity" in Sentences



"Information Uniformity" in Sentences



Study 1: OV-to-VO in English and Icelandic

Middle English:

(2) Mi feader & Mi moder for-þi þt ich nule þe
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* (Kroch and Taylor, 2000))

OV-to-VO in English and Icelandic

Historical Icelandic:

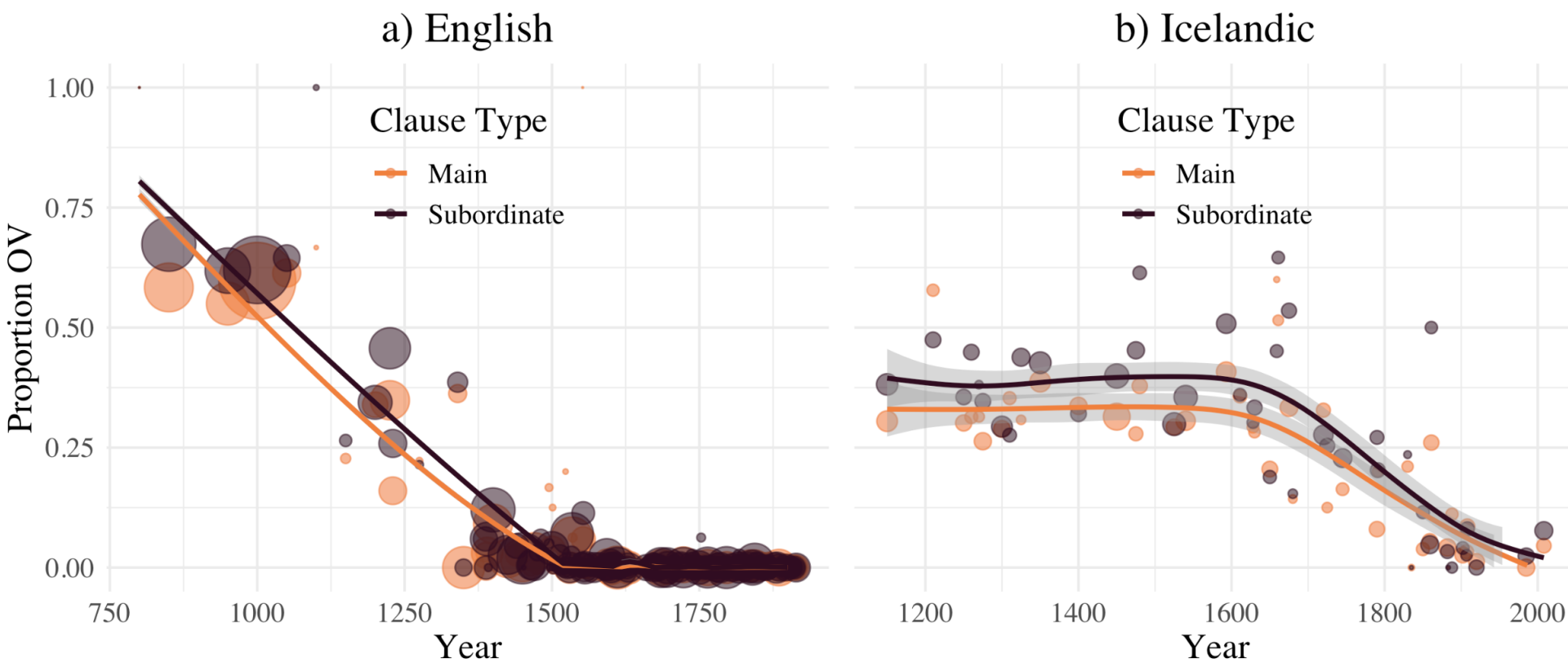
(3) a. . . . og sannleikurinn mun yður frelsa
. . . and the truth will you free
". . . and the truth will set you free."

(*Oddur Gottskálksson's New Testament*, date: 1540; ID 1540.NTJOHN.REL-BIB, 204.662 from *Icelandic Parsed Historical Corpus* (Wallenberg et al., 2011))

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)

OV-to-VO in English and Icelandic



- Note the Constant Rate Effect (CRE), shown for English by Pintzuk and Taylor (2006).

OV-to-VO and Information Theory

Constituent Type	**Average Information Content**
Pronominal DP	low ($\approx$ 11.7 bits)
Nominal DP	HIGH ($>$ 13.7 bits)
Lexical Verb	MID ($\approx$ 13.5 bits)

Sbj Aux Obj V
Sbj Aux V Obj

Hypothesis$_1$: VO is favoured when Sbj and Obj are the same type.
(low MID low , HIGH MID HIGH) **vs** (low low MID, HIGH HIGH MID)

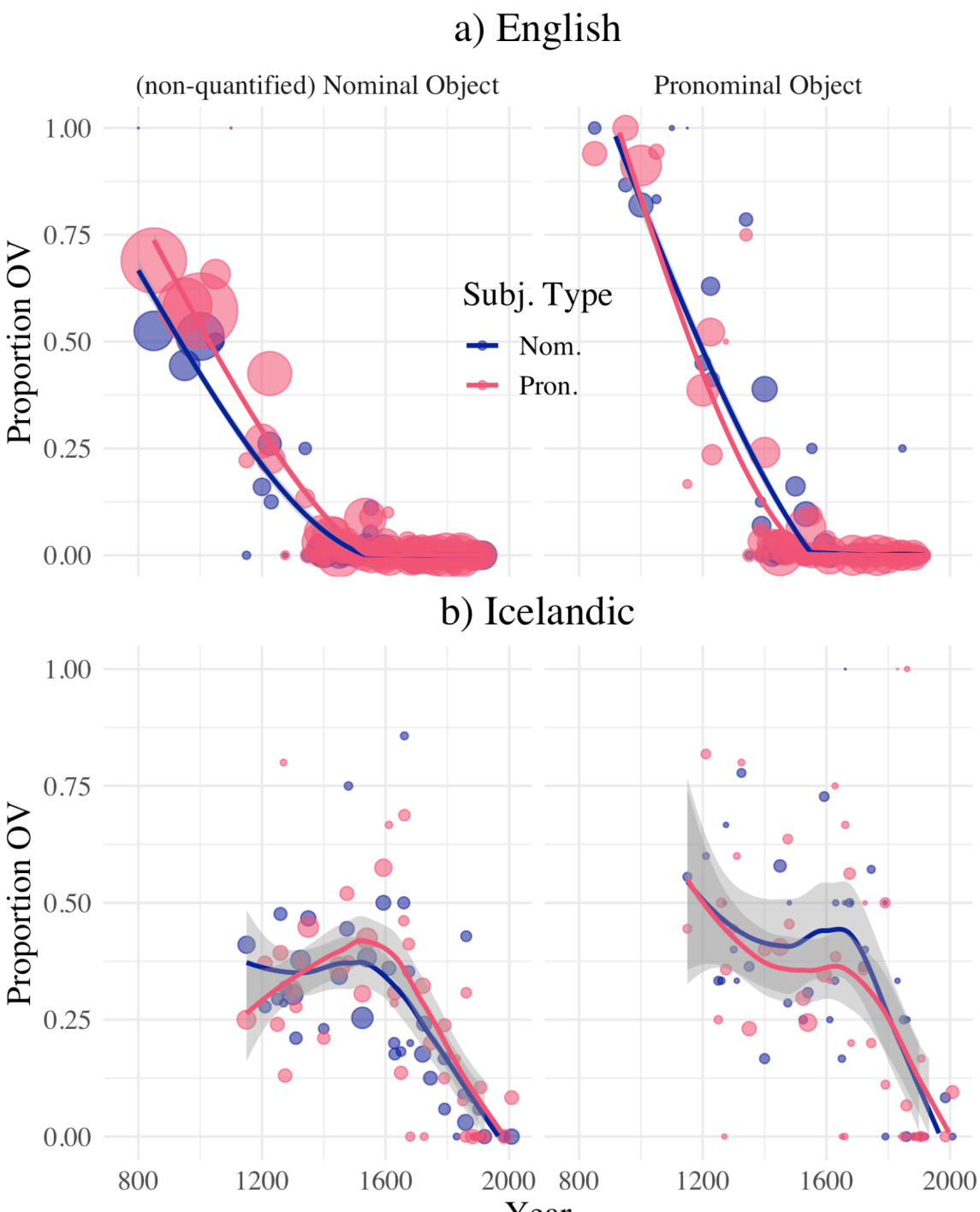
Hypothesis$_2$: OV is favoured when Sbj, Obj are **NOT** the same type.
(low HIGH MID , HIGH low MID) **vs** (low MID HIGH, HIGH MID low)

Hypothesis$_3$: These effects are orthogonal to the change (a CRE).

OV-to-VO and Information Theory

(4) sua sal ye yure sinnes les.
so shall you (**low**) your sins (**HIGH**) lose **(mid)**
"In this way, you will let go of your sins."
(*Rule of St. Benet*, Yorkshire, date: 1425)

(5) þabbes sal quaintelike drahe hir to
the-abbess (**HIGH**) shall wisely draw (**mid**) her (**low**) to hir
herself
(*Rule of St. Benet*, Yorkshire, date: 1425)


a) English
(non-quantified) Nominal Object
Pronominal Object
Proportion OV
1.00
0.75
0.50
0.25
0.00
Subj. Type
Nom.
Pron.
b) Icelandic
800
1200
1600
2000

"DORM": Deviation of the Rolling Mean

- A **summary statistic** for how informationally uniform a sentence is (Cuskley, Bailes & Wallenberg, *Forthcoming*).

low DORM → more uniform
high DORM → more lopsided

- Based on strings of lemmas for Icelandic sentences, due to large number of morphological forms (and some spelling variation).

"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
6.79	**6.15**	10.1	9.25	6.15	10.4
6.47					

"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				

"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			

"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		

"DORM": Deviation of the Rolling Mean

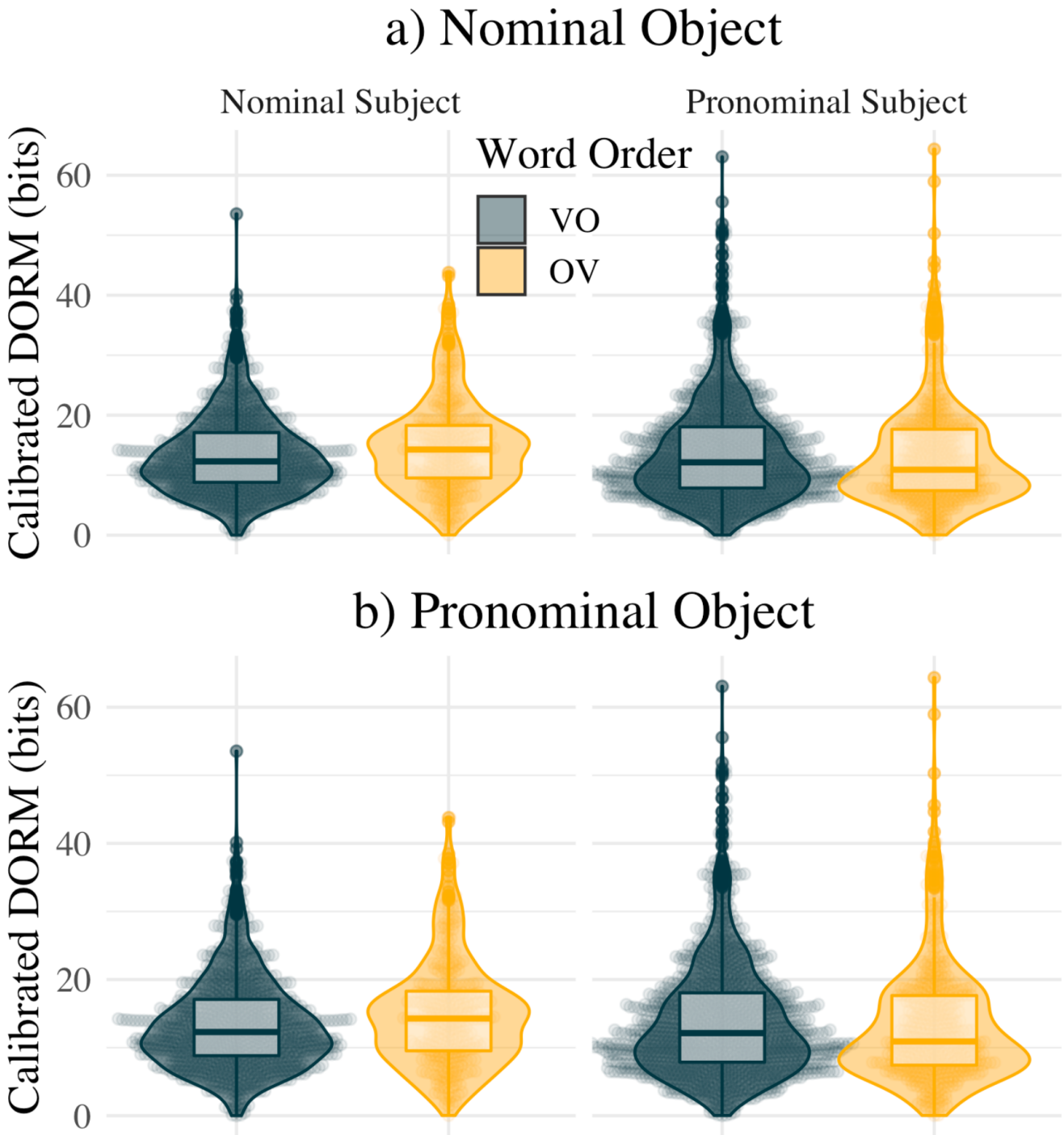
en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	**6.15**	**10.4**
6.47	8.12	9.67	7.70	8.29	

"DORM": Deviation of the Rolling Mean

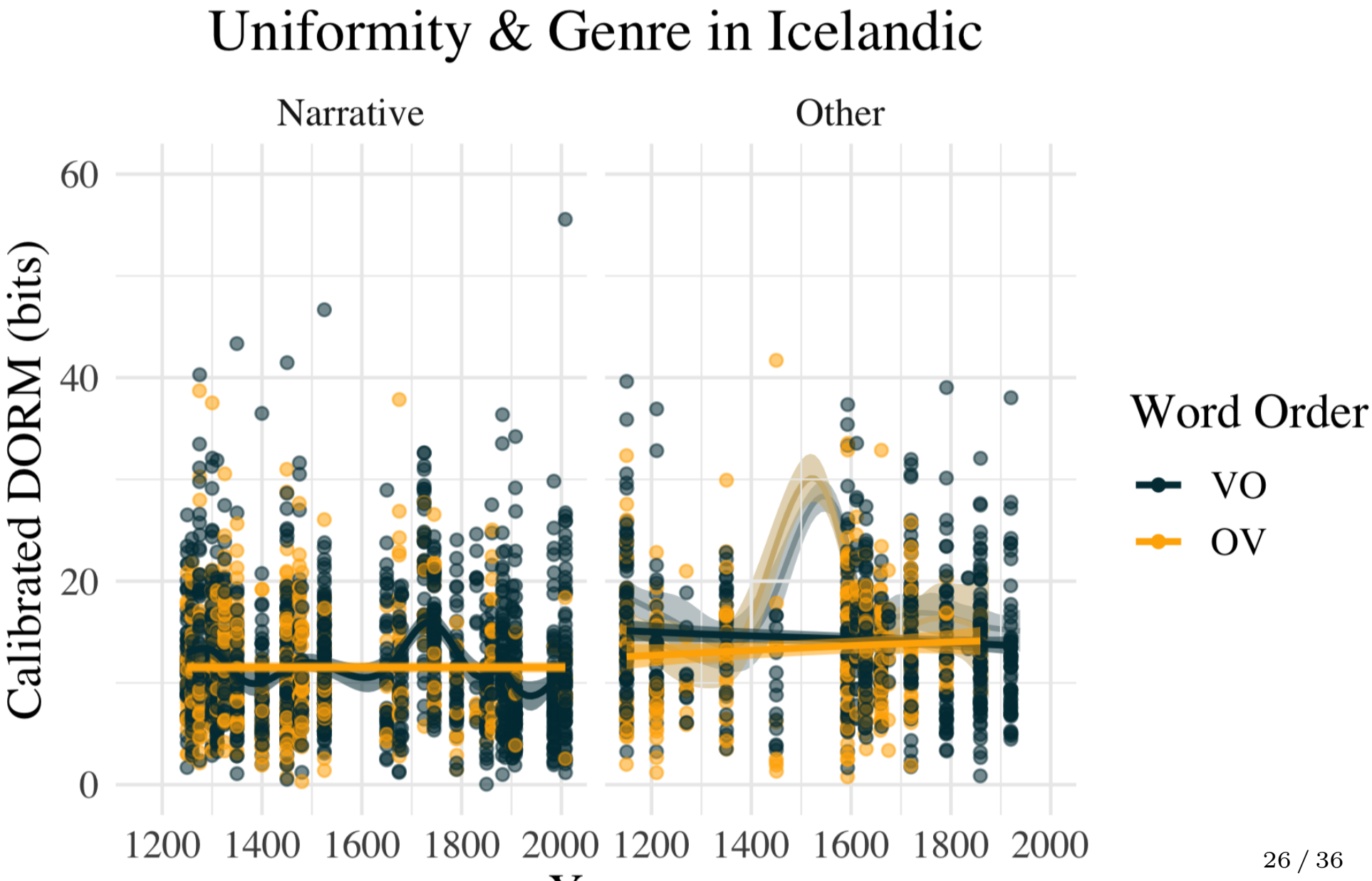
en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	**6.15**	**10.4**
6.47	8.12	9.67	7.70	8.29	

Sample variance of rolling means = 1.33 bits
(plus a further calibration for length and lexical idiosyncracy)

low DORM → more uniform
high DORM → more lopsided


a) Nominal Object
Nominal Subject
Pronominal Subject
Word Order
VO
OV
Calibrated DORM (bits)
60
40
20
0
b) Pronominal Object
Calibrated DORM (bits)
60
40
20
0

What Doesn't Change, Doesn't Change



Conclusions

- It is exciting to be able to predict patterns of language change with this level of precision.

Conclusions

- It is exciting to be able to predict patterns of language change with this level of precision.
- We **predicted and explained** a very subtle Constant Rate Effect on the VO-to-VO changes in both Icelandic and English, which is unexpected without an information theoretic approach.

Conclusions

- It is exciting to be able to predict patterns of language change with this level of precision.
- We **predicted and explained** a very subtle Constant Rate Effect on the VO-to-VO changes in both Icelandic and English, which is unexpected without an information theoretic approach.
- The average uniformity of sentences is constant across the history of Icelandic.

Conclusions

- It is exciting to be able to predict patterns of language change with this level of precision.
- We **predicted and explained** a very subtle Constant Rate Effect on the VO-to-VO changes in both Icelandic and English, which is unexpected without an information theoretic approach.
- The average uniformity of sentences is constant across the history of Icelandic.
- Even while generations of speakers are participating in the OV-to-VO change, they use their syntactic resources to keep a target of information uniformity.

Conclusions

- It is exciting to be able to predict patterns of language change with this level of precision.
- We **predicted and explained** a very subtle Constant Rate Effect on the VO-to-VO changes in both Icelandic and English, which is unexpected without an information theoretic approach.
- The average uniformity of sentences is constant across the history of Icelandic.
- Even while generations of speakers are participating in the OV-to-VO change, they use their syntactic resources to keep a target of information uniformity.
- This complex unconscious planning could be a deep property of the linguistic system (and perhaps the memory system).

Future Directions

- **Replicate:** for historical English in lemmatised versions of the *Penn Parsed Historical Corpora*, and hopefully *York Corpus of Old English Prose* (Taylor et al., 2003)

Future Directions

- **Replicate:** for historical English in lemmatised versions of the *Penn Parsed Historical Corpora*, and hopefully *York Corpus of Old English Prose* (Taylor et al., 2003)
- **Collaboration with Neuroscience:** Information uniformity, episodic memory, and ageing (Wallenberg, Smulders, Cuskley & Read, *TBS to Scientific Reports*)

Future Directions

- **Replicate:** for historical English in lemmatised versions of the *Penn Parsed Historical Corpora*, and hopefully *York Corpus of Old English Prose* (Taylor et al., 2003)
- **Collaboration with Neuroscience:** Information uniformity, episodic memory, and ageing (Wallenberg, Smulders, Cuskley & Read, *TBS to Scientific Reports*)
- **Public Engagement / Impact:** Centre for Life, National Centre for the Written Word

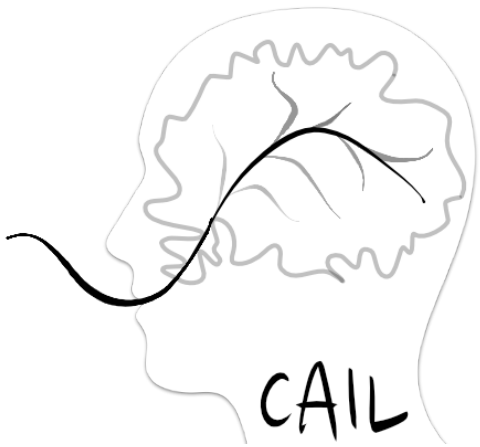
Future Directions

- **Replicate:** for historical English in lemmatised versions of the *Penn Parsed Historical Corpora*, and hopefully *York Corpus of Old English Prose* (Taylor et al., 2003)
- **Collaboration with Neuroscience:** Information uniformity, episodic memory, and ageing (Wallenberg, Smulders, Cuskley & Read, *TBS to Scientific Reports*)
- **Public Engagement / Impact:** Centre for Life, National Centre for the Written Word
- **Crazy idea**: language and "ruin"
(collaboration with York Actuarial Science?)

Acknowledgements

Thanks to Rachael Bailes, Christine Cuskley, and colleagues at the CBE.
This research was funded by ESRC grant ES/T005955/1.

`https://github.com/joelcw/constantentropy`
`https://github.com/joelcw/iceBits`



References I

Aylett, Matthew, and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech* 47:31–56.

Cuskley, Christine, Rachael Bailes, and Joel C. Wallenberg. 2021. Noise resistance in communication: Quantifying uniformity and optimality.

Fenk, August, and Gertraud Fenk. 1980. Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie* 27:402.

Hartley, Ralph VL. 1928. Transmission of information 1. *Bell System technical journal* 7:535–563.

References II

Ingason, Anton Karl, Einar Freyr Sigurðsson, and Joel C. Wallenberg. 2012. Antisocial Syntax: Disentangling the Icelandic VO/OV parameter and its lexical remains. In *Presented at the 14th meeting of the Diachronic Generative Syntax (DIGS) Conference, July 6, 2012*. University of Lisbon.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

Kroch, Anthony S., and Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition, release 4. Size: 1.3 million words.

Levy, Roger P., and Florian T. Jaeger. 2007. Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, 849–856.

References III

Pintzuk, Susan, and Ann Taylor. 2006. The loss of OV order in the history of English. In *Blackwell handbook of the history of English*, ed. A. van Kemenade and B. Los, 247–278. Blackwell.

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27:379–423.

Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. `http://www-users.york.ac.uk/~lang22/YcoeHome1.htm`.

Wallenberg, Joel C, Rachael Bailes, Christine Cuskley, and Anton Karl Ingason. 2021. Smooth signals and syntactic change. *Languages* 6:60.

Wallenberg, Joel C., Anton K. Ingason, Einar F. Sigurðsson, and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.9. Size: 1 million words. URL `http://www.linguist.is/icelandic_treebank`.

Crash course

- The amount of information in a fair coin toss is 1 bit.
- The amount of information in an unfair coin toss with

$$p = \frac{1}{3}, \frac{2}{3}$$

is less, even though less probable events have higher information content.

Statistics: OV-to-VO in English

`OV ~ Clause + zYear + SbjType + ObjType + SbjType*ObjType*zYear`

Term	β	**p-value**
pronSbj:pronObj	-0.66	0.015
nomSbj:nomObj	-0.67	0.01

Slope estimates not significantly non-zero for interaction with Text Date, $0.221 \leq p \leq 0.884$ depending on the argument combinations.

Statistics: OV-to-VO in Icelandic

$$\texttt{OV} \sim \texttt{Clause + zYear + SbjType + ObjType + SbjType*ObjType*zYear}$$

Term	β	**p-value**
pronSbj:pronObj	-0.271	0.085
nomSbj:nomObj	-0.271	0.085
nomSbj:quantObj	-0.554	9.36×10^{-3}

Slope estimates not significantly non-zero for interaction with Text Date, $0.221 \leq p \leq 0.884$ depending on the argument combinations.

Statistics: OV and VO variation in historical Icelandic

```
SentDormUido ~ (1 | TextId) + Year + OV + Clause +
SimpleGenre + ObjType +
SbjType + SbjType * ObjType * OV
```

Term	β	**p-value**
pronSbj:pronObj:OV	2.66	0.014
nomSbj:nomObj:OV	2.66	0.014
pronSbj:nomObj:OV	-2.66	0.014
nomSbj:pronObj:OV	-2.66	0.014

Effect of Text Date on calibrated DORM not significantly different from zero:

$0.524 \leq p \leq 0.579$