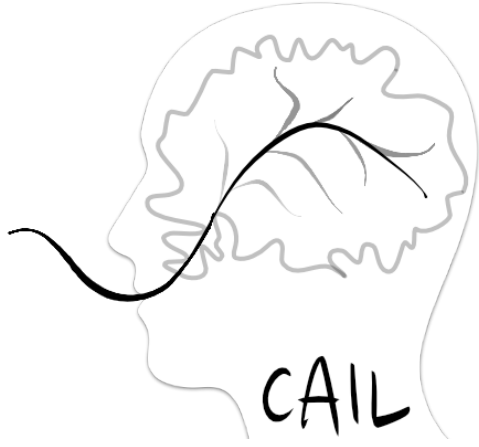
How I learned to stop worrying and love the leave

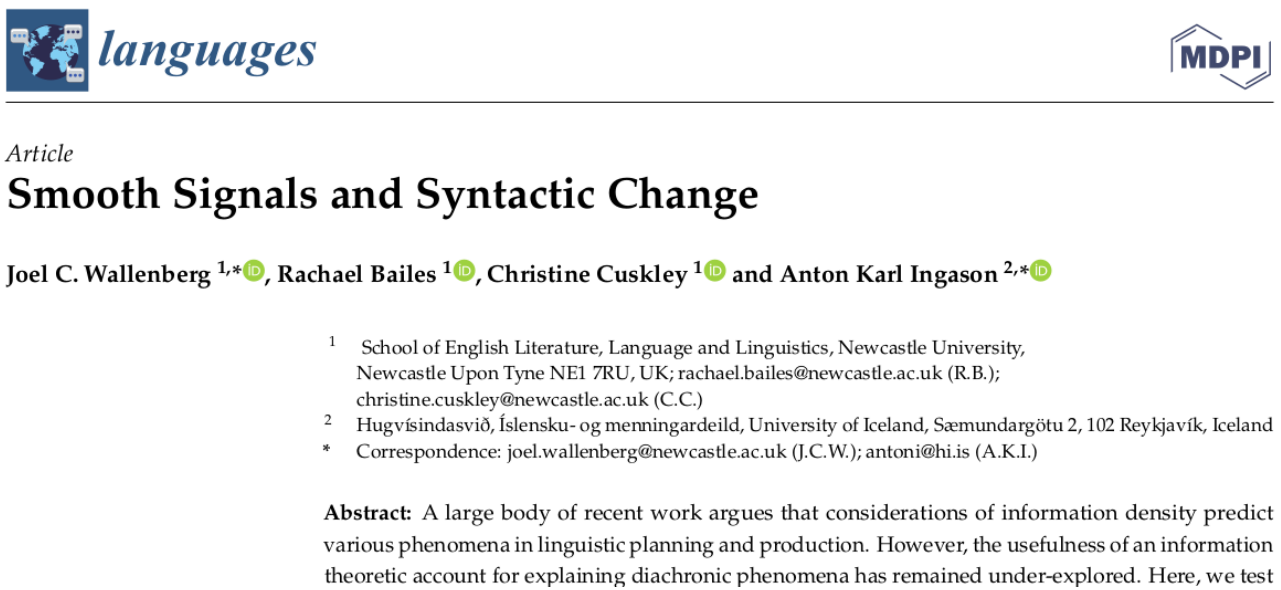
Joel C. Wallenberg
joel.wallenberg@ncl.ac.uk



April 9, 2021

"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.)

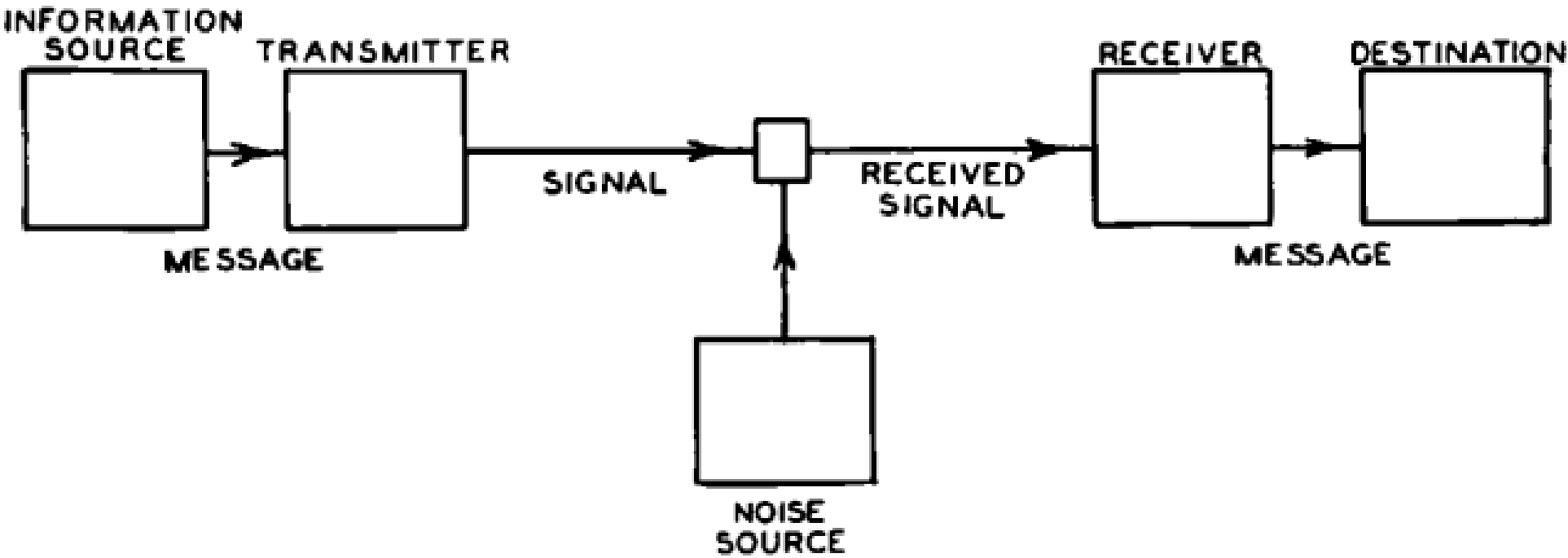
Abstract: A large body of recent work argues that considerations of information density predict various phenomena in linguistic planning and production. However, the usefulness of an information theoretic account for explaining diachronic phenomena has remained under-explored. Here, we test

Wallenberg et al. (2021)

Outline

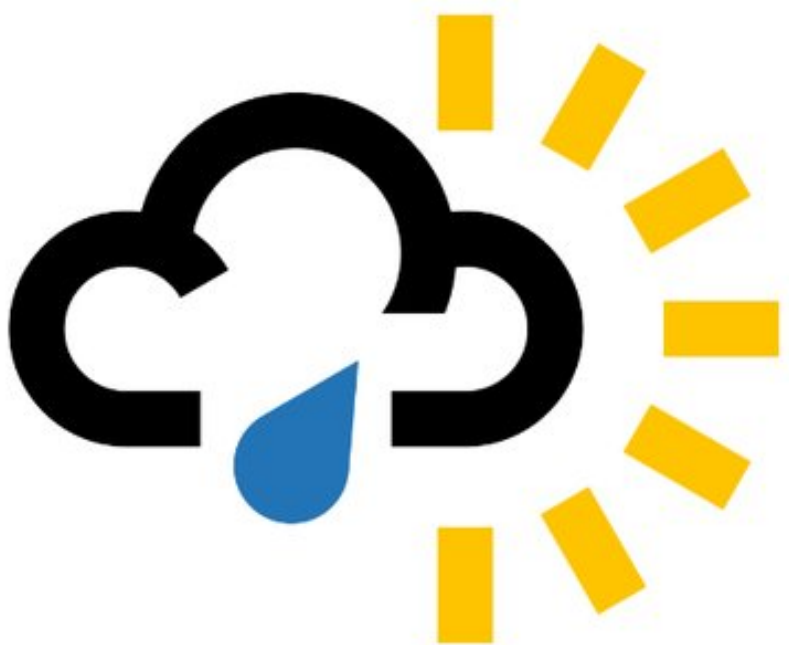
Crash course: Information theory and language

- **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"**!

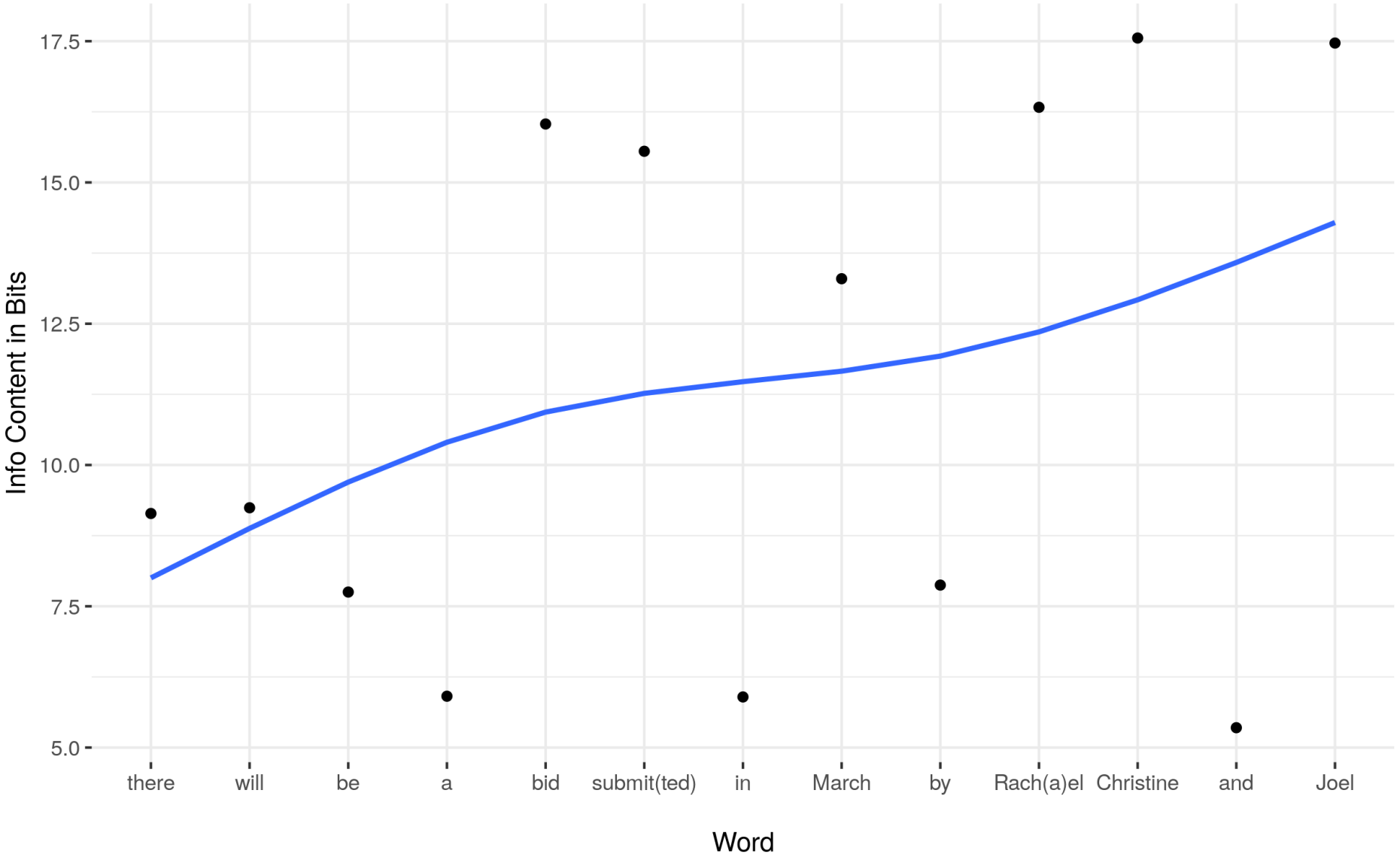
"Uniform Information Density" in language

- Any linguistic unit can be thought of as a signal to the overall content/function of an utterance.
- "noise" is any interference, including: noise, memory, other processing costs, etc.
- Speakers tend to spread information content across utterances as uniformly as possible (Fenk and Fenk 1980; Aylett and Turk 2004; Levy and Jaeger 2007; Cuskley, Bailes & Wallenberg, *Forthcoming*).

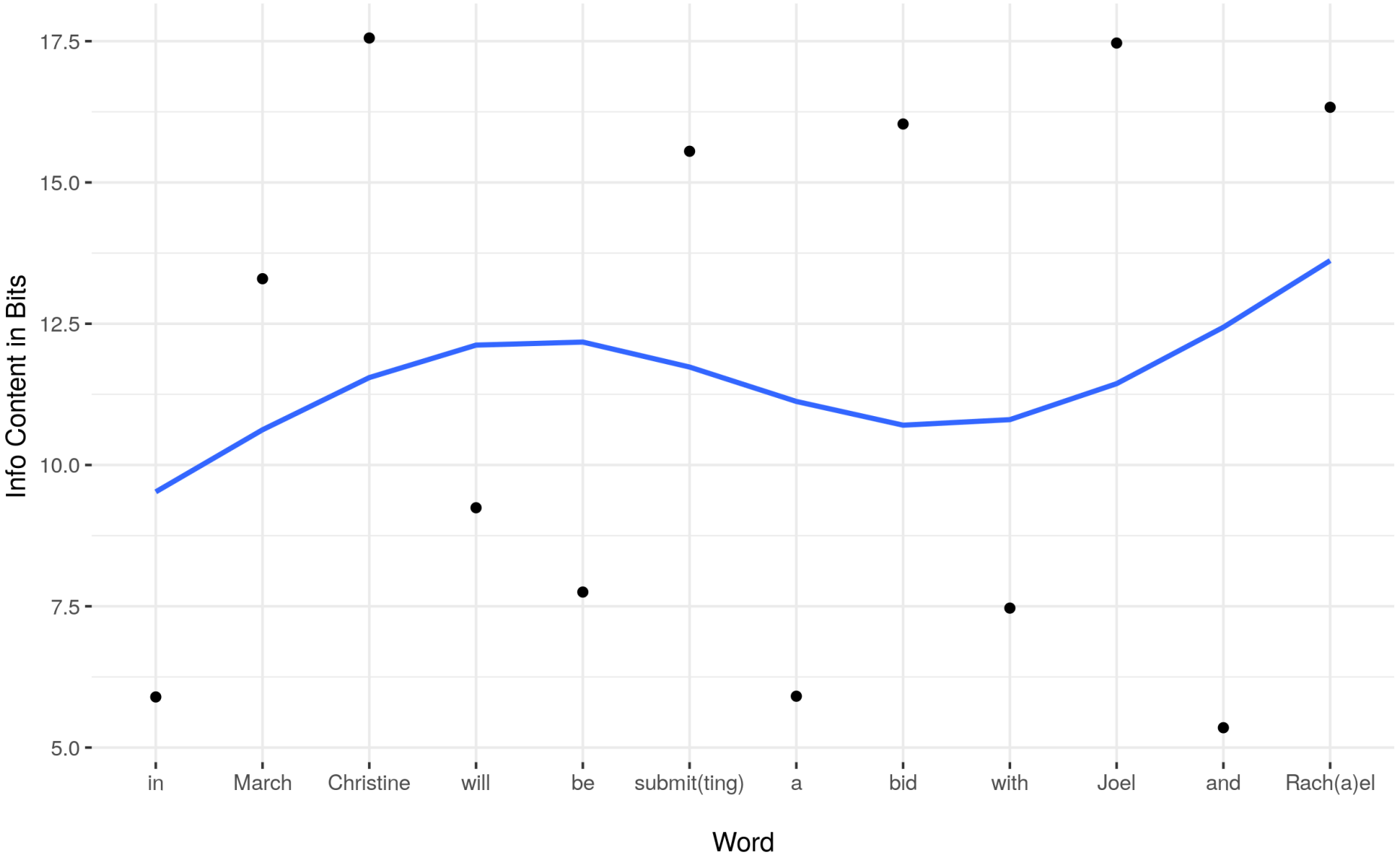
(1) How big is the family [(that) you cook for]?

If *that* is deleted, more information is carried by *you*, so information is more dense.

UID and noise resistance



UID and noise resistance



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:

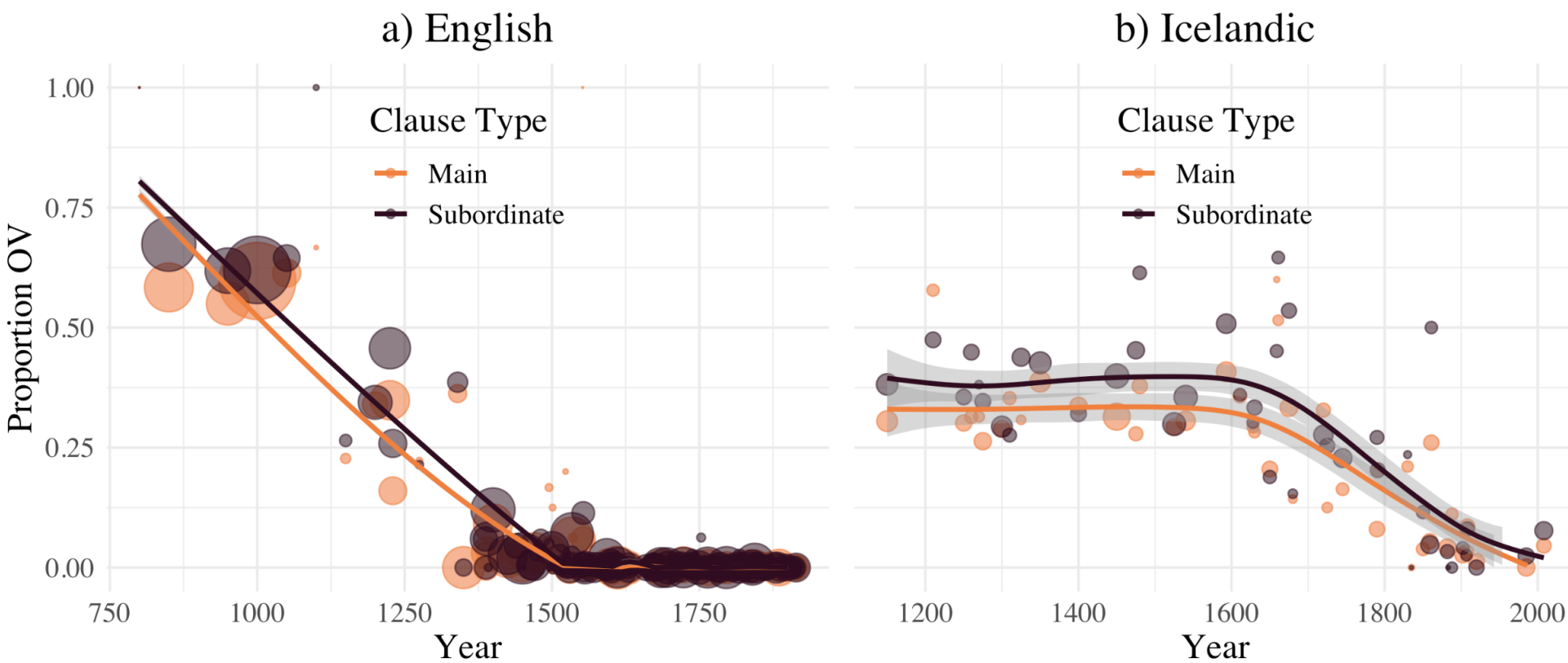
(3) a. . . . og sannleikurinn mun yður frelsa
. . . and the truth will you free
". . . and the truth will set you free."

(*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)

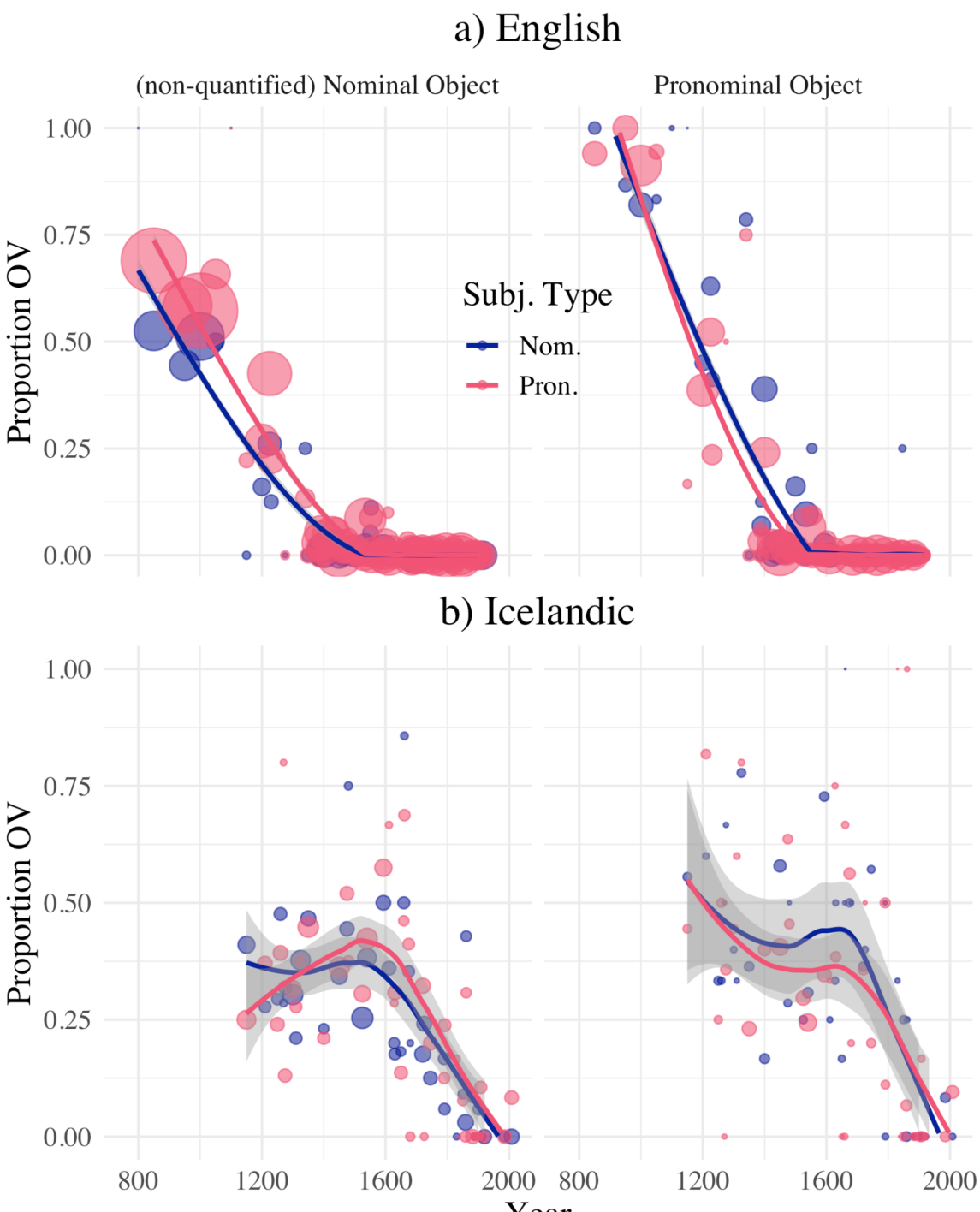
OV-to-VO in English and Icelandic



- Note the Constant Rate Effect (Kroch, 1989), shown for English by Pintzuk and Taylor (2006).

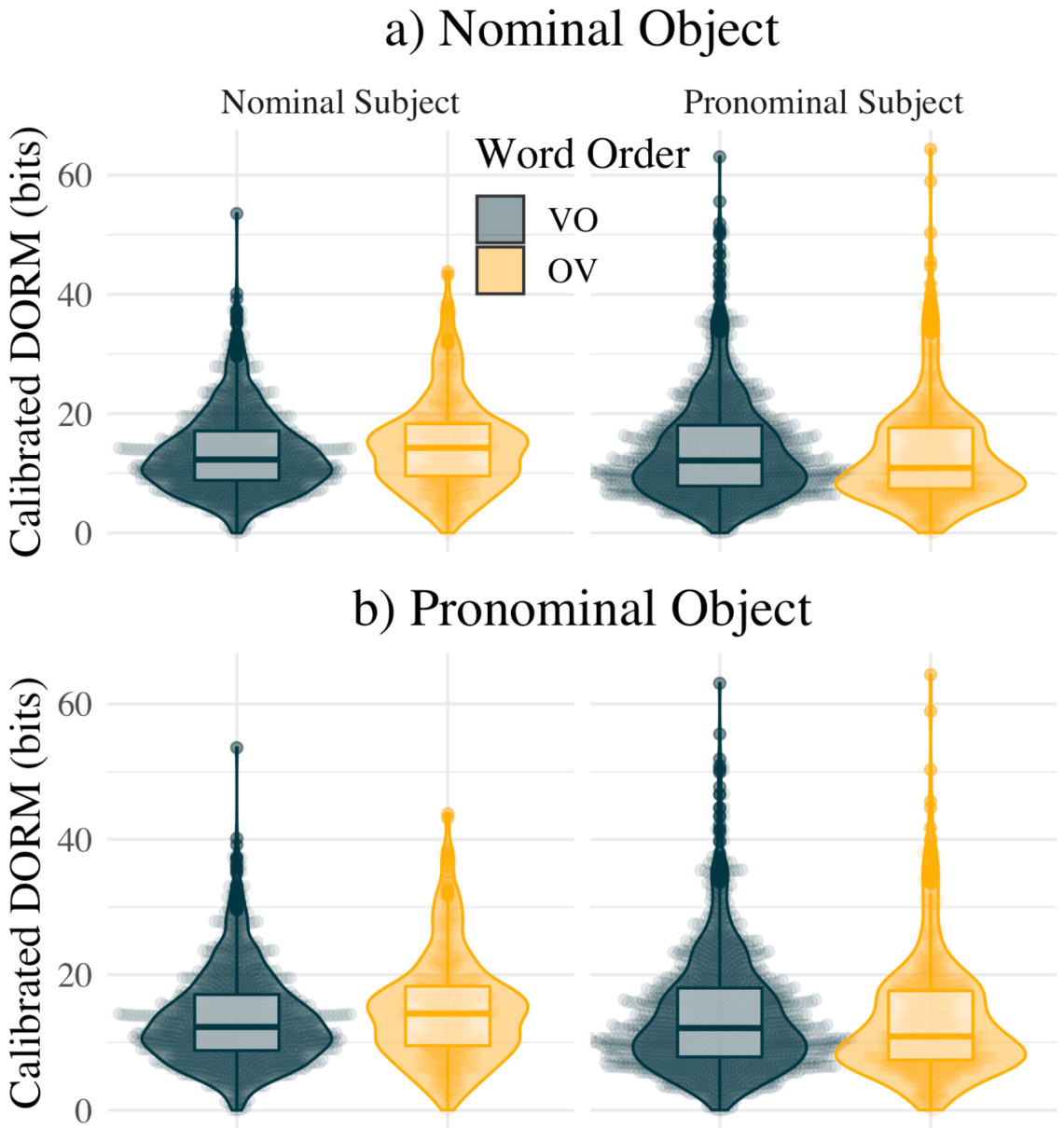
OV-to-VO and Information Theory

1. Pronouns are closed-class and frequent, and so are low information content, (cf. Shannon 1948) in comparison with nominal DPs (e.g., *you* vs. *the politician*). For example, the average information content value for pronouns in the Penn Parsed Corpus of Modern British English (PPCMBE; Kroch et al. 2016) is 11.7 bits.
2. Nominal DP Objects are high information content in general, because any particular DP will be low probability. There are more common noun lexemes than any other part of speech, and so any given noun is lower probability by virtue of belonging to a very large set. The probability of a given noun *combined with* a determiner and/or modifiers is necessarily lower than the probability of the noun alone. For example, the average information content for nouns in PPCMBE is 13.7 bits, and we can reasonably expect the average information content is much higher for nominal DPs, since they can be of arbitrary complexity.
3. Verbs are mid-level in frequency and so also in information content: lower than nominal DPs, on average, but higher than

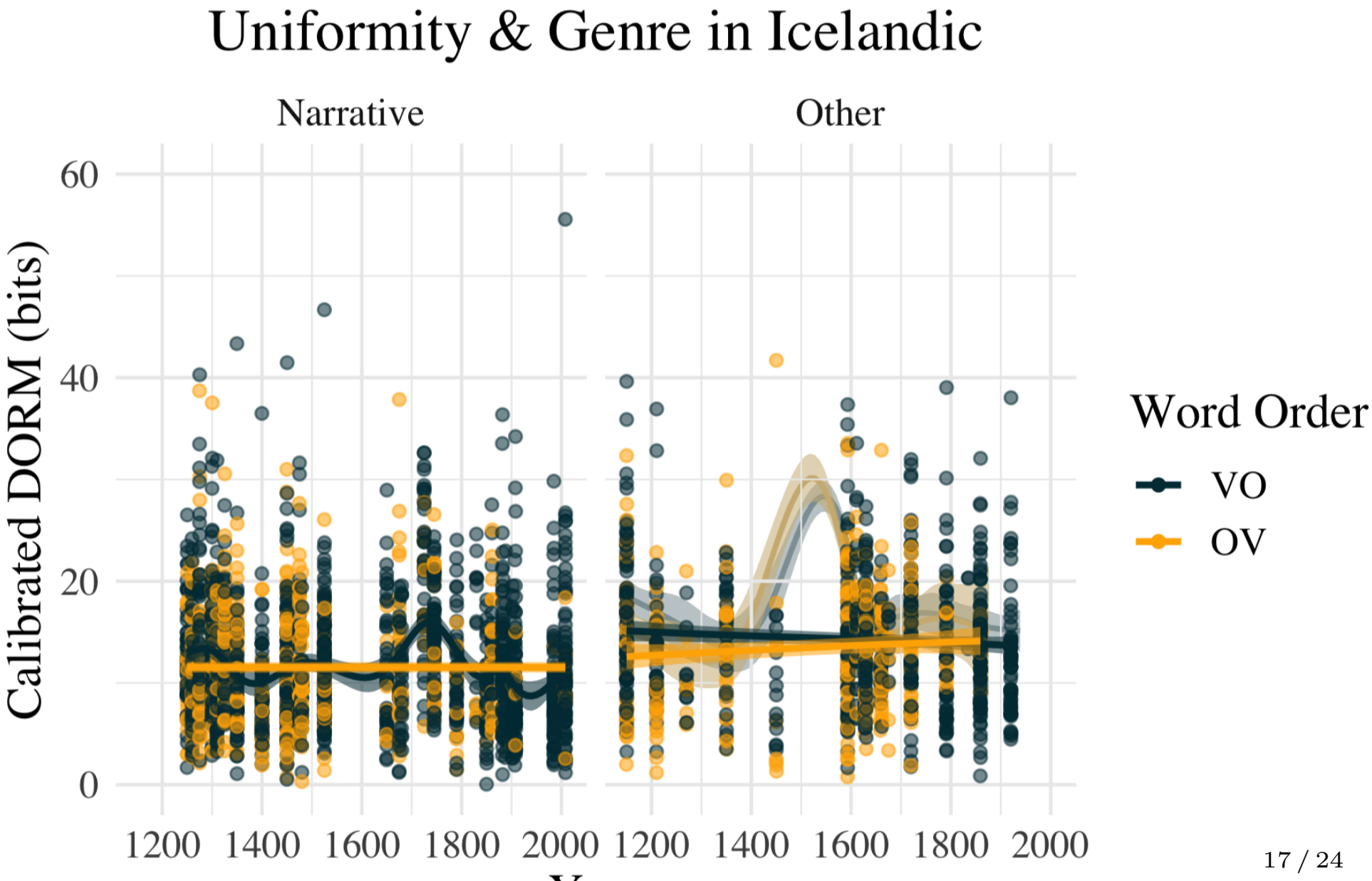

a) English
(non-quantified) Nominal Object
Pronominal Object
Proportion OV
1.00
0.75
0.50
0.25
0.00
Subj. Type
Nom.
Pron.
b) Icelandic
800
1200
1600
2000

DORM

XXX lemmas


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



What Doesn't Change, Doesn't Change (Kroch, Labov, p.c.)

variance XXX

Conclusions

- XXX

Future Directions

- XXX

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.

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

References II

Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani. 2016. Penn Parsed Corpus of Modern British English 2nd edition, release 1. Size $\sim$ 2.8 million words.

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.

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.

References III

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.