

How useful is information theory in predicting patterns of language use?

Joel C. Wallenberg

joel.wallenberg@ncl.ac.uk



May 14, 2019

Overall Points

- **Shameless trick:** I don't really know *how* useful, and it's a limited domain of "language use".
 - But the results are still very surprising, and at a level of detail that is only rivalled by one or two other results in the field.
- **Main idea:** speakers unconsciously manipulate linguistic structure so that their utterances are more resistant to "noise" events that could destroy a whole message.
- "noise" is any interference, including: noise, memory, other processing costs, etc.

Outline

Information Theory

- Crash Course

- Noise resistance simulation

Language Change

- Syntactic change in English

- Syntactic change in Icelandic

Conclusion

Crash course

- The amount of information a sender can theoretically communicate about an event is the uncertainty (“entropy”) the receiver has about the event.
- ?’s formula for information in an event with n discrete outcomes with probabilities $p_1 \dots p_n$:

$$\sum_1^n p_i \log_2 \frac{1}{p_i}$$

- The $\log_2 \frac{1}{p_i}$ part is the *information content* (or “surprisal”) of an outcome.

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.



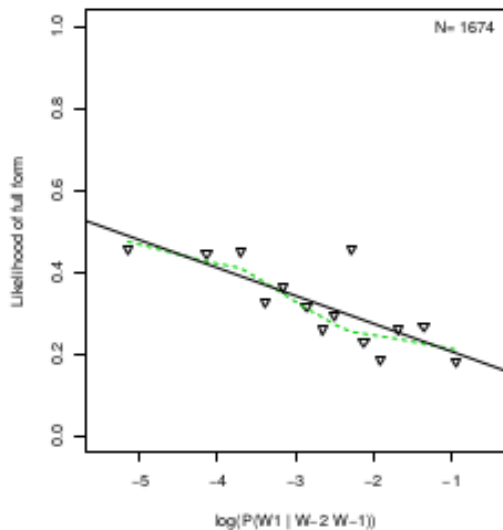
“Uniform Information Density” in language

Speakers spread information content across utterances as uniformly as possible, (possibly) so that utterances are more resistant to noise events (Aylett and Turk, 2004; Jaeger and Levy, 2007; Levy, 2008).

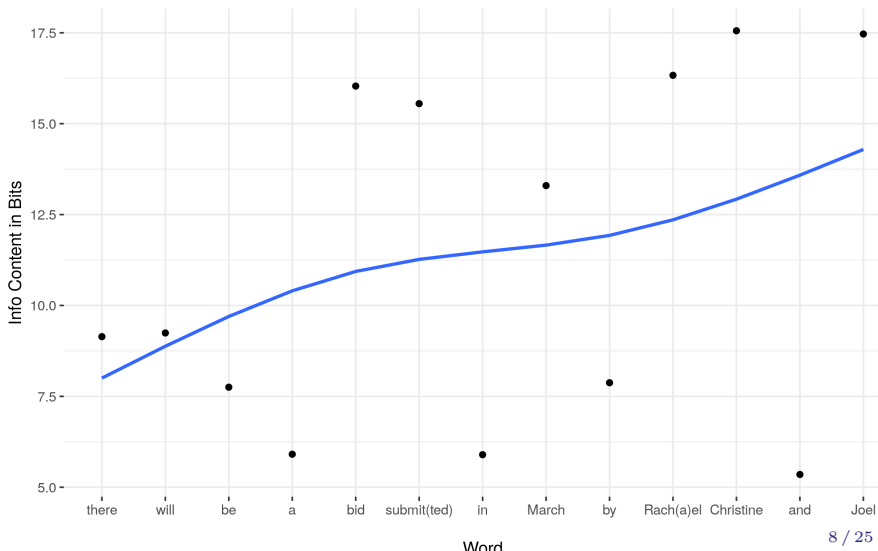
(1) How big is the family [(that) you cook for]?

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

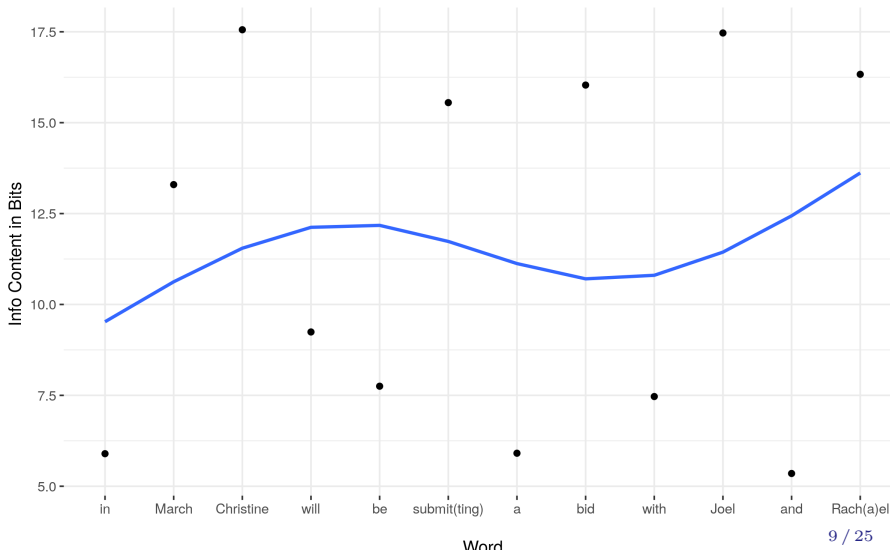
“Uniform Information Density”



UID and noise resistance



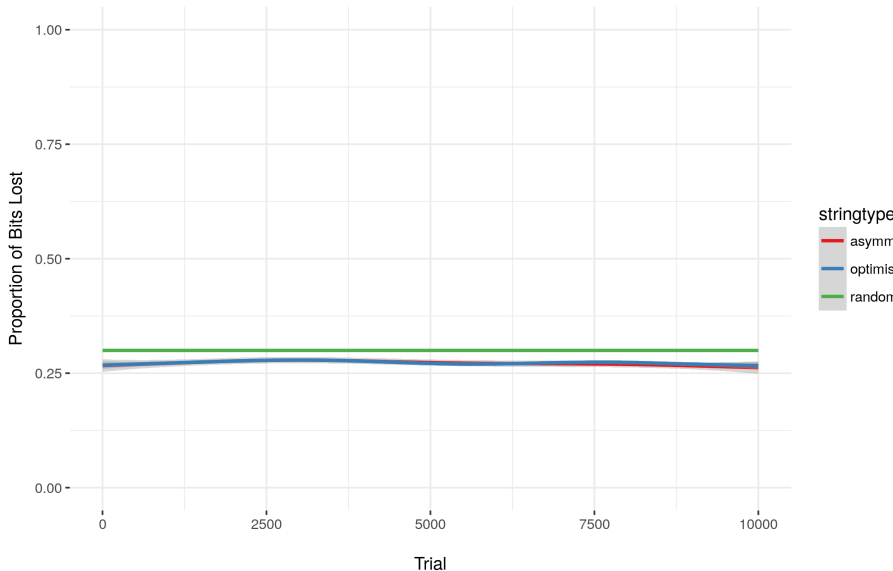
UID and noise resistance



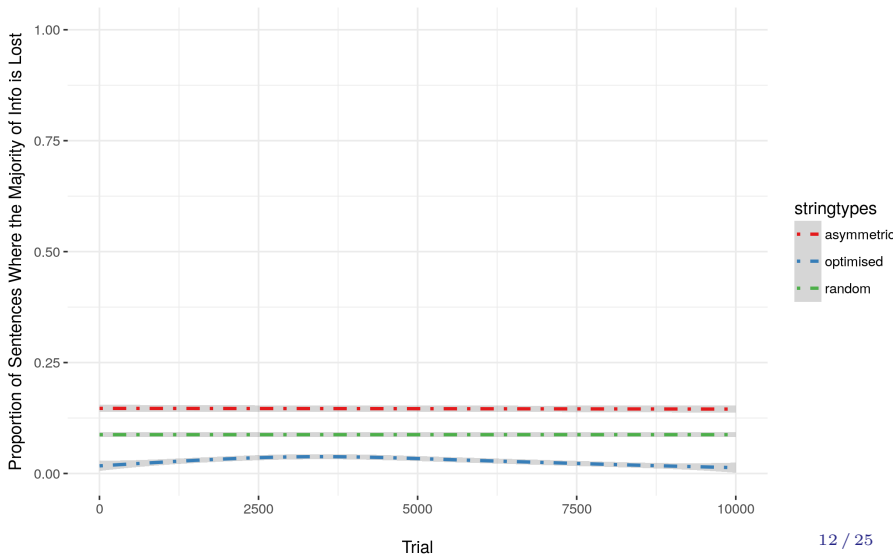
Noise resistance simulation

1. Generate 10-item sequences of probabilities.
2. Order them randomly, by size (“asymmetric”), or hyperdispersed (“optimised”).
3. A noise event randomly destroys a 3-item sub-sequence.
4. See how much information the noise events destroy (over many trials)!

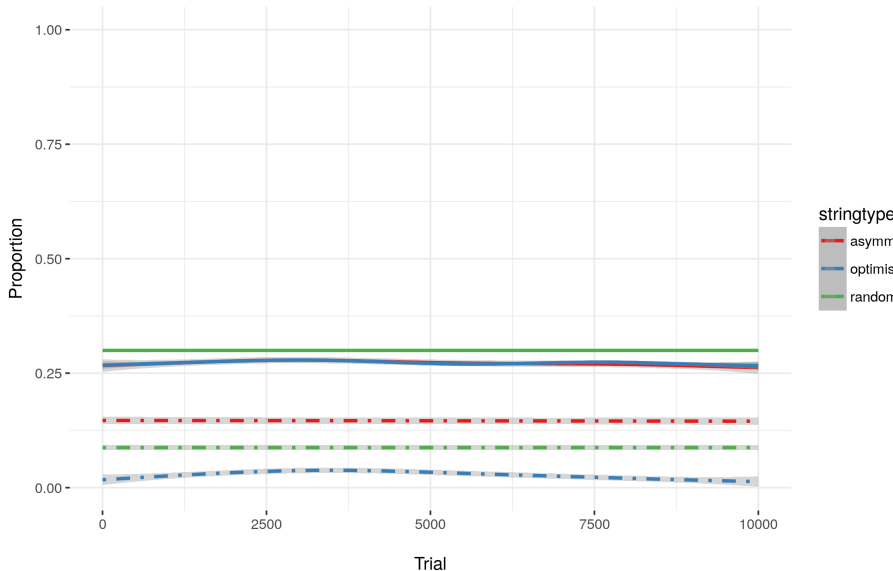
Noise Resistance Simulation: proportion of bits lost



Noise Resistance Sim: proportion of “sentences” where bits lost $> 50\%$ bits in sentence



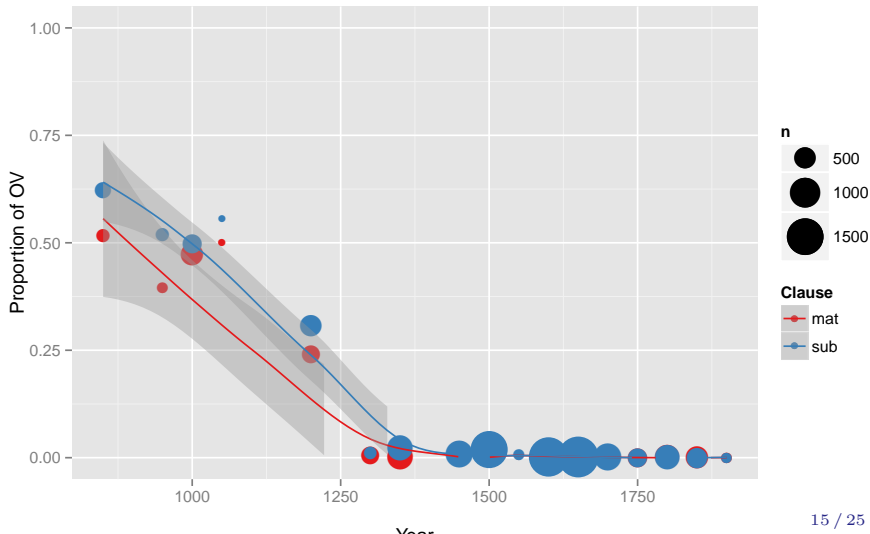
Noise Resistance Sim: bits lost and big bit losses



OV to VO Change in English

- (2) tu mihht ec **gastlike laf** Onn operr wise zarrkenn
 you might also spiritual loaf in another way prepare
 “In this way, you will let go of your sins.”
 (*Ormulum*, Lincoln, date: 1200)
- (3) Ne ma33 he nohht rihht cnawenn **me**
 NEG may he not right know me
 “He may not rightly know me”
 (*Ormulum*, Lincoln, date: 1200)

OV to VO Change in English



Some grammatical preliminaries...

- (4) [*Matrix* Malvina Reynolds implied [*Sub* that you should not like little boxes]*Sub*]*Mat*.
- (5) Malvina Reynolds implied that you should not like them. (**Pronoun Object**)
- (6) Malvina Reynolds implied that you should not like any/many/most/both/... (**Quantified Object**)

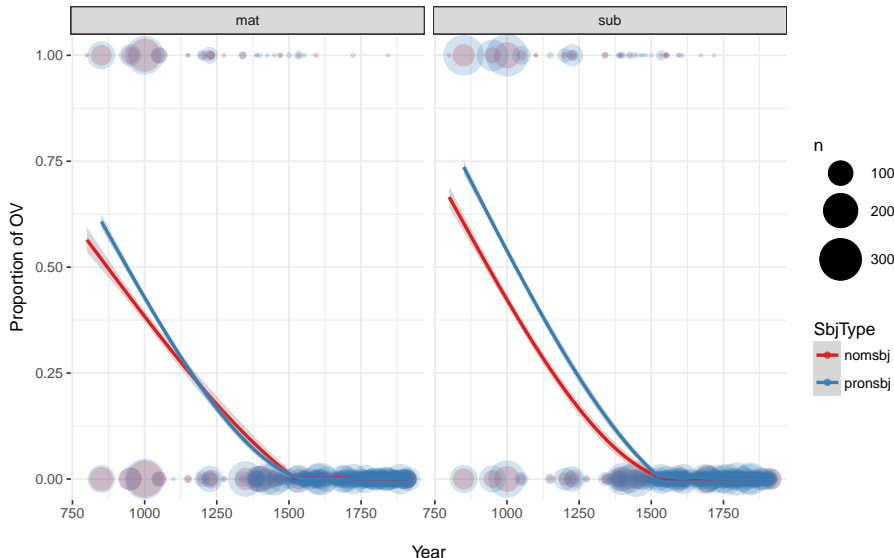
Guestimated information content hierarchy:

Other nominal Obj or Sbj >
Quantified Obj or Sbj , Verbs >
Pronoun Obj or Sbj

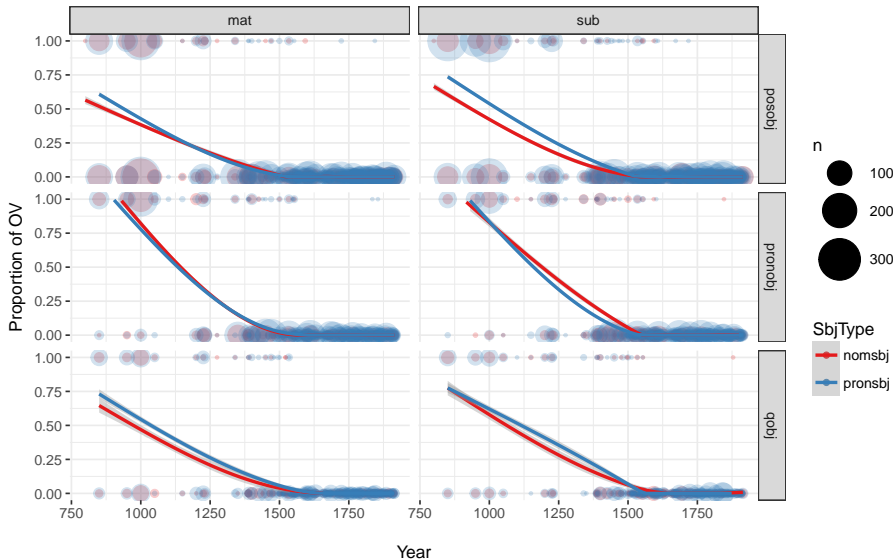
Hypotheses

1. Pronoun subjects should favor OV compared to nominal subjects if the object is nominal: pronSbj-nomObj-V is more uniform in info content than nomObj-nomObj-V, and the latter is less uniform than nomSbj-V-nomObj
2. Pronoun subjects should do the *opposite* if the Object is a pronoun: nomSbj-pronObj-V is more uniform than pronSbj-pronObj-V, and the latter is less uniform than pronSbj-V-pronObj.
3. Quantified Objs should be in-between.
4. All of these effects should be larger in sub clauses than mat clauses, because of the pressure of preceding information.

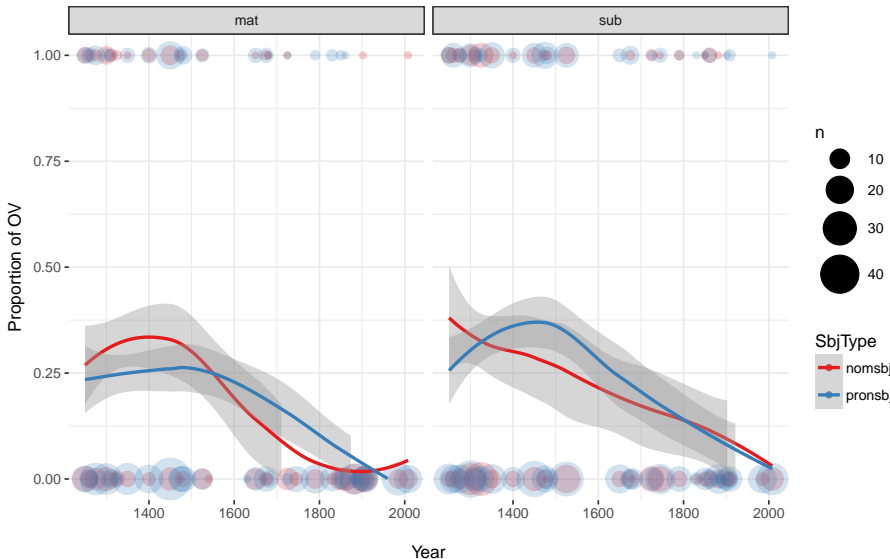
OV: Subj and Clause Type, English, N = 28,580



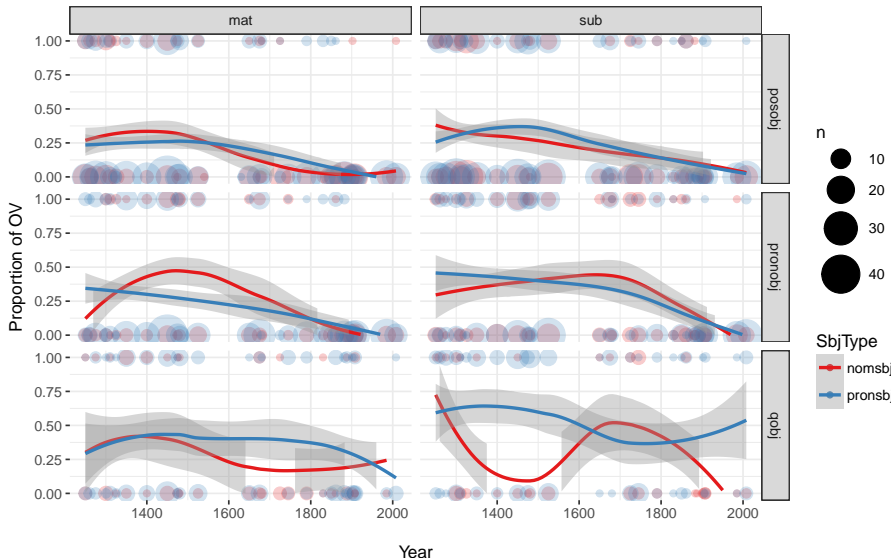
OV: Subj, Obj and Clause Type, English



OV: Subj and Clause Type, Icelandic, N = 1,860



OV: Subj, Obj and Clause Type, Icelandic



Conclusions

- Where speakers have an option, i.e. during language change, their choice responds to the resulting density of the entire utterance.
- Project with Christine Cuskley and Rachael Bailes to find out whether they also respond to hearer and other factors, and whether we can graduate to actual probabilities in the above work.
- MRes student Jack Winter (supervised by Rachael), doing an experiment about rhythm and memory, manipulating information density of symbols in a rhythmic sequence.
- ASD vs. typical speakers

Acknowledgements

Thanks to Rachael Bailes, Christine Cuskley, Tony Kroch,
among others.

<https://github.com/joelcw/constantentropy>



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.
- Jaeger, T Florian, and Roger P Levy. 2007. Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, 849–856.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2004. Penn-Helsinki Parsed Corpus of Early Modern English, release 3. Size 1.8 Million Words.
- Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English. Size ~ 950000 words.

References II

Levy, Roger. 2008. A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing*, 234–243. Association for Computational Linguistics.

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., 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.