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

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

Newcastle University
Centre for Behaviour & Evolution

14th May, 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
   Synctactic change in English
   Synctactic 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.

• Shannon (1948)’s formula for information in an event with $n$ discrete outcomes with probabilities $p_1...p_n$:

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

• 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

![Graph showing proportion of bits lost over trials](image)
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 oberr wise ʒarrkenn 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 maʒʒ 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

The diagram illustrates the proportion ofOV clauses from 1000 to 1750, indicating a decrease over time. The year is plotted on the x-axis, and the proportion of OV clauses is plotted on the y-axis. The data points are color-coded to indicate different clauses: red for mat and blue for sub. The trend shows a significant decrease in the proportion of OV clauses as the year progresses.
Some grammatical preliminaries...

(4) \([M_{atrix} \text{ Malvina Reynolds implied } [S_{ub} \text{ that you should not like little boxes }]_{Sub} ]_{Mat}\).

(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 favour OV more than nominal subjects will, if the object is also nominal: pronSbj-nomObj-V is more uniform in info content than pronSbj-V-nomObj or nomSbj-nomObj-V, and the latter are less uniform than nomSbj-V-nomObj.

2. Pronoun subjects should do the opposite, i.e. favour more VO, if the Object is also a pronoun: pronSbj-V-pronObj is more uniform than pronSbj-pronObj-V or nomSbj-V-pronObj, and the latter are less uniform than nomSbj-pronObj-V.

3. Quantified Objs should be in-between for all effects.

4. All of these effects should be larger in sub clauses than mat clauses, because of the pressure that preceding information (in the earlier clause) creates.
OV: Subj and Clause Type, English, N = 28,580

The diagram shows the proportion of OV (Object-Verb) constructions over time for different subject types (nominal and pronominal) in English data from 750 to 1750. The x-axis represents the years, and the y-axis shows the proportion of OV constructions. The data points are color-coded and sized to indicate frequency, with larger and darker points representing higher frequencies.
OV: Subj, Obj and Clause Type, English

The diagram shows the proportion of OV (Object Verb) constructions over time, with different types of subjects and objects. The x-axis represents the year, ranging from 750 to 1750, and the y-axis represents the proportion of OV constructions. The data is presented for two categories: mat and sub, which likely refer to different subtypes or categories of clauses.

The SbjType legend indicates two types of subjects: nomsbj (nominal subject) and pronsbj (pronominal subject). The colors and markers used in the graph help to distinguish between these types over the years.

The graph provides a visual representation of how the proportion of OV constructions changes with respect to time and subject type, offering insights into language change over historical periods.
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


References II

