Production and Communicative Efficiency

• Is production affected by considerations about efficient communication?
Noisy channel theorem

• Recall: Shannon’s channel coding theorem places limits (upper and lower) on transmission rates through noise

→ Ideal coder should aim for transmission rates at channel capacity $C$, but not higher
Uniform Information Density

*Given a choice*, speakers prefer to keep the amount of information transmitted per unit signal uniform.


⇒ Speakers should spread more information over more signal (where there is a choice)
Boundedly rational Bayesian inference over a noisy channel
Information Density & Auxiliary Contraction


Brain and Cognitive Sciences, University of Rochester
Efficient Morpho-Syntactic Production

Pres. Clinton did *n’t/not* have ...
Estimating the information carried by a contractible element

Information theoretic definition of Shannon information content

\[
\begin{align*}
I(\text{NOT} \mid \text{context}) &= -\log p(\text{NOT} \mid \text{context}) \\
&= -\log p(\text{NOT} \mid "Clinton did ") \\
&= -\log \left[ p("not" \mid "Clinton did ") + p("n't" \mid "Clinton did ") \right]
\end{align*}
\]

Use trigram model to estimate probability (backoff)
Efficient Morpho-Syntactic Production

Pres. Clinton did *n’t/not* have ...
Data

• Extracted from a corpus of spontaneous American English speech (Switchboard, 800k sentences in 650 dialogues)
  – Only cases that are contractible in American English are included (e.g. not “I have/*’ve a car”.).

  – HAVE: e.g. ’d vs. had (>2,400 contractible cases)
  – NOT: n’t vs. not (> 5,000 contractible cases)
  – BE: e.g. ’s vs. is (> 9,000 contractible cases)
Analysis

• Mixed logit model to analyze when speakers’ choose **full** over **contracted** forms depending on the information carried by it.

\[
\text{logit}[p(\text{full})] = \ln \frac{p(\text{full})}{p(\text{reduced})} = -\beta \log p(\text{NOT} | w_{i-1}) - \beta \log p(\text{NOT} | w_{i+1}) + X_{\text{Controls}} \beta_{\text{Controls}} + Zb
\]

• Simultaneously controlling for:
  – Position in intonational phrase
  – Complexity of *upcoming* material
  – Complexity of *host* (e.g. pronominality, number of words)
  – Speech rate and fluency (e.g. presence of filled pauses)
  – Social effects (gender, education)
  – Random effects for individual differences
Replicated for

\{\text{WAS, WERE, AM, ARE, IS, WILL}\}  \quad \{\text{HAD, HAS, HAVE}\}

\begin{align*}
\text{Counts} & : 537, 504, 470, 436, 403, 370, 336, 302, 269, 235, 202, 168, 135, 102, 68, 34, 1 \\
\text{Counts} & : 169, 158, 148, 138, 127, 116, 106, 96, 85, 74, 64, 54, 43, 32, 22, 12, 1
\end{align*}

Information content of BE: $I(\text{BE} | \text{preceding context})$

Information content of HAVE: $I(\text{HAVE} | \text{preceding context})$
Summary

• Cases that would otherwise be more information dense are less likely to be contracted.

→ Predicted by Uniform Information Density ...

and other accounts that predict a trade-off between the amount of signal provided and the redundancy of the linguistic unit (e.g. negation) in its context.
‘Choices’ at many levels in production

<table>
<thead>
<tr>
<th>Level</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utterance level:</td>
<td>Move the triangle to the left.</td>
</tr>
<tr>
<td></td>
<td>Select the triangle. Move it to the left.</td>
</tr>
<tr>
<td>Phrasal level:</td>
<td>She already ate (dinner)</td>
</tr>
<tr>
<td></td>
<td>She stabbed him (with a knife).</td>
</tr>
<tr>
<td>Word level:</td>
<td>I read a book (that) she wrote.</td>
</tr>
<tr>
<td>Morphological level:</td>
<td>I’ve have gone there.</td>
</tr>
<tr>
<td>Phonological level:</td>
<td>t/d-deletion; final cluster reduction; vowel weakening</td>
</tr>
<tr>
<td>Phonetic level:</td>
<td>formant energies, F1/F2 ratio, speech rate</td>
</tr>
</tbody>
</table>
Information = -\log p(CC | ctxt) + -\log p(\text{onset} | CC, ctxt)
Results
Information at RC onset = \(\log \frac{1}{p(\text{RC} | \text{ctxt})} + \log \frac{1}{p(\text{they} | \text{RC, ctxt})}\)
Results
The style of life [(that was) chosen by the beatnik generation] is designed to enhance sexual experience.
High Information Environments (Producing Dispreferred Structure)

Information per word throughout discourse in Mandarin Chinese

Information Density & Inter-clausal Planning

Information Density & Resumptive Morphology in Yucatec

Robustly Communicating Core-Meaning outranks Precision?

Degen** and Jaeger* (2011-LSA)

* Linguistics, University of Rochester
* Brain and Cognitive Sciences, University of Rochester
Question

• Are similar biases operating during language acquisition that then induce changes in representations/distributions over representations compared to the input language?

• Such biases would also provide an explanation as to how the observed correlations between information density and speakers’ preferences in production arise (via lexicalized grammaticalization)
Information – signal trade-off in acquisition

Fedzechkina, Jaeger & Newport (2011, in prep)

Brain and Cognitive Sciences,
University of Rochester

The following slides are omitted since they contain unpublished materials. Thank you for understanding.
Conclusion

• In **incremental production**, we observe a bias to provide more linguistic signal where information density would otherwise be high, thereby lowering information density.

• During the **acquisition** we find the same/a similar bias at work: Learners condition the use of linguistic form in such a way that the unexpected is more likely to be marked by more linguistic signal.
On language ‘utility’: processing complexity and communicative efficiency
T. Florian Jaeger¹ and Harry Tily¹


Redundancy and reduction: Speakers manage syntactic information density
T. Florian Jaeger
Thank you!

Many thanks to our collaborators: Masha Fedzechkina, Austin Frank, Lissa Newport and Nathanial Smith

Funding: NSF 0953870, NIH 1R01HD065829, ARL Cognition & Neuroergonomics Collaborative Technology Alliance to RL; NSF Grant BCS-0844472; Provost’s Award, Wilmot Award, Alfred P. Sloan Fellowship to TFJ