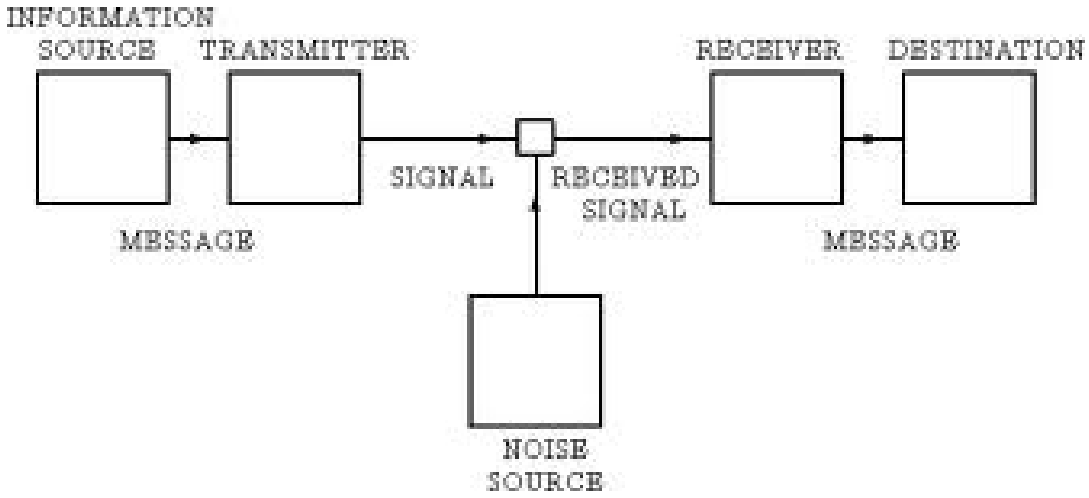
Production and Communicative Efficiency

- 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

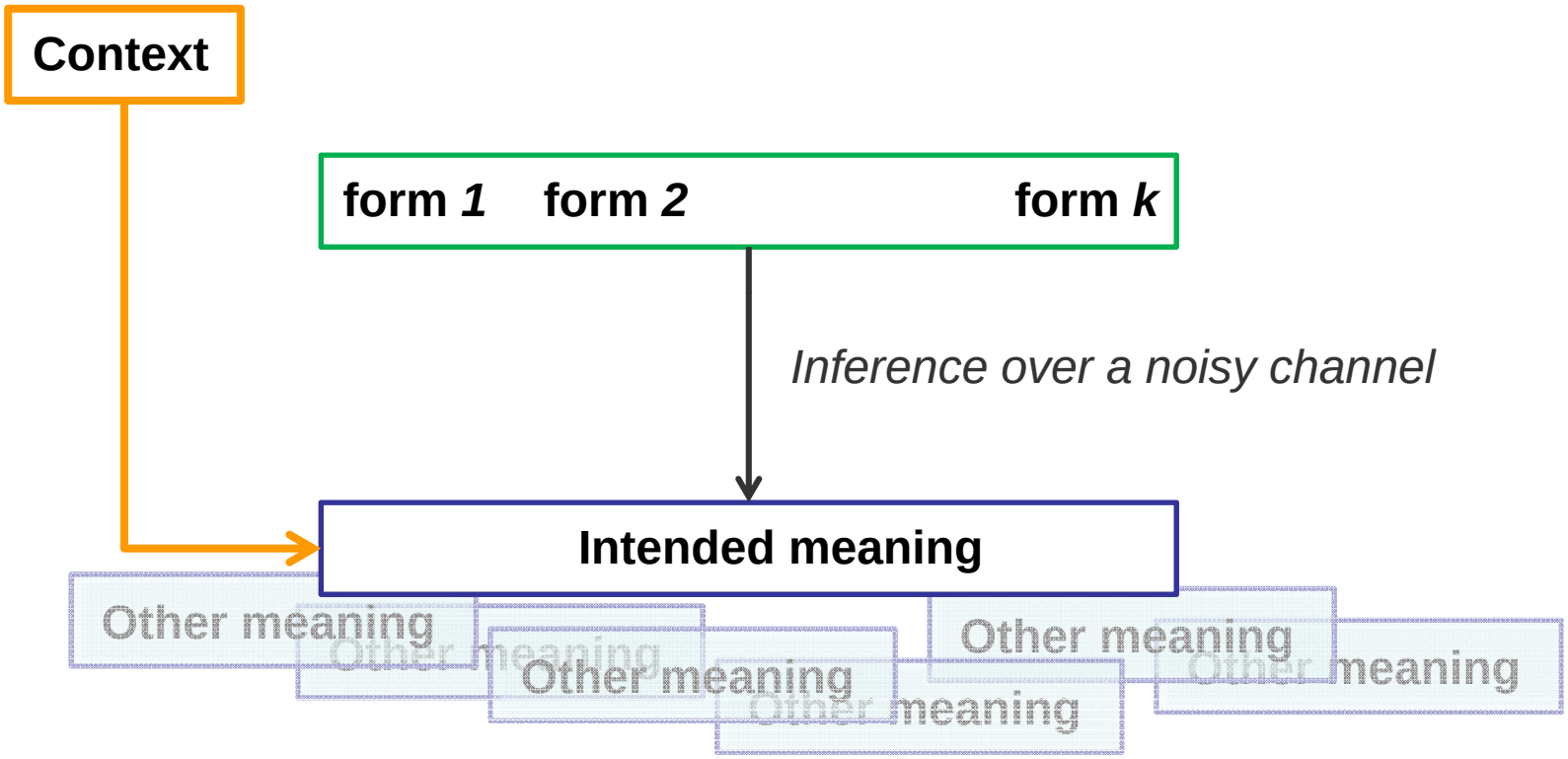
Uniform Information Density

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

Jaeger (2006, in prep), Levy & Jaeger (2007), based on Genzel & Charniak (2002) and Aylett & Turk (2004)

→ Speakers should spread more information over more signal (where there is a choice)

Boundedly rational Bayesian inference over a noisy channel



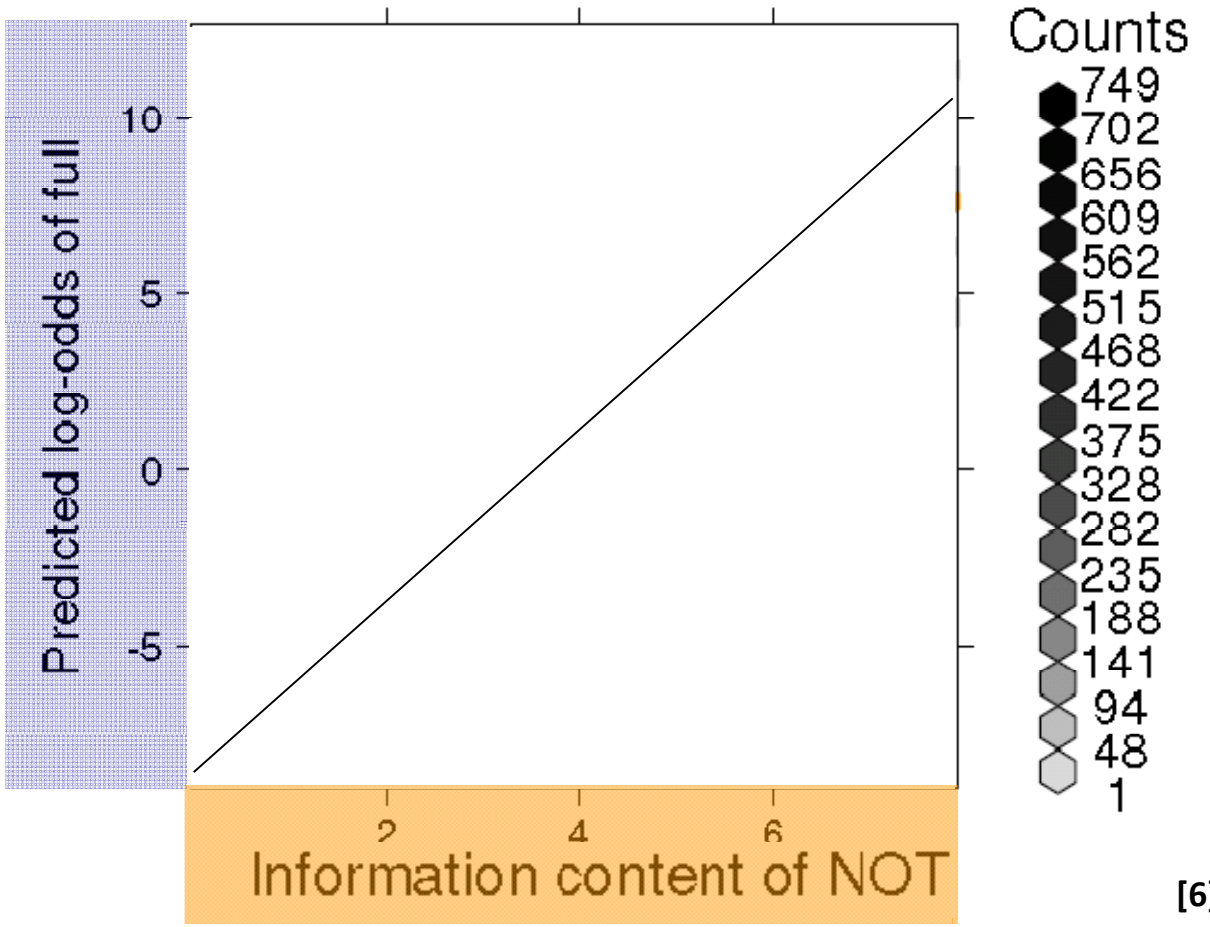
Information Density & Auxiliary Contraction

Frank & Jaeger (2007-AMLaP; 2008-CUNY; 2008-CogSci; in prep)

Brain and Cognitive Sciences, University of Rochester

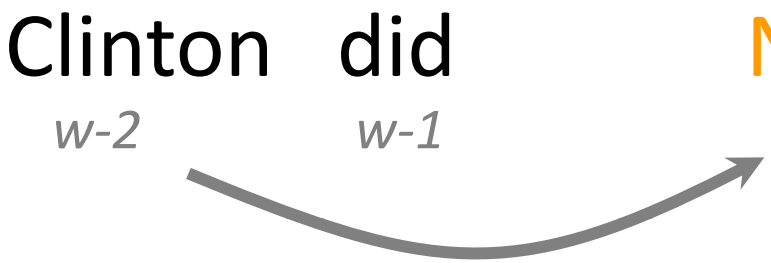
Efficient Morpho-Syntactic Production

Pres. Clinton did ***n't/not*** *have …*



[6]

Estimating the information carried by a contractible element



I(NOT | context)

Information theoretic definition of Shannon information content

=

-log p(NOT | context)

Use trigram model to estimate probability (backoff)

≈

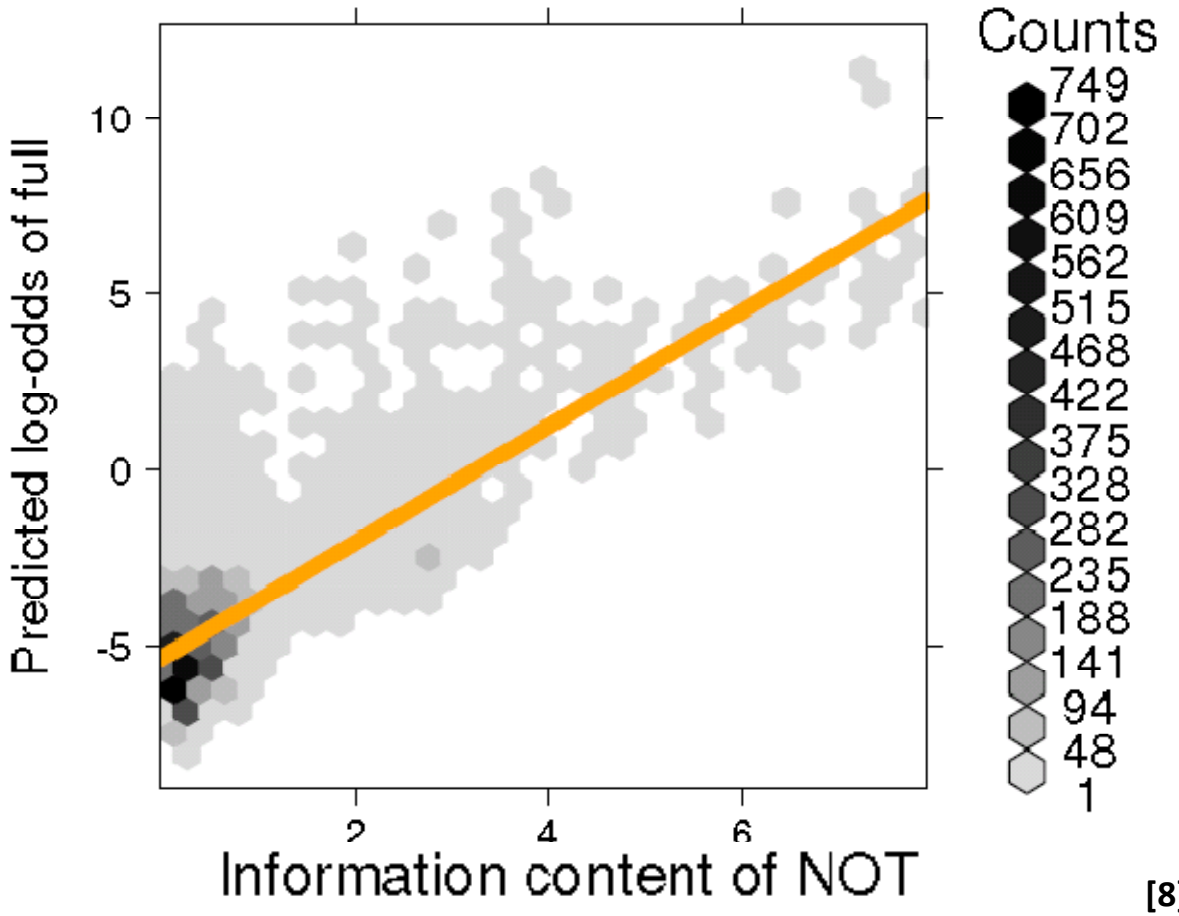
-log p(NOT | “*Clinton did* ”)

=

-log [p(“*not*” | “*Clinton did* ”) + p(“*n’t*” | “*Clinton did* ”)]

Efficient Morpho-Syntactic Production

Pres. Clinton did ***n't/not*** *have …*



[8]

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

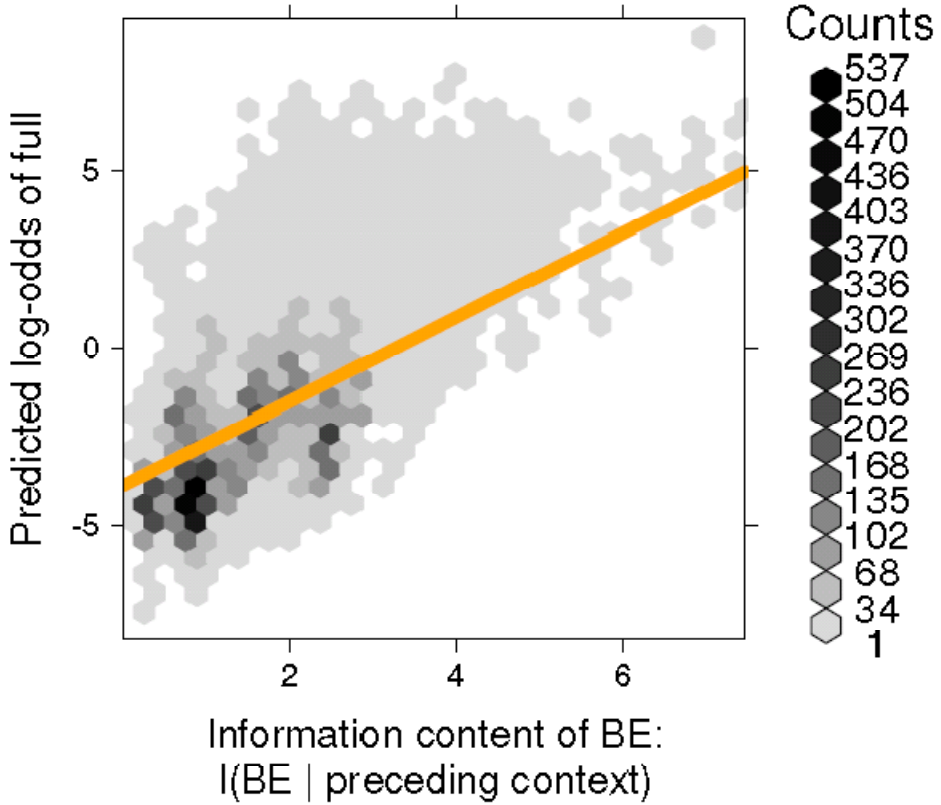
- Mixed logit model to analyze when speakers' choose **full over contracted** forms depending on the information carried by it.

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

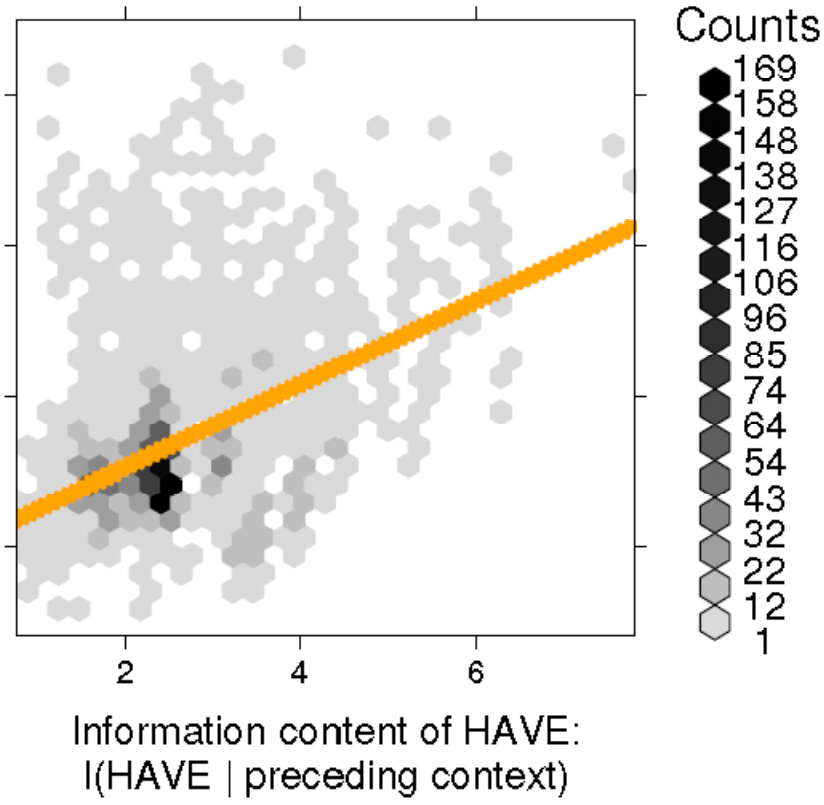
- Simultaneously controlling for:
  - Position in intonational phrase
  - Complexity of *upcoming* material
  - Complexity of *host* (e.g. pronominality, number of words)
  - Speech rate and fluency (e..g presence of filled pauses)
  - Social effects (gender, education)
  - Random effects for individual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

Replicated for

{WAS,WERE, AM, ARE, IS, WILL}



{HAD, HAS, HAVE}



Summary

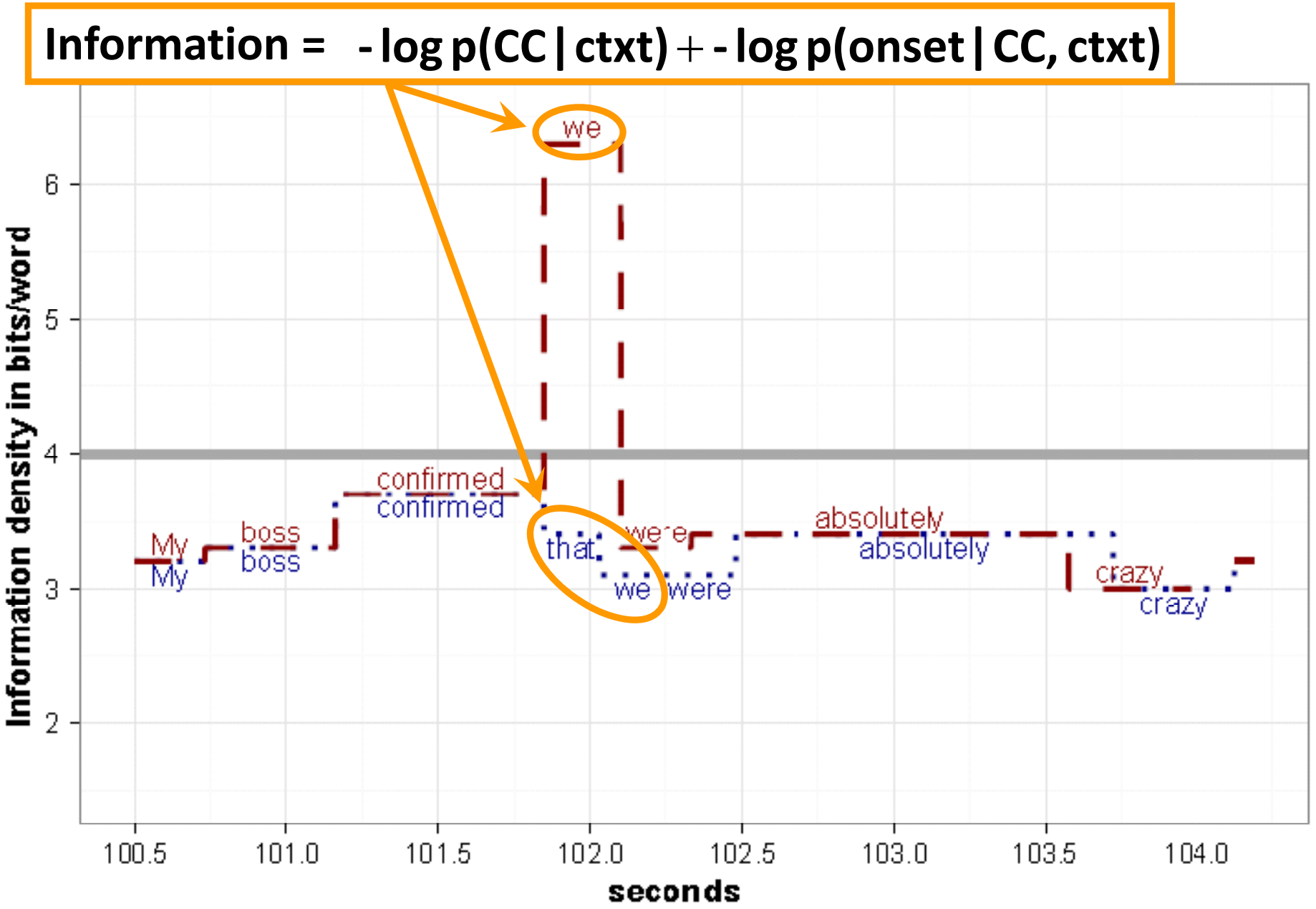
- Cases that would otherwise be more information dense are less likely to be contracted.

→ Predicted by Uniform Information Density …

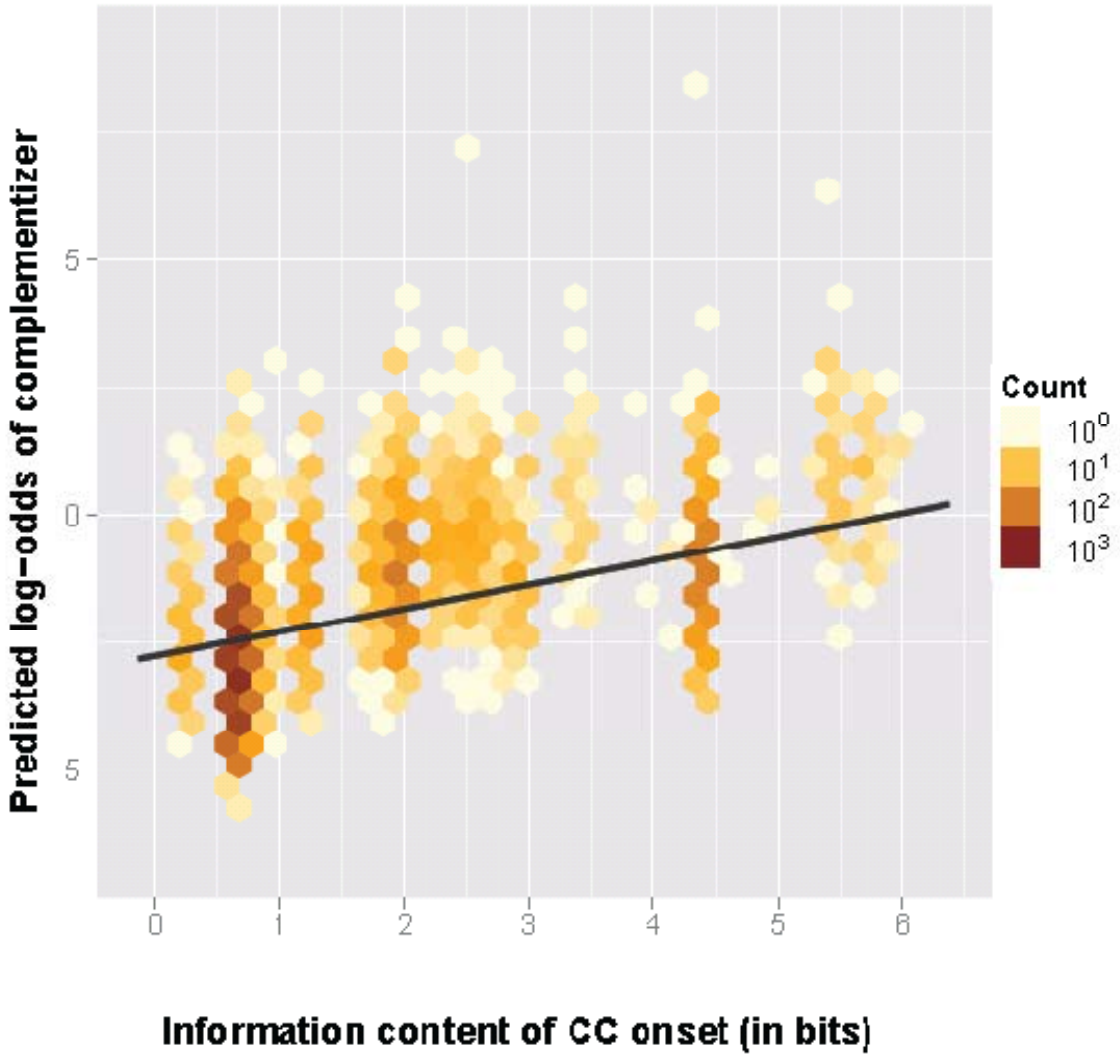
and other accounts that predict a trade-off between the amount of signal provided and the redundancy of the linguistic unit (e.g. negation) in its context.

‘Choices’ at many levels in production

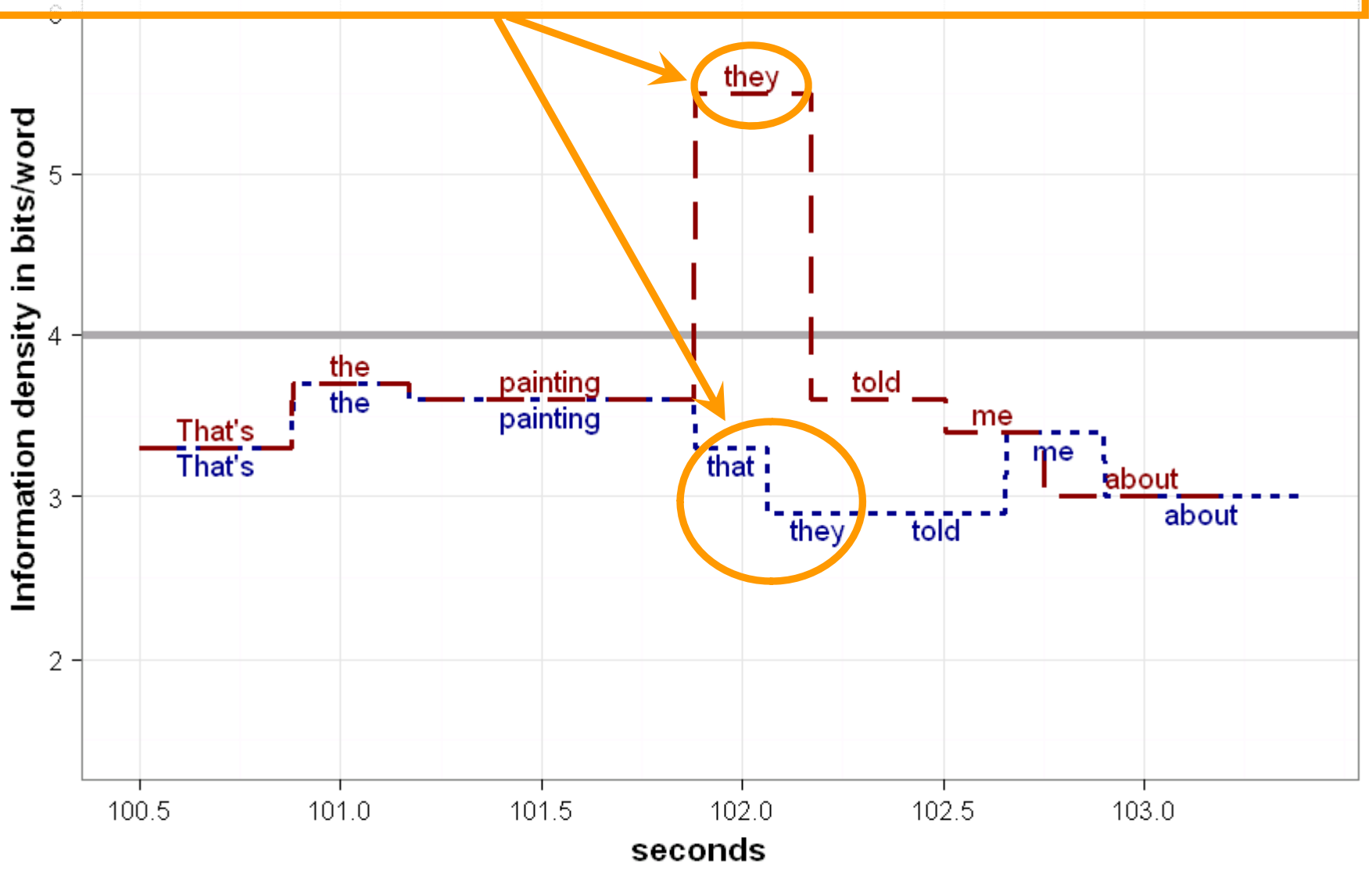
Utterance level:	*Move the triangle to the left.*
	Select the triangle. Move it to the left.
Phrasal level:	*She already ate (dinner)*
	She stabbed him (with a knife).
Word level:	*I read a book (that) she wrote.*
Morphological level:	*I’ve\have gone there.*
Phonological level:	*t/d*-deletion; final cluster reduction; vowel weakening
Phonetic level:	formant energies, F1/F2 ratio, speech rate


Information = - log p(CC | ctxt) + - log p(onset | CC, ctxt)
Information density in bits/word
6
5
4
3
2
100.5
101.0
101.5
102.0
102.5
103.0
103.5
104.0
seconds
we
confirmed
confirmed
My
My
boss
boss
that
were
we
absolutely
absolutely
crazy
crazy

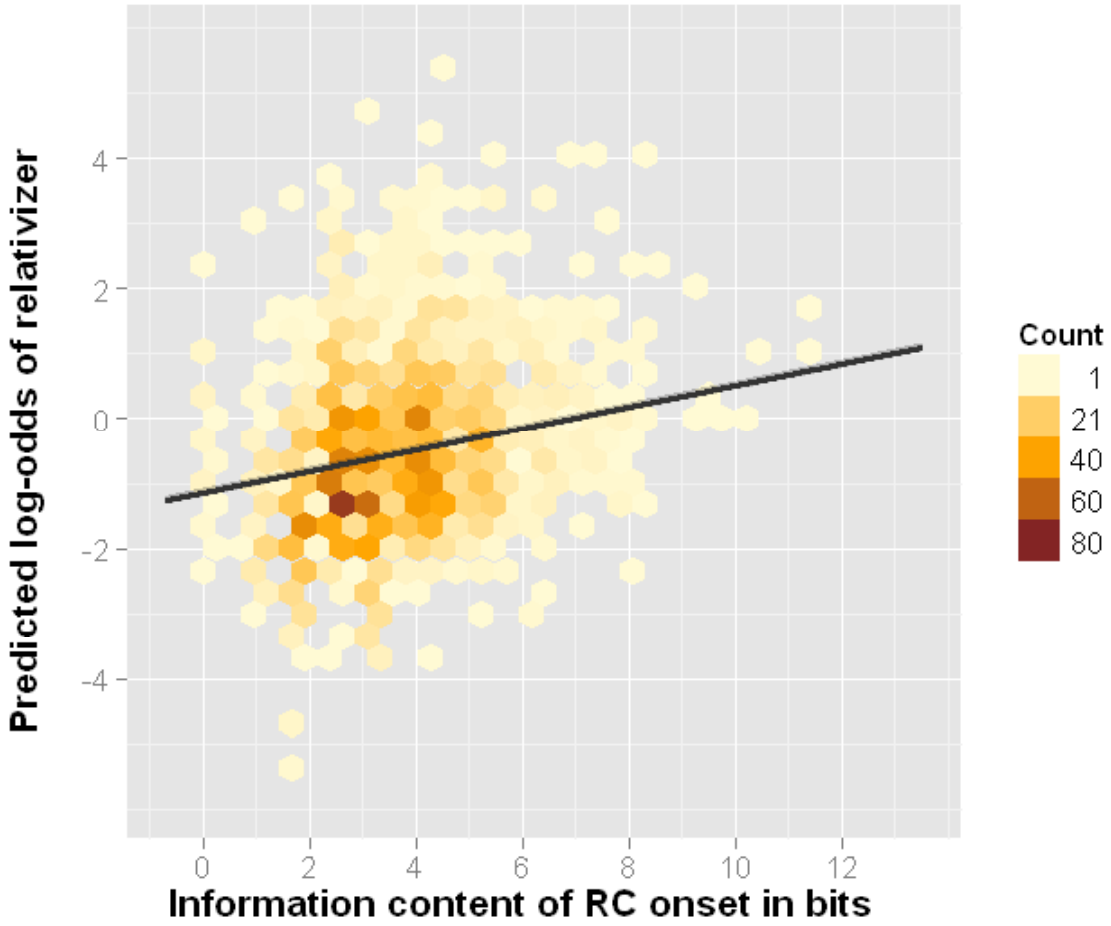
Results


Predicted log−odds of complementizer
5
0
5
0
1
2
3
4
5
6
Information content of CC onset (in bits)
Count
10^0
10^1
10^2
10^3

$$\text{Information at RC onset} = \log\frac{1}{p(\text{RC}\,|\,\text{ctxt})} + \log\frac{1}{p(\textit{they}\,|\,\text{RC, ctxt})}$$

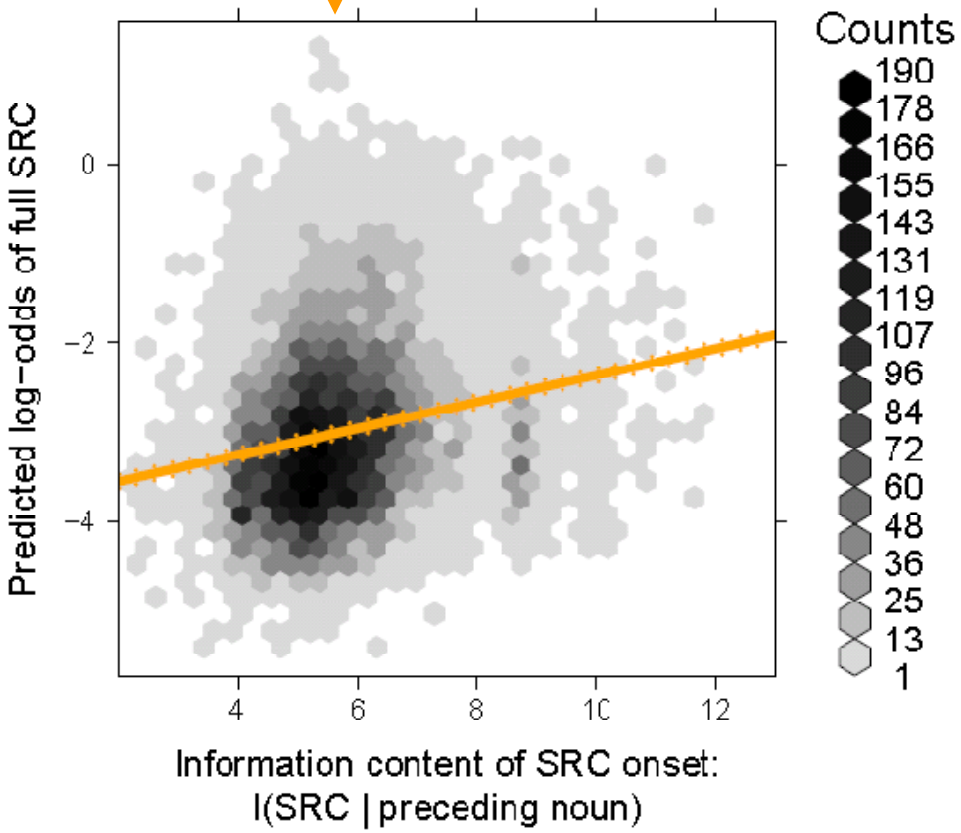


Results


Predicted log-odds of relativizer
4
2
0
-2
-4
0
2
4
6
8
10
12
Information content of RC onset in bits
Count
1
21
40
60
80

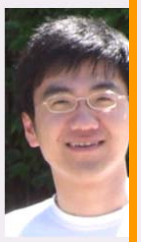
SRC reduction

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



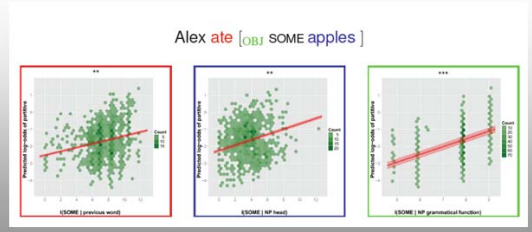
Information Density & Inter-clausal Planning

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



[65]

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 the 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.

wires.wiley.com/cogsci

Focus Article

On language 'utility': processing complexity and communicative efficiency

T. Florian Jaeger[1]* and Harry Tily[2]

Volume 2, May/June 2011

Cognitive Psychology 61 (2010) 23–62

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

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