ACT-R&BIG DATA SYMPOSIUM

CRACKING LANGUAGE WITH ACT-R:

HOW DOES LANGUAGE MIX WITH COGNITION? CHALLENGES FOR THE MODELING COMMUNITY

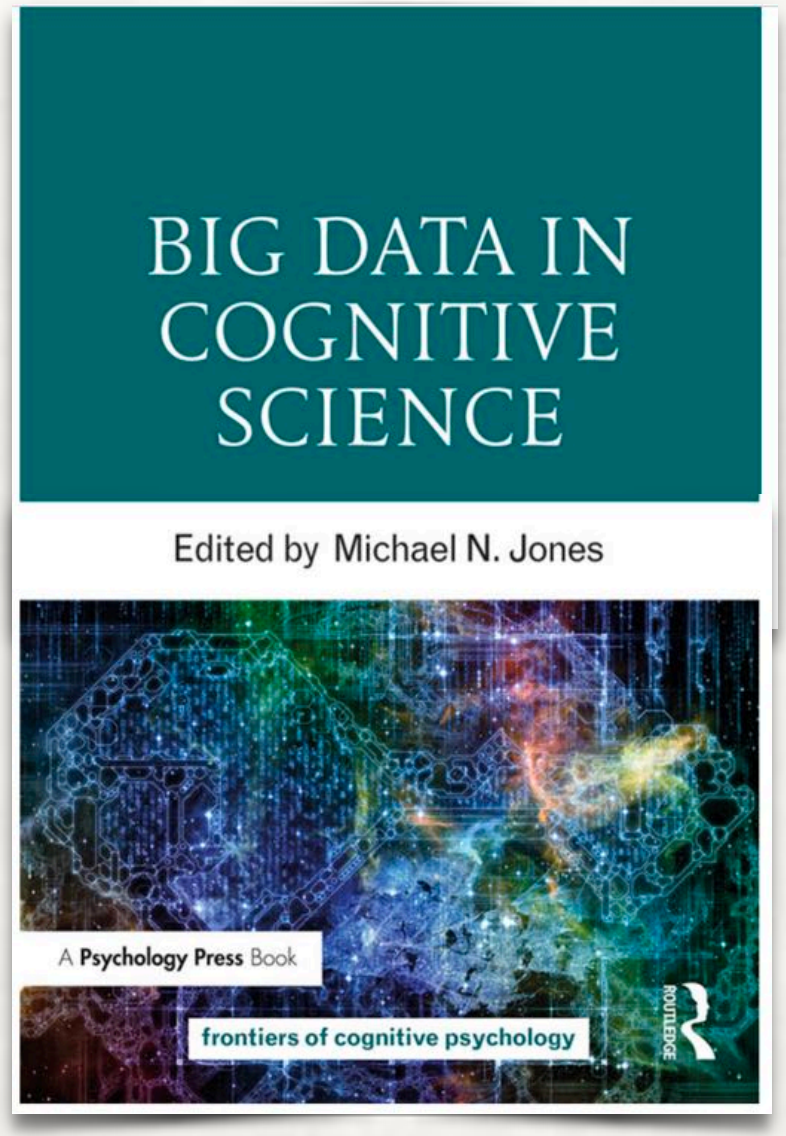
DAVID REITTER, JEREMY COLE,
MATTHEW A. KELLY, YANG XU
PENN STATE

COGNITIVE SCIENCE

COMPUTER SCIENCE

BIG DATA IN COGNITIVE SCIENCE

MICHAEL N. JONES (ED.), ROUTLEDGE, 2017



SOME BIG QUESTIONS

THAT MODELERS MIGHT ADDRESS WITH BIG DATA

- How does cognition give rise to language learning and language processing?
  - What are general, and which are language-specific operations and representations necessary for human language processing?
  - Can we learn, through language, about memory?
- Can language datasets inform our model of (declarative) memory?
  - Decay, spreading activation, and language change
  - Patterns of similarity through distributed representations

DETECTING MEMORY EFFECTS IN LANGUAGE

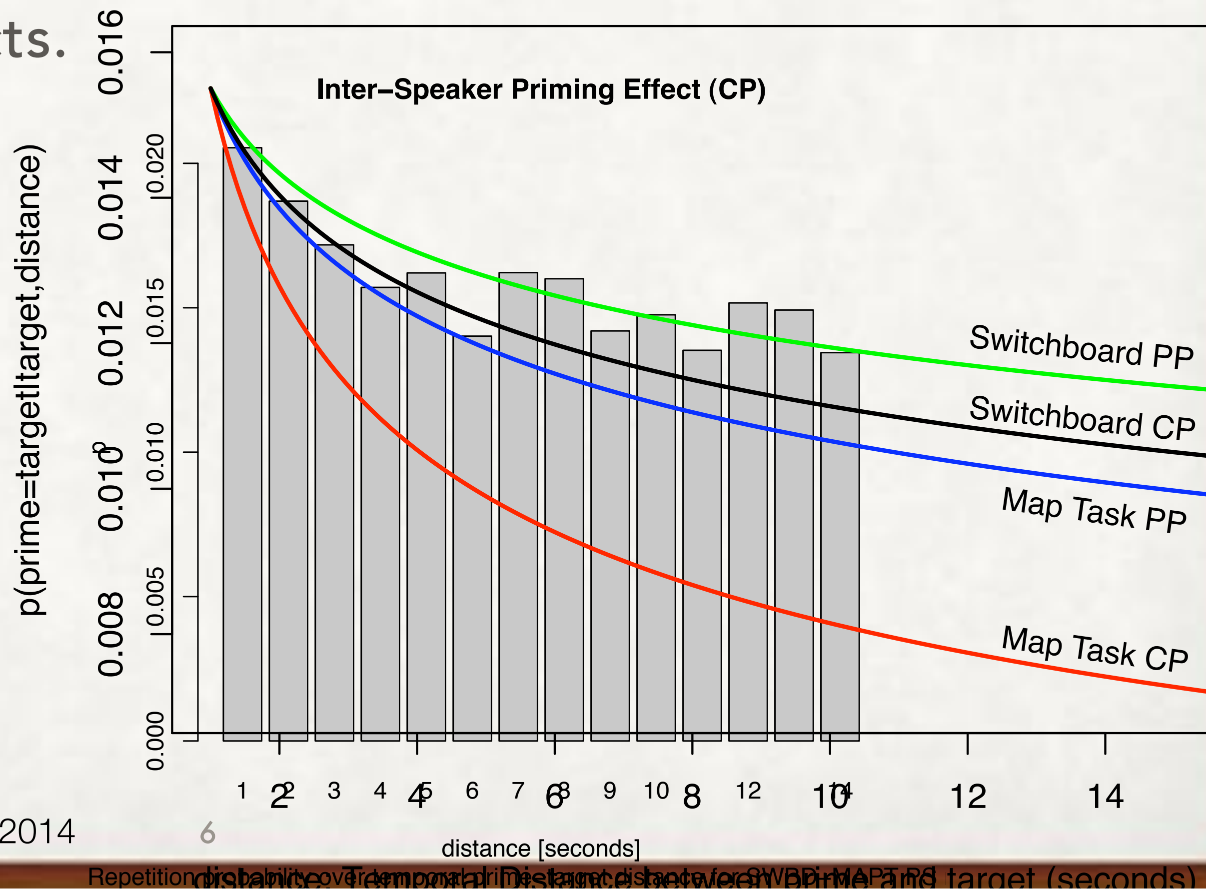
- Sentence structure is influenced by recent context:
  Structural priming causes syntactic choices to be repeated.
  (Bock 1986, Pickering&Branigan 1998, etc.)
- Syntactic repetition is associated with increased task success
  (where the task relies on mutual understanding).
  (Reitter&Moore 2014)
  - This can be seen in small corpora (few MB)

STRUCTURAL PRIMING (SHORT-TERM)

Stronger decay = more priming
in task-oriented dialogue
compared to spontaneous conversation
Compatible with Interactive Alignment Theory

Control for lexical repetition, disfluencies, frequency effects. By-turn analysis.

Effect of Prime-Target Distance

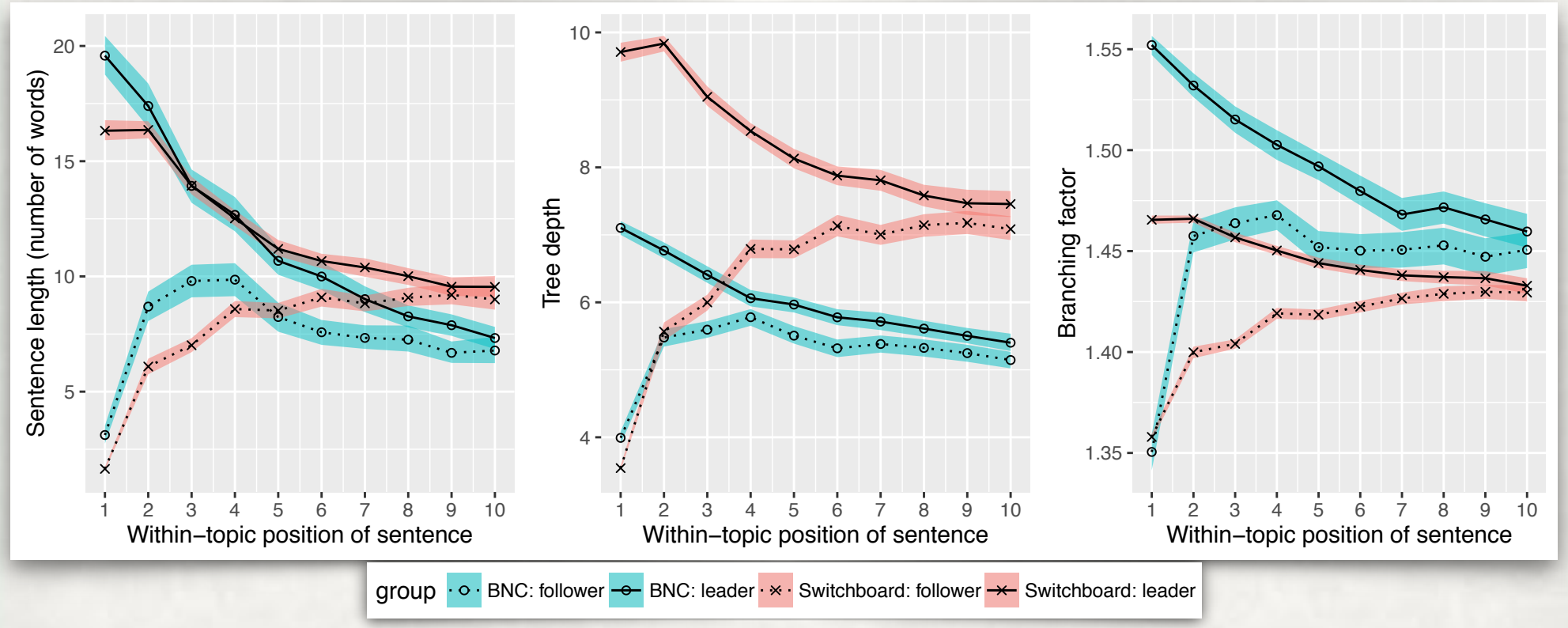


Reiter & Moore, Journal of Memory and Language, 2014

SYNTACTIC COMPLEXITY CONVERGES

XU & REITTER, ACL, 2016

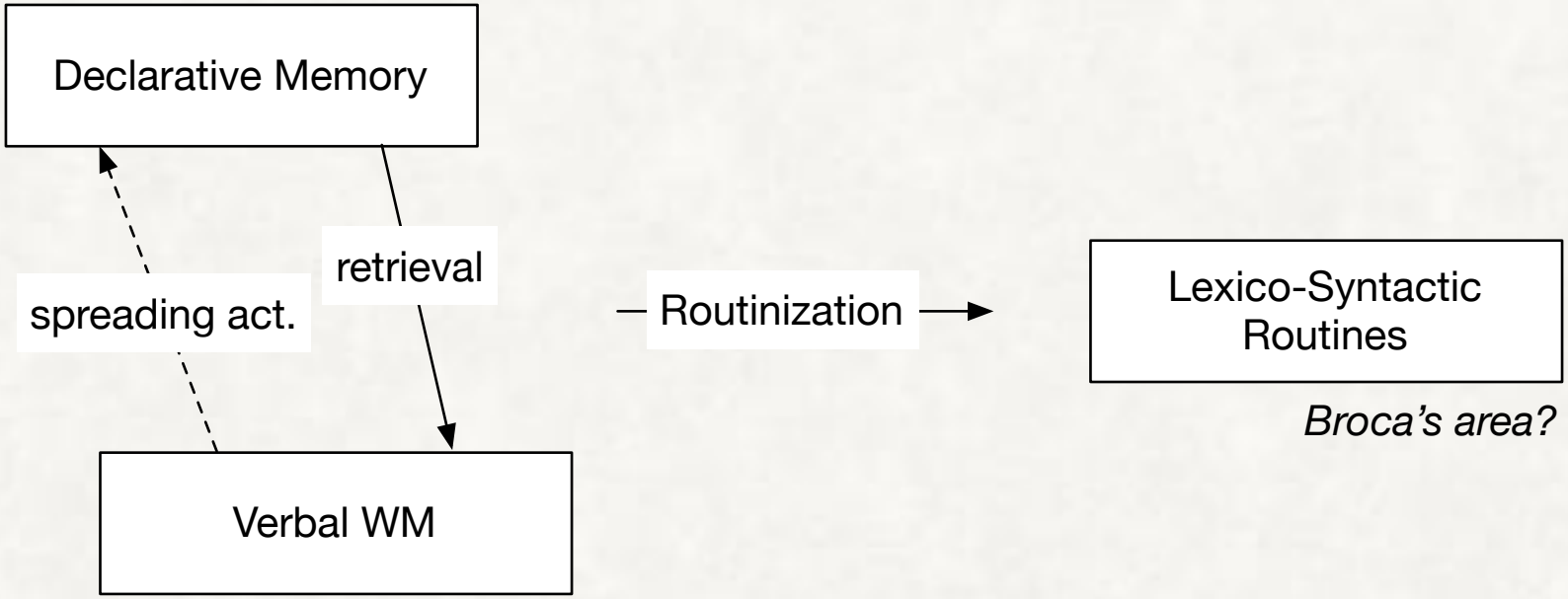
- Adaptation or accommodation appears to be strategic, at certain levels: syntactic complexity converges towards that of one's conversation partner (X&R 2016)
  - Seen in large corpora of written dialogue (100s MB)



CHALLENGE: HOW DOES LANGUAGE EXISTS IN THE MIND?

GENERAL COGNITION AND PSYCHOLINGUISTICS

Implicit Processes in Language Production and Comprehension



Reitter, Keller & Moore 2011 (ACT-R)

Cole & Reitter, ICCM 2017

Kelly, Reitter, & West, ICCM 2017

Chang, Dell, Bock 2006 (Connectionist)

ACT-R MODEL OF LANGUAGE PRODUCTION

COLE&REITTER, ICCM 2017

Automatic generation of semantic representation from corpus sentences.

Learning of syntactic rules (DM, productions) necessary to process these sentences.

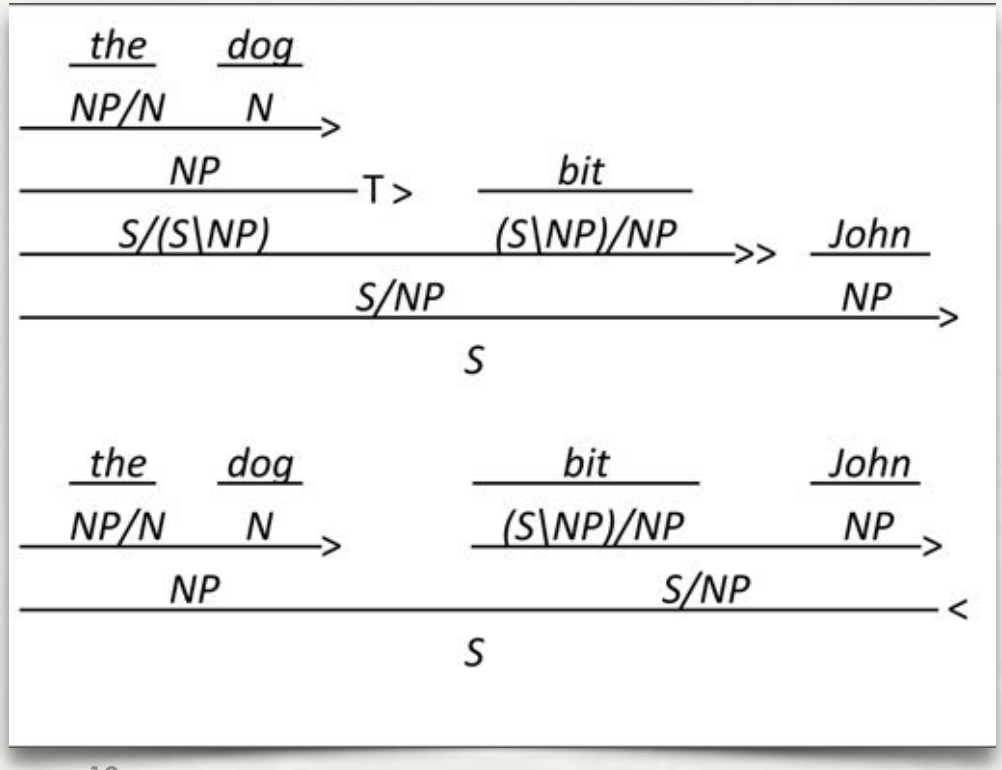
How well can a model recover the corpus sentences?

BIG-DATA ACT-R MODEL OF LANGUAGE PRODUCTION

Grammatical Encoding: words are combined into sentences. Lexical Retrieval: Words are retrieved (and thus, enter working memory). How do constraints on working memory affect this process?

incremental model:

Non-incremental model:



*less effective than incremental ****

*also: more WM <-> more effective ****

WHAT OUR ACT-R MODEL CAN DO (TODAY)

1000 sentences

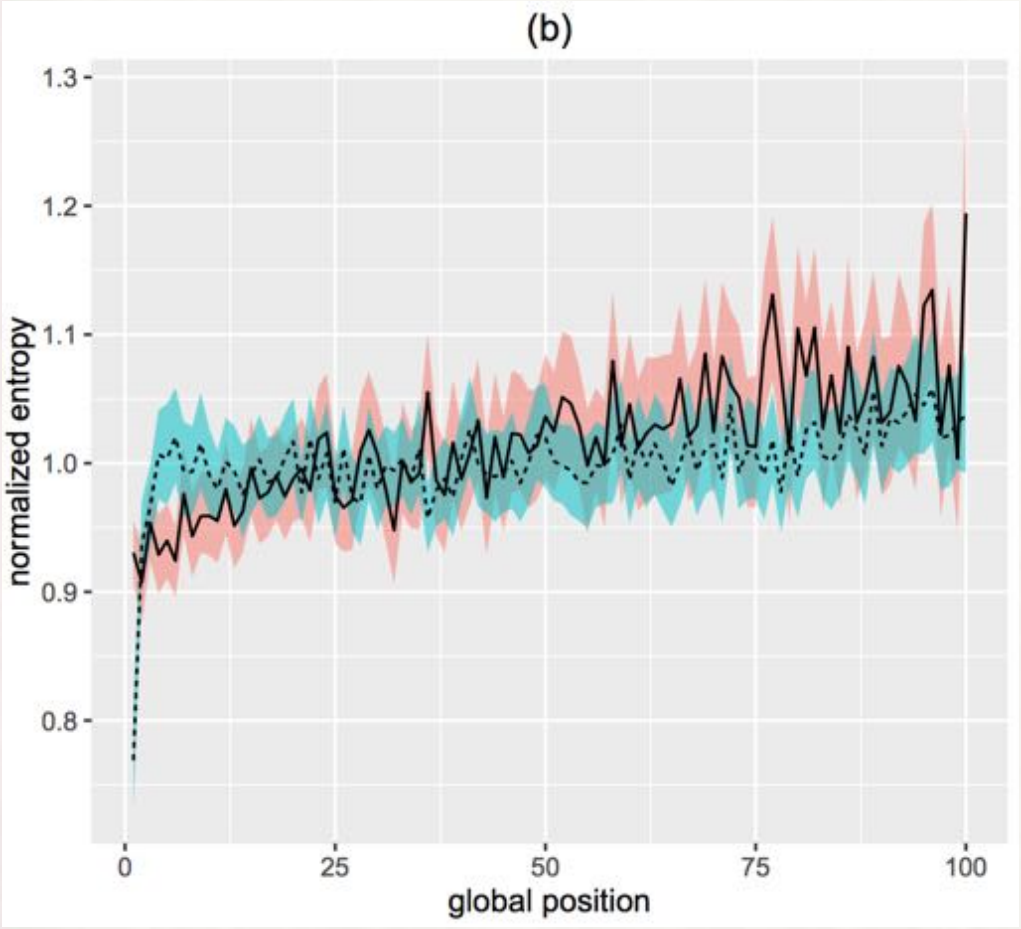
Contrast different model variants (high/low WM, decay, incrementality)

language processing vs. general cognition

large-scale routinization of syntactic/semantic knowledge possible (in the future)

AN INFORMATION-THEORETIC ACCOUNT OF DIALOGUE

Yang Xu
ACL 2016
Xu&Reitter, Cognition (to appear)



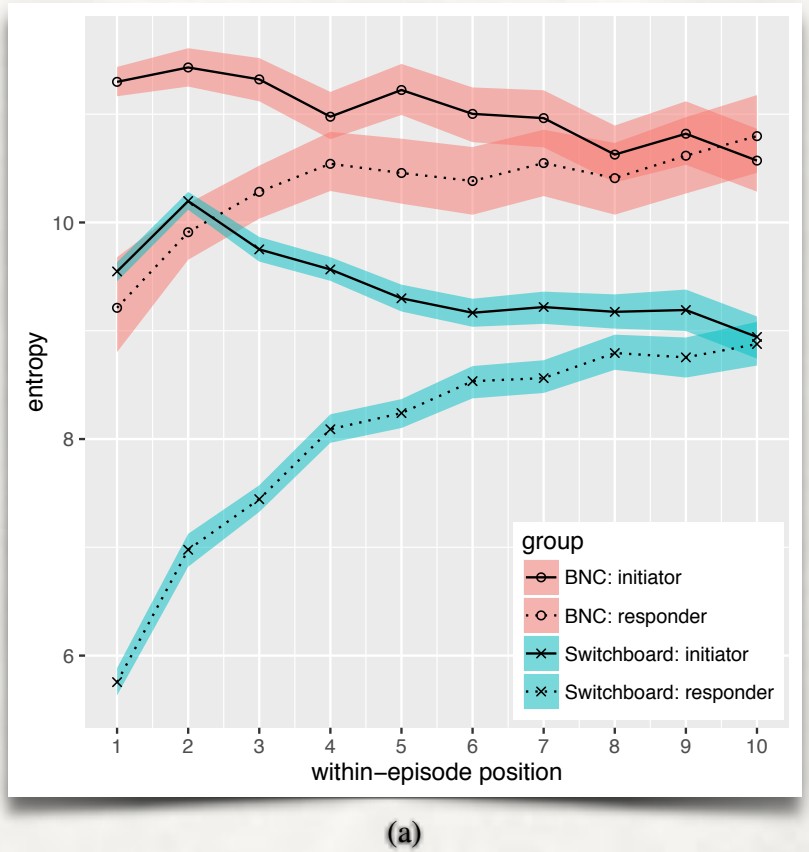
$$H(w_1 \dots w_n) = -\frac{1}{n} \sum_{w_i \in W} \log P(w_i | w_1 \dots w_{i-1})$$

- Entropy approximates information content: the more predictable language is, the more lower its entropy
- Information increases within documents, paragraphs (Genzel&Charniak, 2002; Keller, 2004)
- Speakers may strive to compensate for high-information content in one part of a sentence (o turn) by reducing it elsewhere ("Uniform Information Density", Levy&Jaeger, 2007): The average entropy (entropy rate) tends to be constant.

CHALLENGE: INFORMATION DENSITY

WHY AND HOW DO WE DISTRIBUTE INFORMATION?

- Entropy: indicates un-predictability of an unseen message
  (Prediction here does not use immediate preceding context)
  - Proxy for **information density**
- **Normally increases over time**
- But pattern is converging among speakers in dialogue
- Language seems to be distributed to equalize information density (**Uniform Information Density, Jaeger&Levy 2007**)
  - Hypothesis: Behavior in general is also distributed to avoid spikes in information load
  - Why?



(a)

Yang Xu, ACL 2016
Xu&Reitter, Cognition (to appear)