At CoNLL-2021

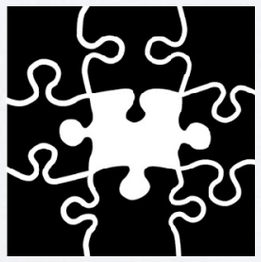
NLPitch - ILLC - 26 October 2021

Analysing Human Strategies of Information Transmission as a Function of Discourse Context

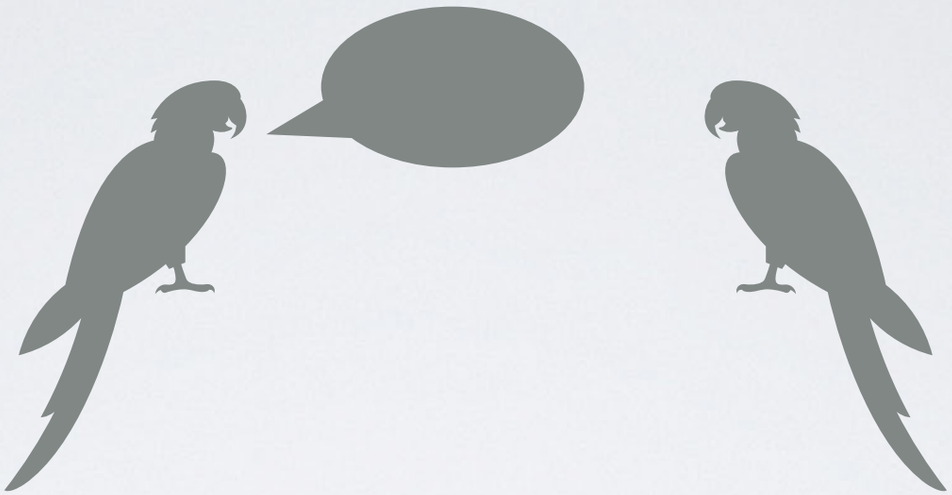
Mario Giulianelli and Raquel Fernández

Institute for Logic, Language and Computation
University of Amsterdam

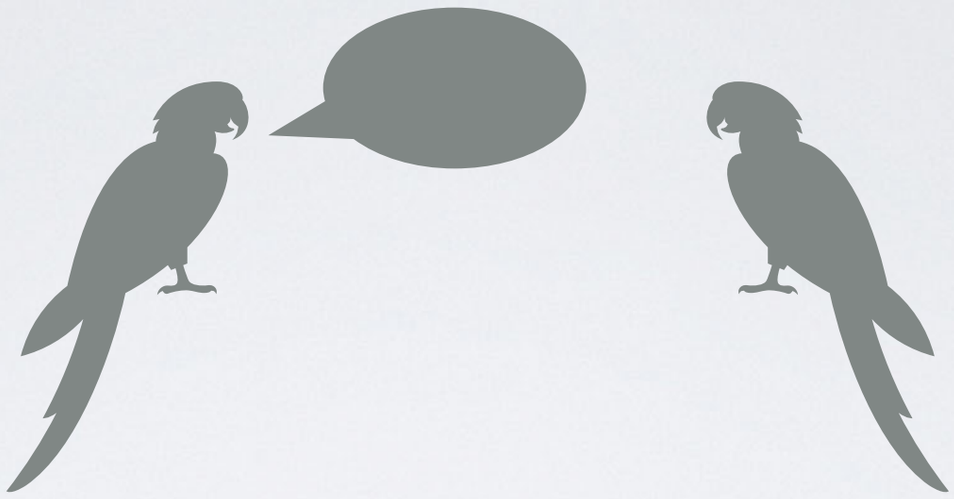
{m.giulianelli|raquel.fernandez}@uva.nl



Collaborative effort in language production

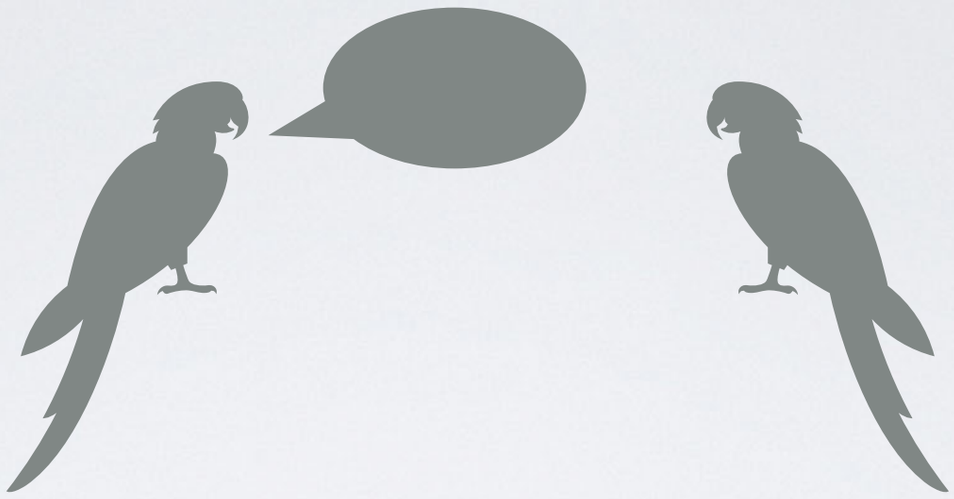


Collaborative effort in language production



PROCESSING EFFORT

Collaborative effort in language production



PROCESSING EFFORT

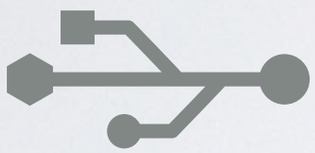
Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

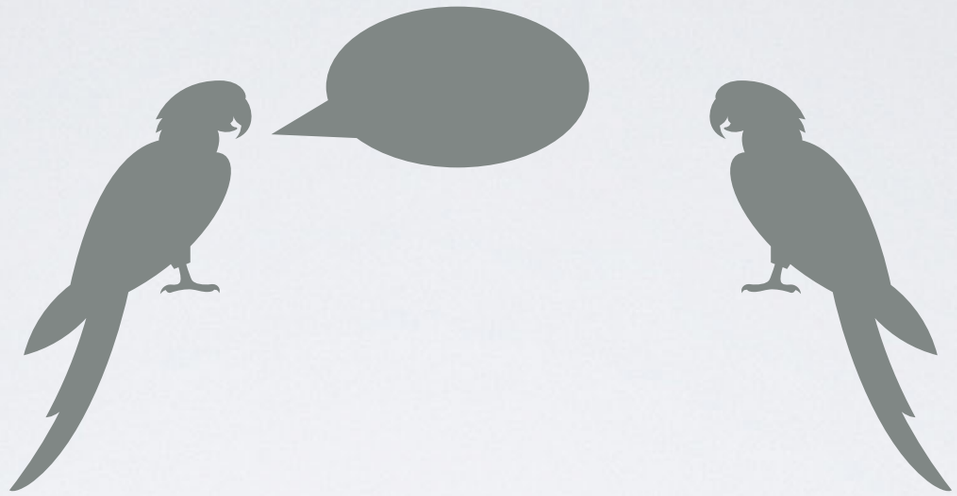
Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central of Illinois Inc., a bank holding company based in Sterling, earned $2 million.

Collaborative effort in language production

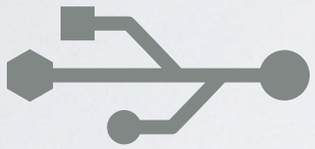
PRODUCTION
EFFORT

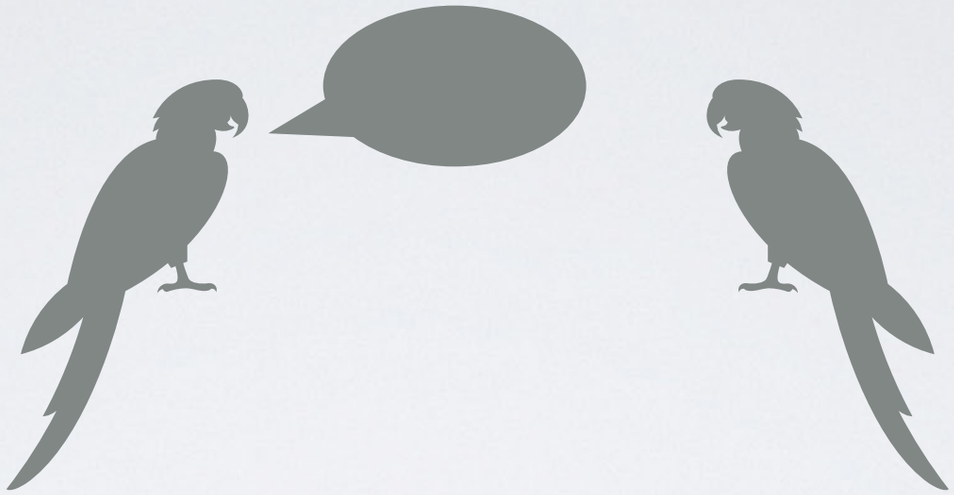




Collaborative effort in language production

PRODUCTION EFFORT



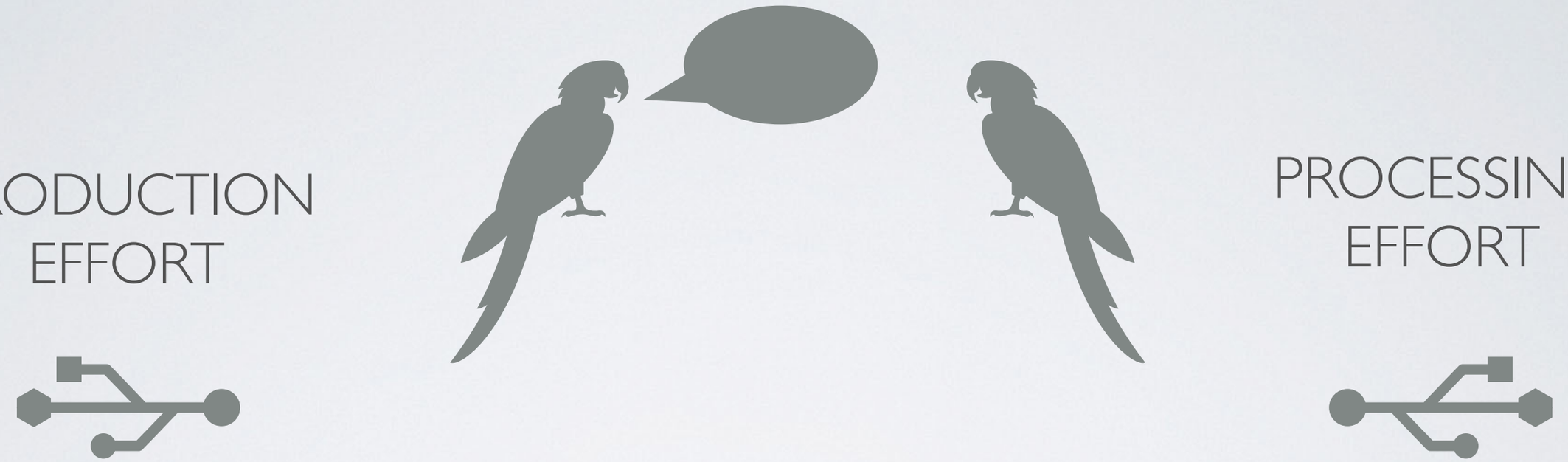


Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Collaborative effort in language production



PRODUCTION EFFORT

PROCESSING EFFORT

Information Theory
(Shannon, 1948)

Information-theoretic measures

Information-theoretic measures

$H(S)$ Shannon information content

Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Information-theoretic measures

$H(S)$ Out-of-context information content

Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Information-theoretic measures

$H(S)$ Out-of-context information content

$I(S;C)$ Mutual information

Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Information-theoretic measures

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Information-theoretic measures

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content

Amcore Financial Inc. said it agreed to acquire Central of Illinois Inc. in a stock swap.

Shareholders of Central, a bank holding company based in Sterling, will receive Amcore stock equal to 10 times Central's 1989 earnings, Amcore said.

For the first nine months of 1989, Central earned $2 million.

Entropy Rate Constancy

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content

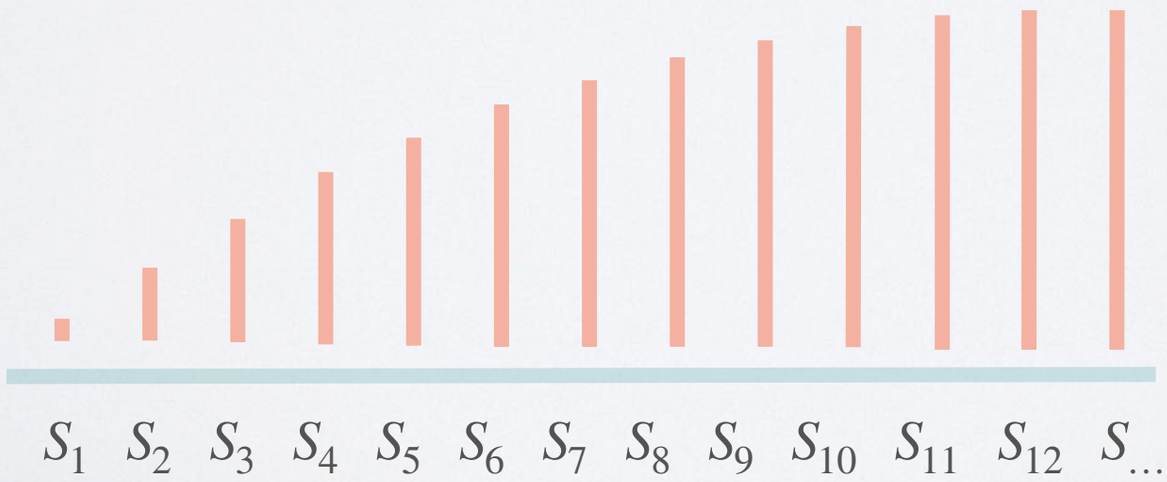
(Genzel & Charniak, 2002)

Entropy Rate Constancy

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content



$I(S;C)$

$H(S \mid C)$

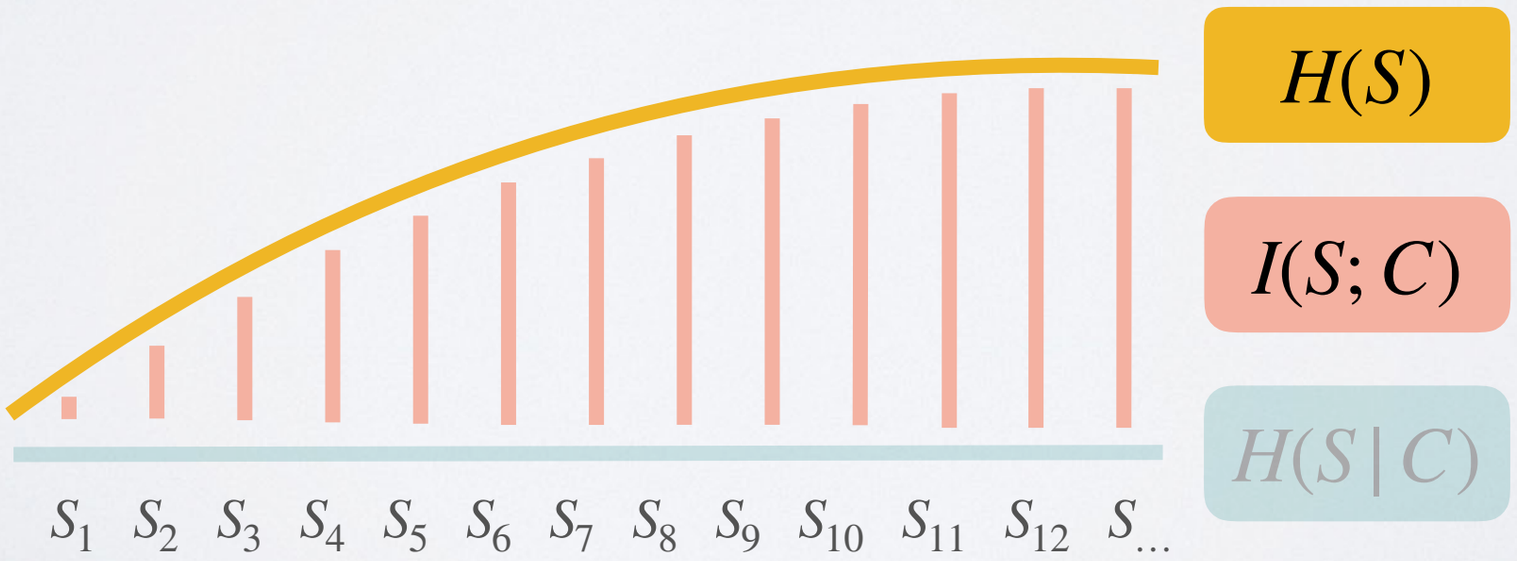
(Genzel & Charniak, 2002)

Entropy Rate Constancy

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content



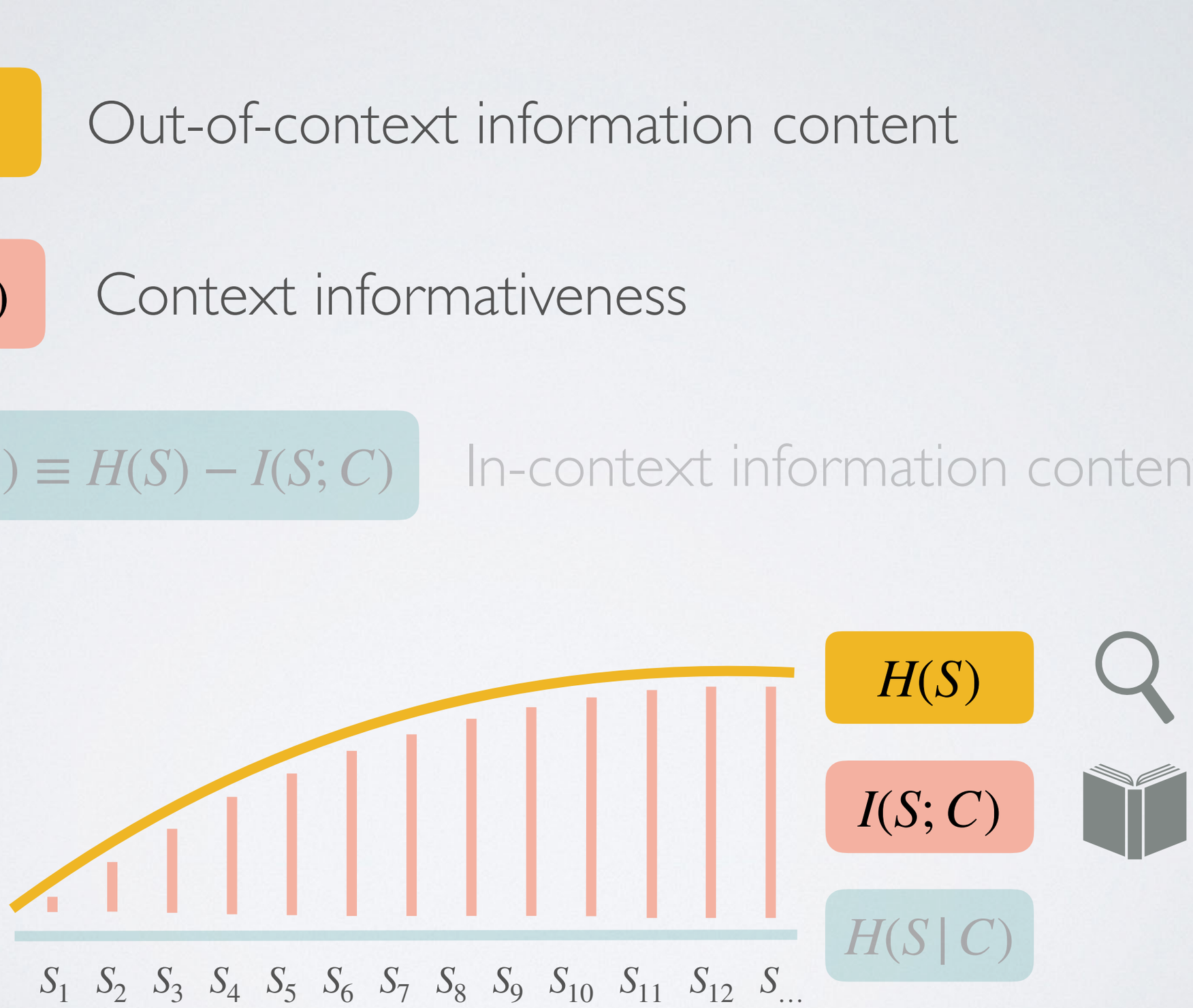
(Genzel & Charniak, 2002)

Entropy Rate Constancy

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content



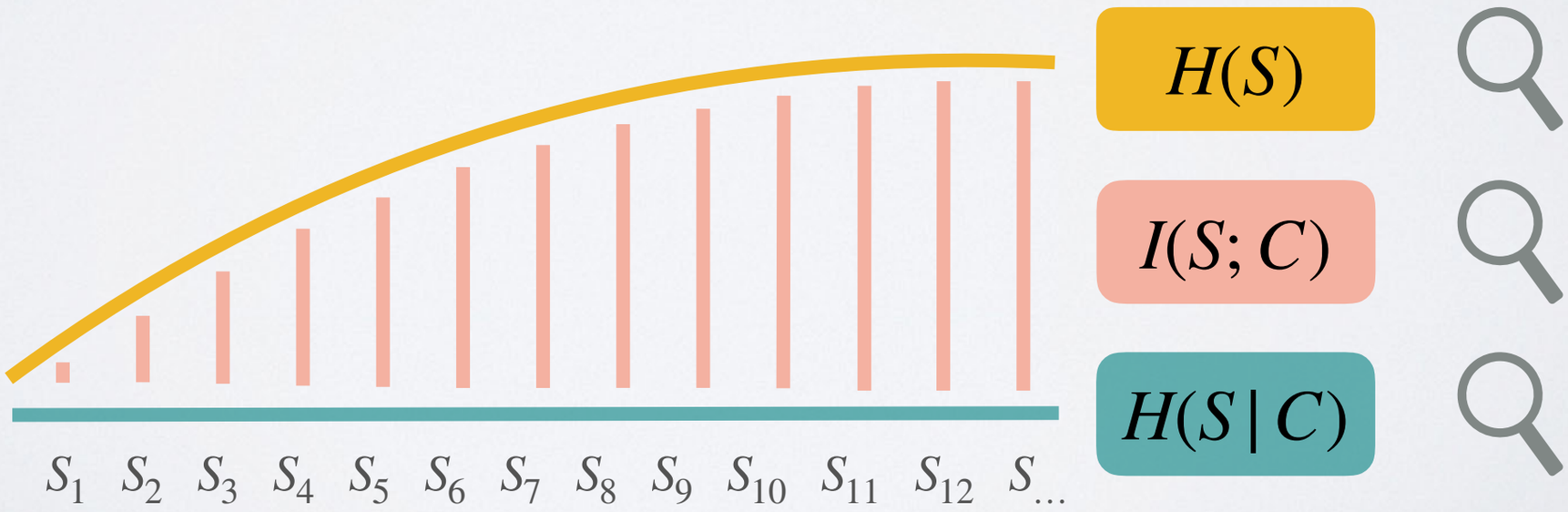
(Genzel & Charniak, 2002)

Entropy Rate Constancy

$H(S)$ Out-of-context information content

$I(S;C)$ Context informativeness

$H(S \mid C) \equiv H(S) - I(S;C)$ In-context information content



The current study

Estimates of information content

$$H(S) = -\log_2(S) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1})$$

Estimates of information content

$$H(S) = -\log_2(S) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1})$$

$$H(S \mid C) = -\log_2 P(S \mid C) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1}, C)$$

Estimates of information content

$$H(S) = -\log_2(S) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1})$$

$$H(S \mid C) = -\log_2 P(S \mid C) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1}, C)$$

$$I(S; C) \equiv H(S) - H(S \mid C)$$

Estimates of information content

$$H(S) = -\log_2(S) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1})$$

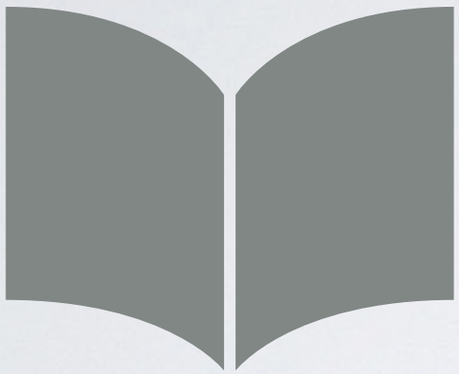
$$H(S \mid C) = -\log_2 P(S \mid C) = -\frac{1}{|S|}\sum_{w_i \in S} \log_2 P(w_i \mid w_1, \ldots, w_{i-1}, C)$$

$$I(S; C) \equiv H(S) - H(S \mid C)$$

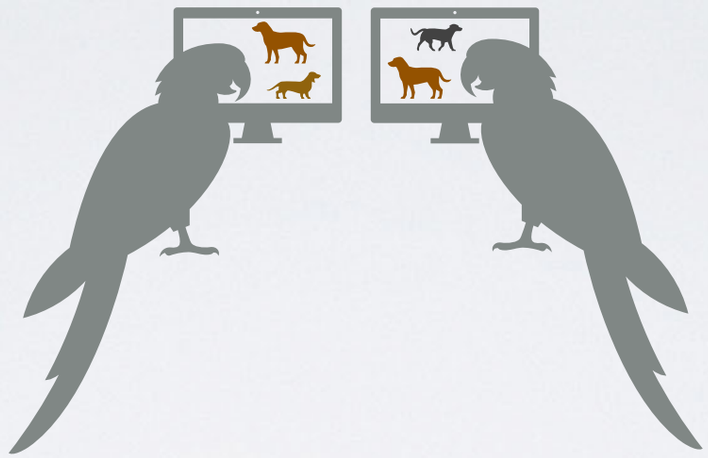
$P(w_i \mid \ldots)$ estimates obtained with GPT-2
fine-tuned on 70% of each target corpus

(30% held-out for analysis)

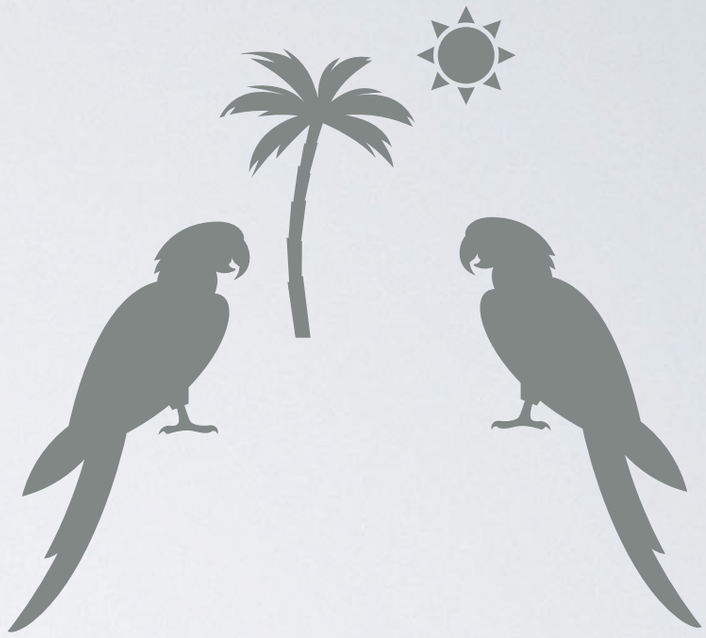
Experimental data



Penn Treebank
newspaper
articles
(Mitchell et al. 1999)

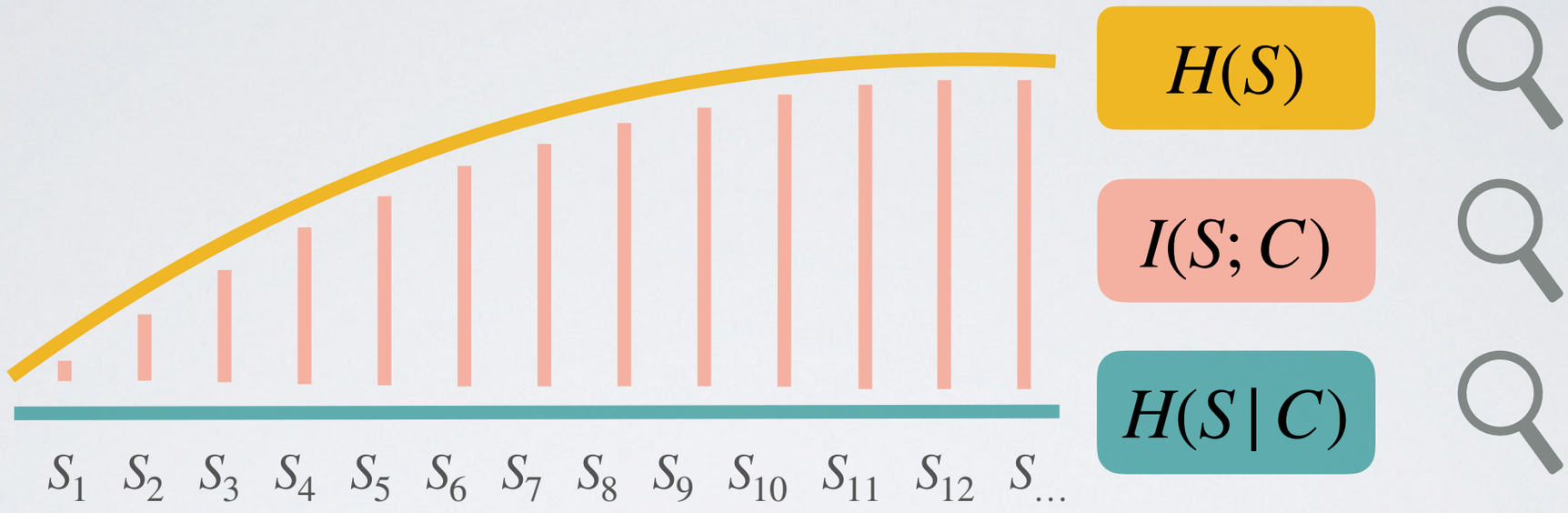


PhotoBook
task-oriented
written dialogues
(Haber et al., 2019)

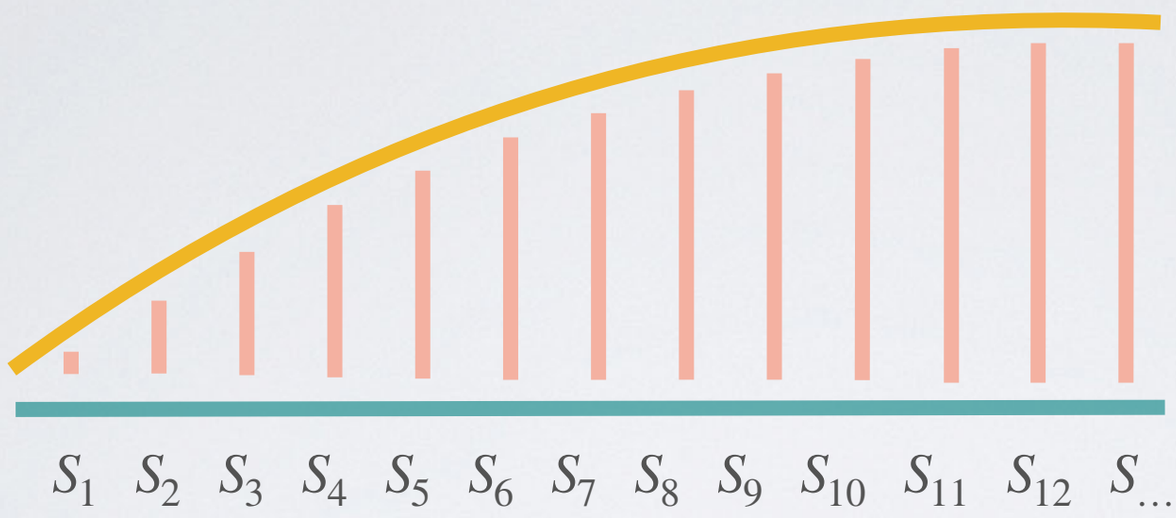


Spoken BNC
open domain
dialogues
(Love et al., 2017)

Experiment 1



Experiment 1



$H(S)$

$I(S;C)$

$H(S \mid C)$

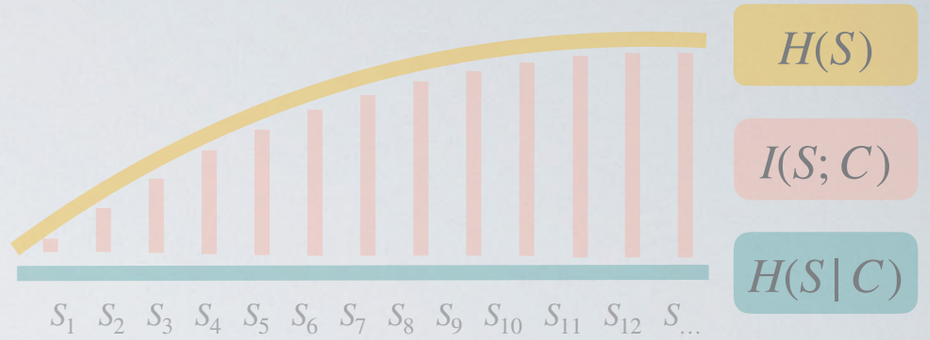
$H(S)$ ~ fixed effects

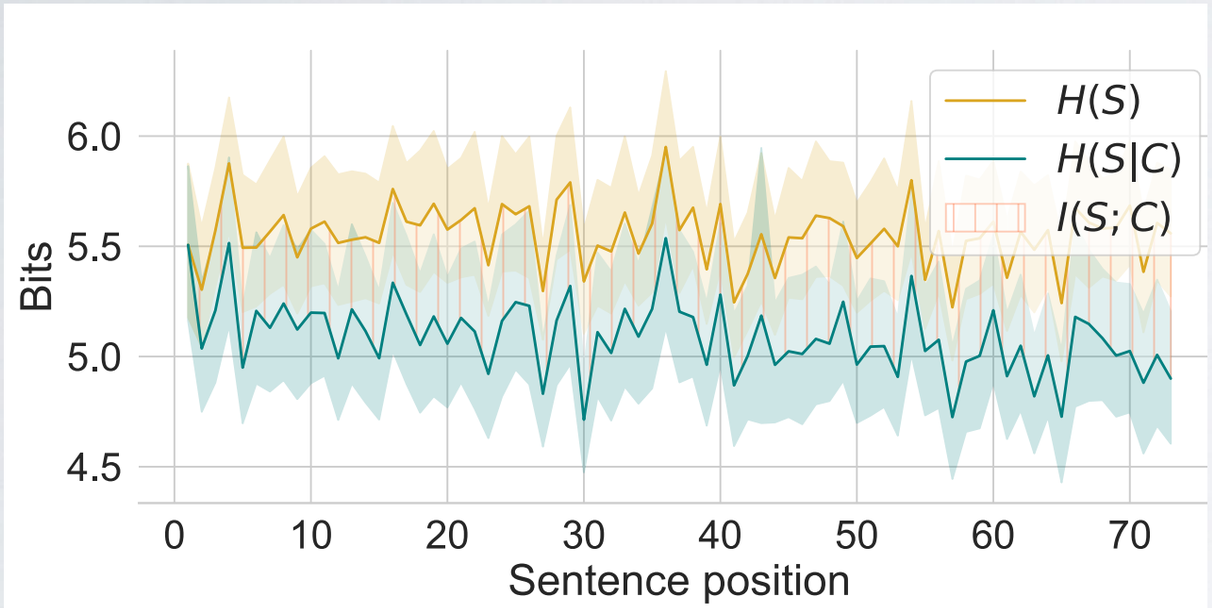
$I(S;C)$ ~ 1 + log sentence position + log sentence length

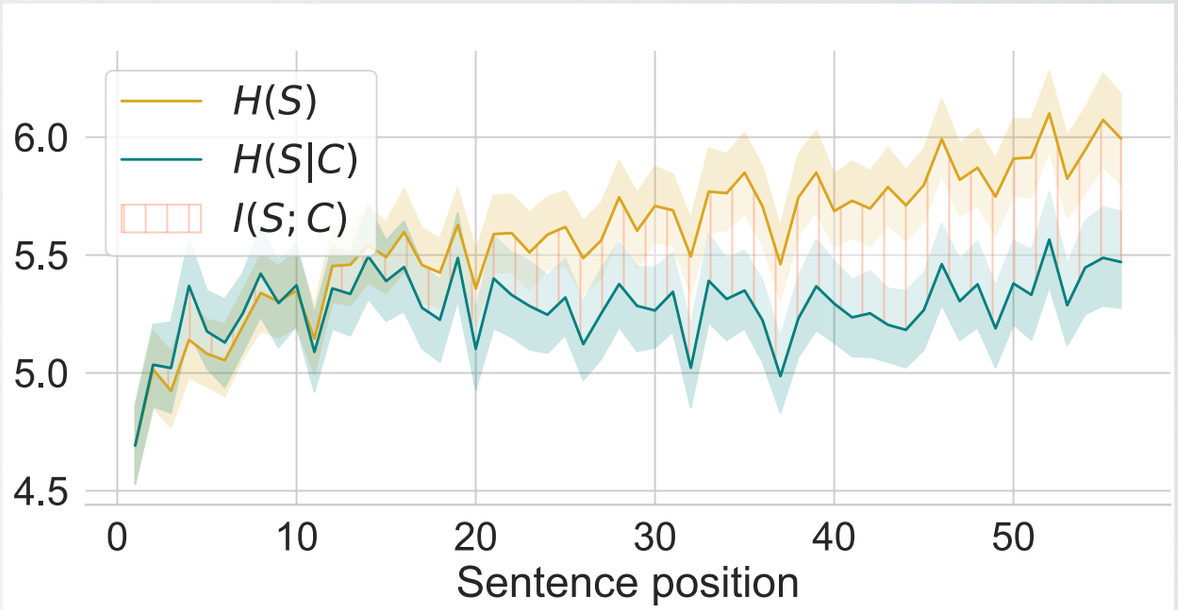
$H(S \mid C)$ ~ + (1 + log sentence position + log sentence length | doc_id)

random intercept and slope
grouped by text / dialogue

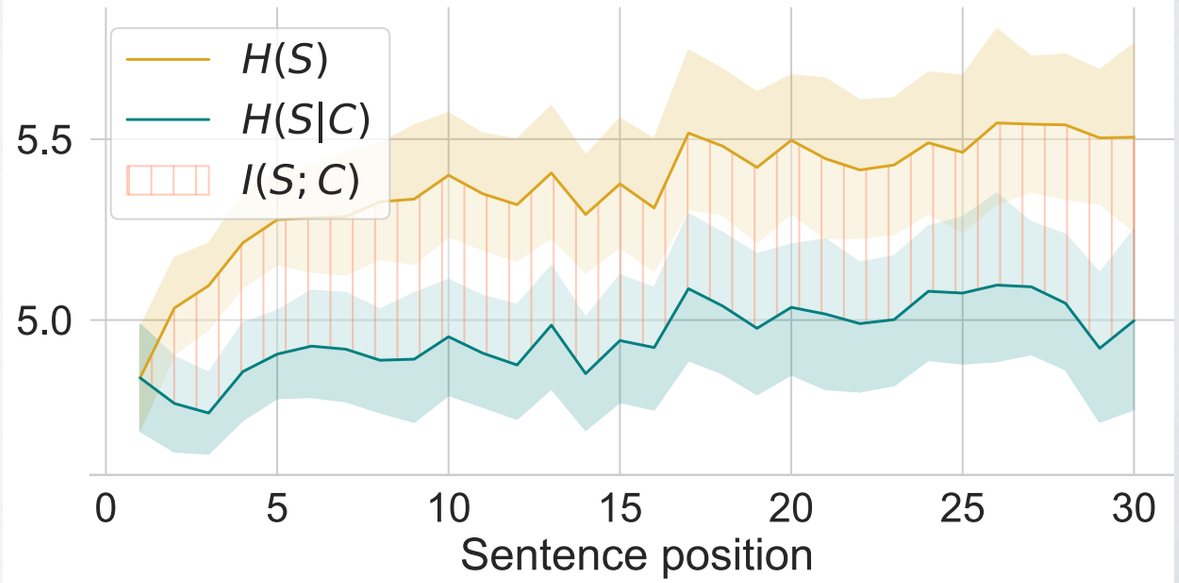
Experiment 1



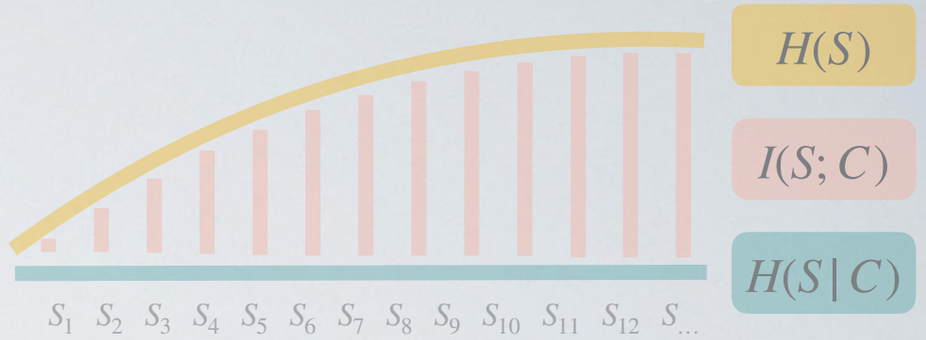


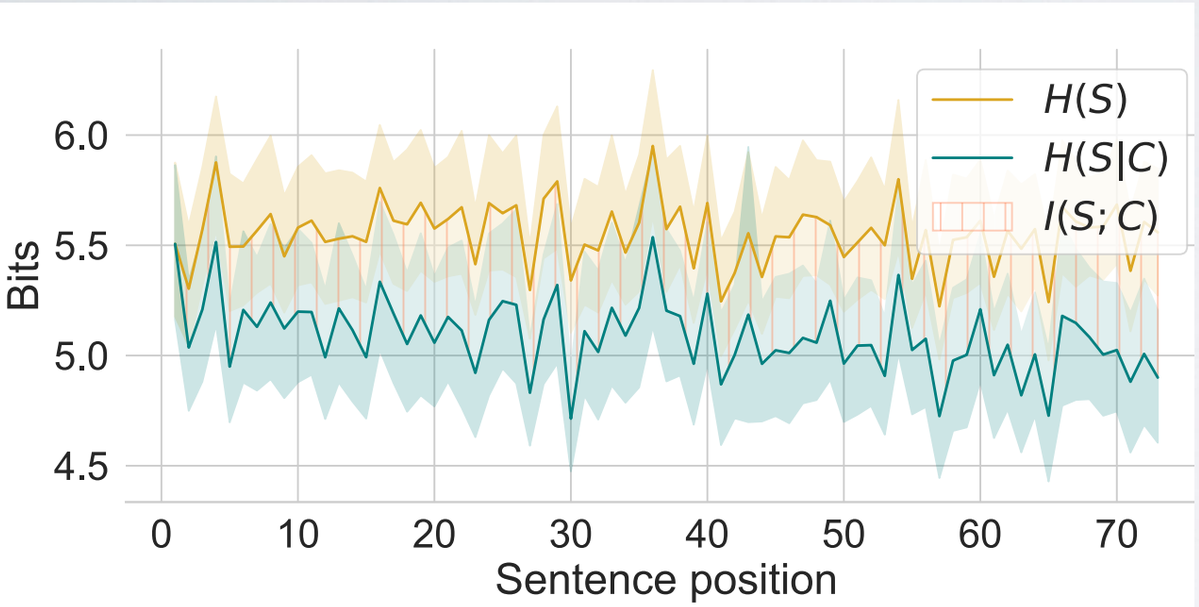


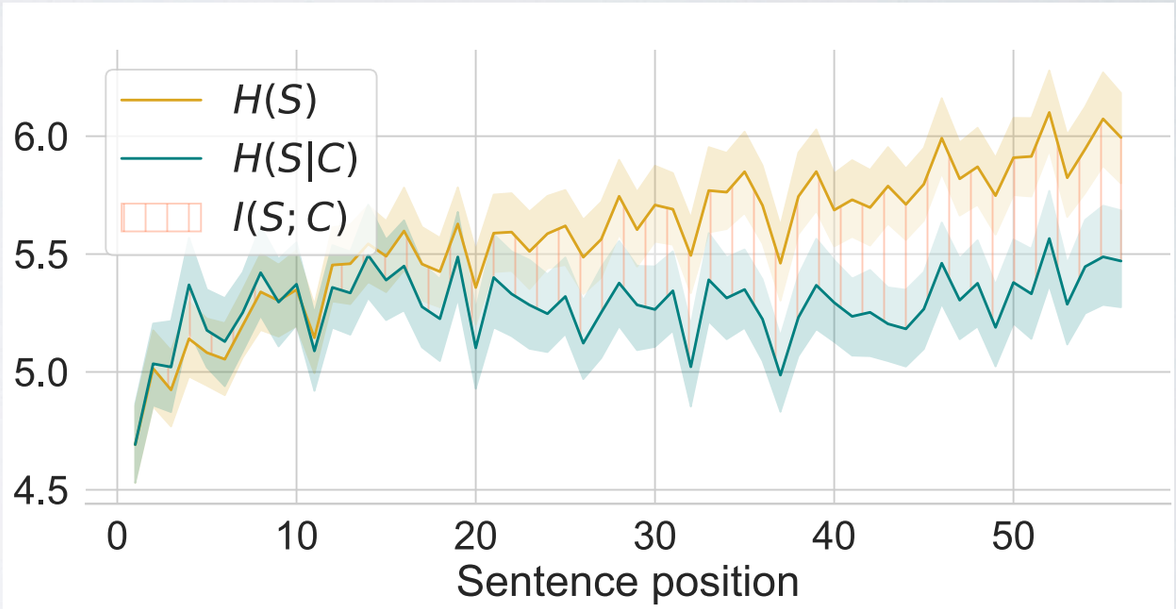
Effect of sentence position

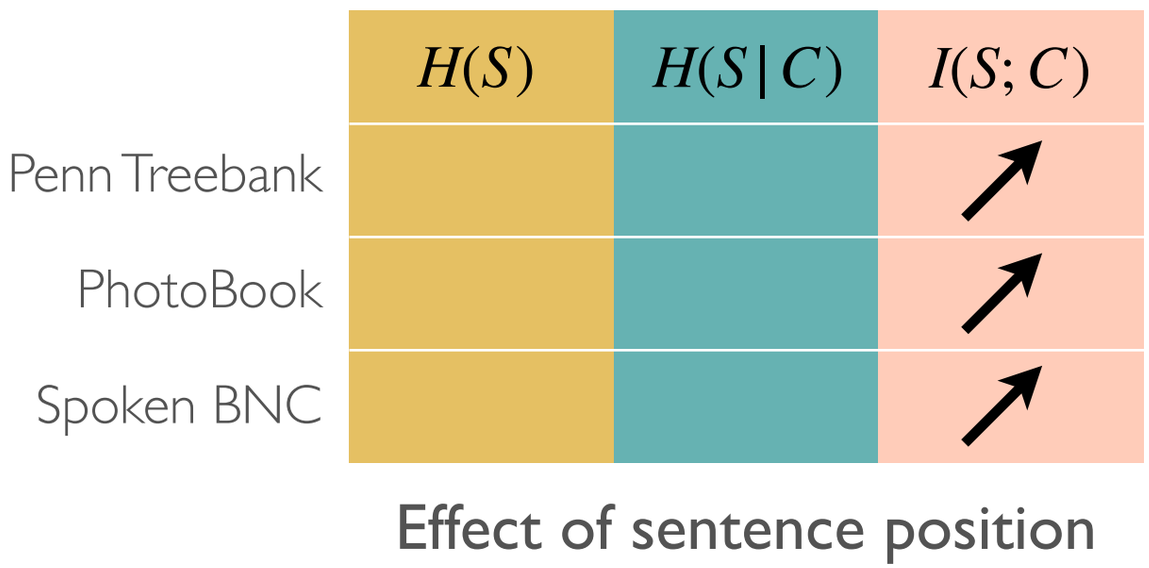


Experiment 1

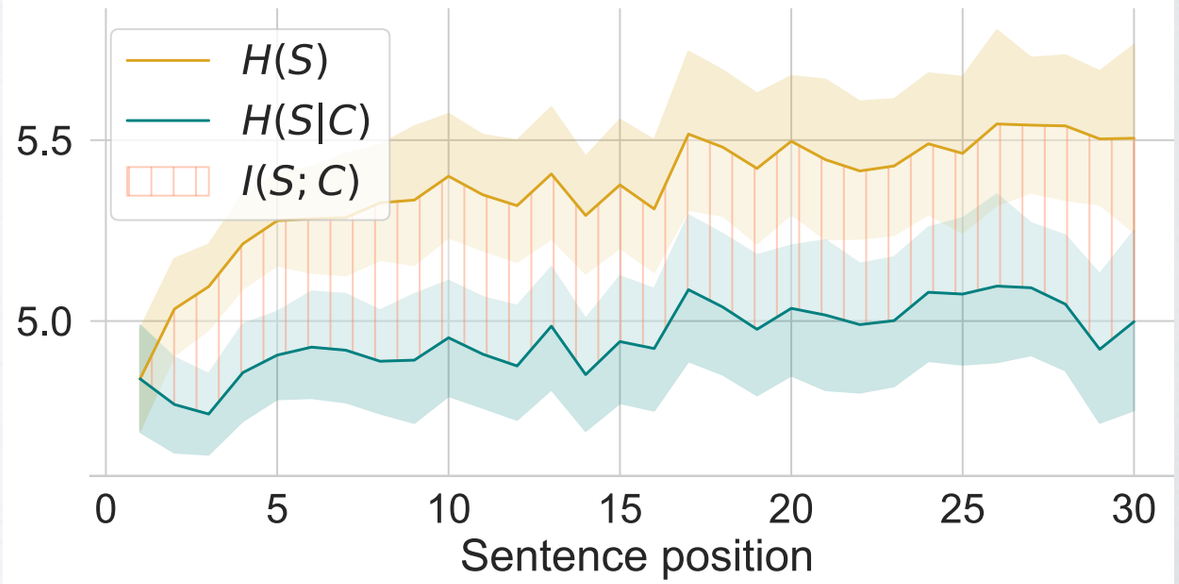




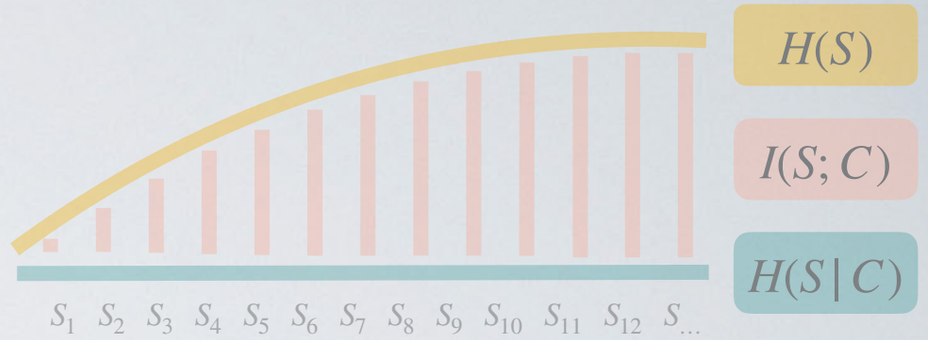


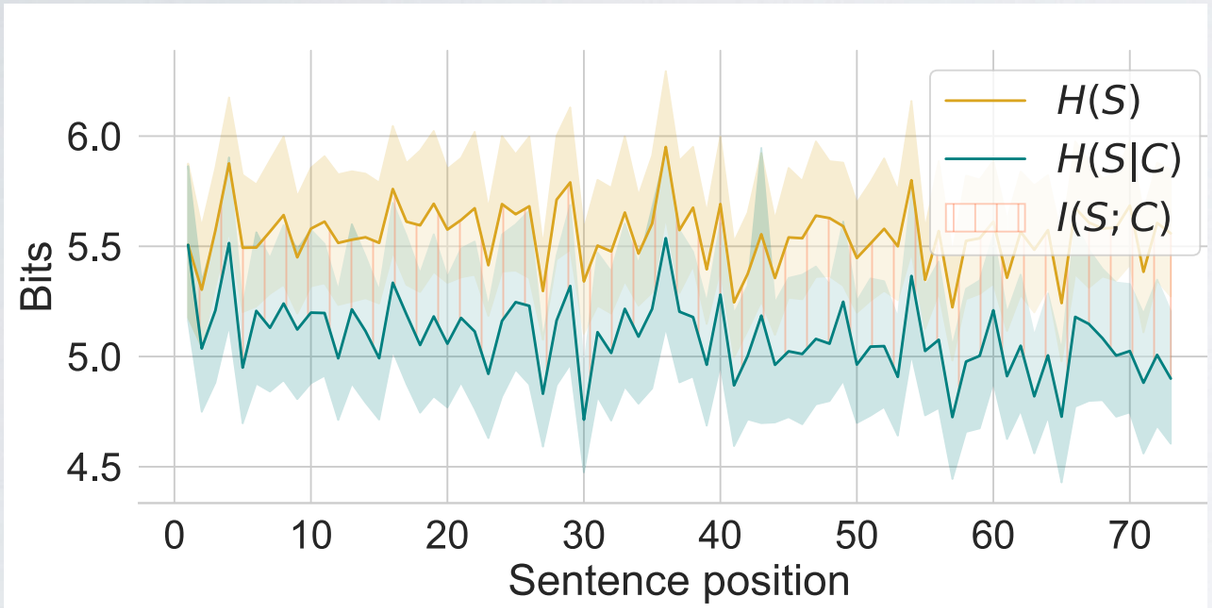


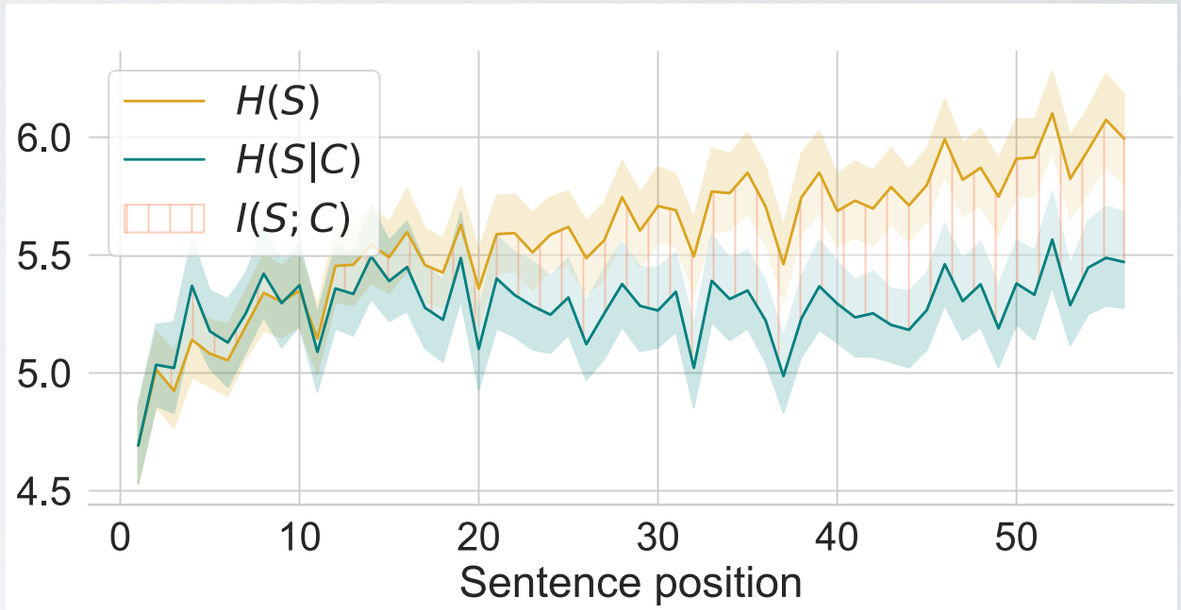
Effect of sentence position

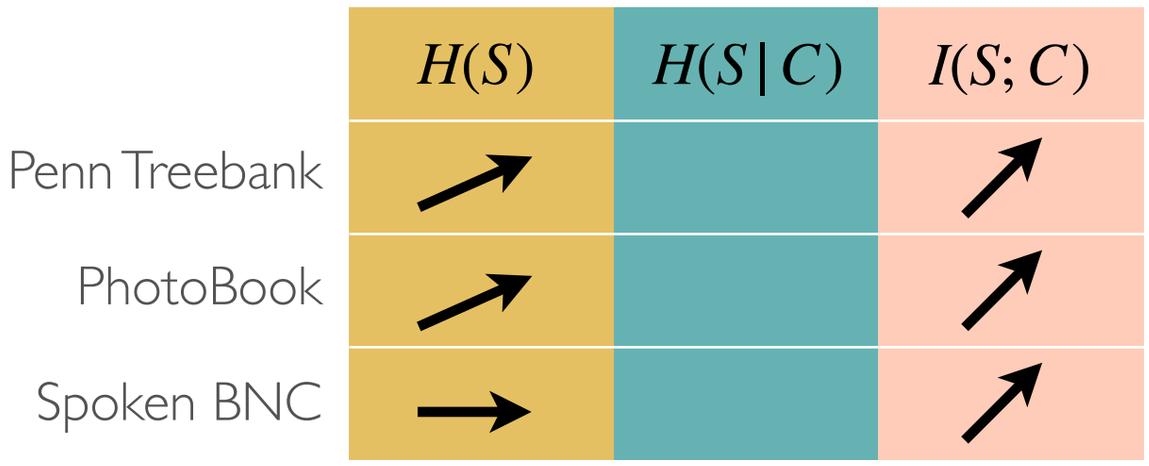


Experiment 1

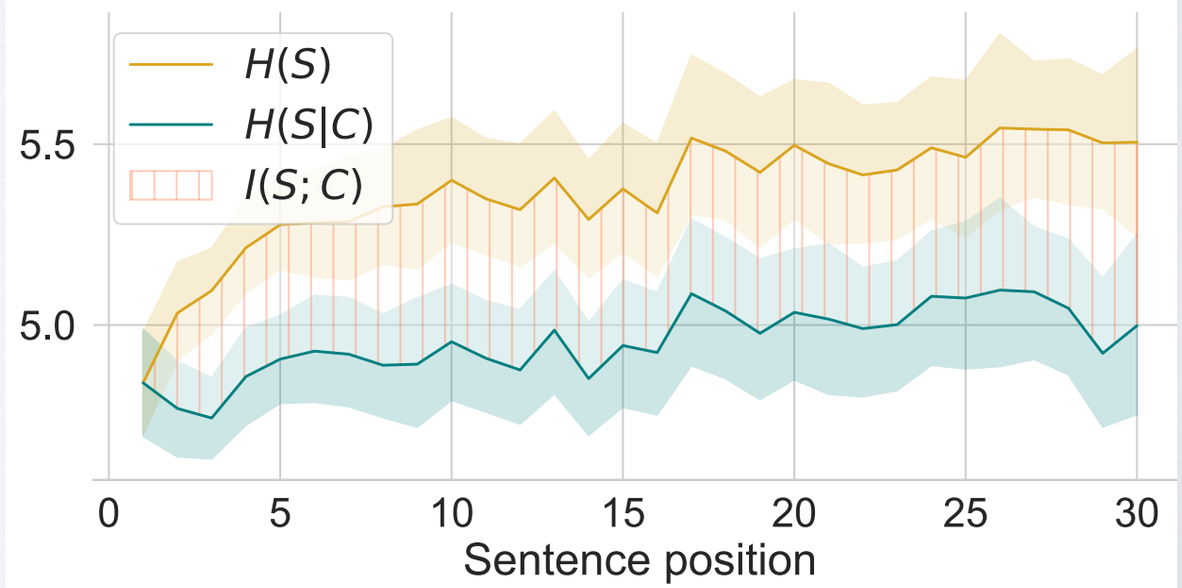




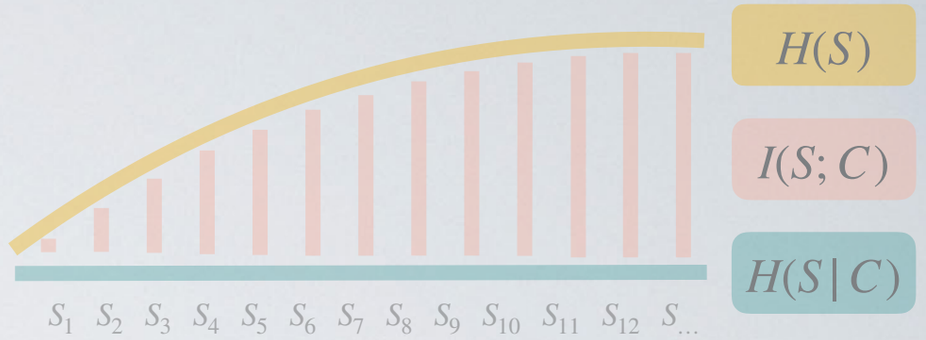


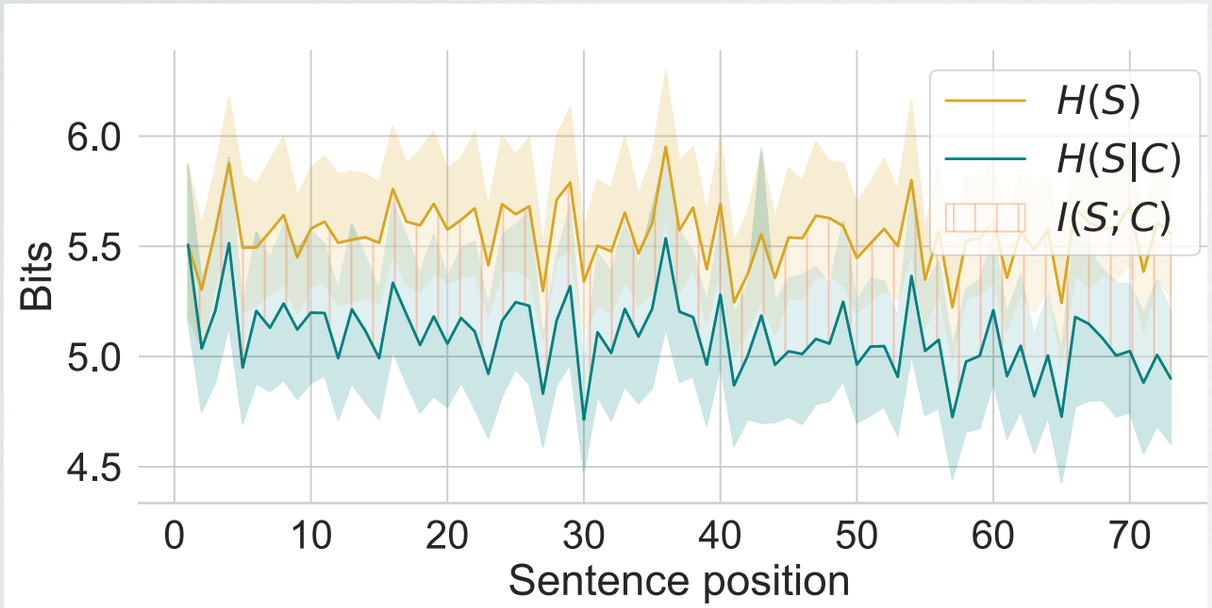


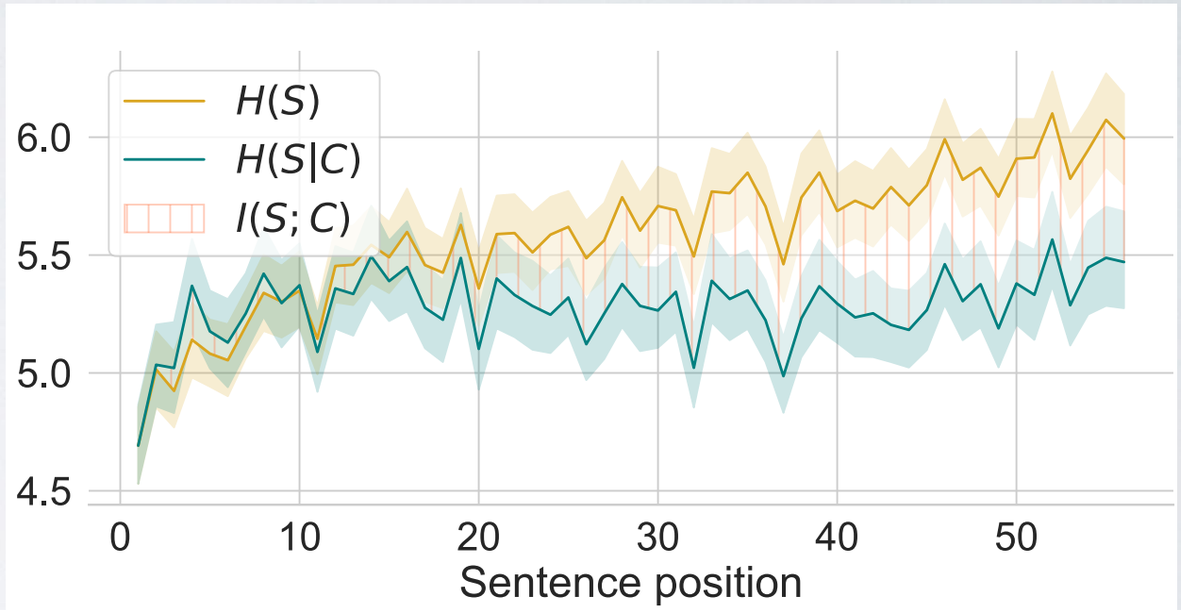
Effect of sentence position

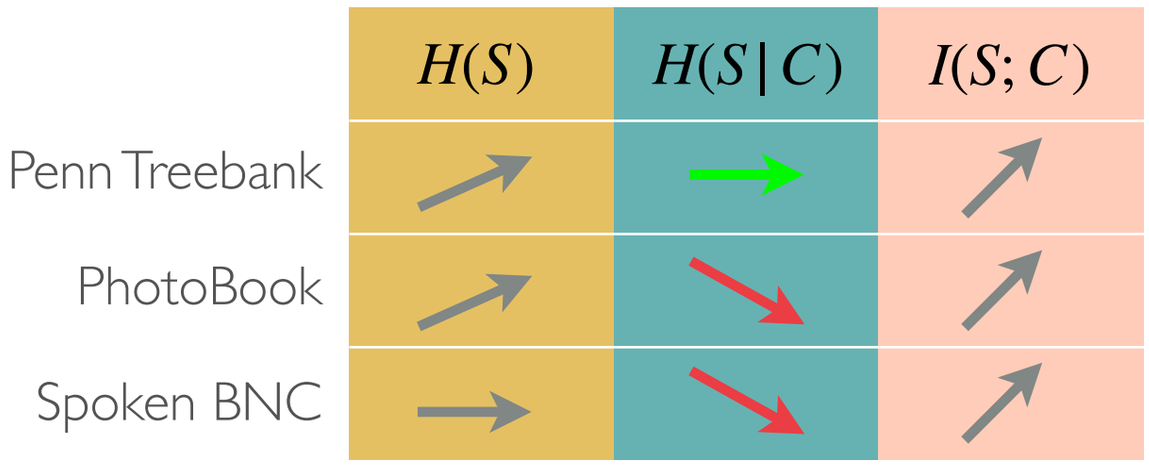


Experiment 1

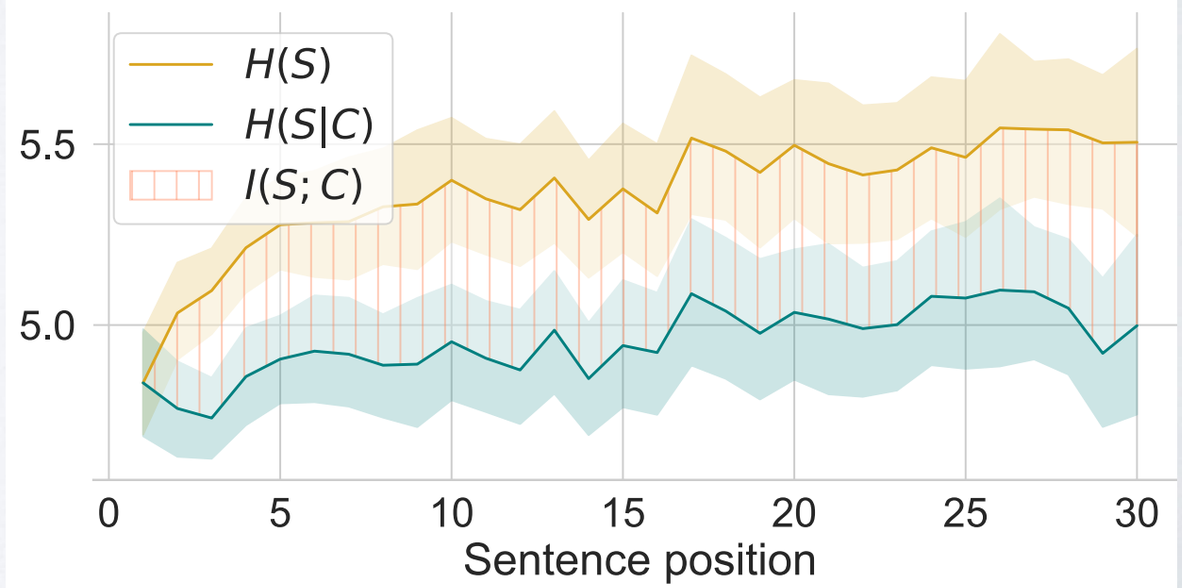








Effect of sentence position



$H(S \mid C)$

Experiment 2

Entropy Rate Constancy

(Genzel & Charniak, 2002)

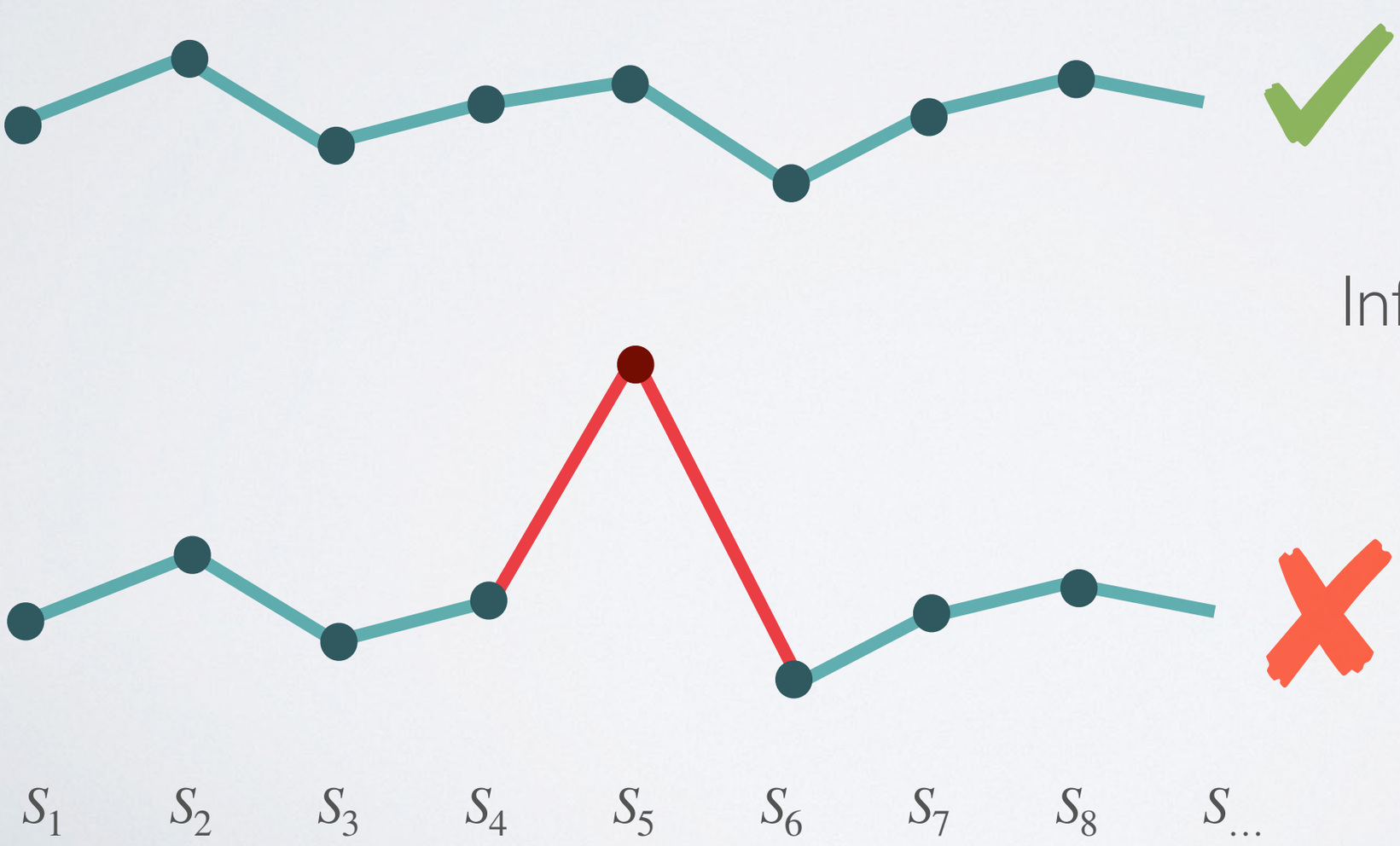
S_1 S_2 S_3 S_4 S_5 S_6 S_7 S_8 $S_{...}$

$H(S \mid C)$

Experiment 2

Entropy Rate Constancy

(Genzel & Charniak, 2002)



Uniform
Information Density

(Aylett & Turk, 2004;
Jaeger & Levy, 2007;
Levy, 2010)

S_1 S_2 S_3 S_4 S_5 S_6 S_7 S_8 $S_{...}$

Experiment 2

Criteria of uniformity (Collins, 2014)

Global centrality

$$-\frac{1}{N}\sum_{i=1}^{N}\left(H\left(S_i \mid C_i\right)-\mu\right)^2$$

Local predictability

$$-\frac{1}{N}\sum_{i=2}^{N}\left(H\left(S_i \mid C_i\right)-H\left(S_{i-1} \mid C_{i-1}\right)\right)^2$$

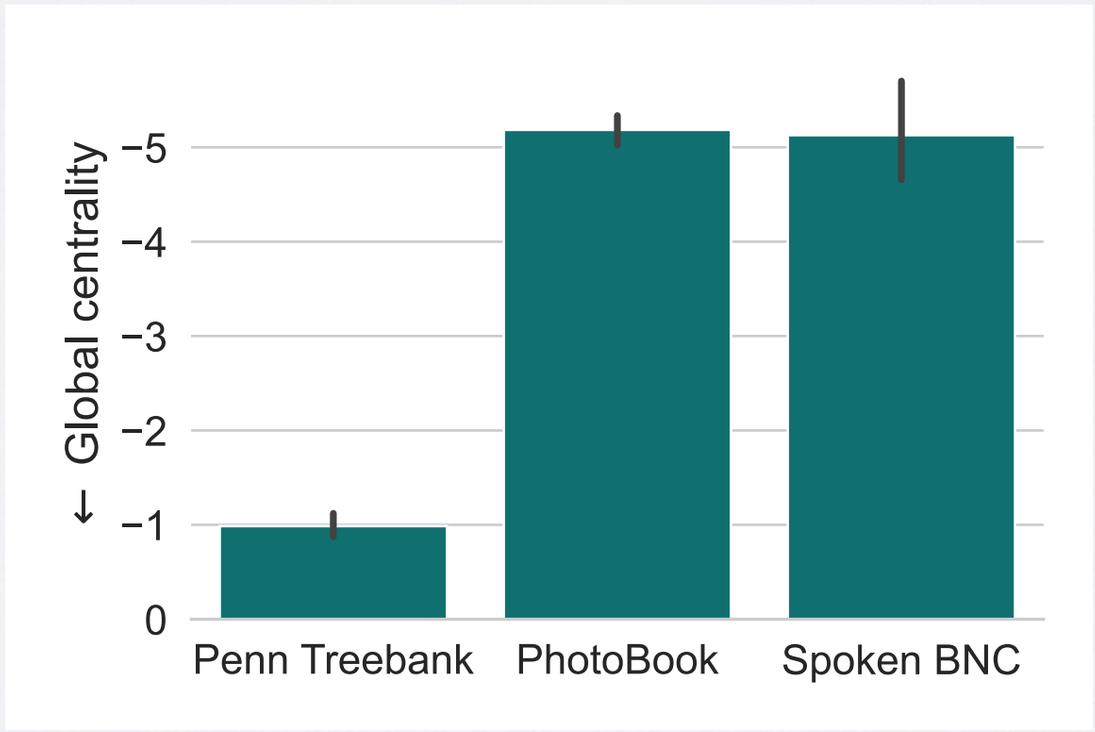
N number of sentences in the text / dialogue

μ average information content in the text / dialogue

Experiment 2

Global centrality

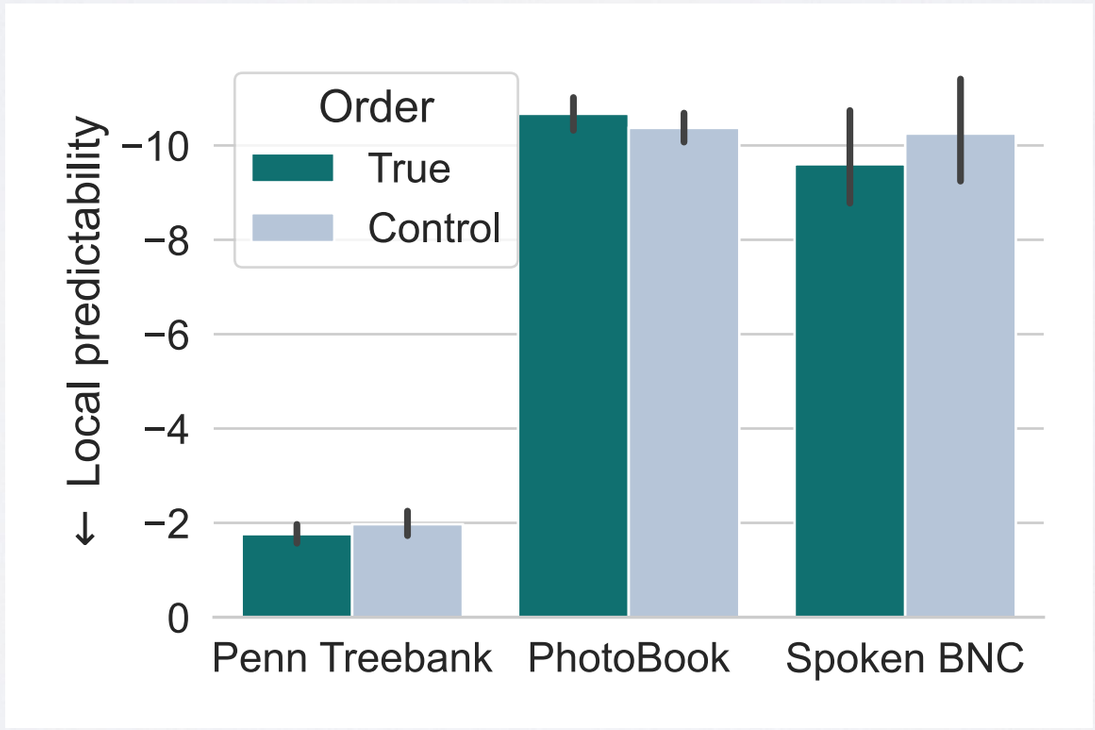
$$-\frac{1}{N}\sum_{i=1}^{N}\left(H\left(S_i \mid C_i\right)-\mu\right)^2$$



Experiment 2

Local predictability

$$-\frac{1}{N}\sum_{i=2}^{N}\left(H\left(S_i \mid C_i\right) - H\left(S_{i-1} \mid C_{i-1}\right)\right)^2$$



Summary

We have examined some central tenets of the classic information-theoretic model of communication.

We have used language models to obtain information content estimates for sentences *within their discourse context*.

Summary

We have examined some central tenets of the classic information-theoretic model of communication.

We have used language models to obtain information content estimates for sentences *within their discourse context.*

Findings

In newspaper articles information content remains stable, as predicted by the Entropy Rate Constancy principle.

This is not the case for spoken open domain dialogues, nor for written task-oriented dialogues.

Global uniformity is a more faithful criterion than local uniformity.

Discussion

Identifying the relevant components of discourse context

→ *Giulianelli, Sinclair, Fernández - EMNLP 2021*

Single vs. multiple addressees

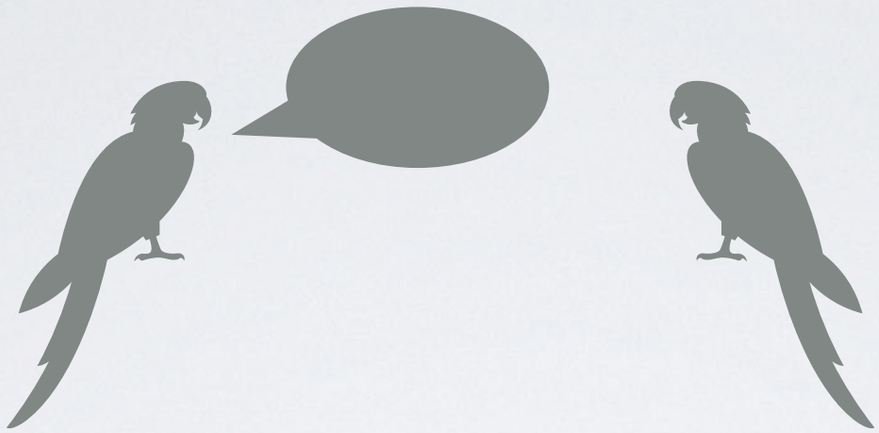
Production and comprehension effort conflated
in a single estimate

Architectures and training objectives to emulate the
organisation of information density found in human data

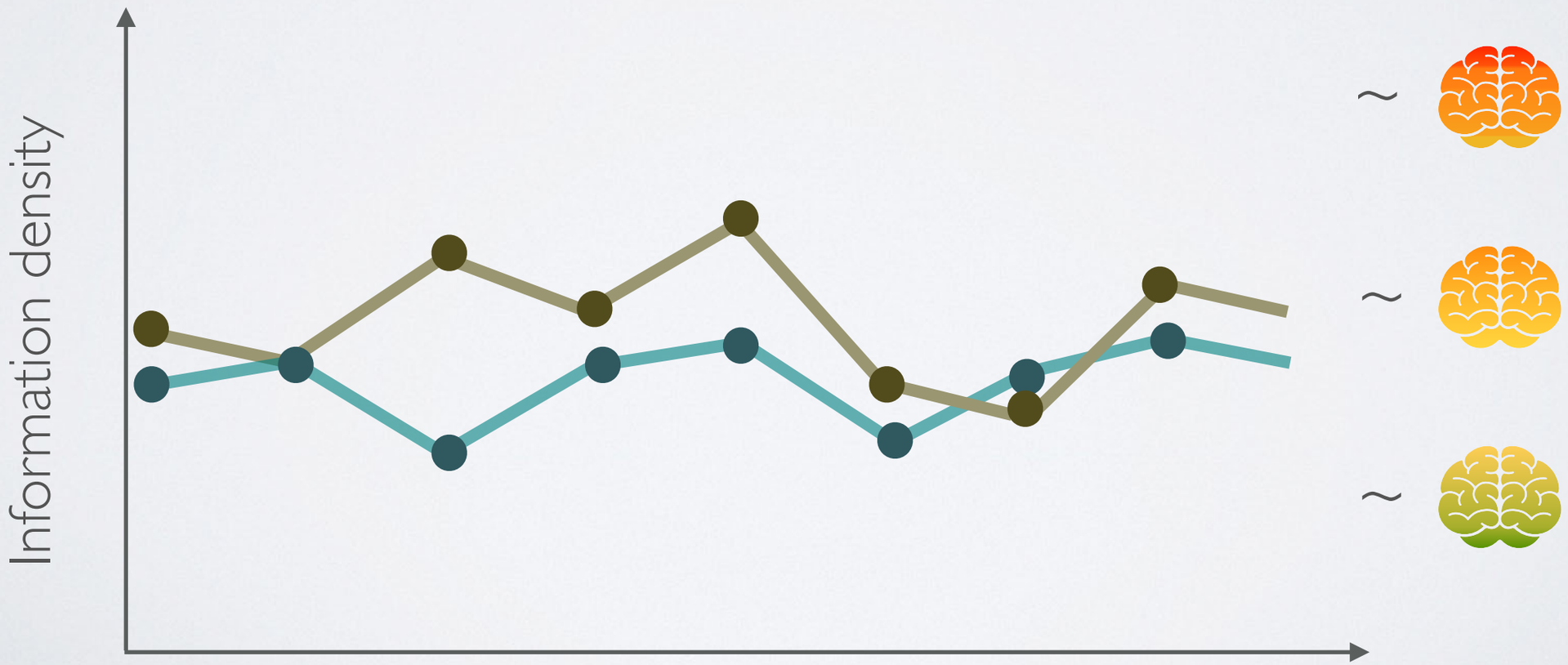
APPENDIX

Uniform Information Density

(Aylett & Turk, 2004; Jaeger & Levy, 2007)

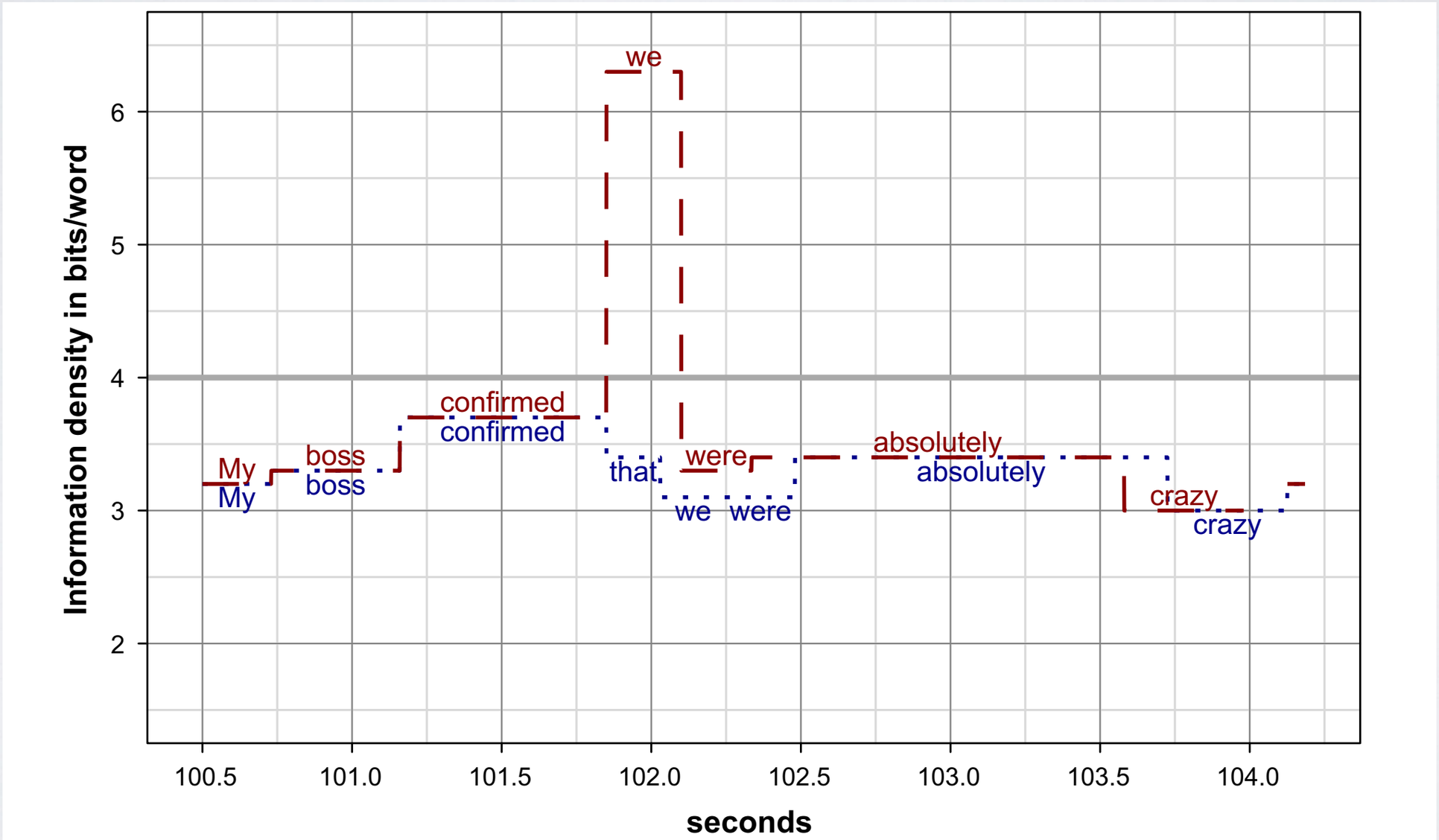


PROCESSING EFFORT



Uniform Information Density

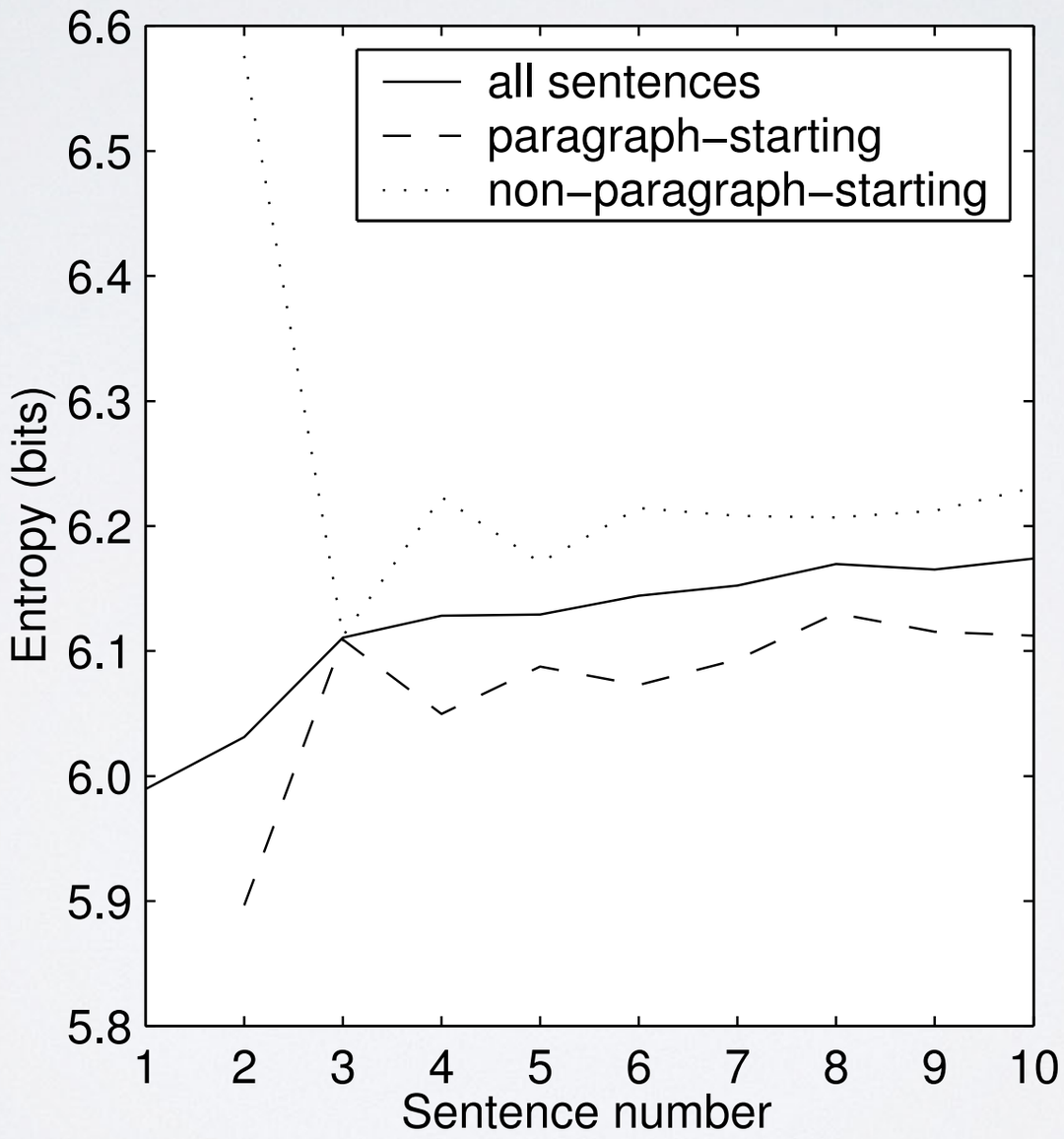
Example 1: Syntax
Complementiser *that*-mentioning (e.g., Jaeger , 2010)



(Jaeger , 2010; Figure 1a)

Uniform Information Density

Example 2: Discourse
Entropy rate constancy (Genzel & Charniak, 2002, 2003)



(Genzel & Charniak, 2003; Figure 1)

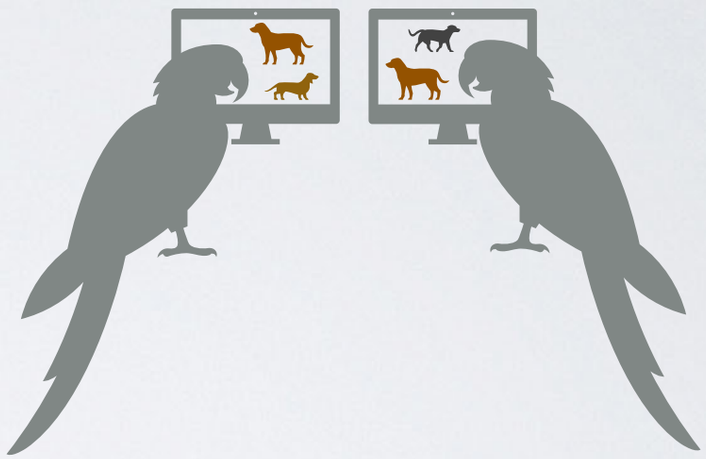
Task-oriented dialogue corpora

Round 3

B	pink sweater woman and man w/umbrella
A	yes
B	statue man with umbrella
A	y
B	guy in black suit with 2 plaid blue umbrellas
A	no

Round 4

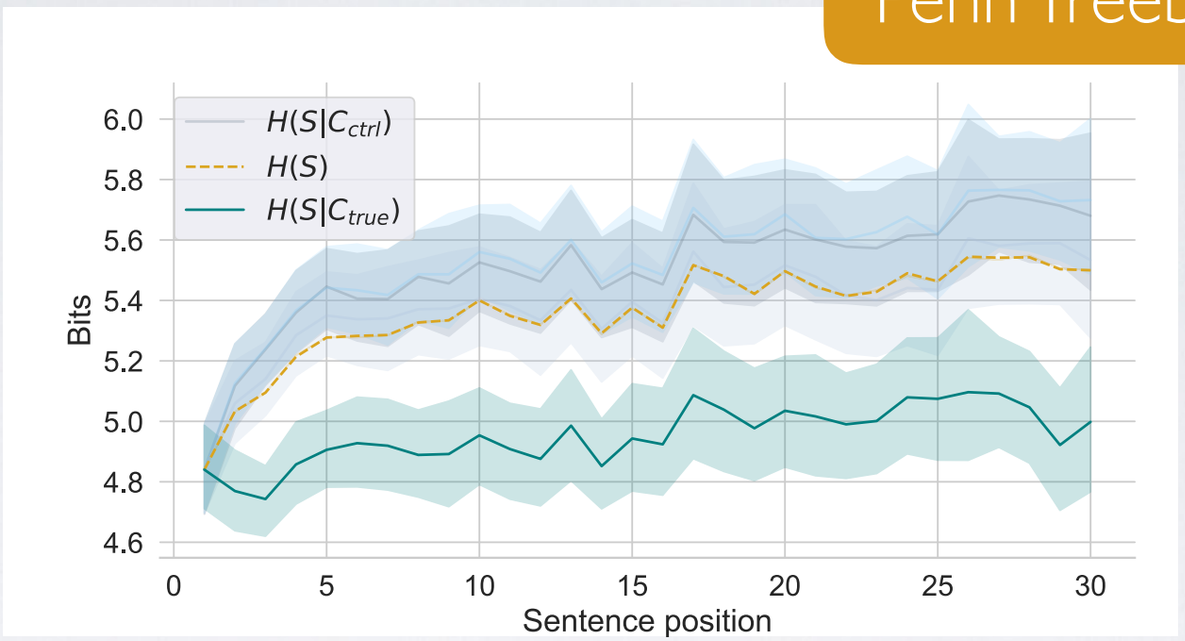
B	guy in black suit with 2 plaid blue umbrellas
A	no
B	lady in pink and guy with umbrella
A	no
B	statue/umbralla again
A	no



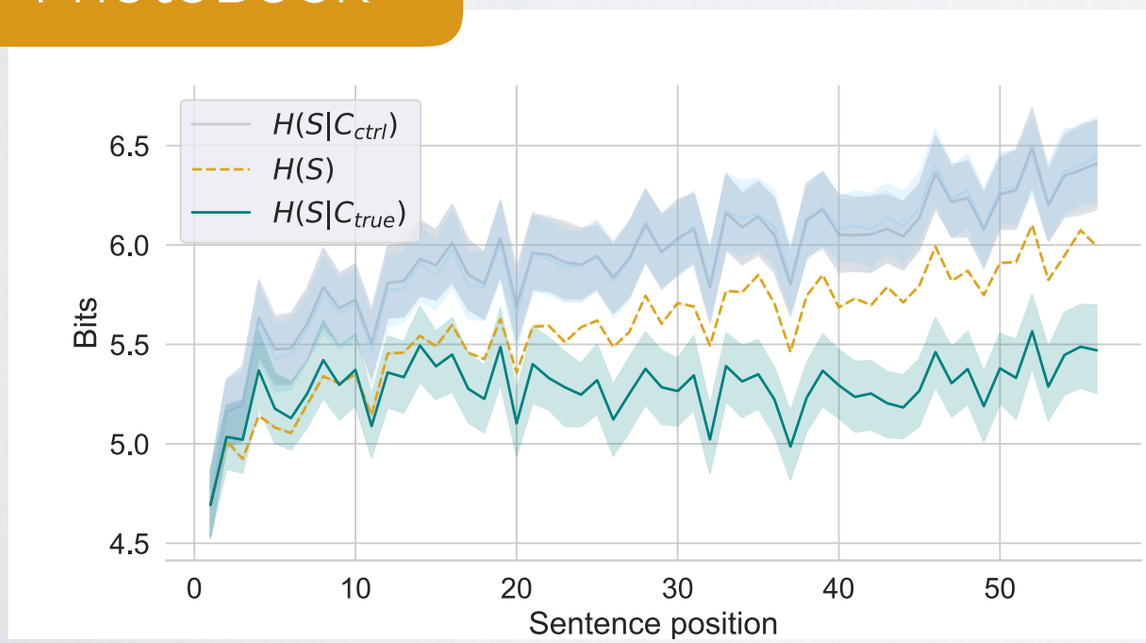
PhotoBook
written cooperative reference game
(Haber et al., 2019)

Control runs

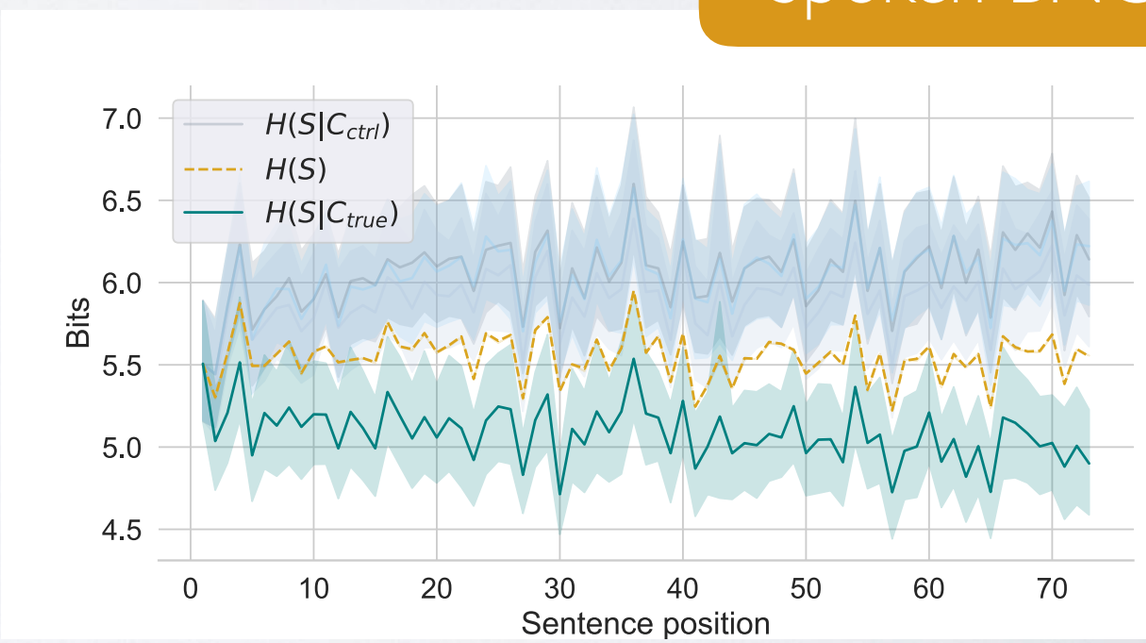
Penn Treebank



PhotoBook

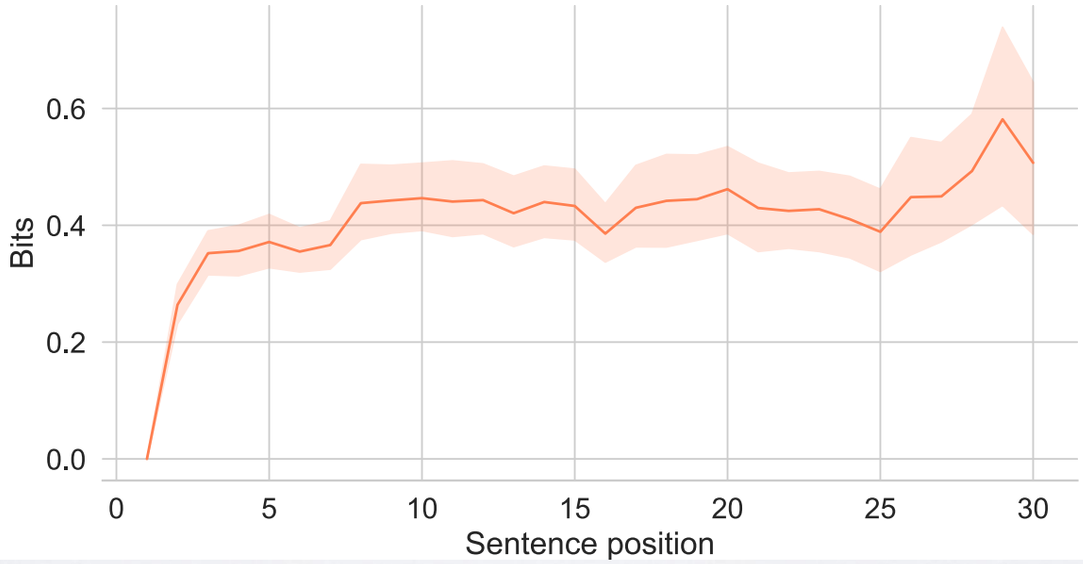


Spoken BNC

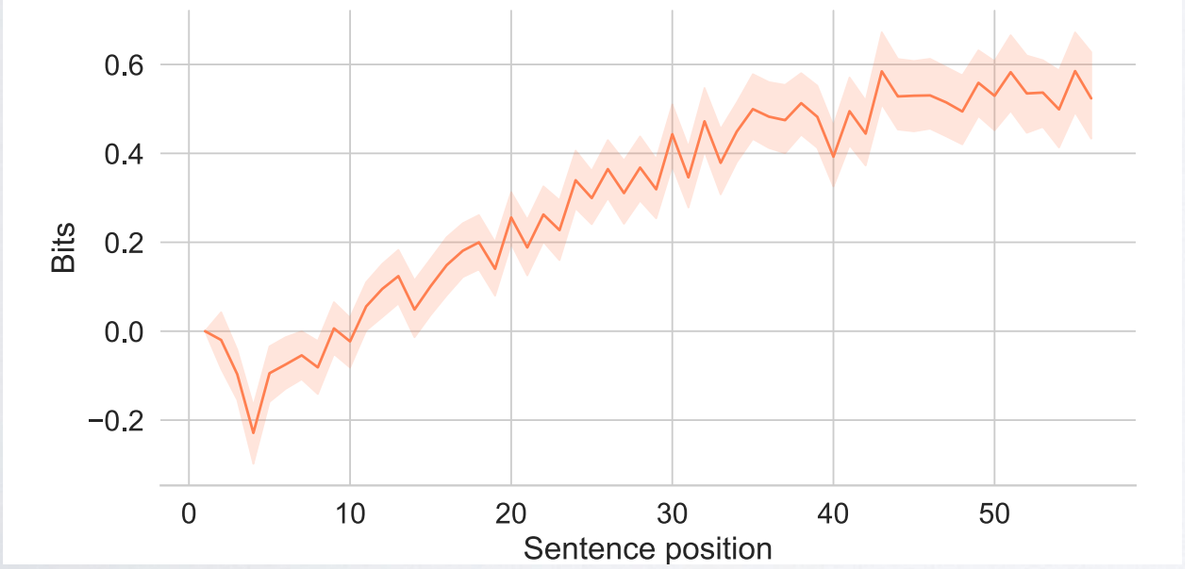


Context informativeness

Penn Treebank



PhotoBook



Spoken BNC