Information-Theoretic Storage Cost in Sentence Comprehension

Kohei Kajikawa
Georgetown

Shinnosuke Isono
NINJAL

Ethan G. Wilcox
Georgetown

←Our paper 👀

Background

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The __ __ __ __ __

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat __ __ __ __

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat ___ ___ ___

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on __ __

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the __

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the mat

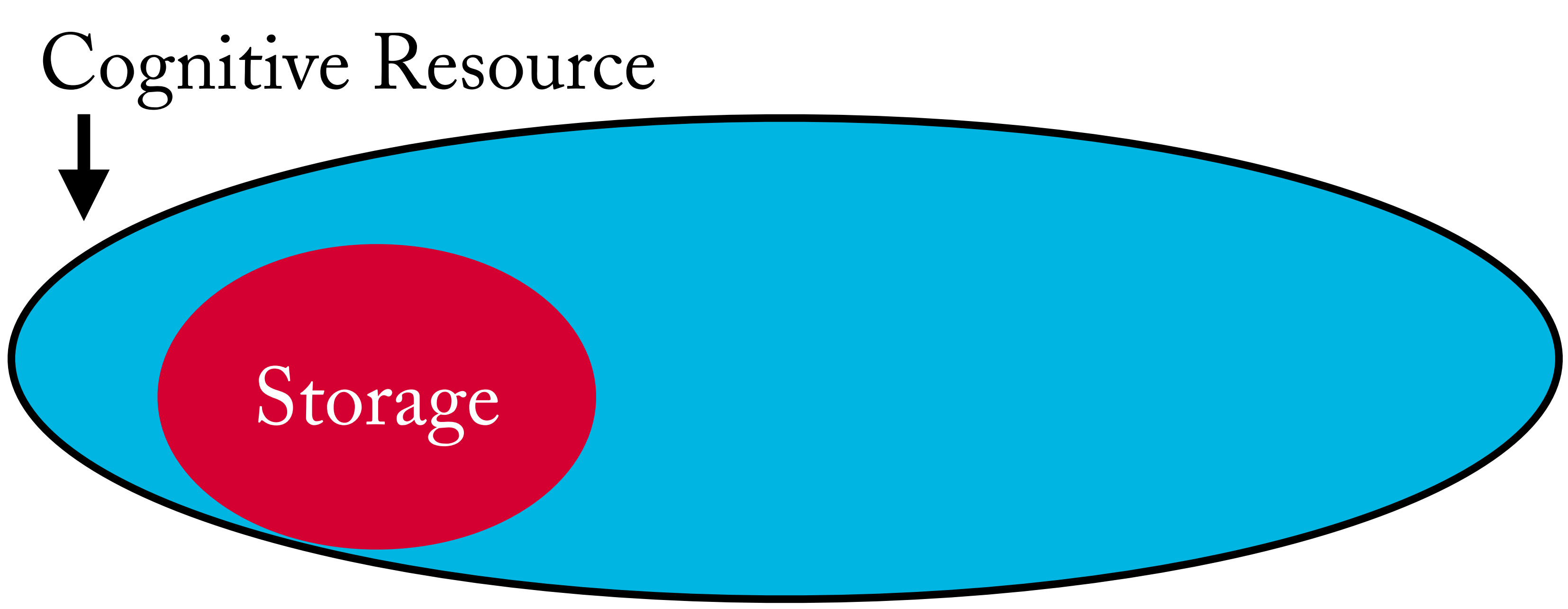
- Since working memory capacity is limited, maintaining these expectations incurs a cognitive cost.

Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the mat

- Since working memory capacity is limited, maintaining these expectations incurs a cognitive cost.
  - Accumulated predictions consume memory resources and create a processing bottleneck [Just & Carpenter 1992]

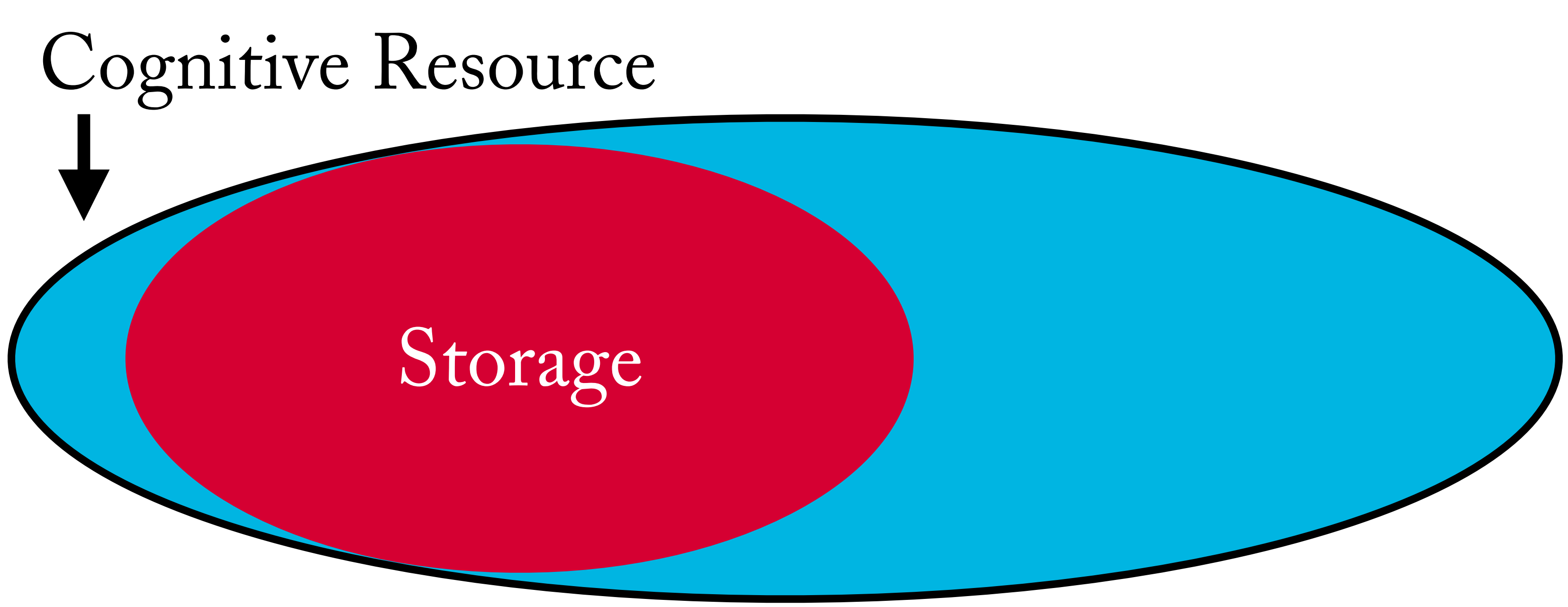


Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the mat

- Since working memory capacity is limited, maintaining these expectations incurs a cognitive cost.
  - Accumulated predictions consume memory resources and create a processing bottleneck [Just & Carpenter 1992]

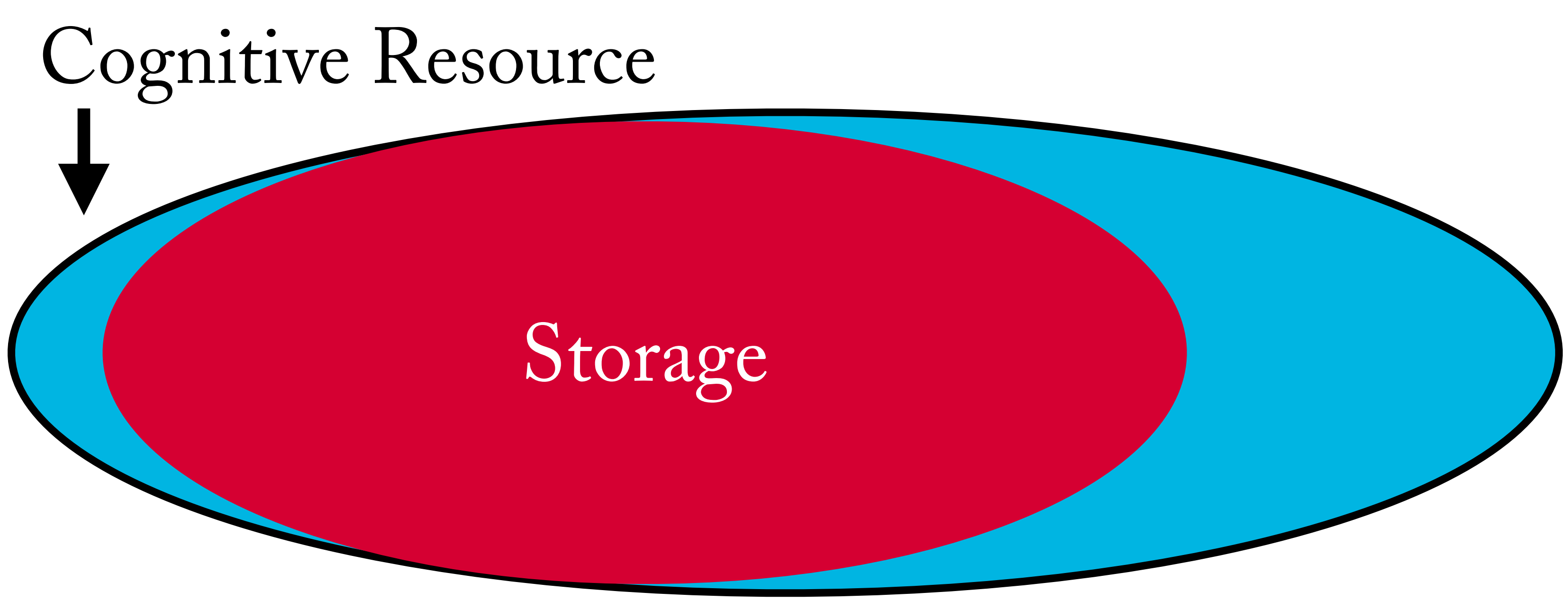


Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the mat

- Since working memory capacity is limited, maintaining these expectations incurs a cognitive cost.
  - Accumulated predictions consume memory resources and create a processing bottleneck [Just & Carpenter 1992]

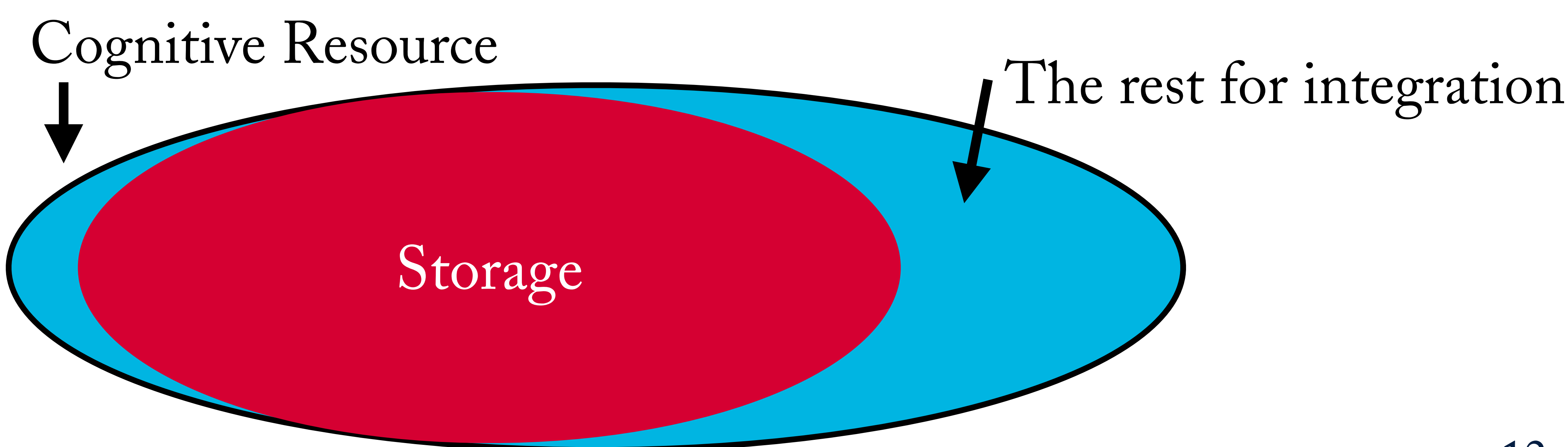


Storage Cost during Comprehension

- Comprehenders maintain incoming words to anticipate future input.

The cat sat on the mat

- Since working memory capacity is limited, maintaining these expectations incurs a cognitive cost.
  - Accumulated predictions consume memory resources and create a processing bottleneck [Just & Carpenter 1992]



Storage Cost during Comprehension

- E.g., center-embedding sentences are hard to process
  [Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

Center-embedded structure:

The reporter [who the senator [who Mary met] attacked] ignored the president

Storage Cost during Comprehension

- E.g., center-embedding sentences are hard to process
  [Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

Center-embedded structure:

The reporter [who the senator [who Mary met] attacked] ignored the president

Right-branching structure:

Mary met the senator [who attacked the reporter [who ignored the president]]

Storage Cost during Comprehension

- E.g., center-embedding sentences are hard to process
  [Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

Center-embedded structure:

The reporter [who the senator [who Mary met] attacked] ignored the president

Right-branching structure:

Mary met the senator [who attacked the reporter [who ignored the president]]

🤔 How much do we memorize the contextual information?

Storage Cost during Comprehension

- E.g., center-embedding sentences are hard to process
  [Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

Center-embedded structure:

The reporter [who the senator [who Mary met] attacked] ignored the president

Right-branching structure:

Mary met the senator [who attacked the reporter [who ignored the president]]

🤔 How much do we memorize the contextual information?
How do we quantify the information in working memory?

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
    - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].

(👀Upcoming talk later today)

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].
  (👀Upcoming talk later today)
- Limitations of grammar-based storage cost metrics:

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

---

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].
  (👀Upcoming talk later today)
- Limitations of grammar-based storage cost metrics:
  1. **Theory-Dependent**: Relies on specific syntactic theories/assumptions.

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].
  (👀Upcoming talk later today)
- Limitations of grammar-based storage cost metrics:
  1. **Theory-Dependent**: Relies on specific syntactic theories/assumptions.
  2. **Fixed-Cost**: Assumes all syntactic elements consume the same cost.

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].

(👀Upcoming talk later today)

- Limitations of grammar-based storage cost metrics:
  1. **Theory-Dependent**: Relies on specific syntactic theories/assumptions.
  2. **Fixed-Cost**: Assumes all syntactic elements consume the same cost.
  3. **Deterministic**: Insensitive to temporal structural ambiguity.

Grammar-based Storage Cost

- Dependency Locality Theory (DLT; [Gibson 2000])
  - Defines storage costs as **the number of predicted syntactic heads** required to complete the current input as a grammatical sentence.

5 storage cost

The reporter who the senator who **Mary** met *t* attacked *t* ignored the president

1 storage cost

Mary met the senator who *t* attacked the reporter who *t* ignored the president

- Empirically supported by controlled experiments [e.g., Chen et al. 2005; Nakatani & Gibson 2010] and naturalistic reading [Isono et al. 2025; Isono and Kajikawa 2026].
(👀Upcoming talk later today)
- Limitations of grammar-based storage cost metrics:
  1. **Theory-Dependent**: Relies on specific syntactic theories/assumptions.
  2. **Fixed-Cost**: Assumes all syntactic elements consume the same cost.
  3. **Deterministic**: Insensitive to temporal structural ambiguity.

➡ We address all of them by the *information-theoretic* reformulation.

Proposal

From Syntactic to Information Storage

- Grammar-based storage cost:
    - Counts the number of predicted syntactic elements.

From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.

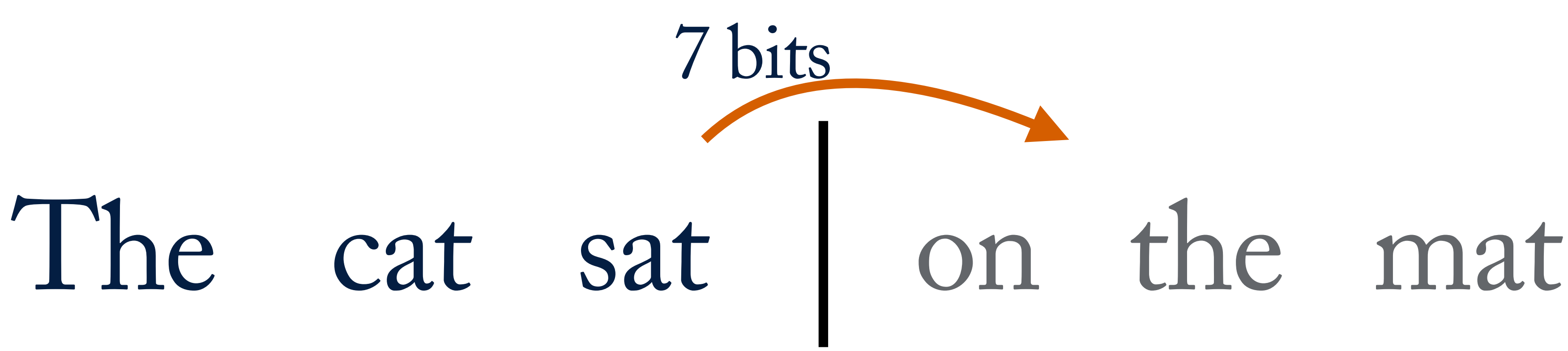
From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.

The cat sat | on the mat

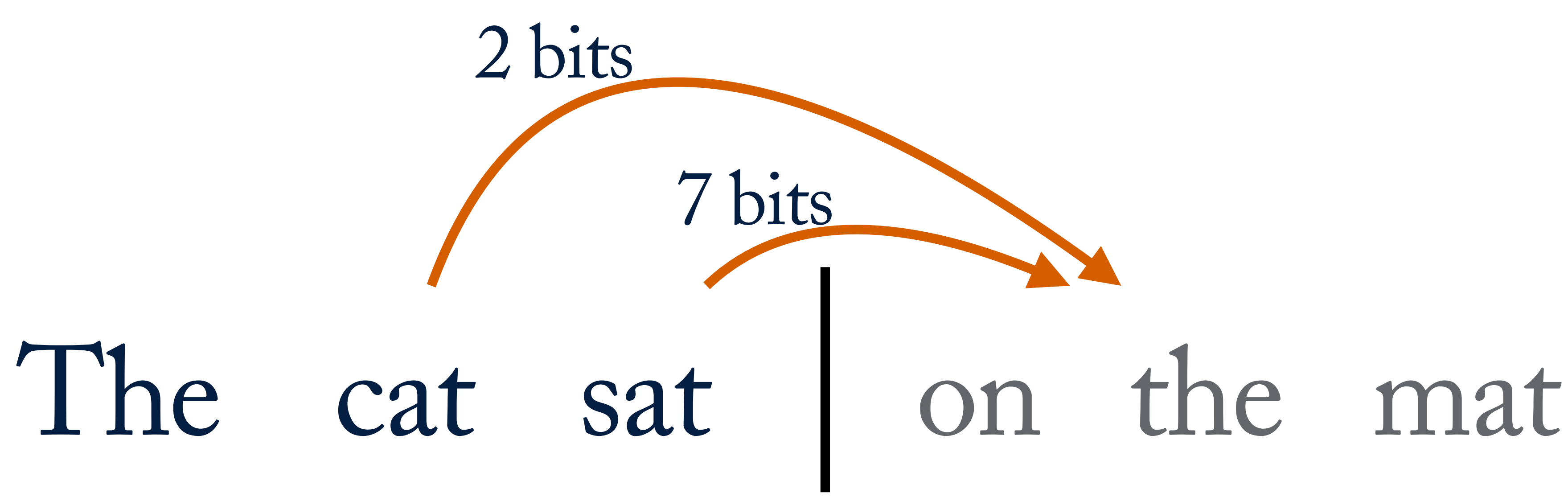
From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.



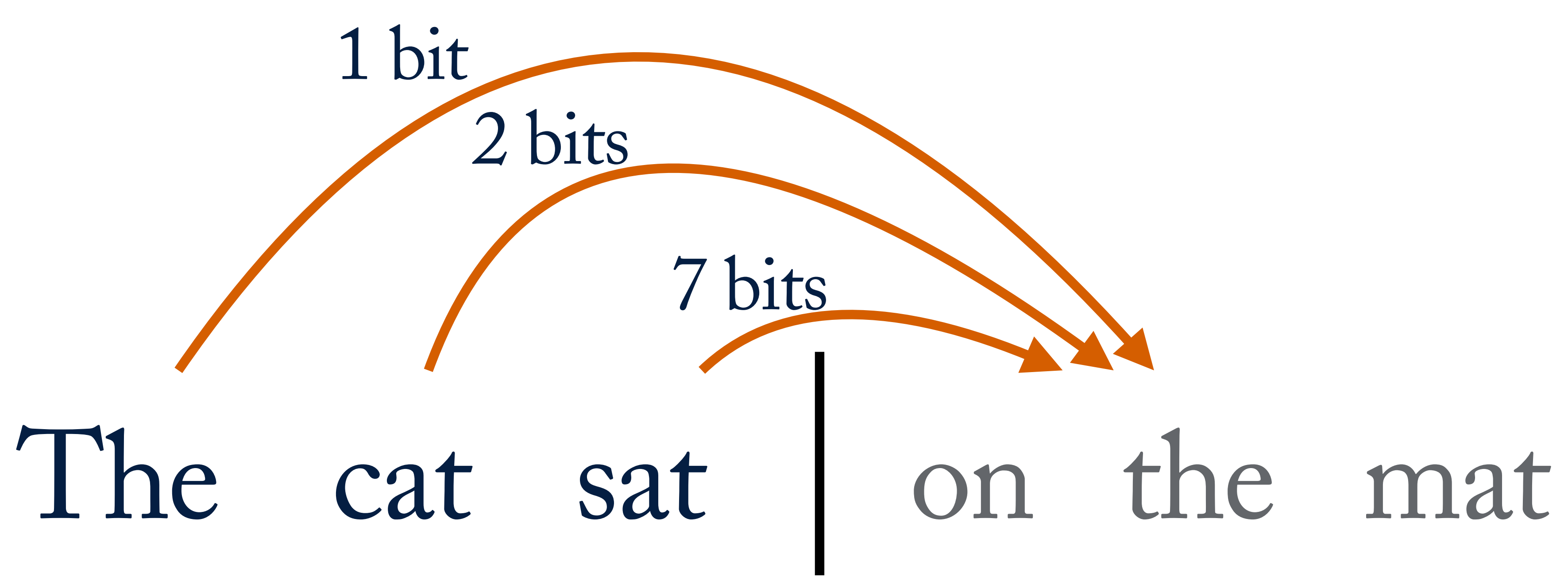
From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.



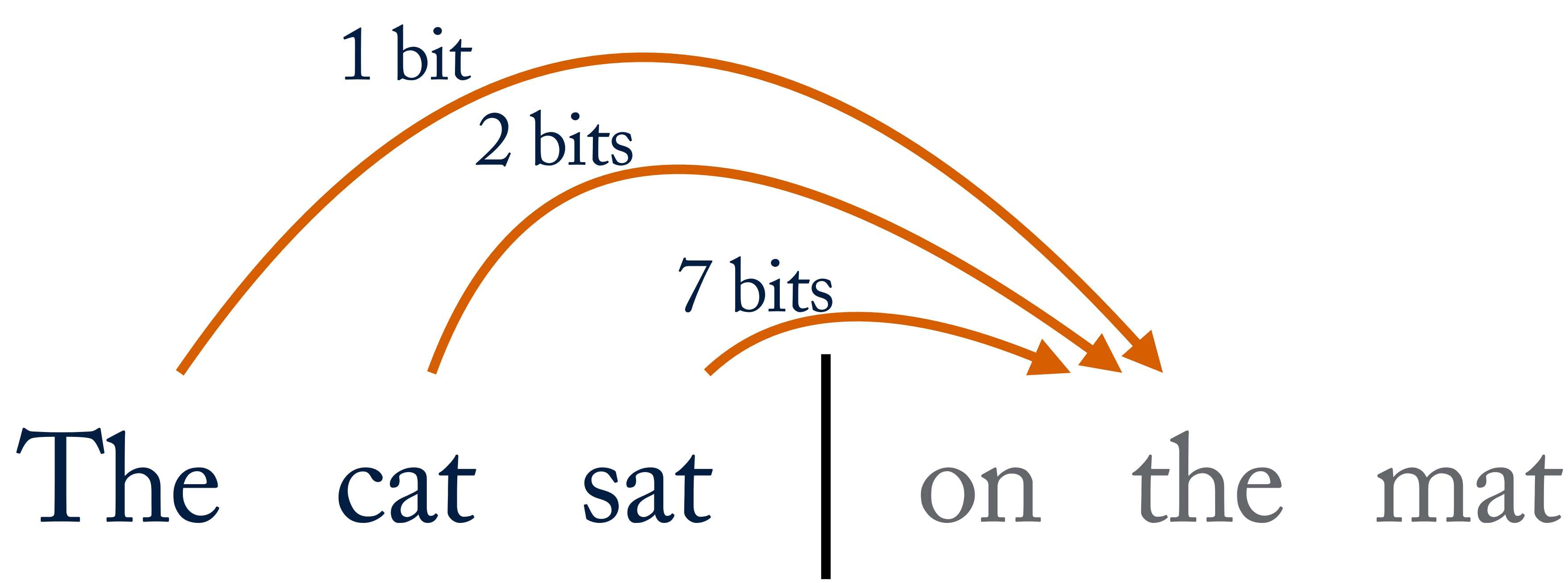
From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.



From Syntactic to Information Storage

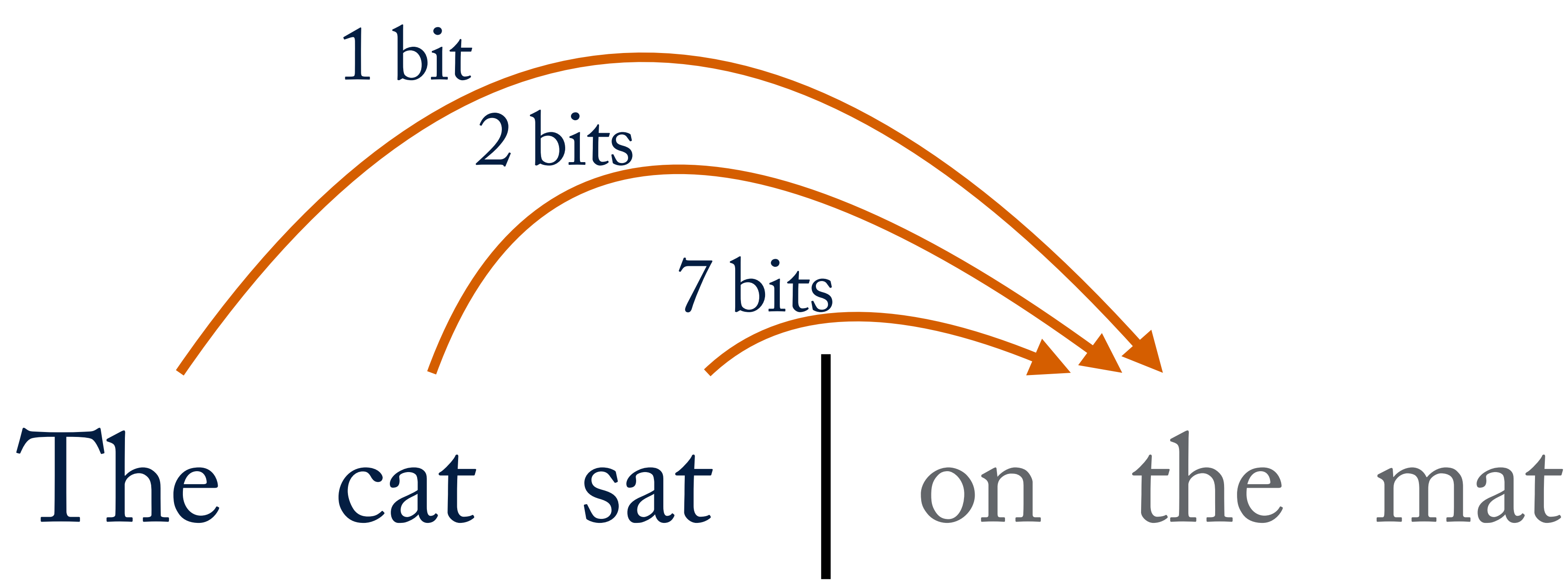
- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.



- Information storage at **on** is:

From Syntactic to Information Storage

- Grammar-based storage cost:
  - Counts the number of predicted syntactic elements.
- Our proposal:
  - Quantifies the amount of information content that the words in context carry about future words.



- Information storage at **on** is: 1 + 2 + 7 = 10 bits

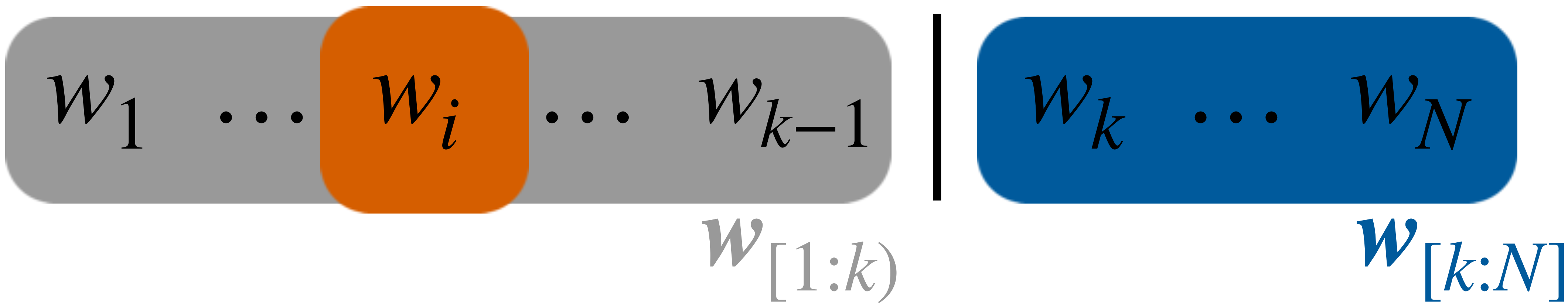
Information-theoretic Storage

More formally …

$$w_1 \ \dots \ w_i \ \dots \ w_{k-1} \ \Big| \ w_k \ \dots \ w_N$$

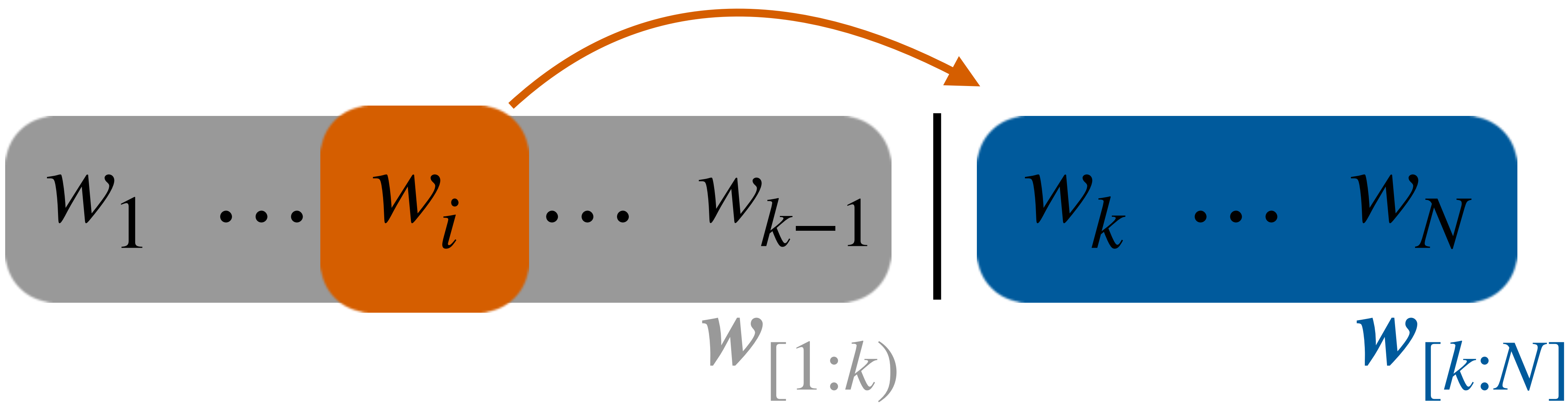
$$\boldsymbol{w}_{[1:k)} \qquad \boldsymbol{w}_{[k:N]}$$

Information-theoretic Storage



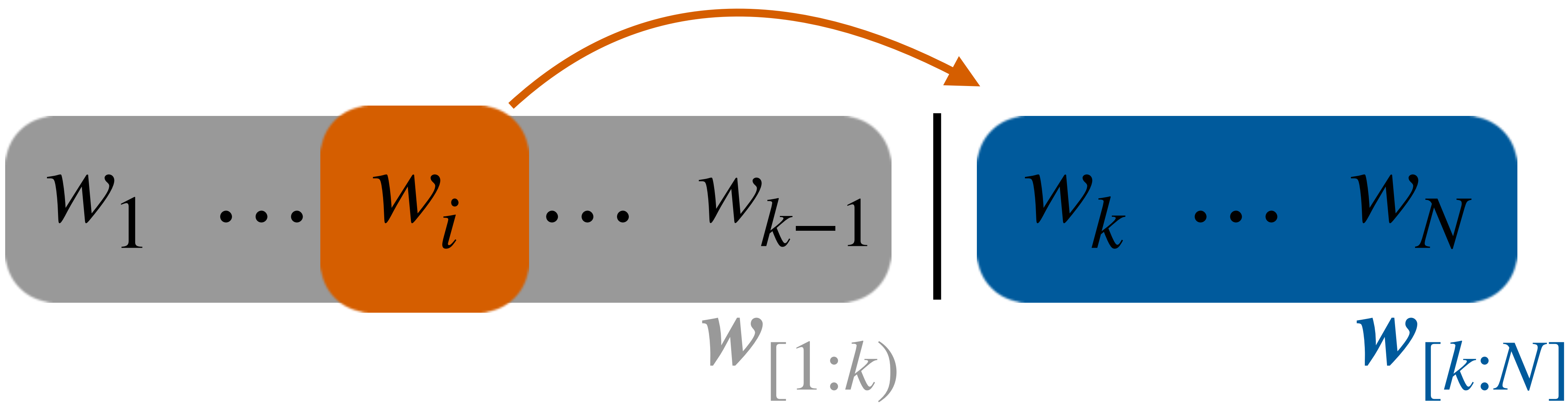
Information-theoretic Storage

$$\mathrm{pmi}(w_i \,;\, \boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})$$



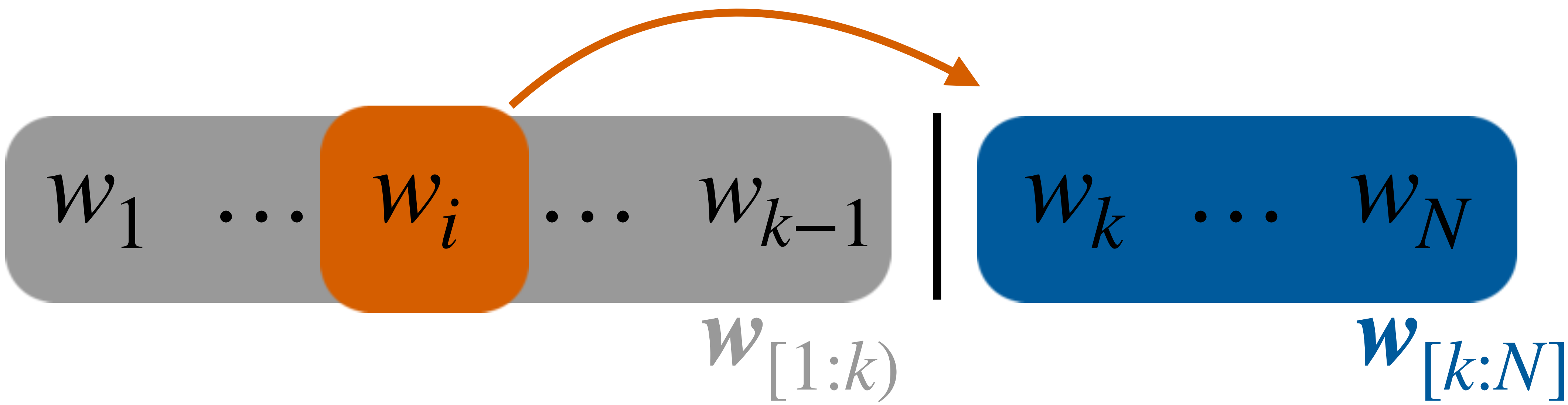
Information-theoretic Storage

$$\mathbb{E}_{\boldsymbol{w}_{[k:N]}|\boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i\,;\,\boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$



Information-theoretic Storage

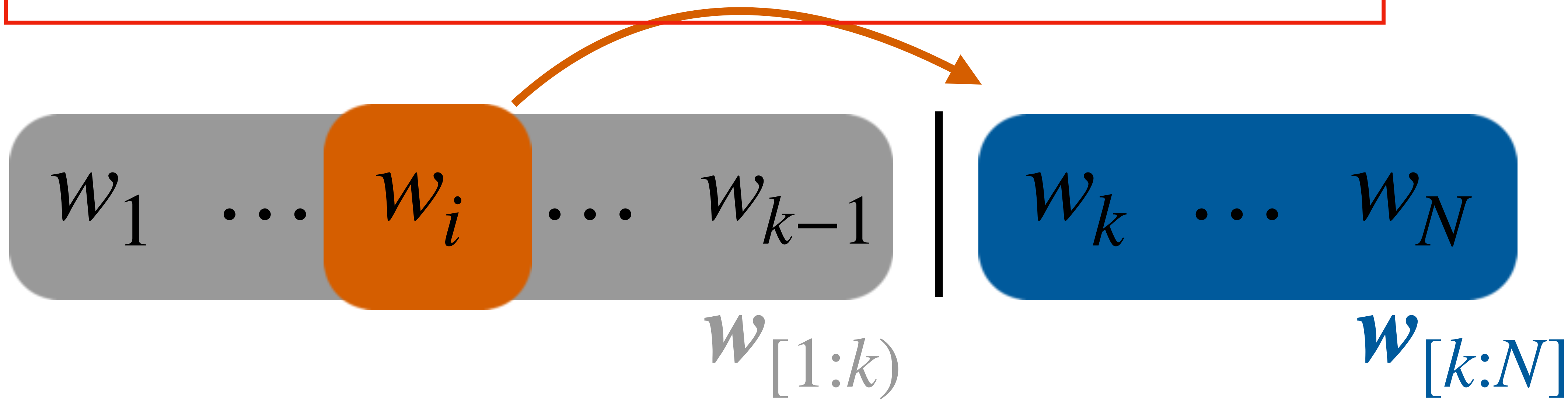
$$\mathbb{E}_{\boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i \, ; \, \boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$



$$\mathrm{InfoStor}_k := \sum_{i=1}^{k-1} \mathbb{E}_{\boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i ; \boldsymbol{w}_{[k;N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$

Information-theoretic Storage

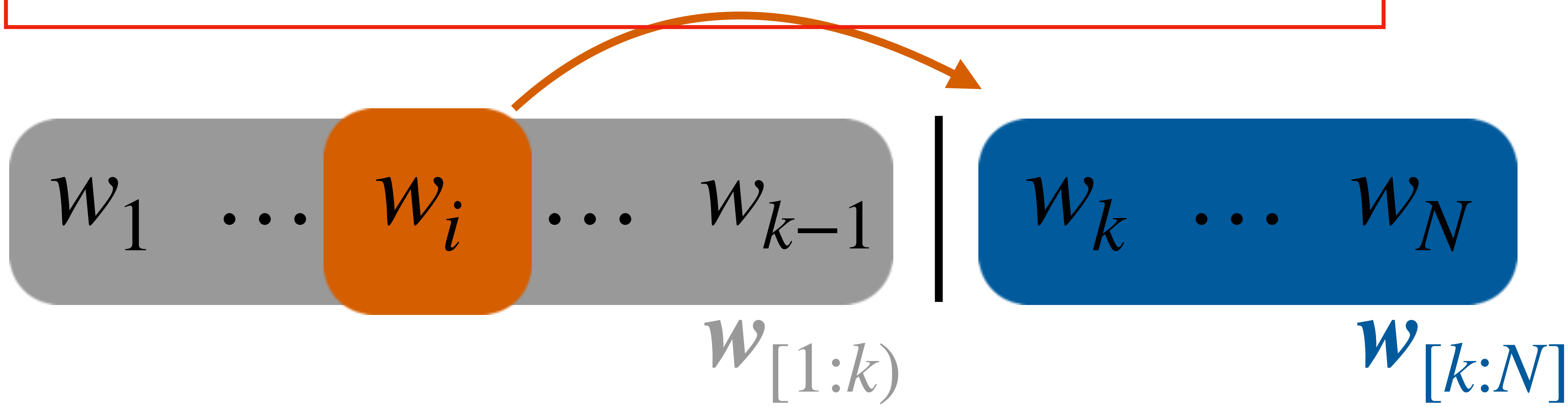
$$\mathbb{E}_{\boldsymbol{w}_{[k:N]}|\boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i \,;\, \boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$



$$\mathrm{InfoStor}_k := \sum_{i=1}^{k-1} \mathbb{E}_{\boldsymbol{w}_{[k:N]}|w_{[1:k)}}\left[\mathrm{pmi}(w_i; \boldsymbol{w}_{[k;N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$

Information-theoretic Storage

$$\mathbb{E}_{\boldsymbol{w}_{[k:N]}|\boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i\, ;\, \boldsymbol{w}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$



$$\mathrm{InfoStor}_k := \sum_{i=1}^{k-1} \mathbb{E}_{\boldsymbol{w}_{[k:N]}|\boldsymbol{w}_{[1:k)}}\left[\mathrm{pmi}(w_i; \boldsymbol{w}_{[k;N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$

- We estimated the expected PMI via BERT [Devlin+ 2019; Hoover+ 2021].

👀Please refer to our paper for the estimation details and the results of other BERT-family models.

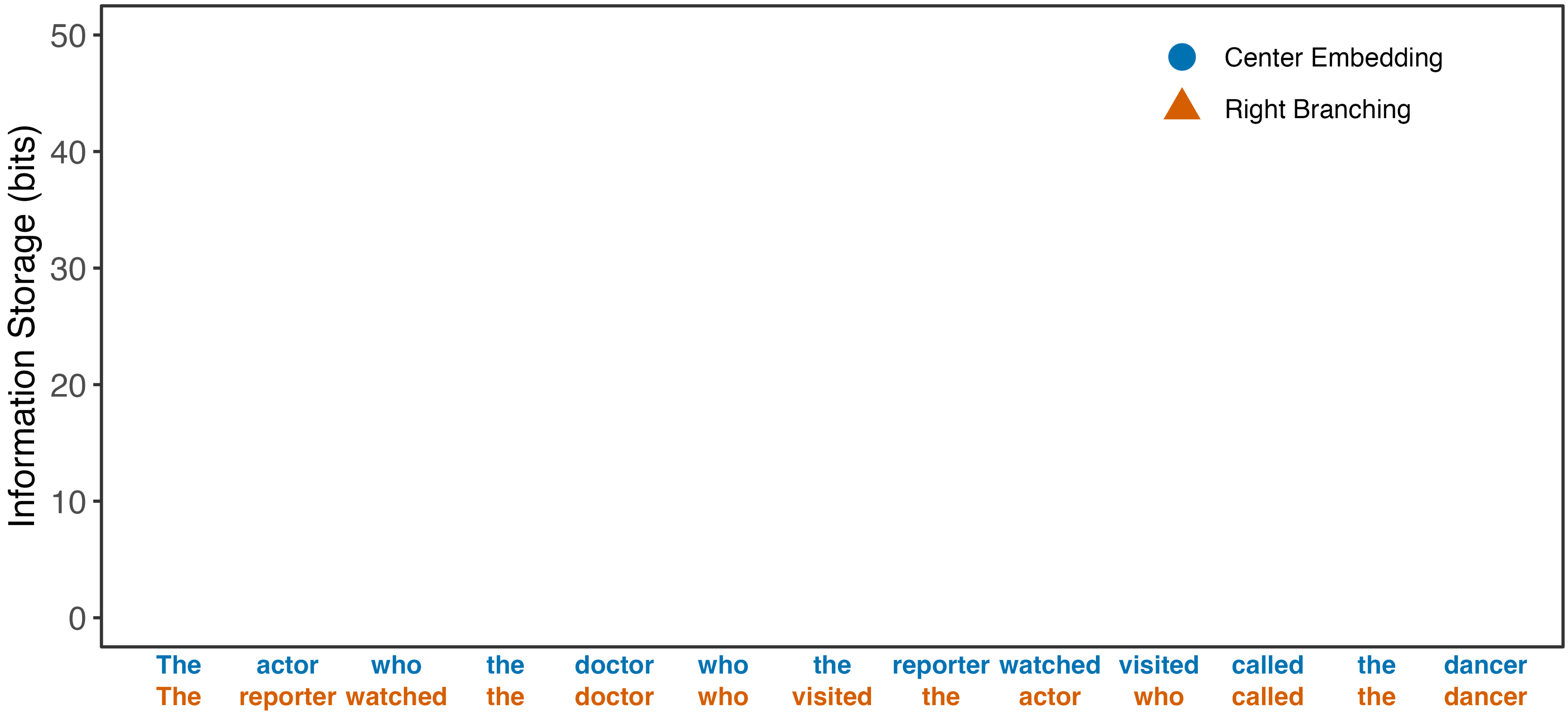
Experiments

1. Case Study: Center-embedding

- Processing difficulty: center-embedded > right-branching

1. Case Study: Center-embedding

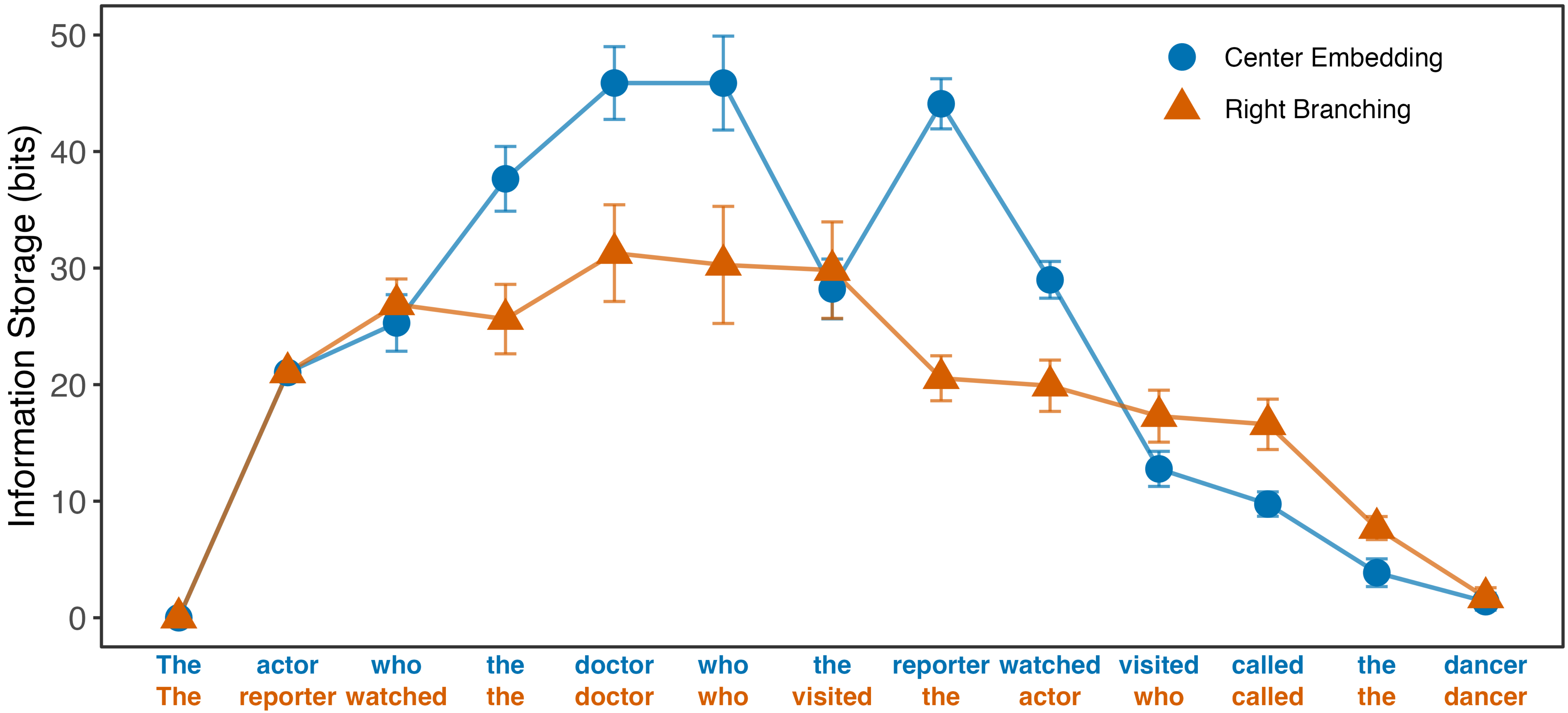
- Processing difficulty: center-embedded > right-branching
  - We randomly generated these sentences (30 for each).
  - Estimate word-by-word information storage.



👀See our paper on relative clause assymetry!!

1. Case Study: Center-embedding

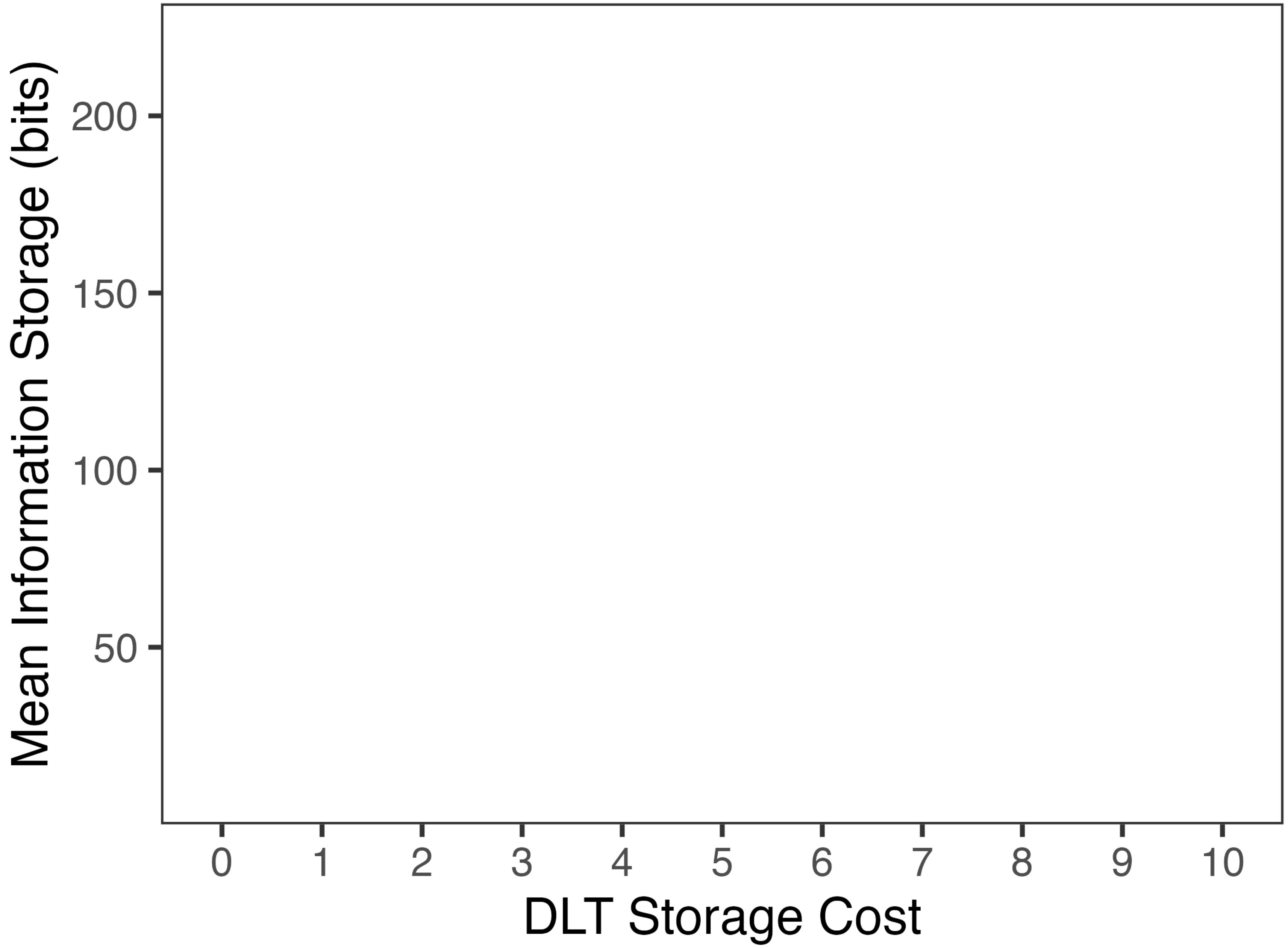
- Processing difficulty: center-embedded > right-branching
  - We randomly generated these sentences (30 for each).
  - Estimate word-by-word information storage.



👀See our paper on relative clause assymetry!!

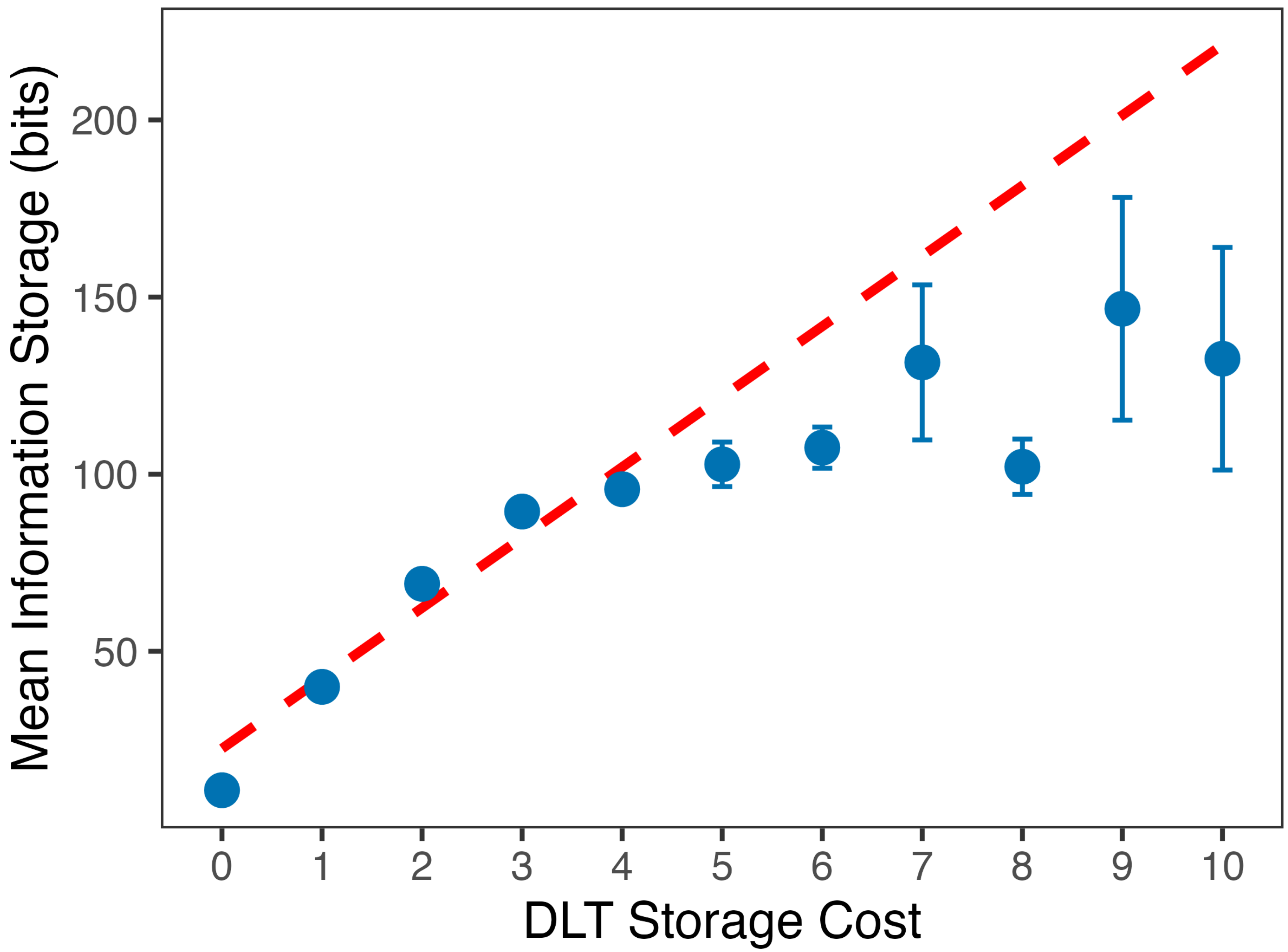
2. Correlation with DLT

- Is our Information Storage correlated with DLT storage?
  - Test in UD_English-GUM [Zeldes 2017] (233,926 words)



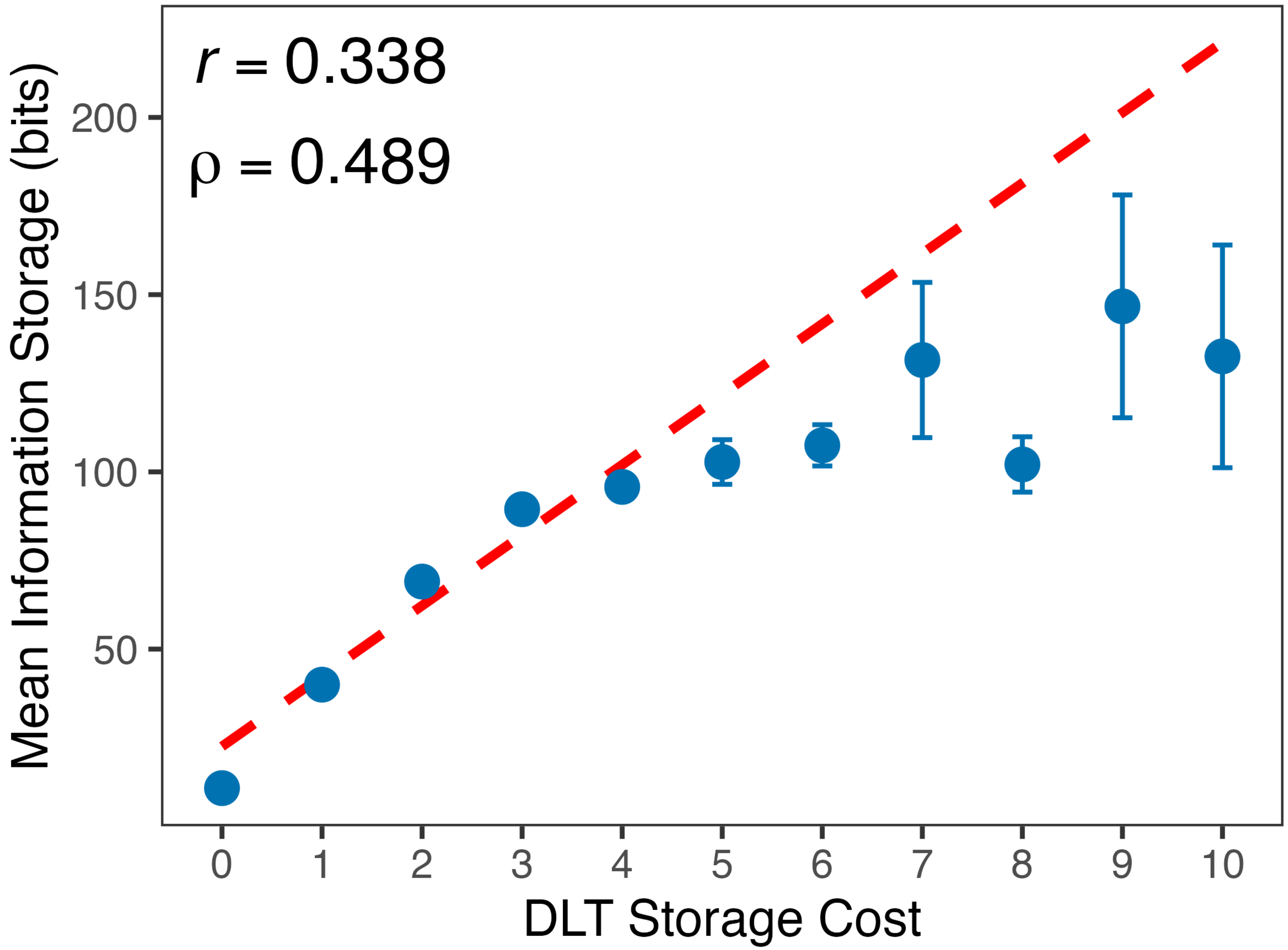
2. Correlation with DLT

- Is our Information Storage correlated with DLT storage?
  - Test in UD_English-GUM [Zeldes 2017] (233,926 words)



2. Correlation with DLT

- Is our Information Storage correlated with DLT storage?
  - Test in UD_English-GUM [Zeldes 2017] (233,926 words)



3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

from GPT-2 ↙

Baseline: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}\,(\mathrm{position}(w) + \mathrm{len}(w) + \mathrm{freq}(w) + \mathrm{surp}(w))$

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

Target: $RT(w) \sim f_{\text{linear}}(\,\textbf{storage}(w) + \text{position}(w) + \text{len}(w) + \text{freq}(w) + \text{surp}(w)\,)$

from GPT-2

Baseline: $RT(w) \sim f_{\text{linear}}(\text{position}(w) + \text{len}(w) + \text{freq}(w) + \text{surp}(w)\,)$

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

Target: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}$ (**storage**(w) + position(w) + len(w) + freq(w) + surp(w))

Δ_{LL} = loglik(**target**) - loglik(**baseline**)

from GPT-2

Baseline: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}$ (position(w) + len(w) + freq(w) + surp(w))

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

Target: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}$ (**storage**(w) + position(w) + len(w) + freq(w) + surp(w))

Δ_{LL} = loglik(**target**) - loglik(**baseline**)

from GPT-2

Baseline: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}$ (position(w) + len(w) + freq(w) + surp(w))

  - 10-fold cross-validation for estimating Δ_{LL} on held-out test data

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

Target: $RT(w) \sim f_{\text{linear}}$ (**storage**(w) + position(w) + len(w) + freq(w) + surp(w))

Δ_{LL} = loglik(**target**) - loglik(**baseline**)

from GPT-2

Baseline: $RT(w) \sim f_{\text{linear}}$ (position(w) + len(w) + freq(w) + surp(w))

- 10-fold cross-validation for estimating Δ_{LL} on held-out test data
- Δ_{LL} is greater than zero => storage predicts reading time

3. Naturalistic Reading-time Analysis

- Can Information Storage predict naturalistic reading time?
  - Does it explain reading-time variance above and beyond DLT storage?
    - We use Universal Dependencies parses for computing DLT storage.

Target: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}\,(\, \mathbf{storage}(w) + \mathrm{position}(w) + \mathrm{len}(w) + \mathrm{freq}(w) + \mathrm{surp}(w)\,)$

$$\Delta_{LL} = \mathrm{loglik}(\mathbf{target}) - \mathrm{loglik}(\mathbf{baseline})$$

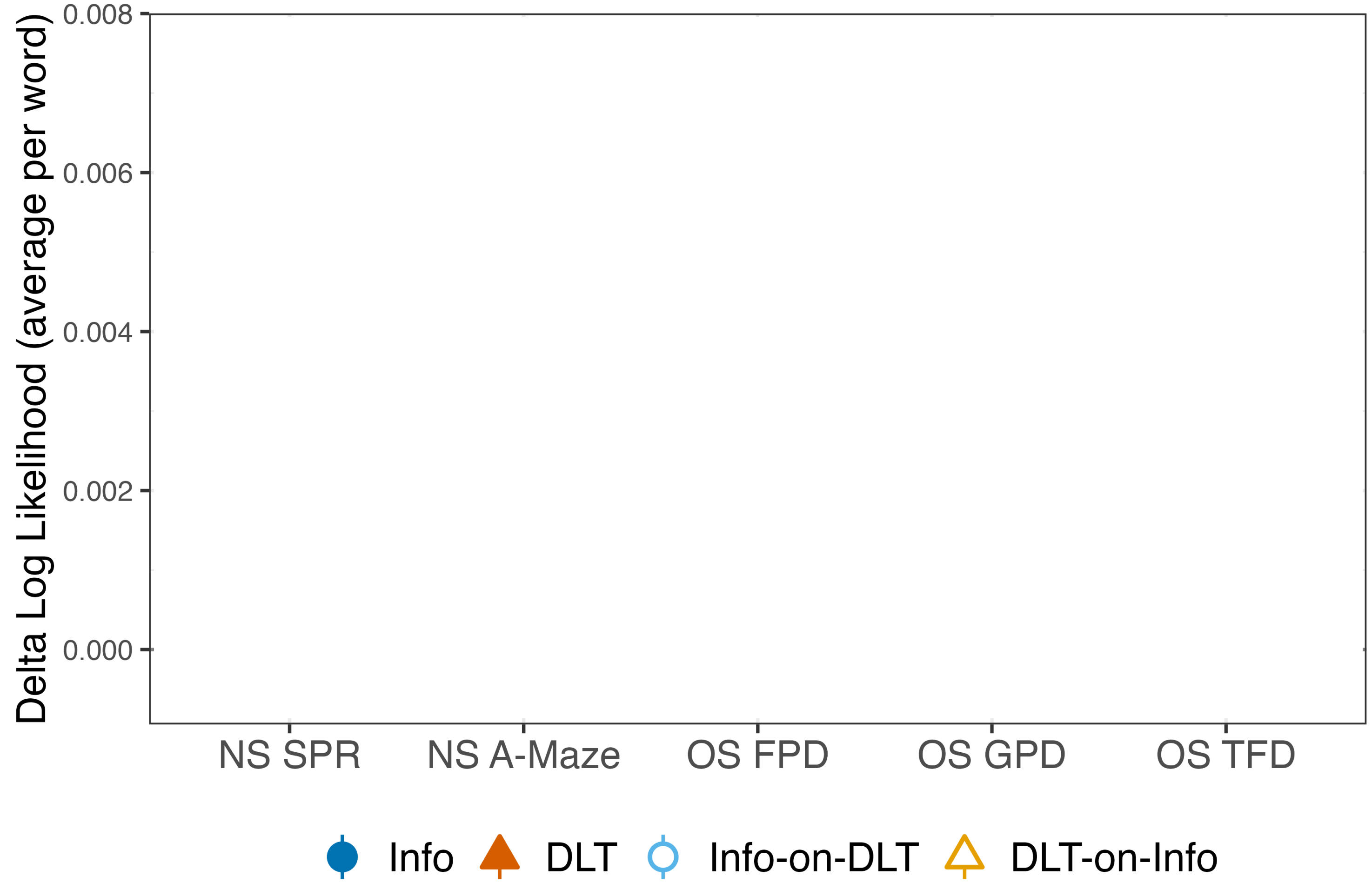
from GPT-2

Baseline: $\mathrm{RT}(w) \sim f_{\mathrm{linear}}\,(\mathrm{position}(w) + \mathrm{len}(w) + \mathrm{freq}(w) + \mathrm{surp}(w)\,)$

- 10-fold cross-validation for estimating Δ_{LL} on held-out test data
- Δ_{LL} is greater than zero => storage predicts reading time

- Test on the large-scale English datasets (<#subject>)
  - *Natural Stories* [Futrell+ 2021]
    - Self-paced reading (181) [Futrell+ 2021]
    - A-maze reading (95) [Boyce & Levy 2023]
  - *OneStop* [Berzak+ 2025]
    - Eye-tracking (180)

3. Naturalistic Reading-time Analysis



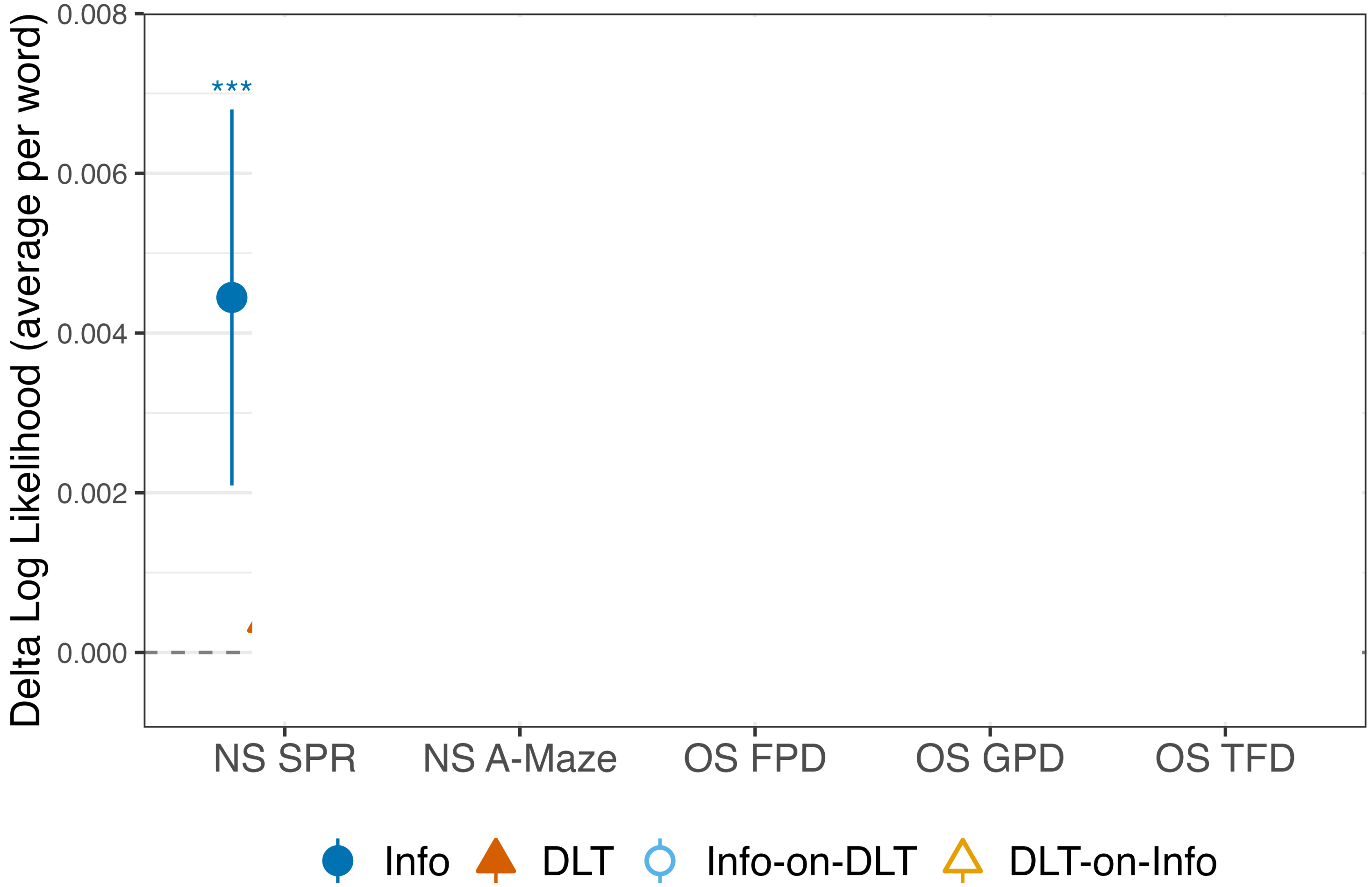
Just information storage cost

How much *more* reading-time variance is explained above DLT

NS=Natural Stories; OS=OneStop
SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

👀See our paper on the **coefficients**!!

3. Naturalistic Reading-time Analysis



Just information storage cost

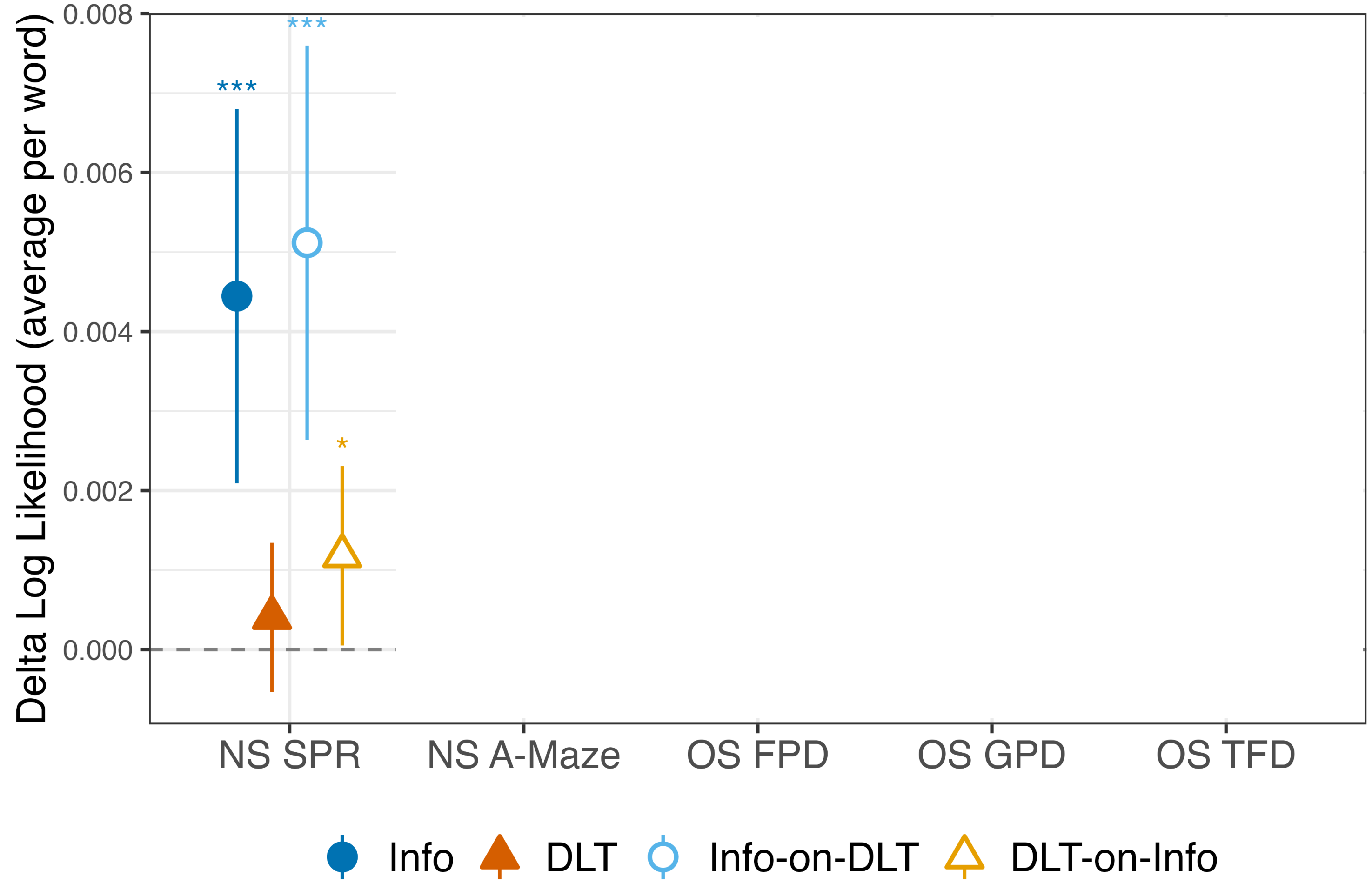
How much *more* reading-time variance is explained above DLT

NS=Natural Stories; OS=OneStop
SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

- Information storage significantly predicts human reading time.

👀See our paper on the **coefficients**!!

3. Naturalistic Reading-time Analysis

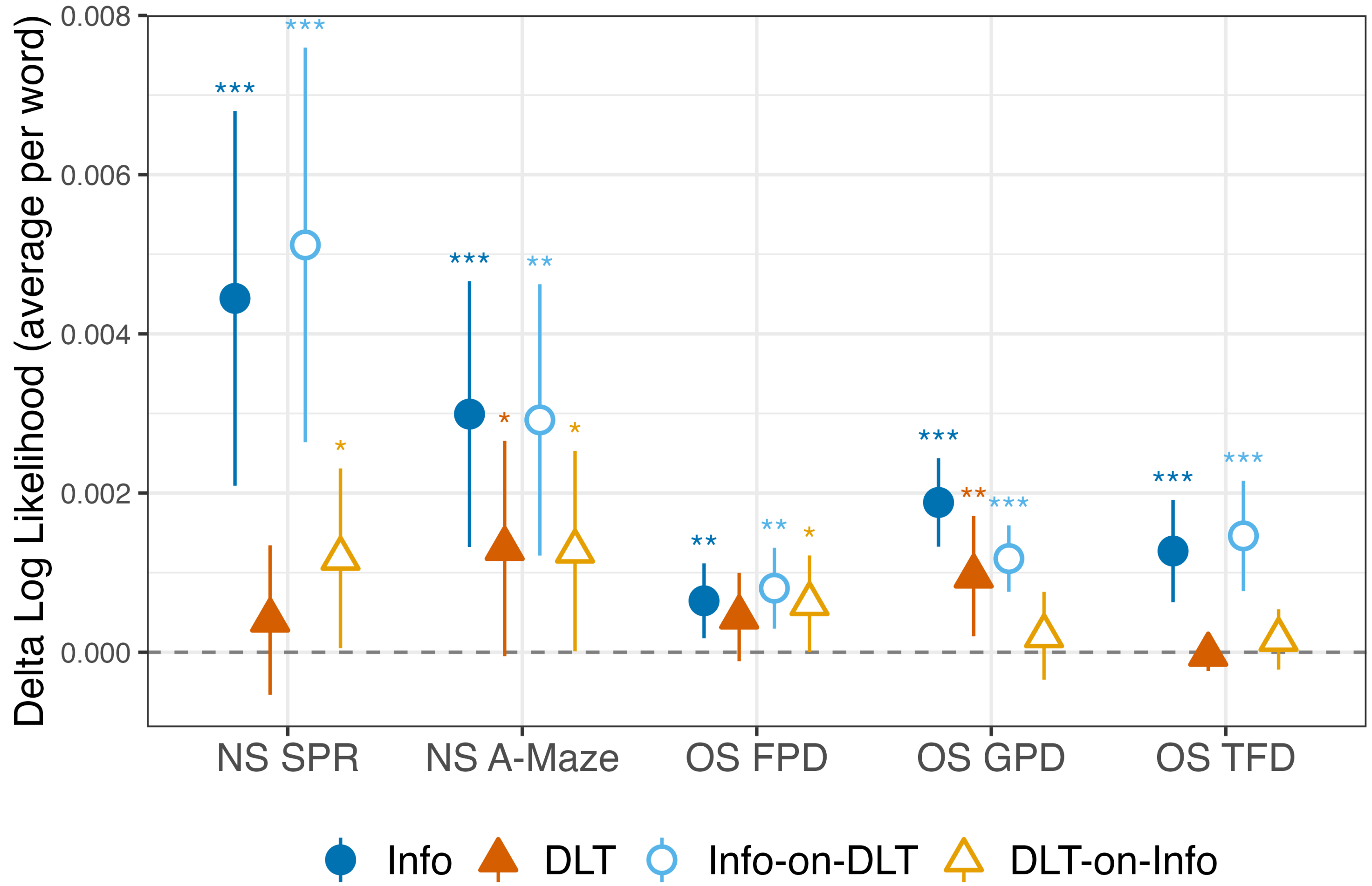


NS=Natural Stories; OS=OneStop
SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

- Information storage significantly predicts human reading time.
- But it does not explain all of the variance that DLT captures.

👀See our paper on the **coefficients**!!

3. Naturalistic Reading-time Analysis



Just information storage cost

How much *more* reading-time variance is explained above DLT

NS=Natural Stories; OS=OneStop
SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

- Information storage significantly predicts human reading time.
- But it does not explain all of the variance that DLT captures.

👀See our paper on the **coefficients**!!

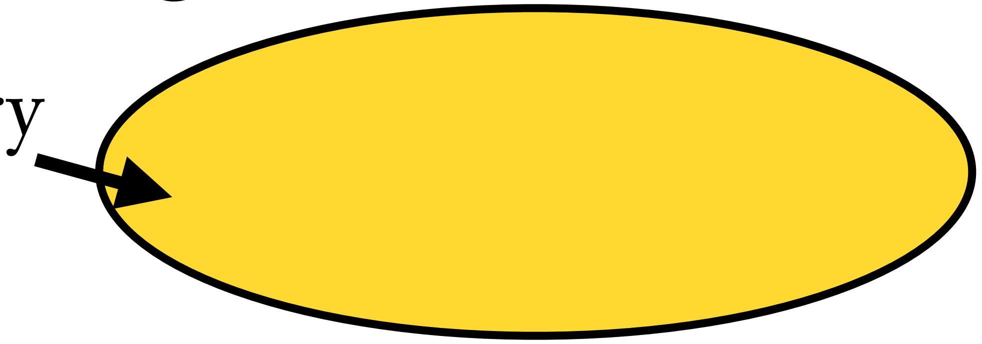
Discussion & Implications

- Operationalized storage cost as (half-pointwise) mutual information between past and future.

Discussion & Implications

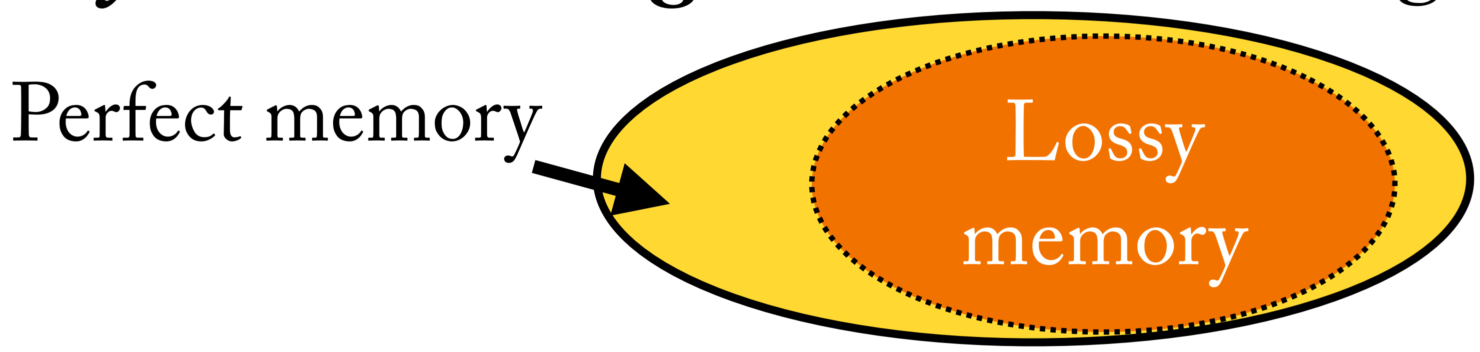
- Operationalized storage cost as (half-pointwise) mutual information between past and future.
  - Our framework provides a formal foundation to explore ***lossy-context storage*** under limited cognitive resources.

Perfect memory



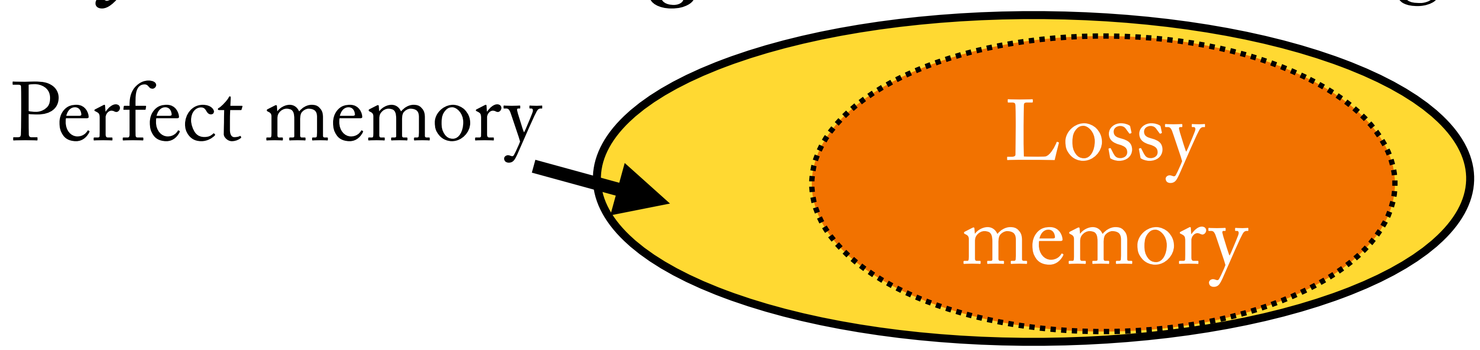
Discussion & Implications

- Operationalized storage cost as (half-pointwise) mutual information between past and future.
  - Our framework provides a formal foundation to explore ***lossy-context storage*** under limited cognitive resources.



Discussion & Implications

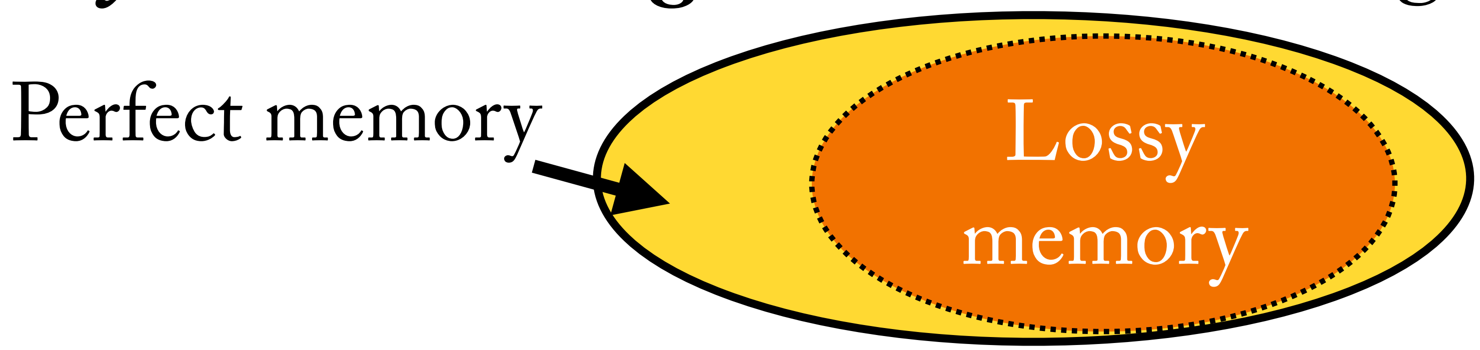
- Operationalized storage cost as (half-pointwise) mutual information between past and future.
  - Our framework provides a formal foundation to explore ***lossy-context storage*** under limited cognitive resources.



- Information/DLT storage capture distinct sources of reading-time variance.

Discussion & Implications

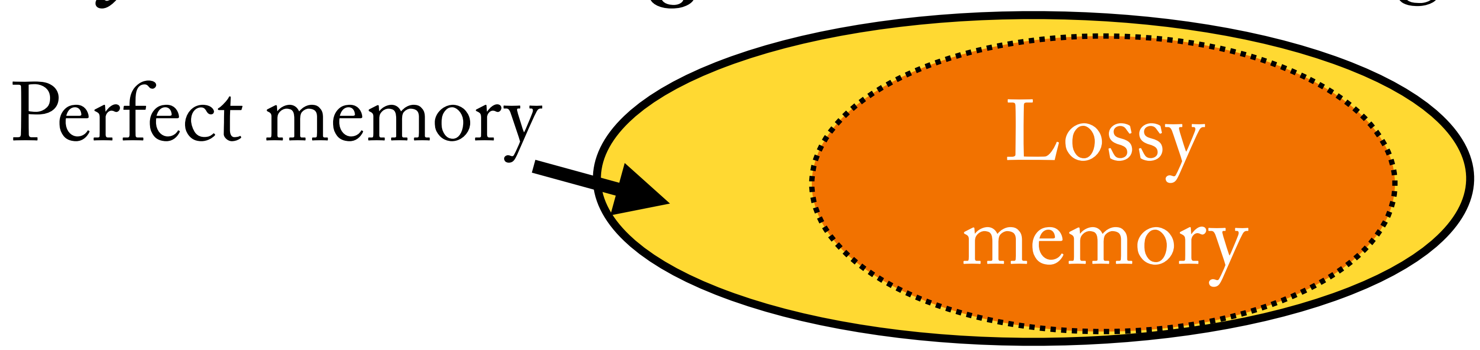
- Operationalized storage cost as (half-pointwise) mutual information between past and future.
  - Our framework provides a formal foundation to explore ***lossy-context storage*** under limited cognitive resources.



- Information/DLT storage capture distinct sources of reading-time variance.
  - **Statistical** prediction and **structural** processing operate as distinct, complementary mechanisms [Kajikawa & Isono 2026].

Discussion & Implications

- Operationalized storage cost as (half-pointwise) mutual information between past and future.
  - Our framework provides a formal foundation to explore ***lossy-context storage*** under limited cognitive resources.



- Information/DLT storage capture distinct sources of reading-time variance.
  - **Statistical** prediction and **structural** processing operate as distinct, complementary mechanisms [Kajikawa & Isono 2026].
  - It might be better to assume **structured representation** rather than token sequences.

Thank you!!

- We proposed an ***information-theoretic storage cost***
  - estimated from a neural language model without syntactic annotation
  - recovers well-known processing asymmetries to specific constructions
  - correlates with DLT storage
  - predicts naturalistic reading times

↑Our paper 👀

Appendix

Estimation using BERT

$$\mathbb{E}_{\boldsymbol{w}_{[k:N]}|w_{[1:k)}}\left[\mathrm{pmi}(w_i; \boldsymbol{w}_{[k;N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right]$$

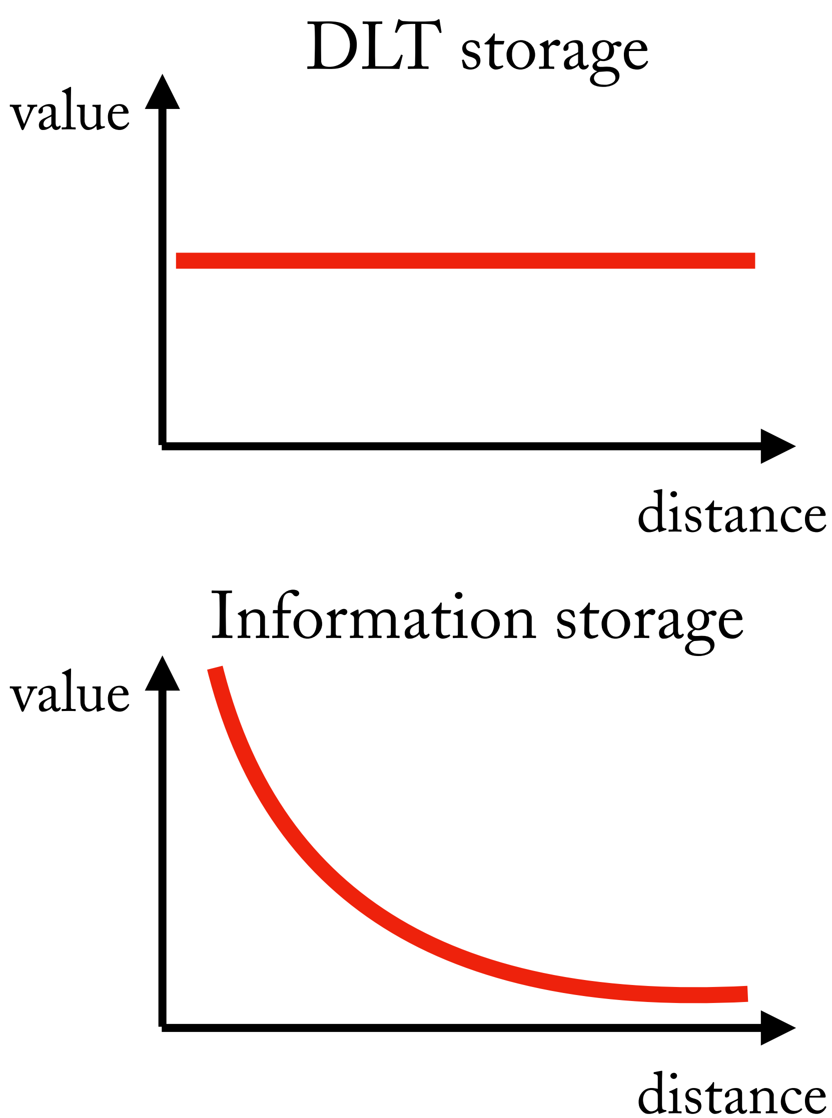
$$= \mathrm{KL}\left(p(\boldsymbol{W}_{[k:N]} \mid \boldsymbol{w}_{[1:k)}) \parallel p(\boldsymbol{W}_{[k:N]} \mid \boldsymbol{w}_{[1:k)\setminus i})\right)$$

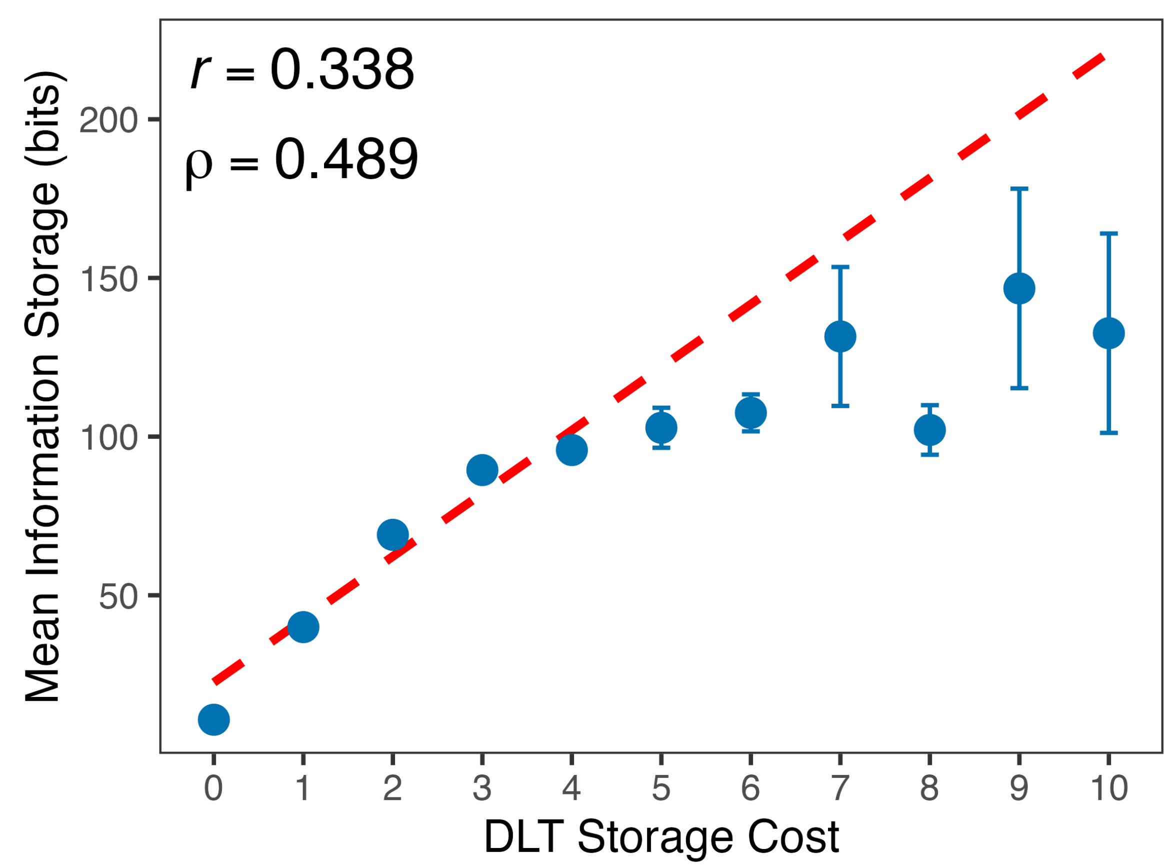
$$\mathbb{E}_W[\mathrm{pmi}(\mathrm{cat}; W \mid \mathrm{The})] = \mathrm{KL}(\ \mathrm{p}(\texttt{[M]} \mid \mathrm{The\ cat}) \mid\mid \mathrm{p}(\texttt{[M]} \mid \mathrm{The}\ \texttt{[M]})\ )$$

👀Please refer to our paper for the estimation details
and the results of other BERT-family models

Why sublinear?

- The sub-linearity likely reflects a difference between the two metrics:
  - DLT storage increases *linearly* as it counts the number of unseen tokens
  - Information storage may not do so because mutual information in natural language decays according to a *power law* over distance [e.g., Debowski 2015; Hahn+ 2021]





Correlations between Predictors

- Pearson correlation matrix of predictors in Natural Stories

	position	wlen	wlen_so	unisurp	unisurp_so	gpt2_surp	gpt2_surp_so	dlt_stor	dlt_stor_so	info_stor	info_stor_so
zone	-0.01	0.01	0.00	0.01	0.00	-0.02	-0.02	0.00	0.00	-0.09	-0.09
position		0.00	0.01	0.00	-0.03	-0.02	-0.04	-0.10	-0.03	-0.12	-0.01
wlen			-0.06	0.67	-0.07	0.50	-0.06	0.03	-0.02	0.08	0.07
wlen_so				-0.07	0.67	-0.07	0.48	0.00	0.14	0.06	0.11
unisurp					-0.07	0.53	-0.05	0.09	0.01	0.07	0.04
unisurp_so						-0.04	0.52	-0.01	0.20	0.02	0.06
gpt2_surp							-0.03	0.11	-0.03	0.01	0.01
gpt2_surp_so								0.03	0.16	0.02	-0.02
dlt_stor									0.61	0.23	0.22
dlt_stor_so										0.21	0.22
info_stor											0.85

- zone: word position in a document
- position: word position in a sentence
- wlen: word length
- unisurp: unigram surprisal
- gpt2_surp: GPT-2 surprisal
- dlt_stor: DLT storage cost
- info_stor: Information storage cost
- _so: spillover region