

人間の文処理に対する 情報理論的アプローチ

梶川康平

Ph.D. student, Dept. of Linguistics,
Georgetown University

NINJALサロン

2026/07/14

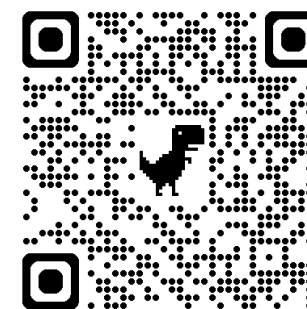


GEORGETOWN UNIVERSITY

自己紹介

- 経歴：

- 2025年8月—：ジョージタウン大学博士課程
(advisor: Prof. Ethan Gotlieb Wilcox)
- 2025年4月—8月：総研大国語研博士課程/DC1
(指導教官：窪田悠介先生)
(← 2023年6月—2025年5月：国語研技術補佐員)



Website

- 研究：

- 計算言語学・心理言語学
 - 文理解を情報理論や統語構造を通して研究しています：
Kajikawa & Isono (2026); Kajikawa, Isono, & Wilcox. (2026)
 - 言語はなぜ今あるような形をしているのか？なども：
Kajikawa et al. (2024; CoNLL Best Paper Award)

宣伝：プレプリント



The Dual Nature of Syntactic Node Count: Facilitating and Inhibiting Sentence Comprehension

Kohei Kajikawa*¹ and Shinnosuke Isono²

¹Department of Linguistics, Georgetown University, 1421 37th Street
NW, Poulton Hall 240, Washington, D.C., 20057-1051, USA.

kk1571@georgetown.edu

²National Institute for Japanese Language and Linguistics, 10-2,
Midoricho, Tachikawa City, Tokyo, 190-8561, Japan.

s-isono@ninjal.ac.jp

Abstract

Neurolinguistic research on syntactic processing has utilized *Node Count*, the number of structure-building steps executed at each word, to investigate how the effort invested in structure building affects brain activity. However, several recent studies cast doubt on the presumed link between Node Count and processing effort; they observe *negative* correlations between Node Count and reading times. Seeking the source of this disagreement, we analyze information-theoretic measures and reading times using three large-scale naturalistic datasets in English across different grammar formalisms: phrase-structure grammar, dependency grammar, and combinatory categorial grammar. This study establishes a dual nature of Node Count: we conclude that it indexes two opposing forces—processing facilitation due to predictability in immediate regions and processing inhibition due to contextual semantic integration in spillover regions. We also demonstrate that Node Count has a unique psycholinguistic value as an index of human prediction, since its facilitatory effect explains reading time variance beyond that captured by a Transformer-based language model. These results highlight the role of structural representations in human language processing and underscore the need for a careful interpretation of the observed correlations between Node Count and brain activity in neurolinguistic studies.

Keywords: Sentence comprehension; Syntactic processing; Node Count; Predictability; Information theory

*Corresponding author.

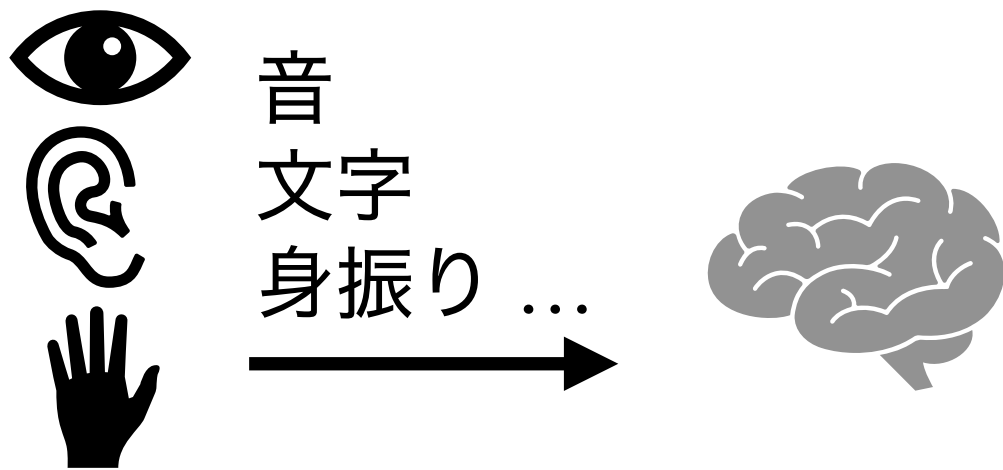
- K. Kajikawa and S. Isono (2026) The Dual Nature of Syntactic Node Count: Facilitating and Inhibiting Sentence Comprehension. *PsyArXiv*. https://osf.io/preprints/psyarxiv/9msby_v2
 - *Keywords:* Sentence comprehension; Syntactic processing; Node Count; Predictability; Information theory

そもそも、文理解における謎とは

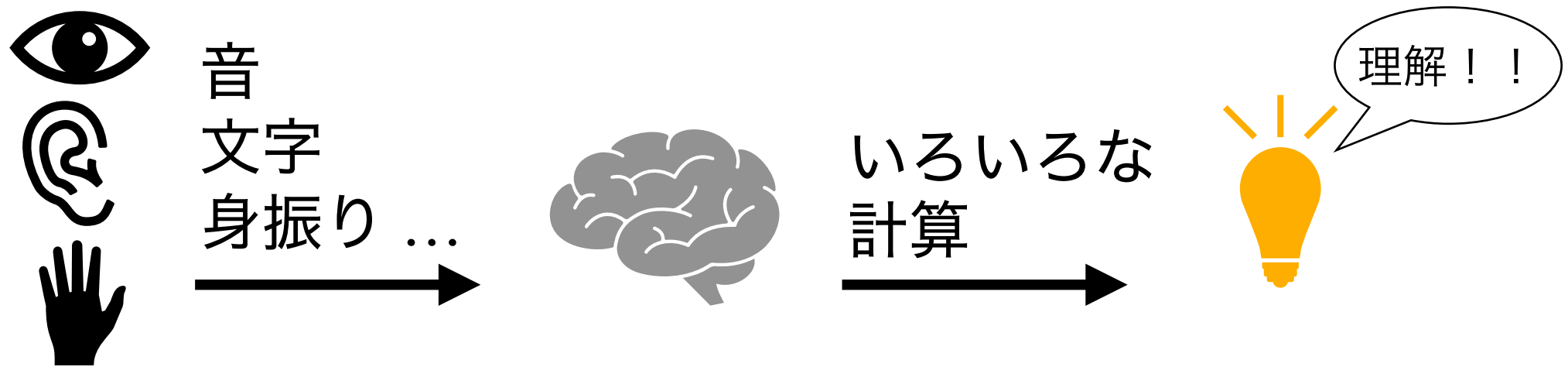


音
文字
身振り ...

そもそも、文理解における謎とは



そもそも、文理解における謎とは



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :
 1. 不完全な入力への頑健さ
 2. 正確な曖昧性解消
 3. 途中段階での推論
 4. 局所的な負荷の変動



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



そもそも、文理解における謎とは

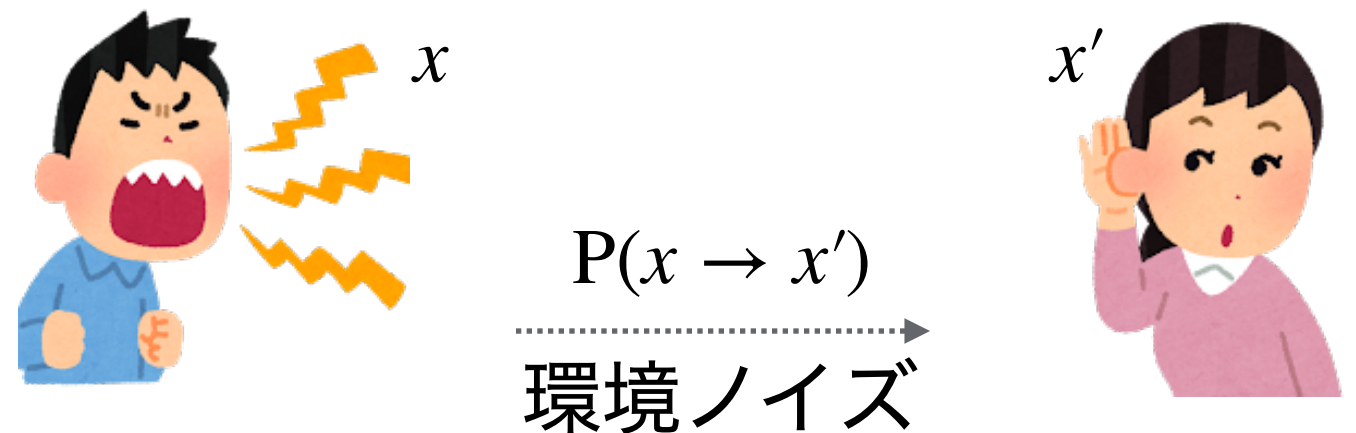
- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



そもそも、文理解における謎とは

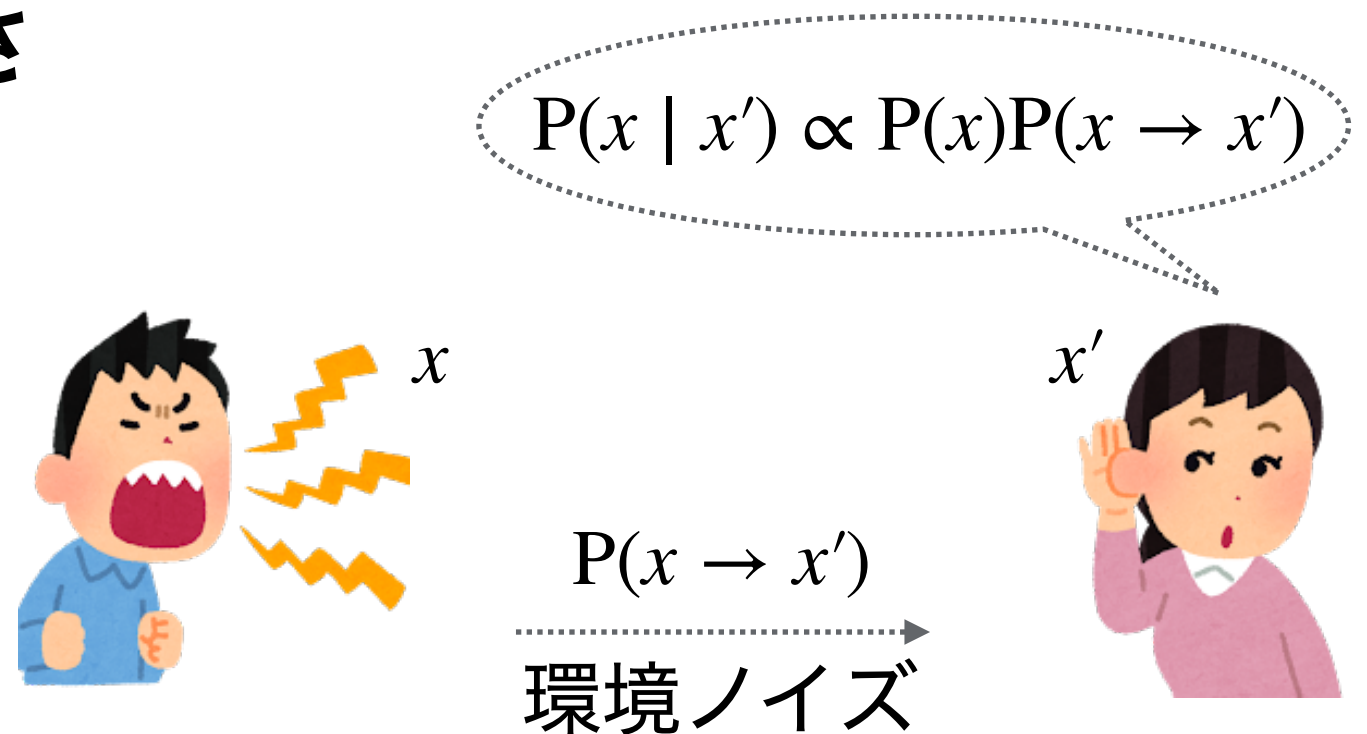
- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :
 1. 不完全な入力への頑健さ
 2. 正確な曖昧性解消
 3. 途中段階での推論
 4. 局所的な負荷の変動



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動

頭が赤い魚を食べる猫



そもそも、文理解における謎とは

- 人間の文理解理論において説明される

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

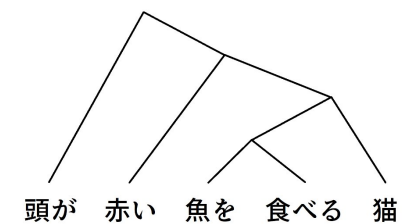
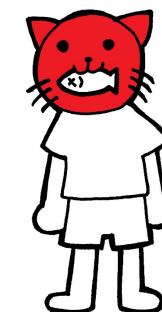
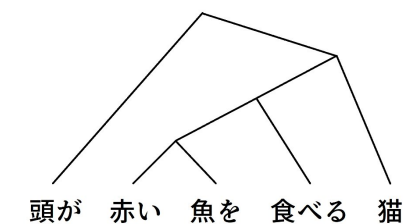
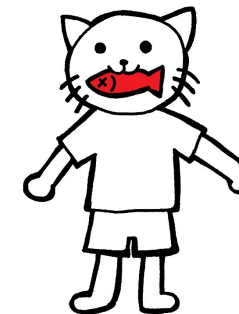
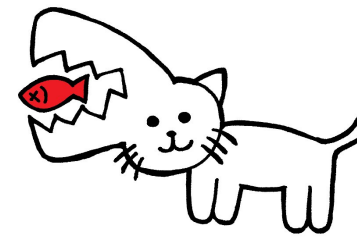
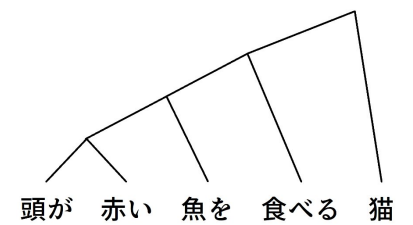
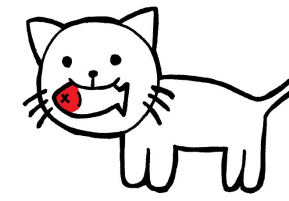
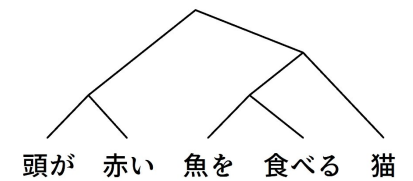
4. 局所的な負荷の変動



音
文字
身振り ...



いる
計算



そもそも、文理解における謎とは

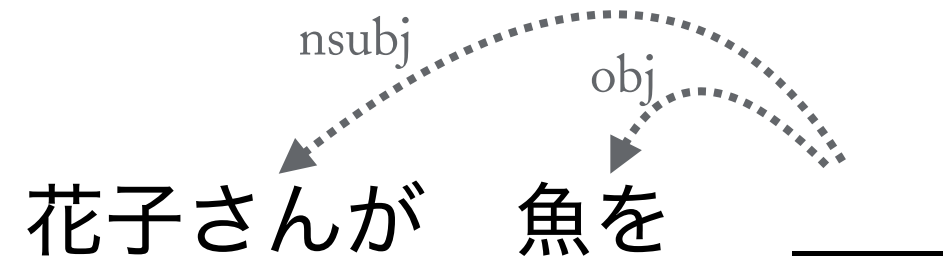
- 人間の文理解理論において説明されるべきもの [Levy 2008] :
 1. 不完全な入力への頑健さ
 2. 正確な曖昧性解消
 - 3. 途中段階での推論**
 4. 局所的な負荷の変動



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ



2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



そもそも、文理解における謎とは

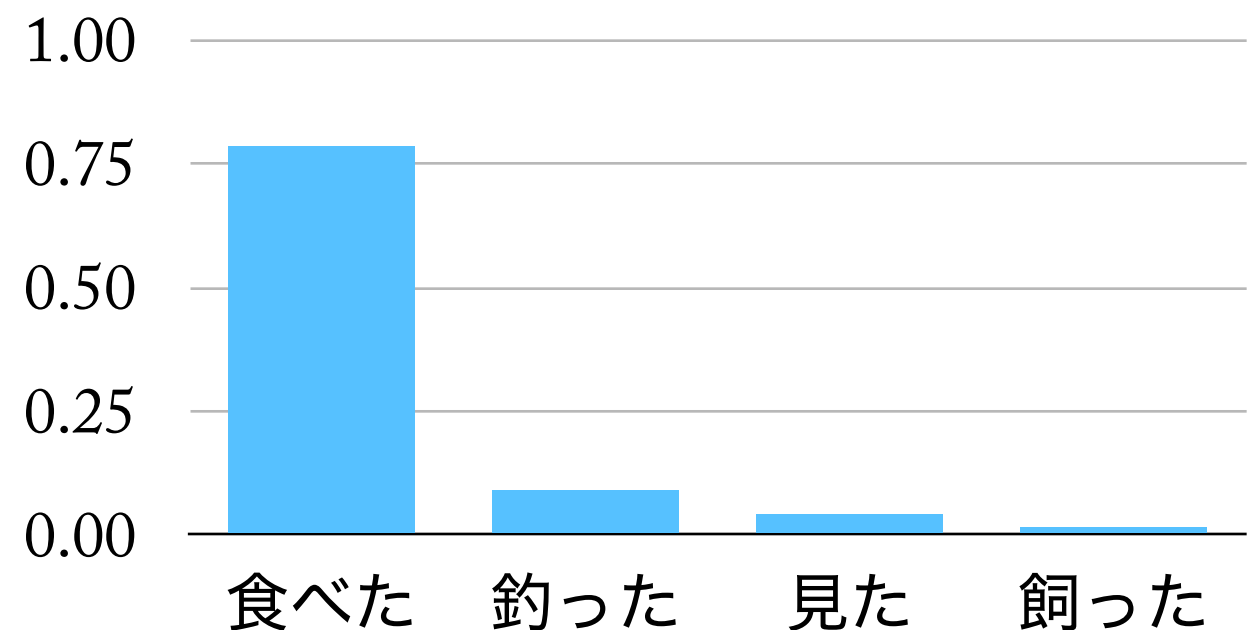
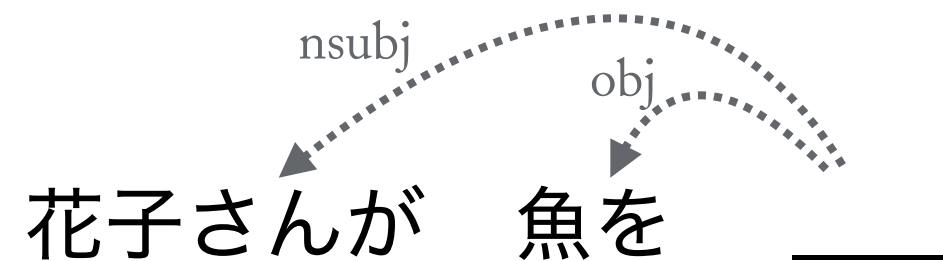
- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



音
文字
身振り ...



いろいろな
計算



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :
 1. 不完全な入力への頑健さ
 2. 正確な曖昧性解消
 3. 途中段階での推論
 4. 局所的な負荷の変動



そもそも、文理解における謎とは

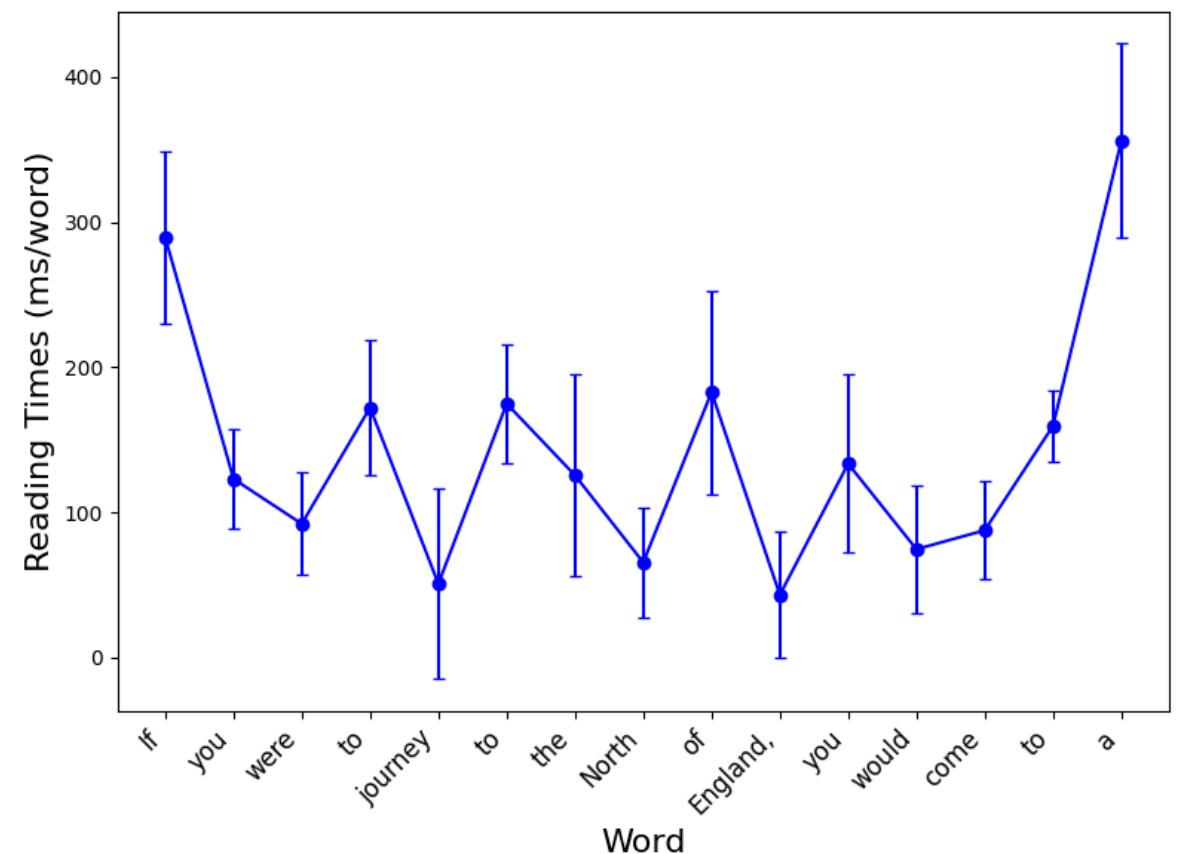
- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力への頑健さ

2. 正確な曖昧性解消

3. 途中段階での推論

4. 局所的な負荷の変動



naturalstories より

<https://github.com/languageit/naturalstories>



音
文字
身振り ...



いろいろな
計算



そもそも、文理解における謎とは

- 人間の文理解理論において説明されるべきもの [Levy 2008] :

1. 不完全な入力

2. 正確な曖昧性

3. 途中段階での

4. 局所的な負荷

🤔 確率的な処理を定式化できたら
嬉しい

🤔 処理する情報の量が測れたら
嬉しい



音

文字
身振り ...



いろいろな
計算



情報の量を測りたい

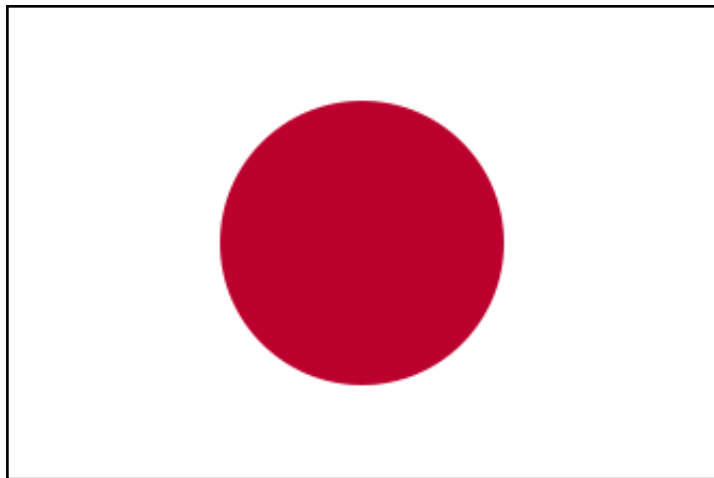
- ある事象 x の情報の情報の量 $h(x)$ は：

$$h(x) := -\log_2 P(x)$$

情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

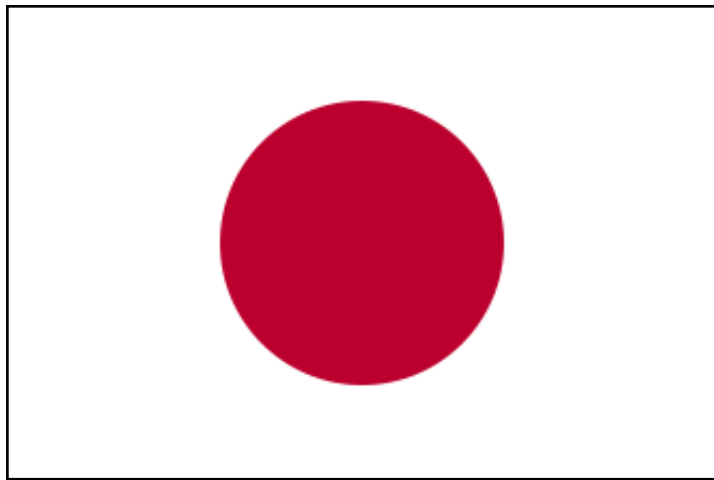
$$h(x) := -\log_2 P(x)$$



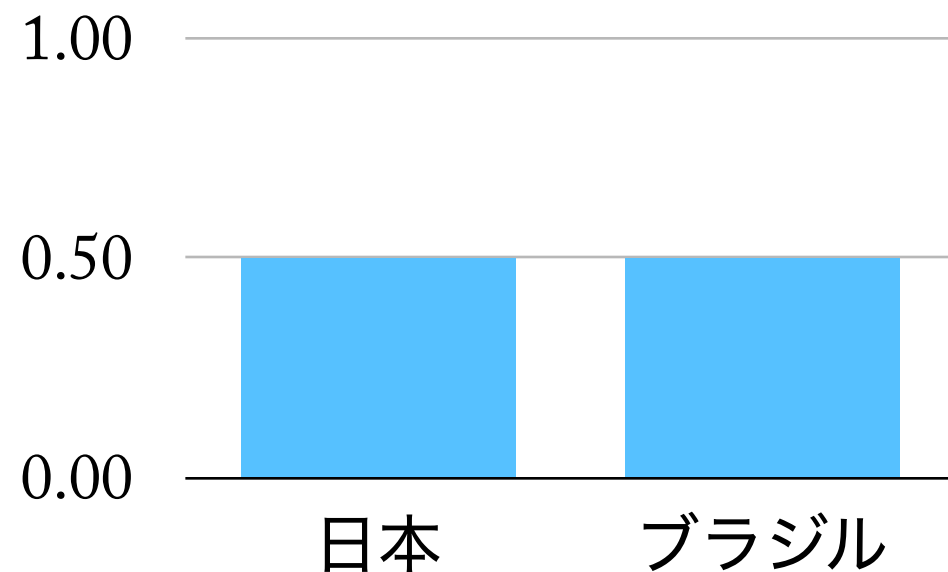
情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

$$h(x) := -\log_2 P(x)$$



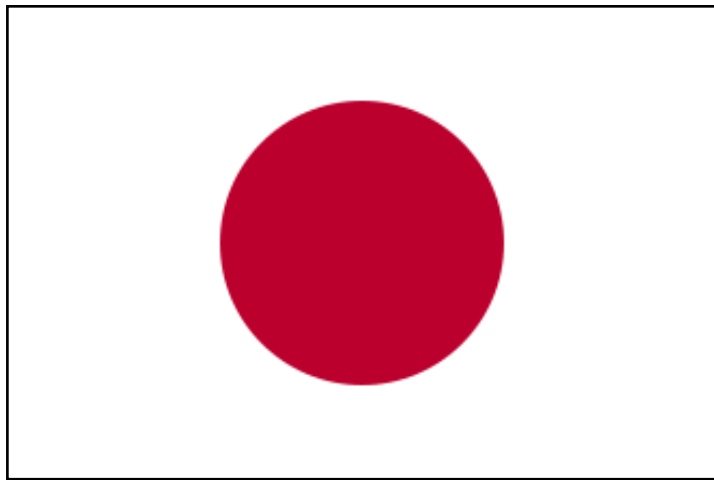
勝利確率



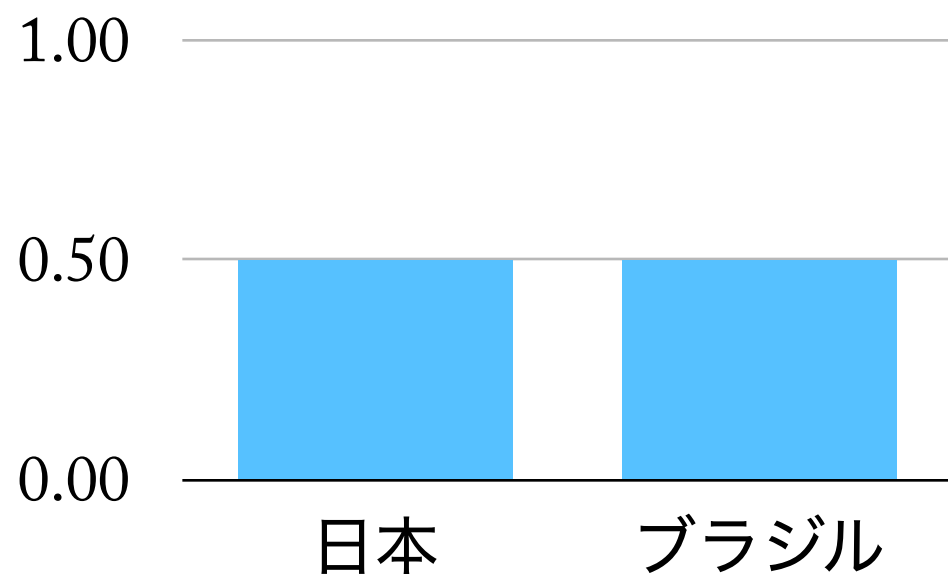
情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

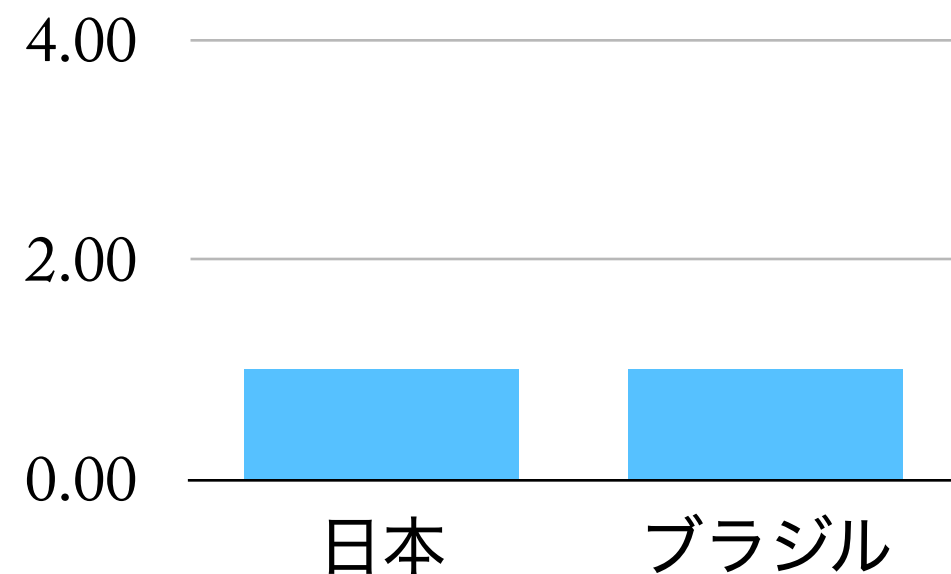
$$h(x) := -\log_2 P(x)$$



勝利確率



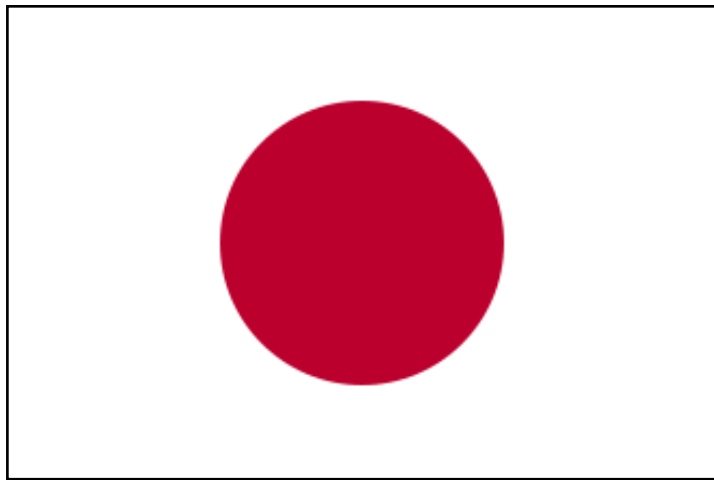
情報量 (単位はbit)



情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

$$h(x) := -\log_2 P(x)$$



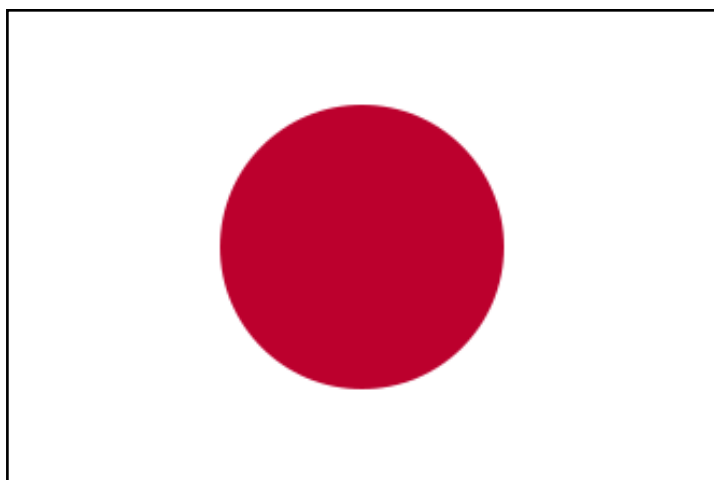
勝利確率



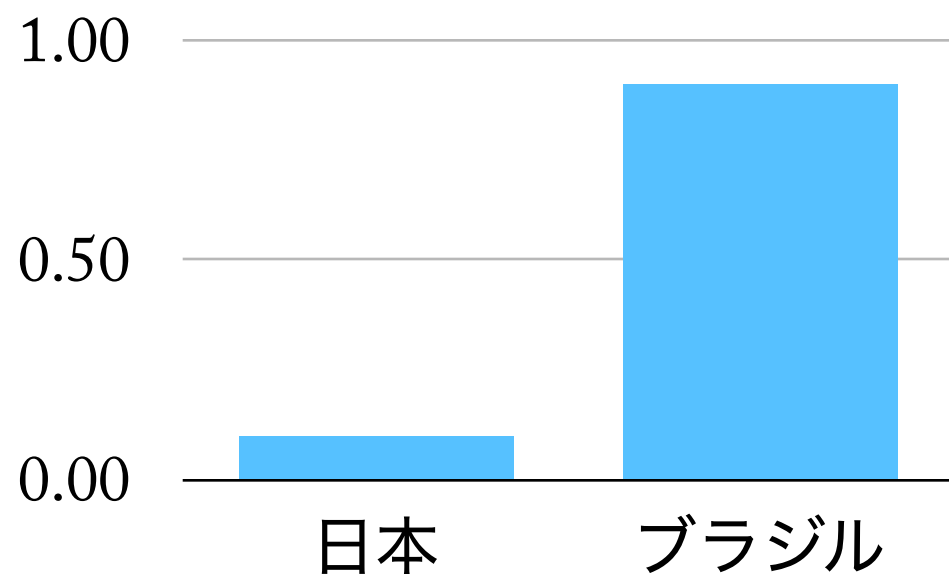
情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

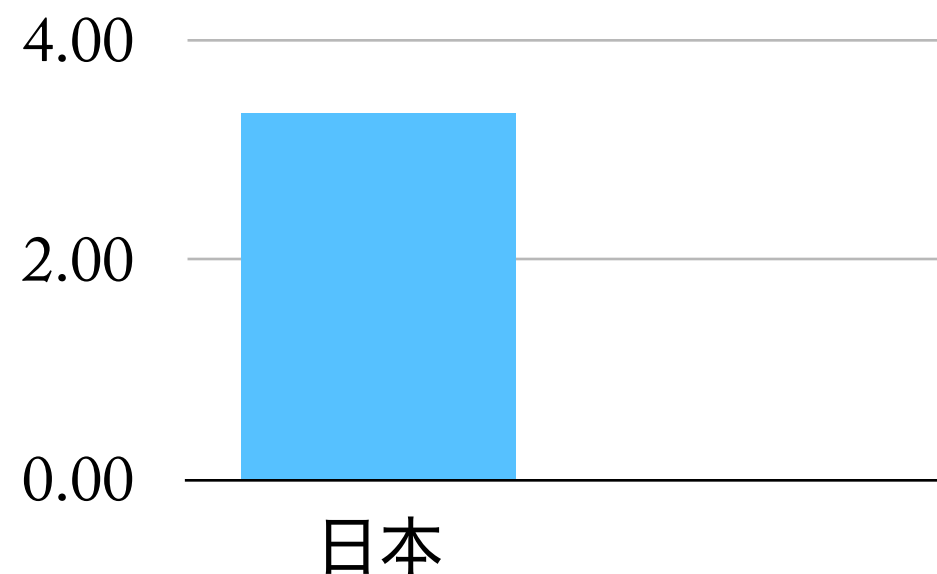
$$h(x) := -\log_2 P(x)$$



勝利確率



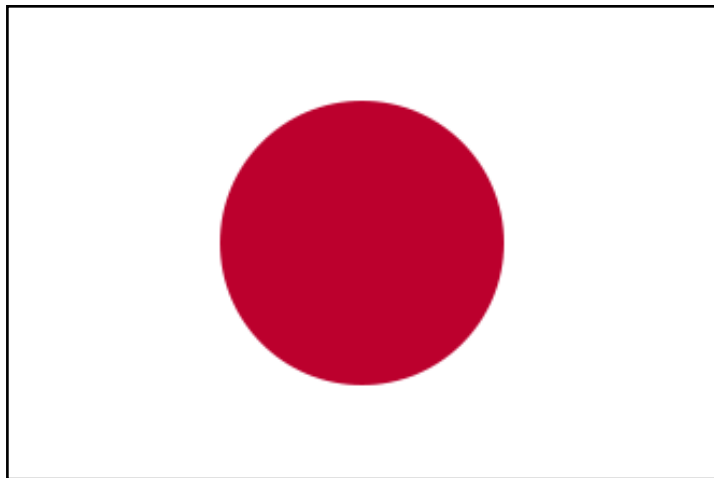
情報量 (単位はbit)



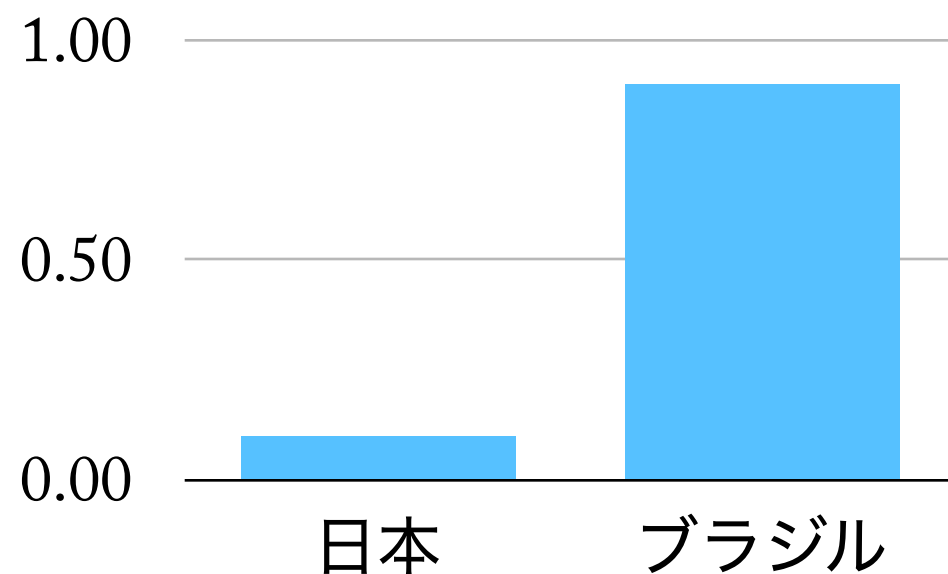
情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

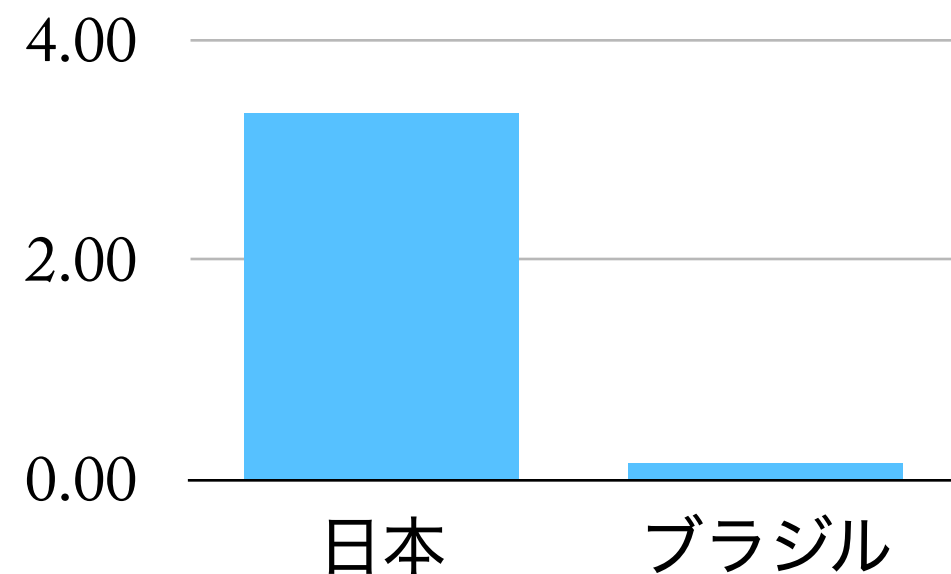
$$h(x) := -\log_2 P(x)$$



勝利確率



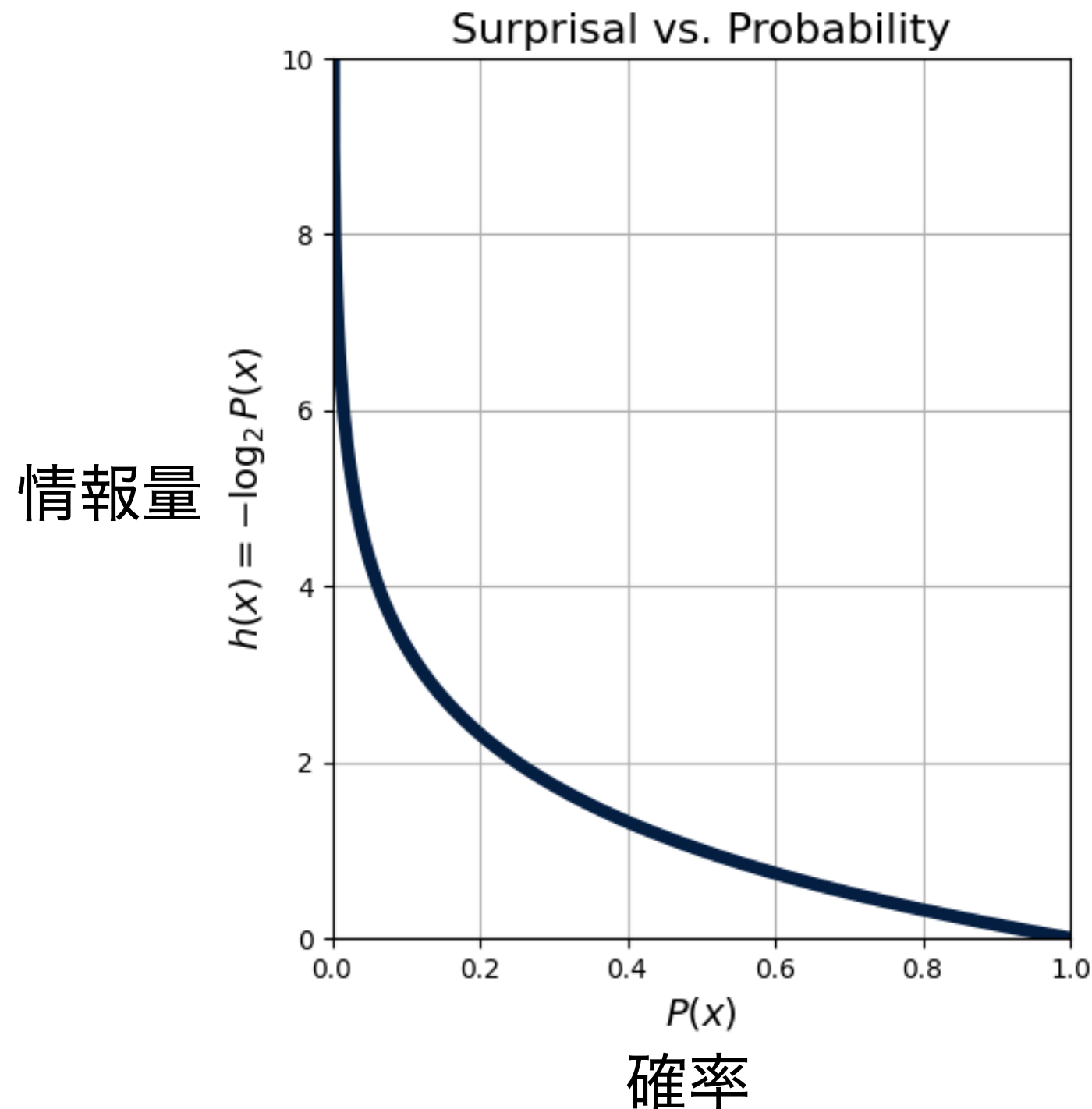
情報量 (単位はbit)



情報の量を測りたい

- ある事象 x の情報の情報の量 $h(x)$ は：

$$h(x) := -\log_2 P(x)$$



情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる

情報の量を測りたい

- 事象 x の情報の量を data-driven で測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)

情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、（従来より正確な）確率推定ができるように

情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、（従来より正確な）確率推定ができるように
今日 7 月 14 日 第 303 回 NINJAL サロン で、...

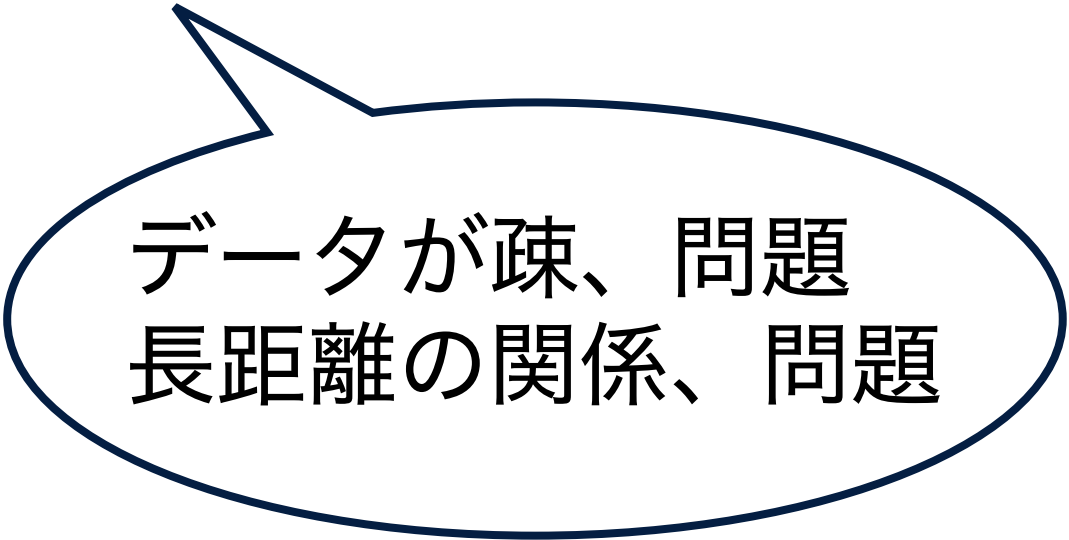
情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、(従来より正確な) 確率推定ができるように
今日 7 月 14 日 第 303 回 NINJAL サロン で、...

データが疎、問題
長距離の関係、問題

情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、(従来より正確な) 確率推定ができるように
今日 7 月 14 日 第 303 回 NINJAL サロン で、...
- (i) 大規模言語データ



データが疎、問題
長距離の関係、問題

情報の量を測りたい

- 事象 x の情報の量をdata-drivenで測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、（従来より正確な）確率推定ができるように
今日 7 月 14 日 第 303 回 NINJAL サロン で、...
- (i) 大規模言語データ
- (ii) 高性能計算機たち

データが疎、問題
長距離の関係、問題

情報の量を測りたい

- 事象 x の情報の量を data-driven で測ることができる
 - 事象 x の確率分布さえ用意できればいい
 - 複数事象 x, y 同士の関係の強さも測ることができる
 - (少なくとも条件確率が用意できるのならば！)
- 近年、（従来より正確な）確率推定ができるように
今日 7 月 14 日 第 303 回 NINJAL サロン で、 ...
 - (i) 大規模言語データ
 - (ii) 高性能計算機たち
 - (iii) （結果としての）大規模言語モデル

情報理論は「良い指標」っぽい

- 人間・言語の振る舞いと（結構）あっていそう
 - サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:
 - $difficulty_i \propto -\log_2 P(w_i \mid \mathbf{w}_{<i})$

情報理論は「良い指標」っぽい

- 人間・言語の振る舞いと（結構）あっていそう

- サブライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:

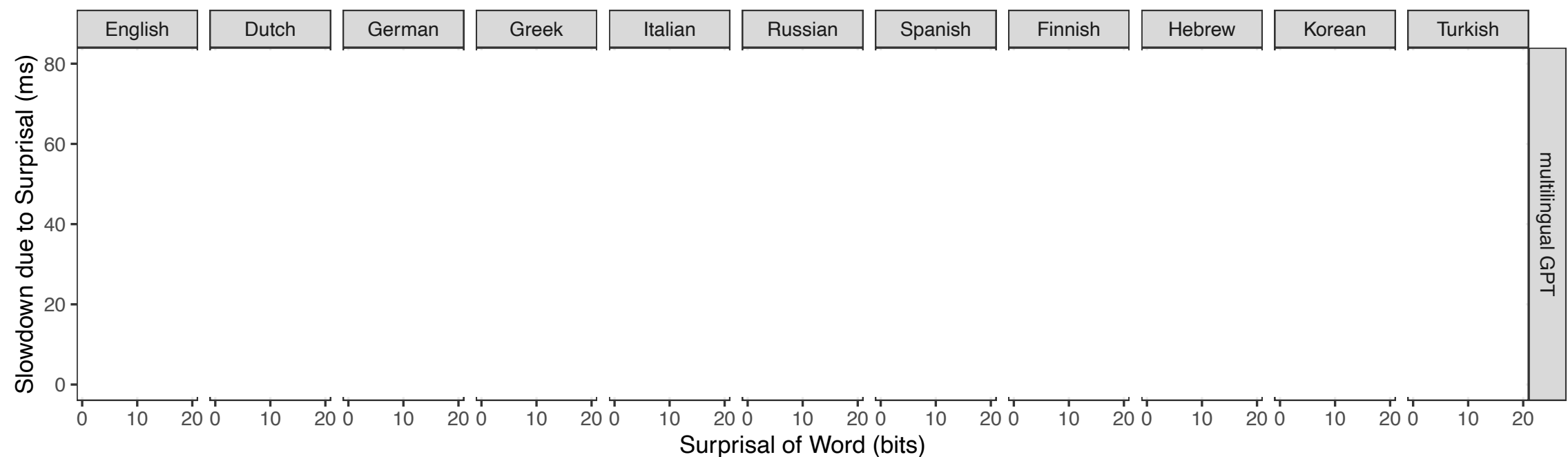
- $difficulty_i \propto -\log_2 P(w_i | w_{<i})$ $w_1, w_2, w_3, \dots, w_{i-1}$ w_i

情報理論は「良い指標」っぽい

- 人間・言語の振る舞いと（結構）あっていそう

- サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:

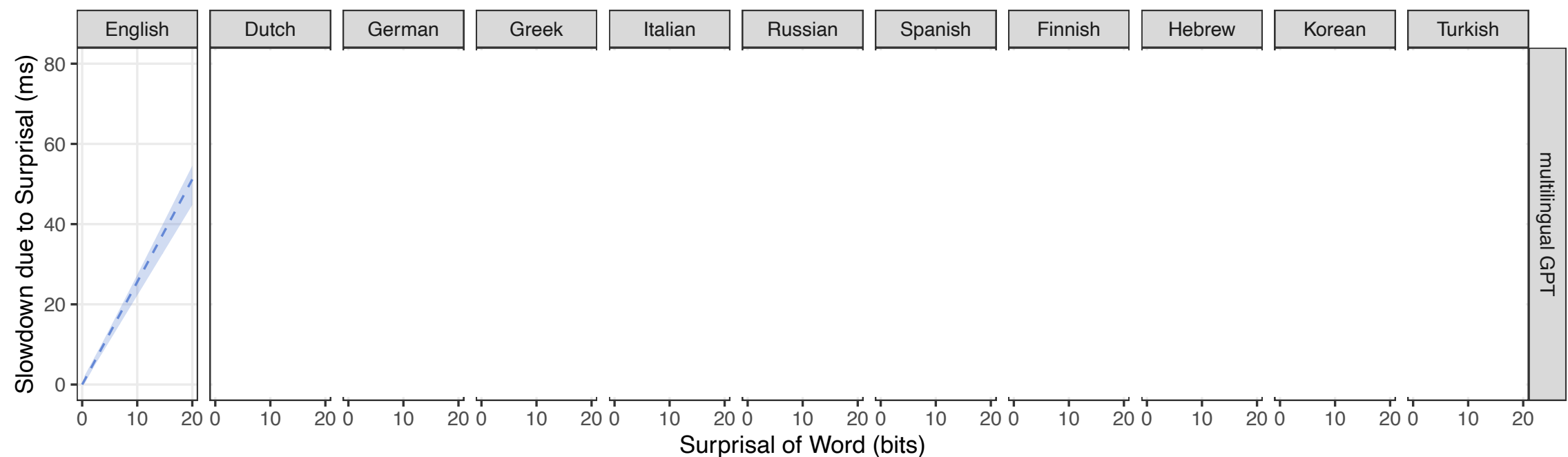
- $difficulty_i \propto -\log_2 P(w_i | w_{<i})$ $w_1, w_2, w_3, \dots, w_{i-1}$ w_i



[Wilcox+ 2023]

情報理論は「良い指標」っぽい

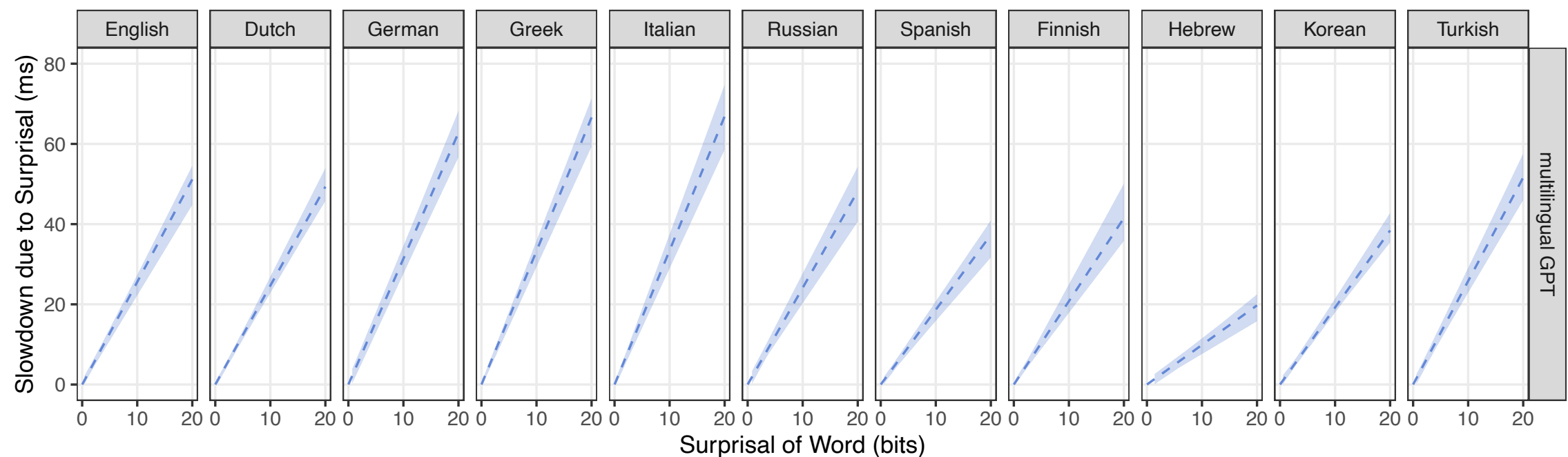
- 人間・言語の振る舞いと（結構）あっていそう
 - サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:
 - $difficulty_i \propto -\log_2 P(w_i | w_{<i})$



[Wilcox+ 2023]

情報理論は「良い指標」っぽい

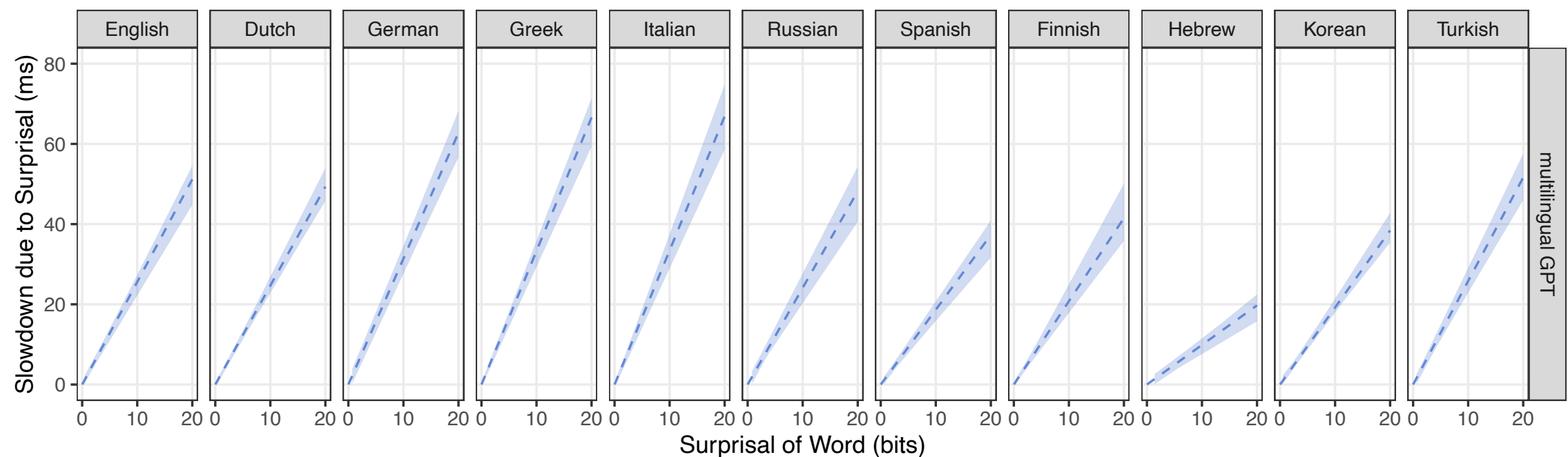
- 人間・言語の振る舞いと（結構）あっていそう
 - サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:
 - $difficulty_i \propto -\log_2 P(w_i | w_{<i})$



[Wilcox+ 2023]

情報理論は「良い指標」っぽい

- 人間・言語の振る舞いと（結構）あっていそう
 - サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:
 - $difficulty_i \propto -\log_2 P(w_i | \mathbf{w}_{<i})$

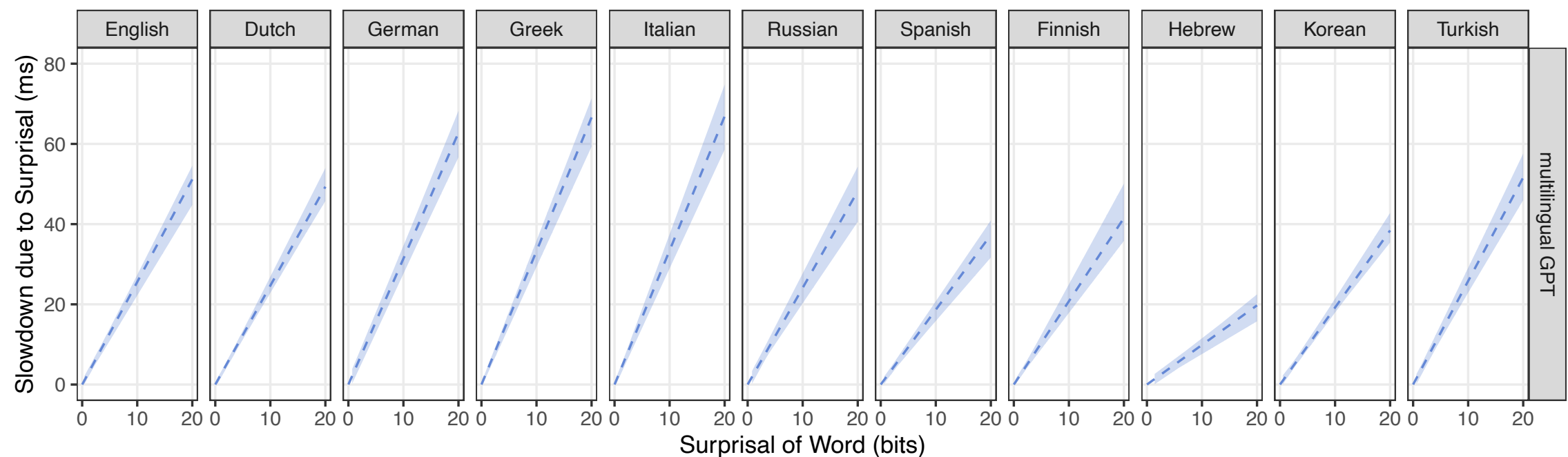


[Wilcox+ 2023]

- 理論中立的な指標

情報理論は「良い指標」っぽい

- 人間・言語の振る舞いと（結構）あっていそう
 - サプライザル理論 [Hale 2001; Levy 2008; Wilcox+ 2023]:
 - $difficulty_i \propto -\log_2 P(w_i | w_{<i})$



[Wilcox+ 2023]

- 理論中立的な指標
 - 特定の言語理論（＝言語の構造に関する理論）にコミットする必要がない
 - その値をどう計算するか、というところがどうでもいい

研究例紹介

研究紹介



Information-Theoretic Storage Cost in Sentence Comprehension

Kohei Kajikawa¹ Shinnosuke Isono² Ethan Gotlieb Wilcox¹

¹Department of Linguistics, Georgetown University, USA

²National Institute for Japanese Language and Linguistics, Japan

Correspondence: kk1571@georgetown.edu

Abstract

Real-time sentence comprehension imposes a significant load on working memory, as comprehenders must maintain contextual information to anticipate future input. While measures of such load have played an important role in psycholinguistic theories, they have largely been formalized using symbolic grammars, which assign discrete, uniform costs to syntactic predictions. This study proposes a measure of processing storage cost based on an information-theoretic formalization, as the amount of information previous words carry about future context, under uncertainty. Un-

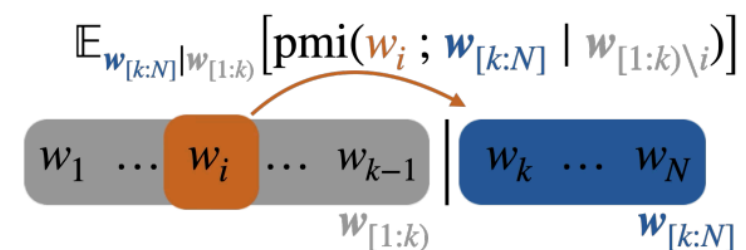


Figure 1: Illustration of the proposed storage cost measure. It quantifies the predictive potential shared between a **target word** w_i and the **future** $w_{[k:N]}$, conditioned on the **remaining context**. Information-theoretic storage cost at w_k is defined as the sum of the predictive potentials across all context words, representing the load of pending information.

- K. Kajikawa, S. Isono, and E.G. Wilcox (2026) Information-theoretic storage cost in sentence comprehension. In *Proc. of CoNLL*. <https://aclanthology.org/2026.conll-main.2/>

情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The _ _ _ _ _



情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat _____



情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat _ _ _



情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on _ _



情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the _



情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the mat

- 作業記憶の容量には限界があるので、予測情報の維持には認知負荷がかかる

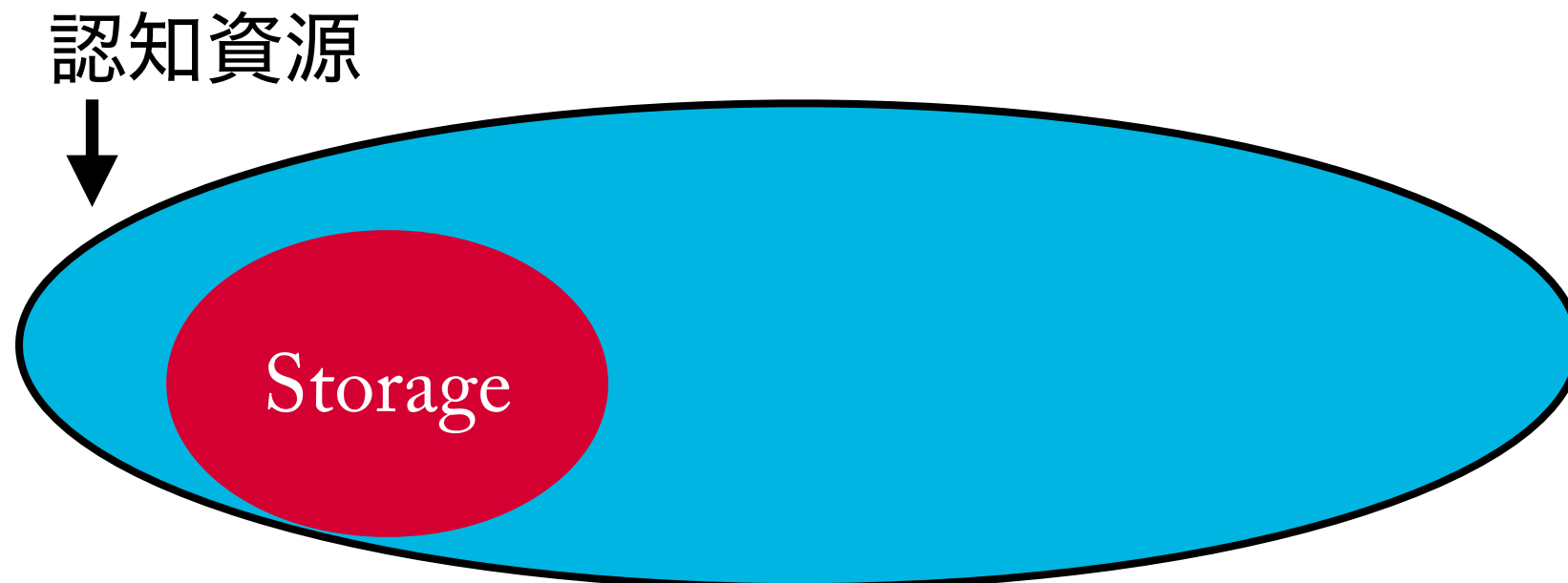
情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the mat

- 作業記憶の容量には限界があるので、予測情報の維持には認知負荷がかかる
 - 予測の累積は記憶のリソースを食い、処理のボトルネックを作る

[Just & Carpenter 1992]



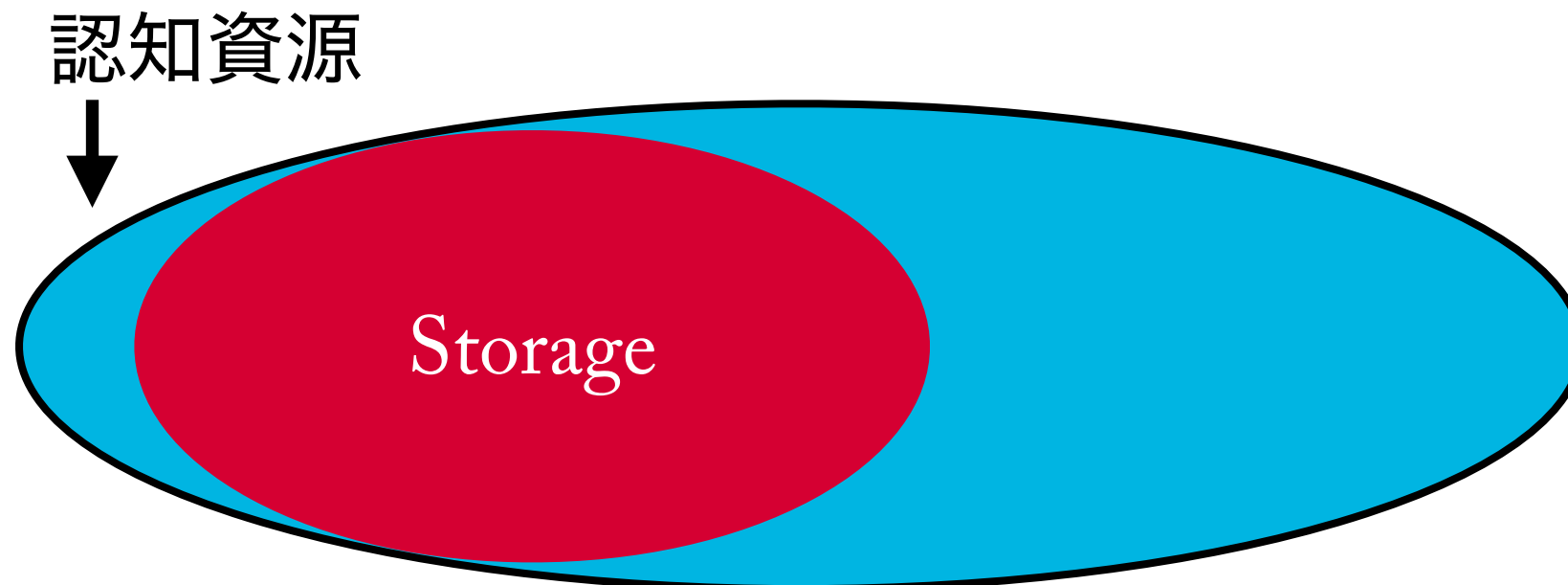
情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the mat

- 作業記憶の容量には限界があるので、予測情報の維持には認知負荷がかかる
 - 予測の累積は記憶のリソースを食い、処理のボトルネックを作る

[Just & Carpenter 1992]



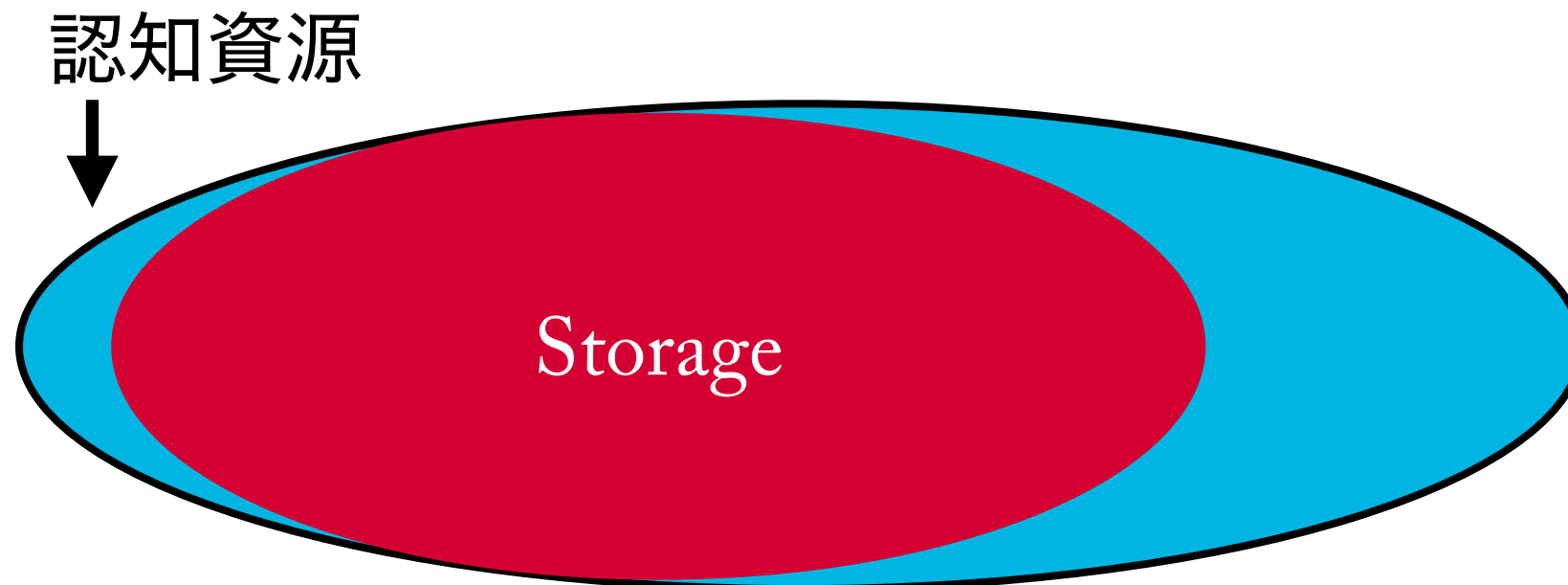
情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the mat

- 作業記憶の容量には限界があるので、予測情報の維持には認知負荷がかかる
 - 予測の累積は記憶のリソースを食い、処理のボトルネックを作る

[Just & Carpenter 1992]



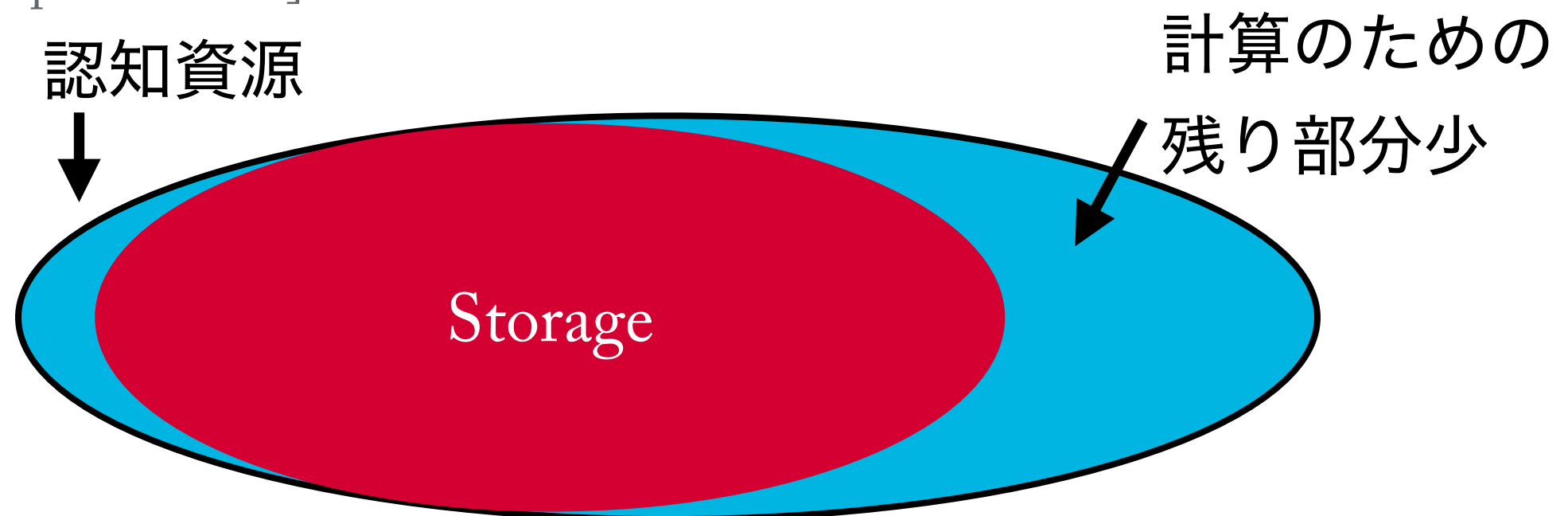
情報の保持にはコストがかかる

- 未来の情報を予測するために文脈情報を保持しておく

The cat sat on the mat

- 作業記憶の容量には限界があるので、予測情報の維持には認知負荷がかかる
 - 予測の累積は記憶のリソースを食い、処理のボトルネックを作る

[Just & Carpenter 1992]



情報の保持にはコストがかかる

- 例えば、中央埋め込み文は処理が大変

[Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

中央埋め込み:

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

情報の保持にはコストがかかる

- 例えば、中央埋め込み文は処理が大変

[Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

中央埋め込み:

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

右枝分かれ:

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

情報の保持にはコストがかかる

- 例えば、中央埋め込み文は処理が大変

[Yngve 1960; Miller & Chomsky 1963; Gibson 1998]

中央埋め込み:

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

右枝分かれ:

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

🤔 では、情報をどのくらい作業記憶に保持している？
どう定量化可能？

文法ベースのstorage

- 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.

文法ベースのstorage

- 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.

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

5 storage cost

1 storage cost

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

文法ベースのstorage

- 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.

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

5 storage cost

1 storage cost

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

- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; [Isono and Kajikawa 2026](#)] の結果からサポート

文法ベースのstorage

- 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.

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



5 storage cost

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



1 storage cost

-
- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; [Isono and Kajikawa 2026](#)] の結果からサポート
 - 文法ベースの storage cost の限界たち :

文法ベースのstorage

- 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.

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



5 storage cost

 1 storage cost
Mary met the senator who t attacked the reporter who t ignored the president

- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; Isono and Kajikawa 2026] の結果からサポート
- 文法ベースの storage cost の限界たち：
 1. 理論依存: 特定の文法理論・想定に依存

文法ベースのstorage

- 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.

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



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



- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; Isono and Kajikawa 2026] の結果からサポート
- 文法ベースの storage cost の限界たち：
 1. 理論依存: 特定の文法理論・想定に依存

文法ベースのstorage

- 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.

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

5 storage cost

1 storage cost

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

- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; [Isono and Kajikawa 2026](#)] の結果からサポート
- 文法ベースの storage cost の限界たち：
 1. 理論依存: 特定の文法理論・想定に依存
 2. 固定コスト: 全ての統語的な要素は同じコストがかかると想定

文法ベースのstorage

- 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.



- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; Isono and Kajikawa 2026] の結果からサポート
- 文法ベースの storage cost の限界たち：
 1. 理論依存: 特定の文法理論・想定に依存
 2. 固定コスト: 全ての統語的な要素は同じコストがかかると想定
 3. 決定論的: 一時的な構造的曖昧性が捉えられない

文法ベースのstorage

- 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.

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



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



- 統制実験 [e.g., Chen et al. 2005; Nakatani & Gibson 2010] や自然な読み時間 [Isono et al. 2025; Isono and Kajikawa 2026] の結果からサポート
 - 文法ベースの storage cost の限界たち：
 1. 理論依存: 特定の文法理論・想定に依存
 2. 固定コスト: 全ての統語的な要素は同じコストがかかると想定
 3. 決定論的: 一時的な構造的曖昧性が捉えられない
- ➡ 情報理論を使って解決を目指す

提案

文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える

文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る

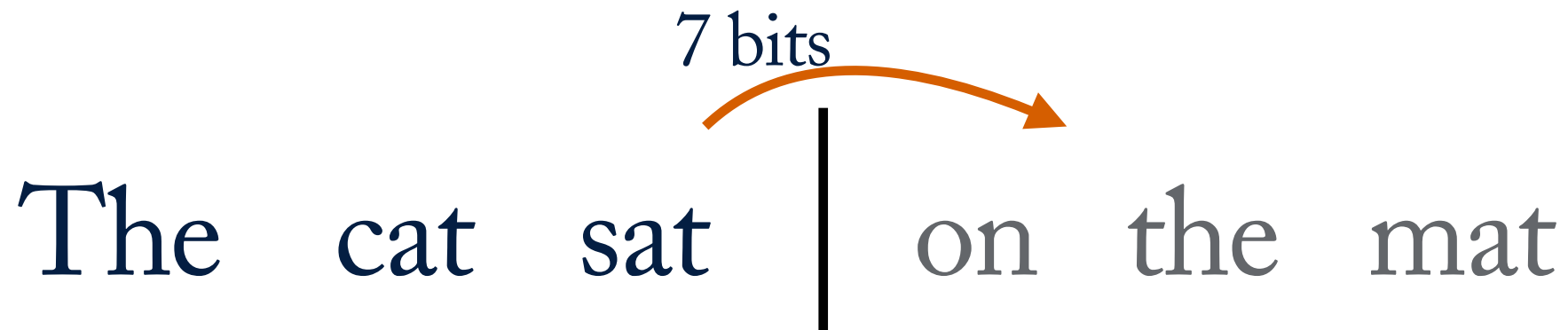
文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る

The cat sat | on the mat

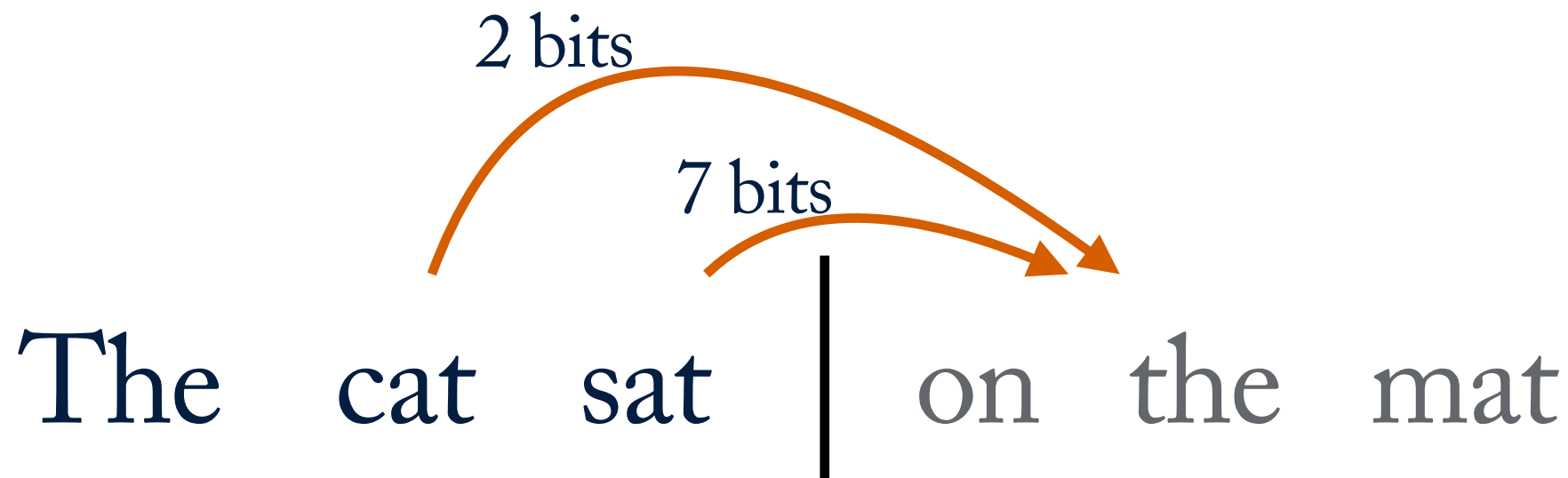
文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る



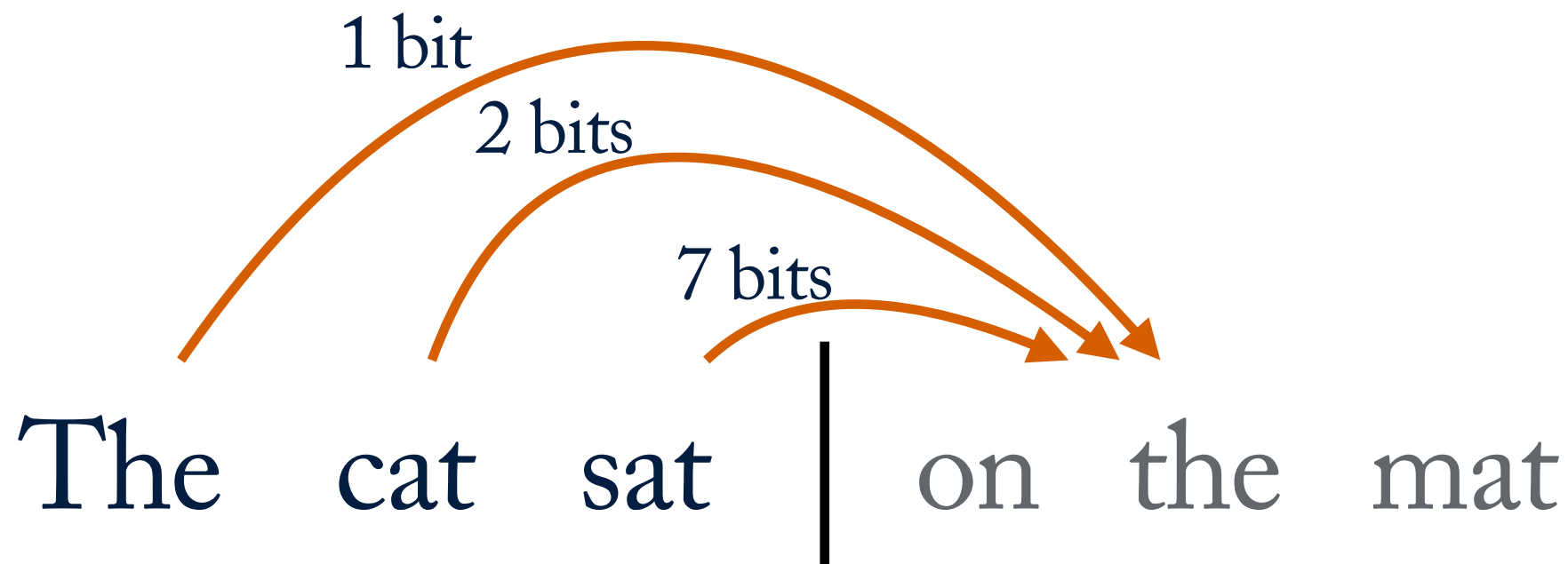
文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る



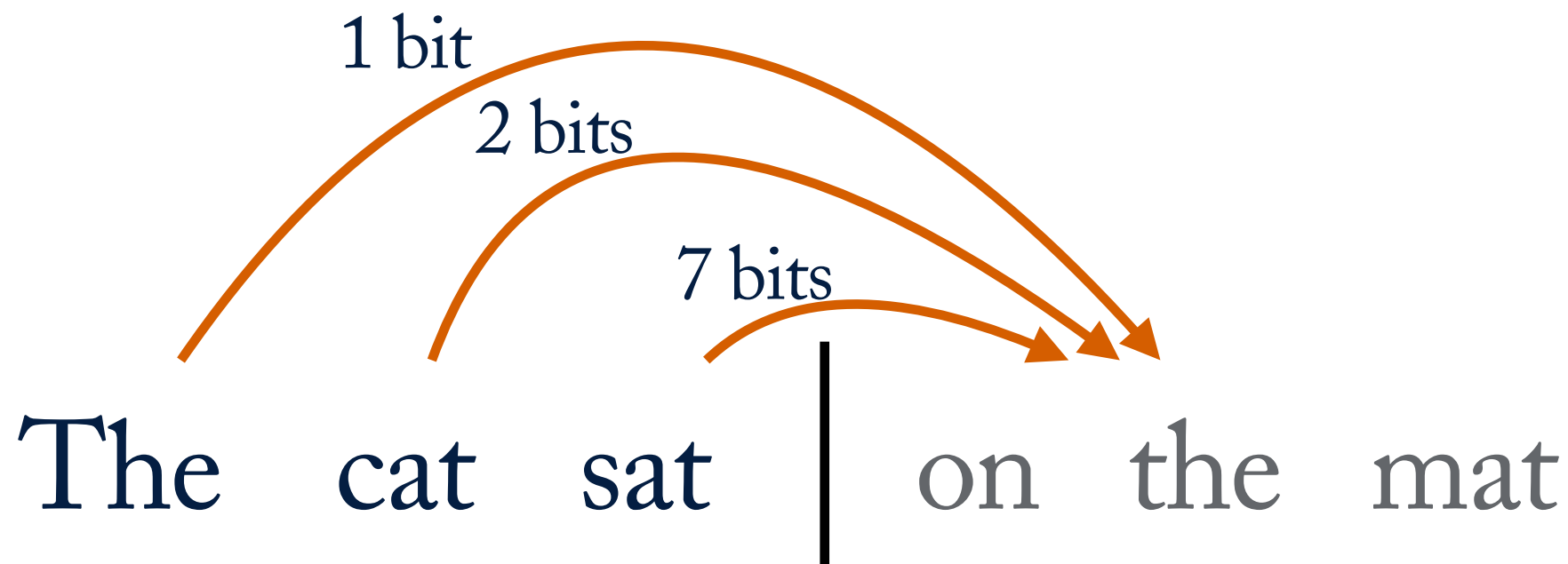
文法から情報理論ベースのstorageへ

- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る



文法から情報理論ベースのstorageへ

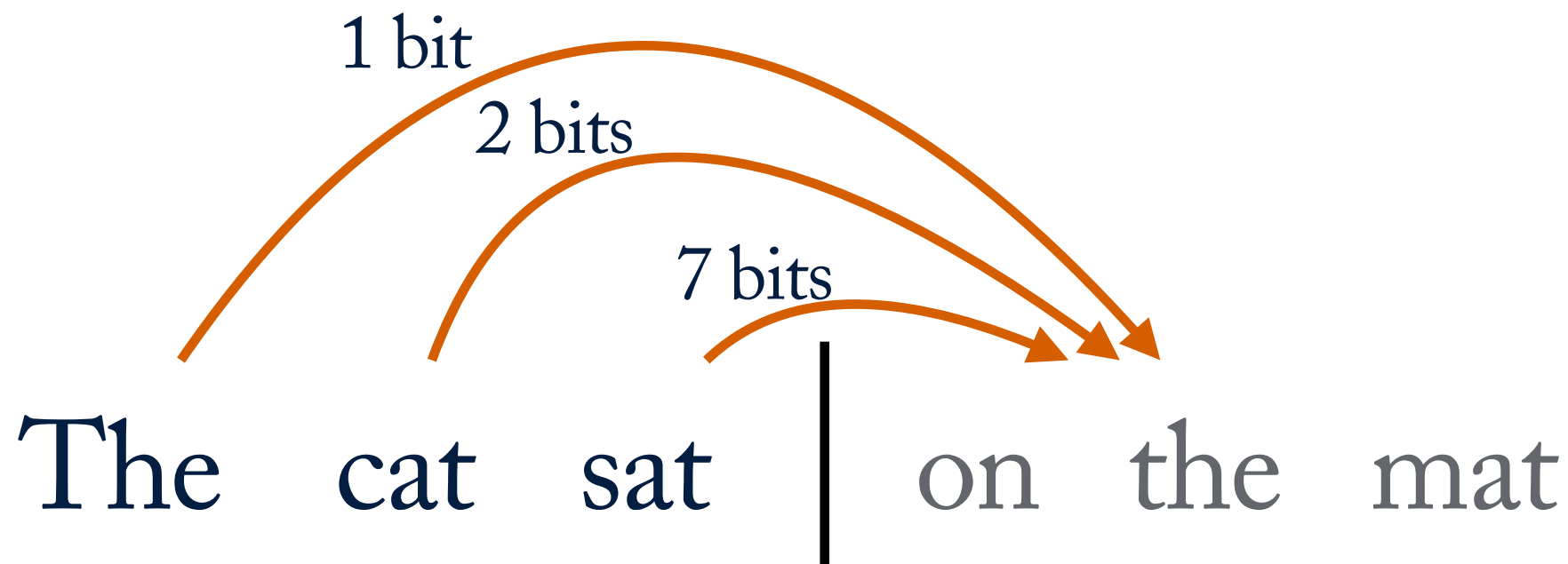
- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る



- **on** での information storage は:

文法から情報理論ベースのstorageへ

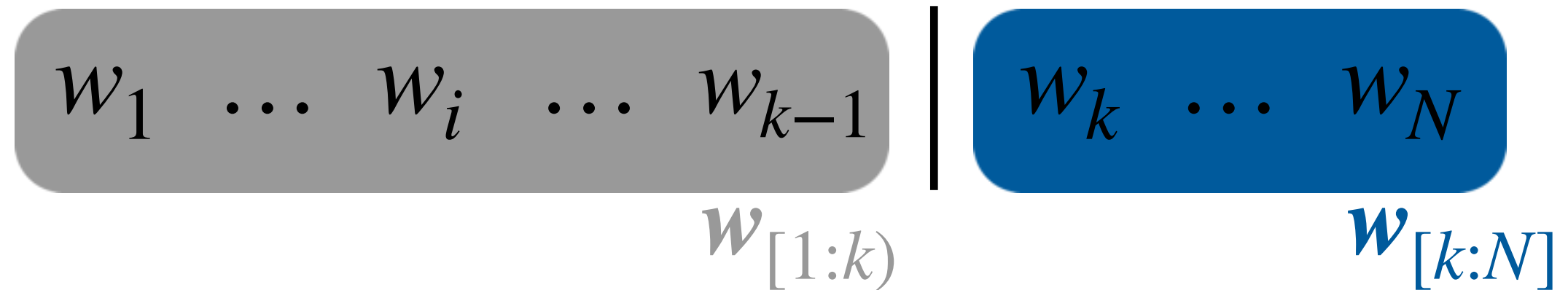
- 文法ベースの storage cost:
 - 予測された統語的要素を数える
- 提案:
 - 文脈中の単語が持つ未来の単語への情報量を測る



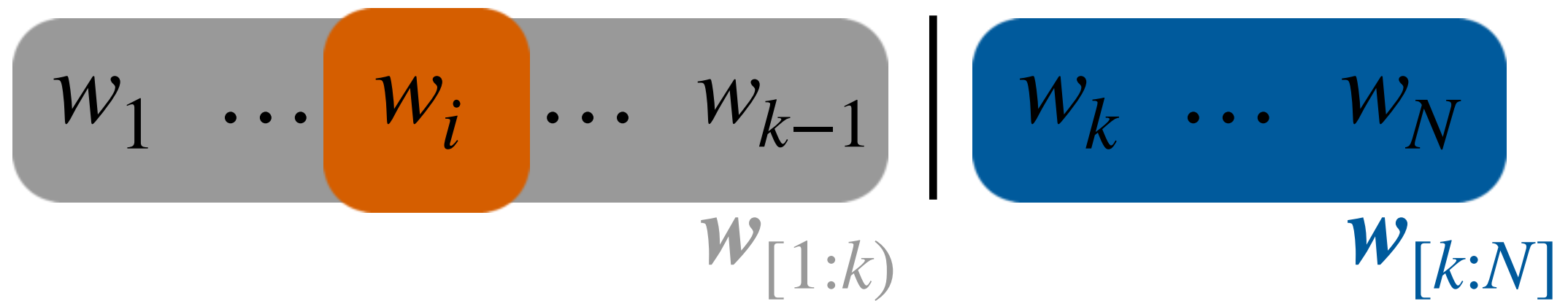
- **on** での information storage は: $1 + 2 + 7 = 10$ bits

情報理論的storage

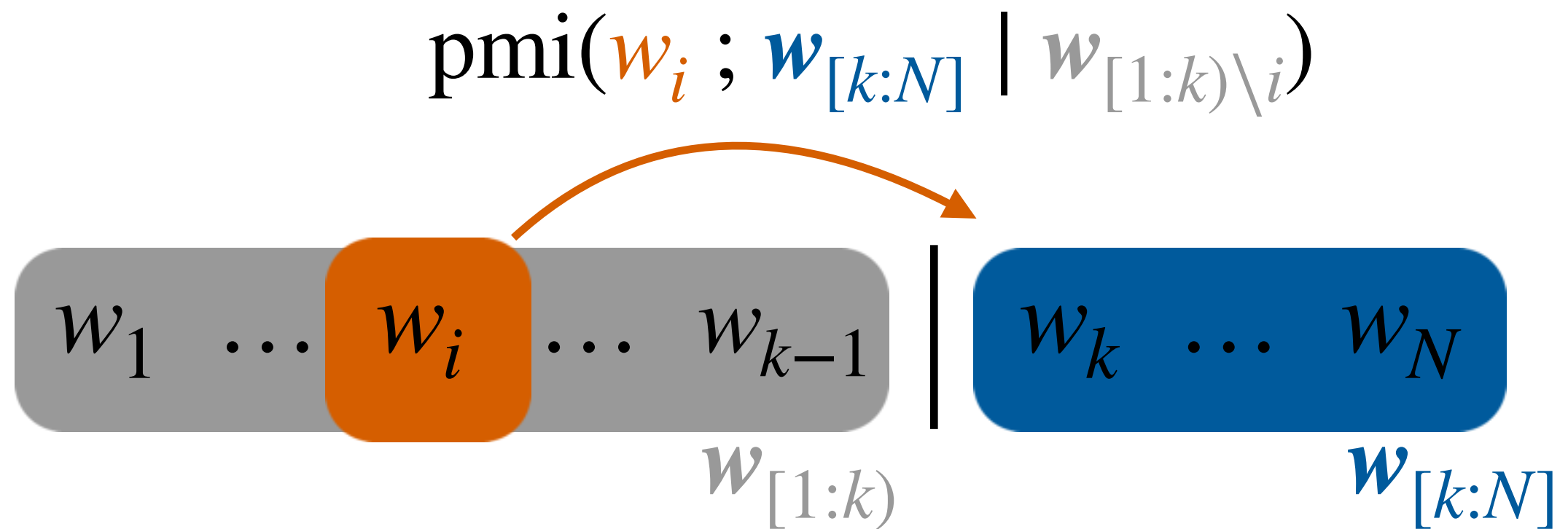
より形式的にいうと ...



情報理論的storage



情報理論的storage

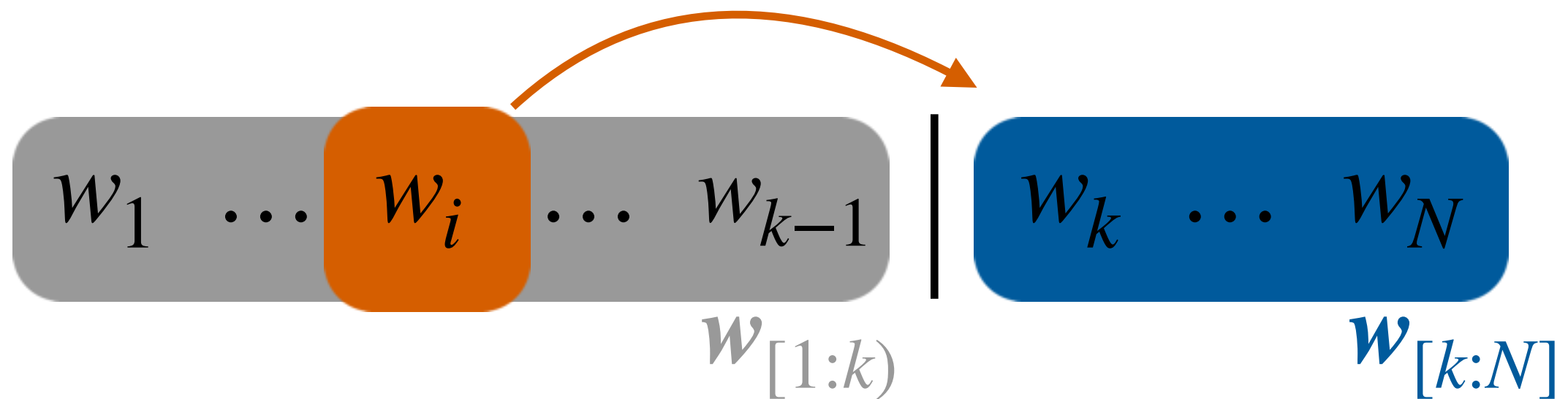


情報理論的storage

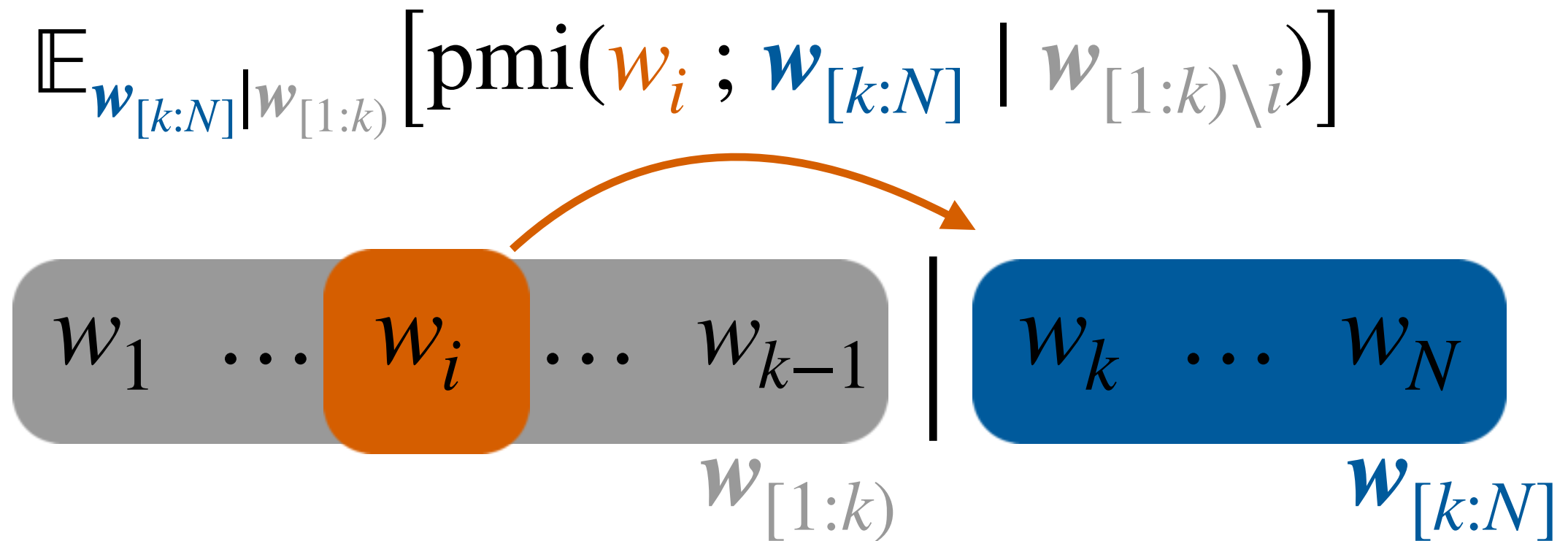
統計的な結びつきの強さの指標

$$\text{pmi}(x; y) := \log \frac{P(x, y)}{P(x)P(y)}$$

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



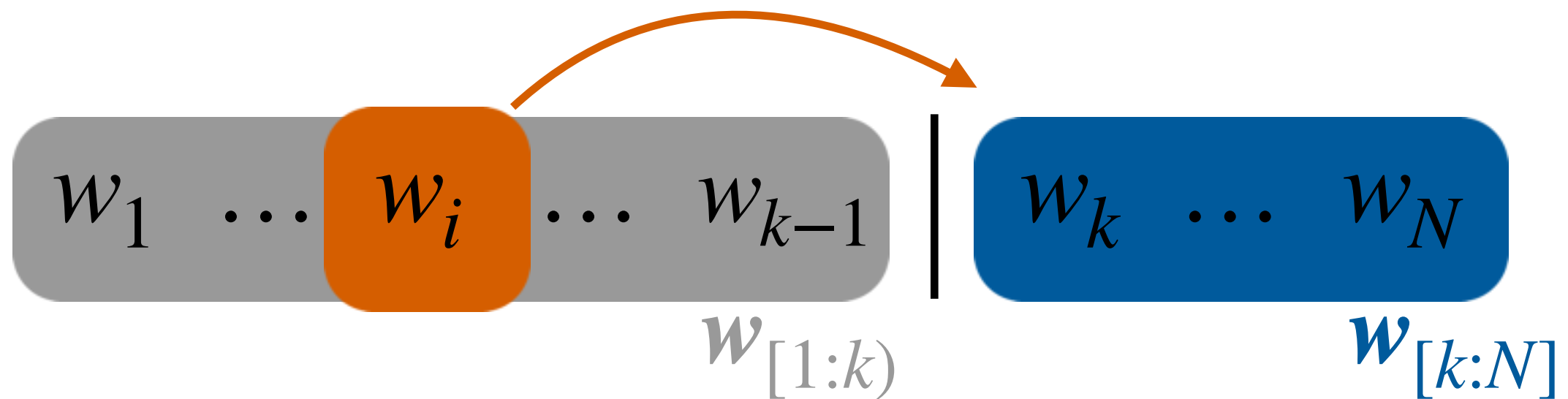
情報理論的storage



情報理論的storage

期待値 あり得る未来全部の平均を取る

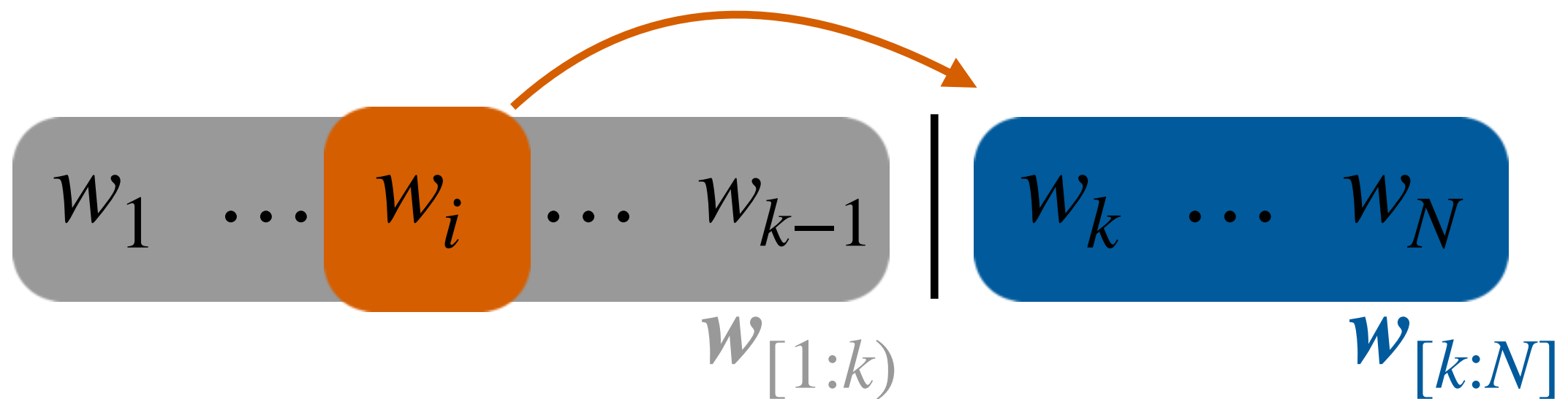
$$\mathbb{E}_{\mathbf{w}_{[k:N]} | \mathbf{w}_{[1:k)}} [\text{pmi}(\mathbf{w}_i; \mathbf{w}_{[k:N]} \mid \mathbf{w}_{[1:k)} \setminus i)]$$



情報理論的storage

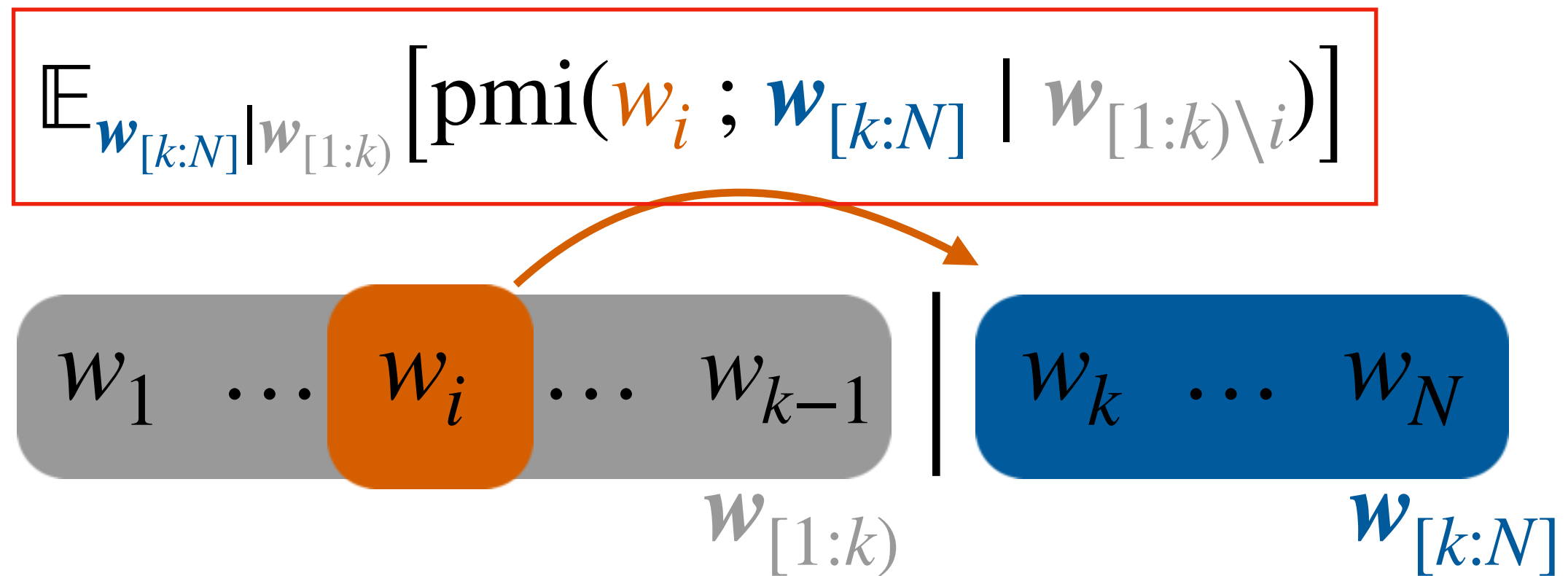
期待値 あり得る未来全部の平均を取る

$$\mathbb{E}_{\mathbf{w}_{[k:N]} | \mathbf{w}_{[1:k)}} [\text{pmi}(w_i; \mathbf{w}_{[k:N]} \mid \mathbf{w}_{[1:k)} \setminus i)]$$



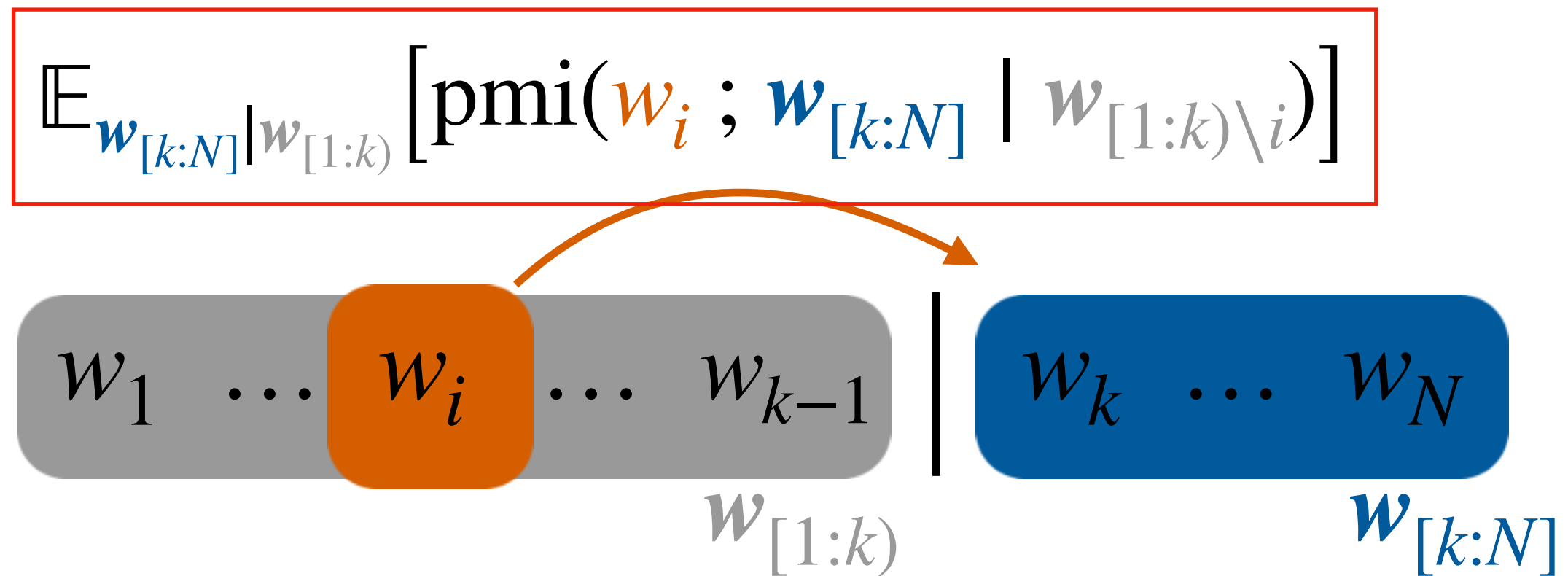
$$\text{InfoStor}_k := \sum_{i=1}^{k-1} \mathbb{E}_{\mathbf{w}_{[k:N]} | \mathbf{w}_{[1:k)}} [\text{pmi}(w_i; \mathbf{w}_{[k:N]} \mid \mathbf{w}_{[1:k)} \setminus i)]$$

情報理論的storage



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

情報理論的storage



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

- PMIの期待値は BERT で推定 [Devlin+ 2019; Hoover+ 2021].

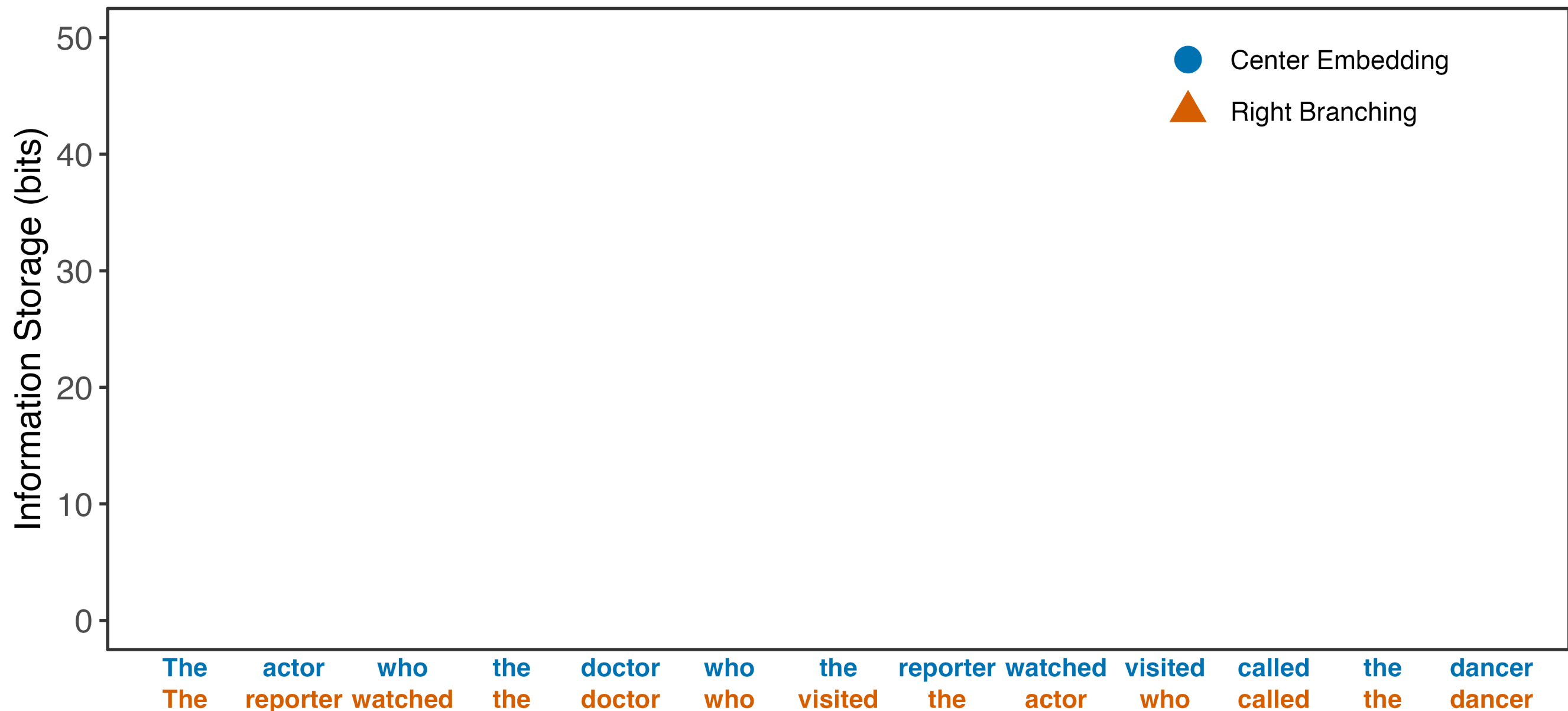
実験

1. ケーススタディ：中央埋め込み

- 処理負荷：中央埋め込み > 右枝分かれ

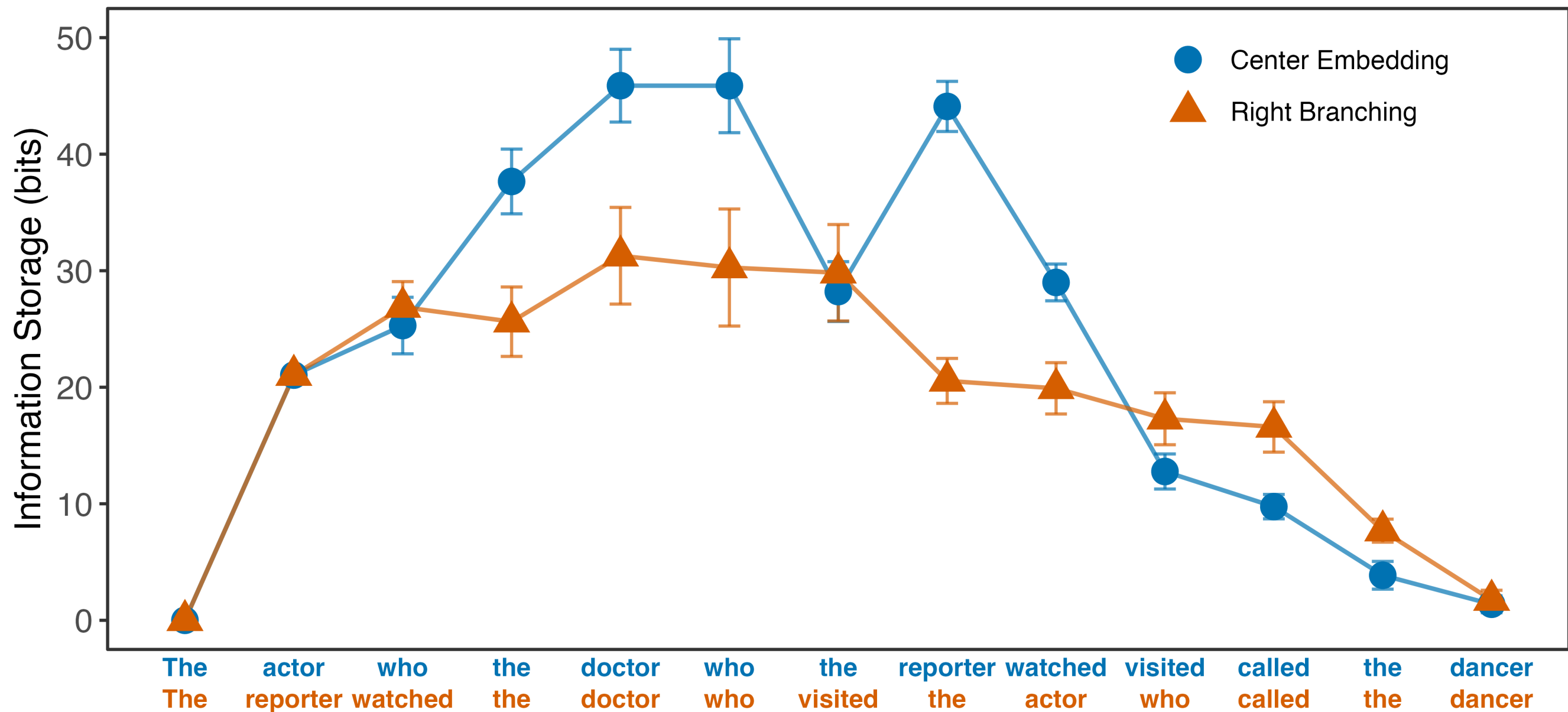
1. ケーススタディ：中央埋め込み

- 処理負荷：中央埋め込み > 右枝分かれ
 - ランダムに30ずつ文を生成
 - 単語ごとの information storage を推定



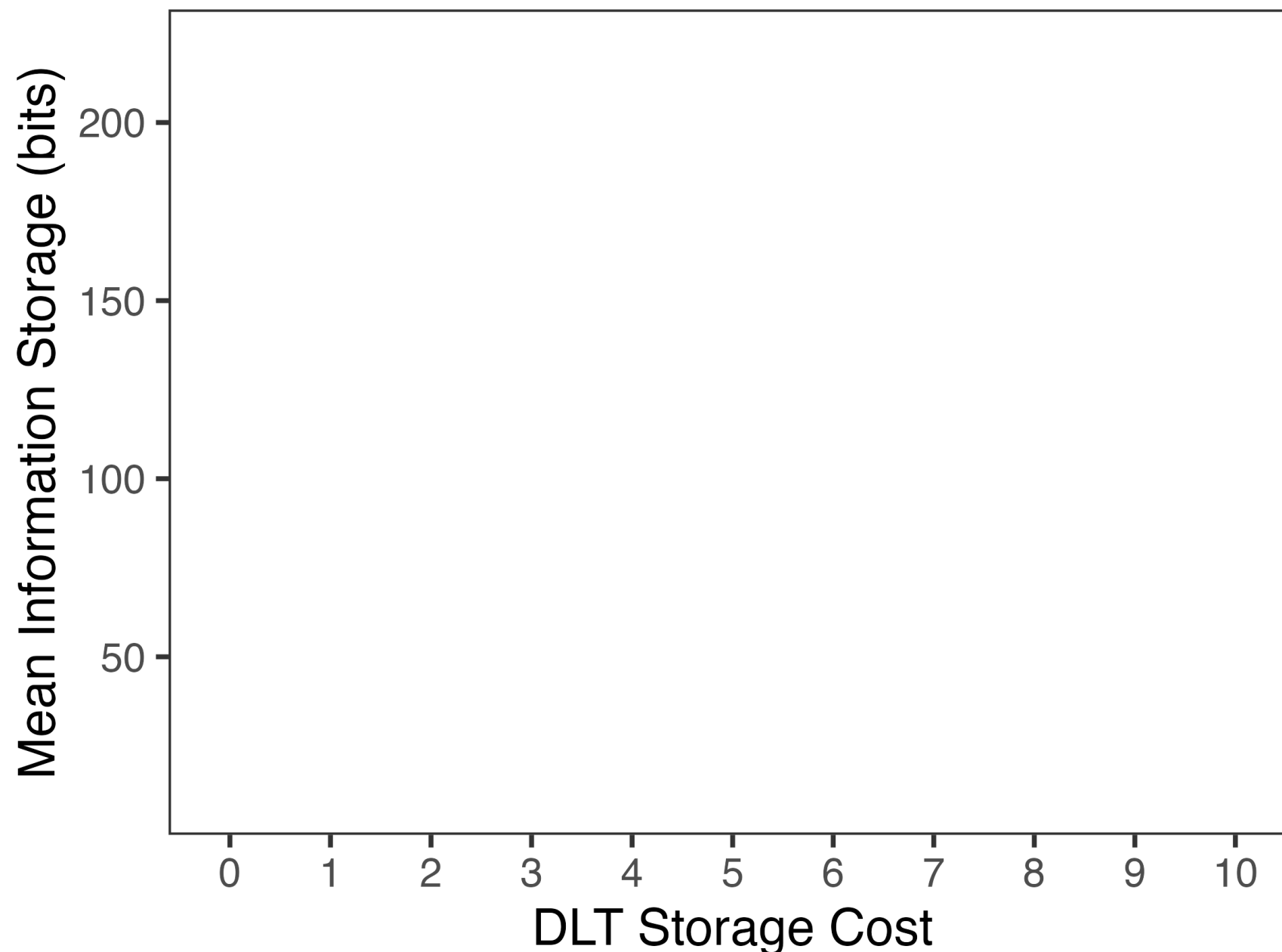
1. ケーススタディ：中央埋め込み

- 処理負荷：中央埋め込み > 右枝分かれ
 - ランダムに30ずつ文を生成
 - 単語ごとの information storage を推定



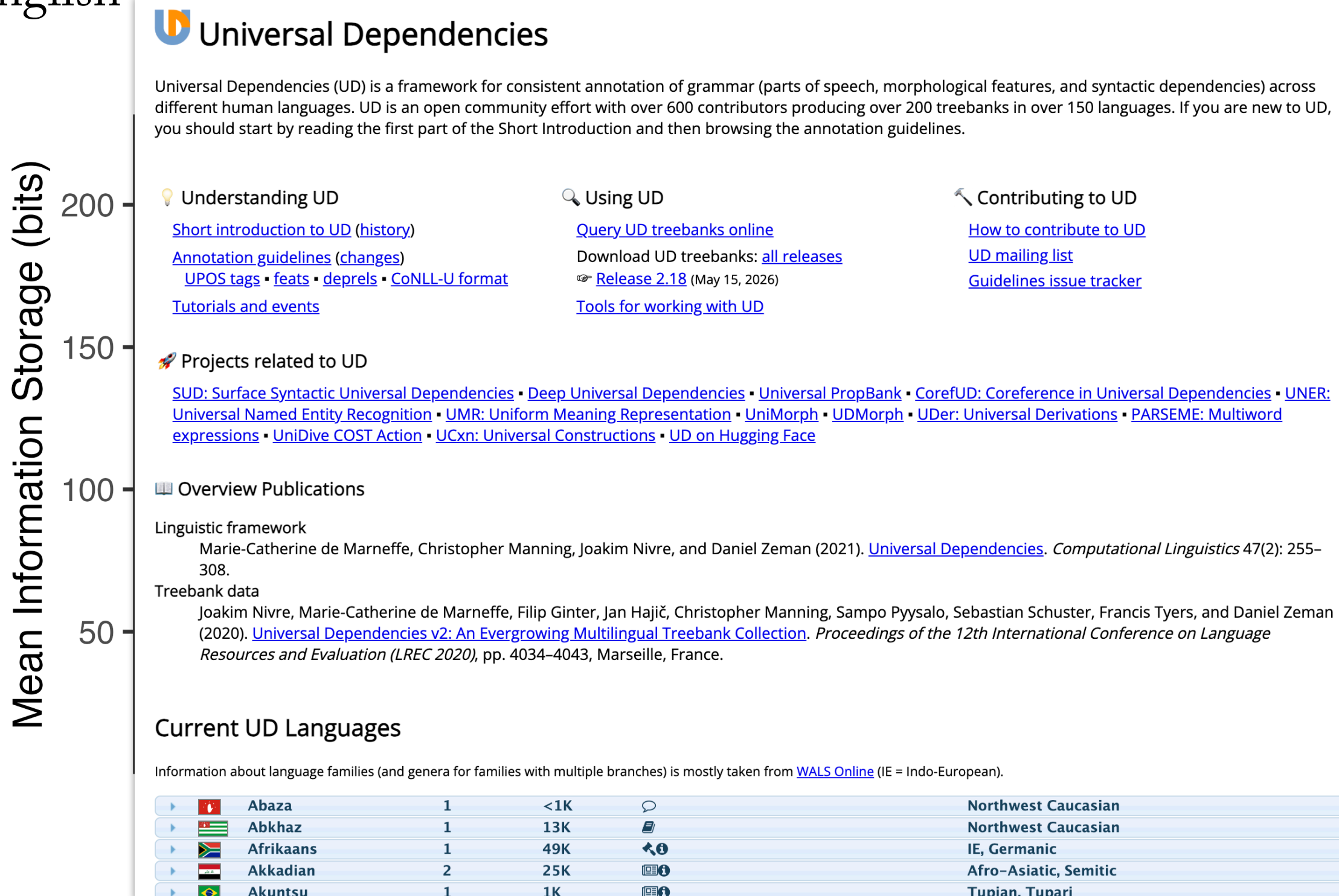
2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GUM [Zeldes 2017] (233,926 words) で確認



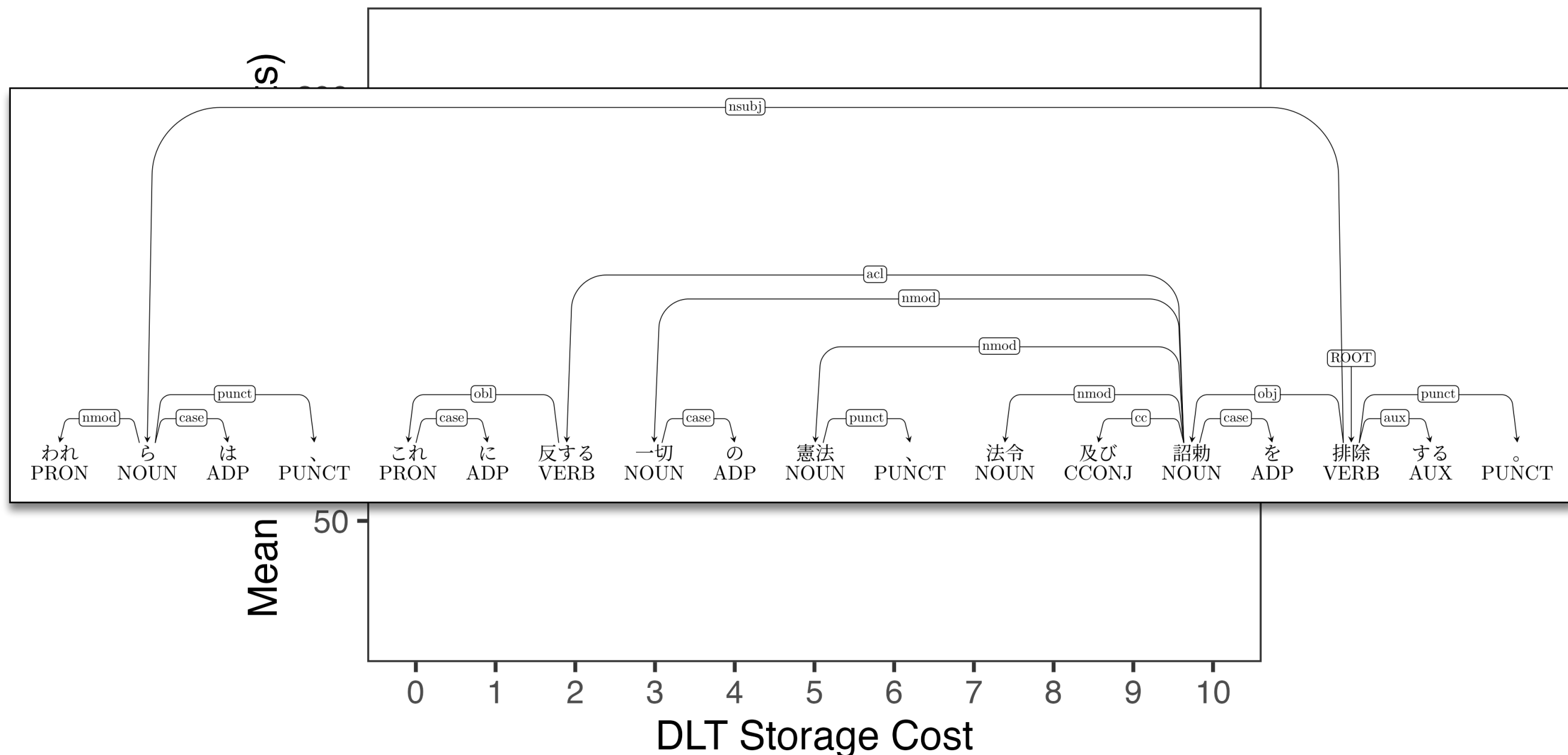
2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GIIM [Zeldes, 2017] (233,926 words) で確認



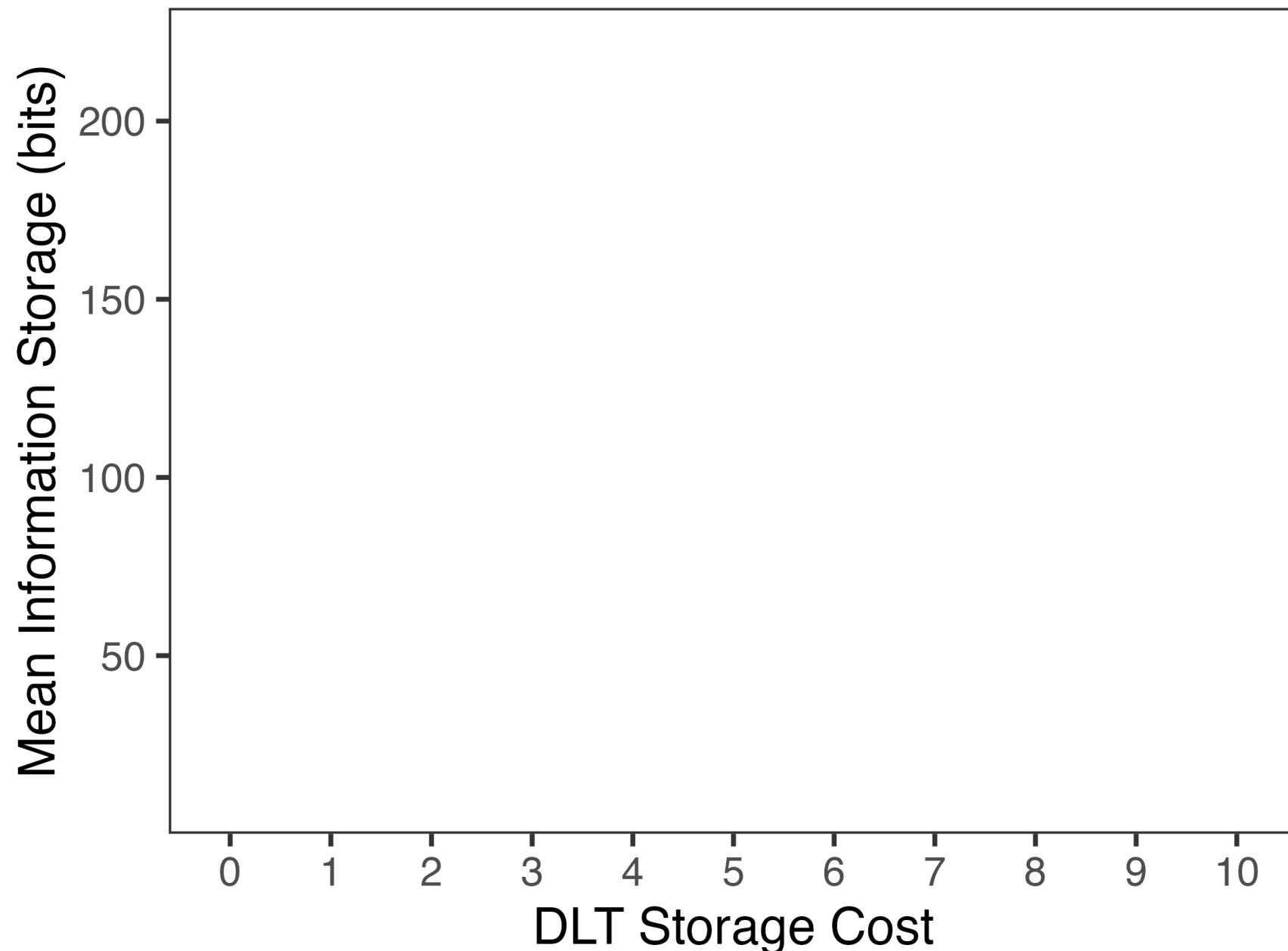
2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GUM [Zeldes 2017] (233,926 words) で確認



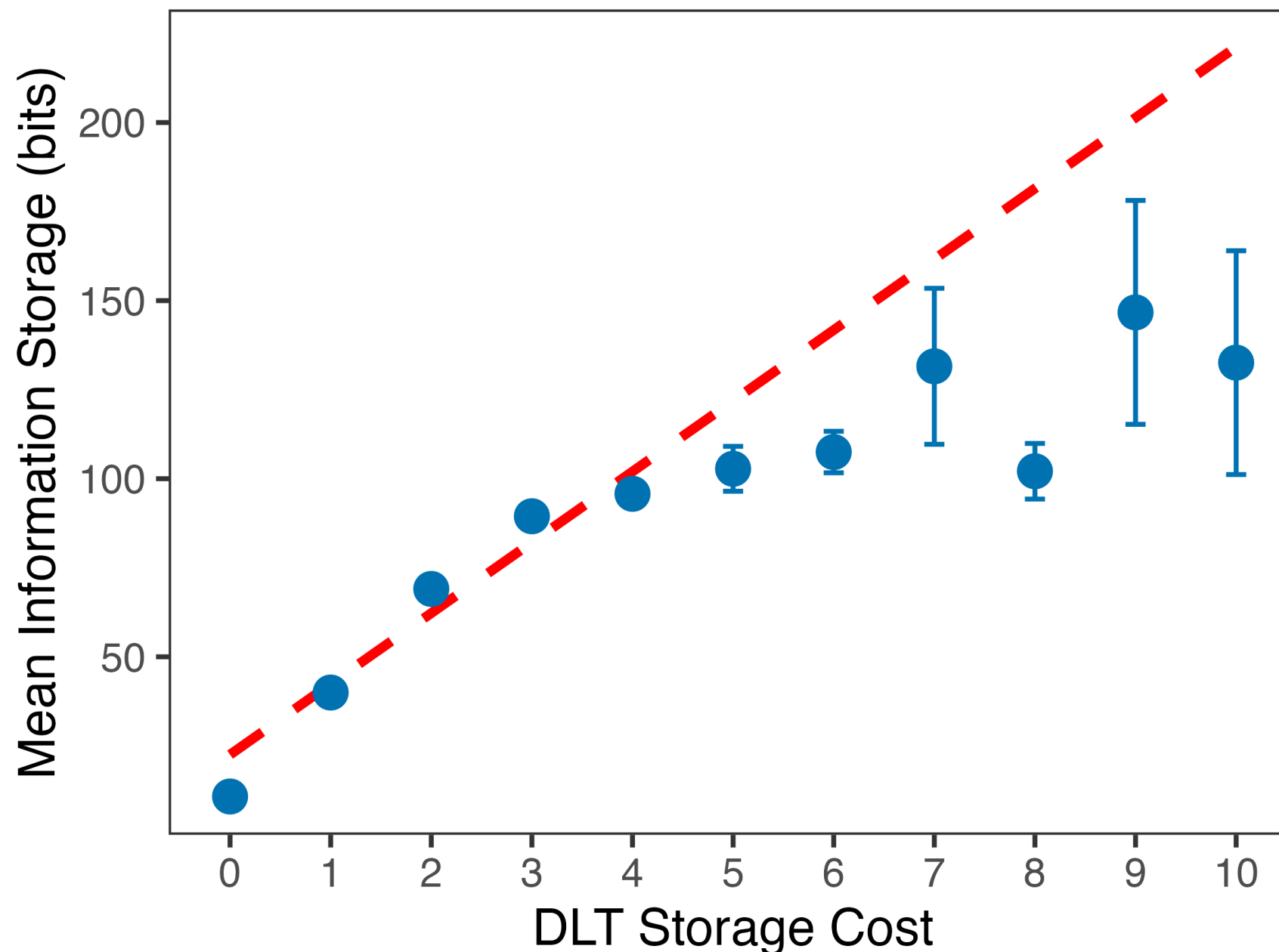
2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GUM [Zeldes 2017] (233,926 words) で確認



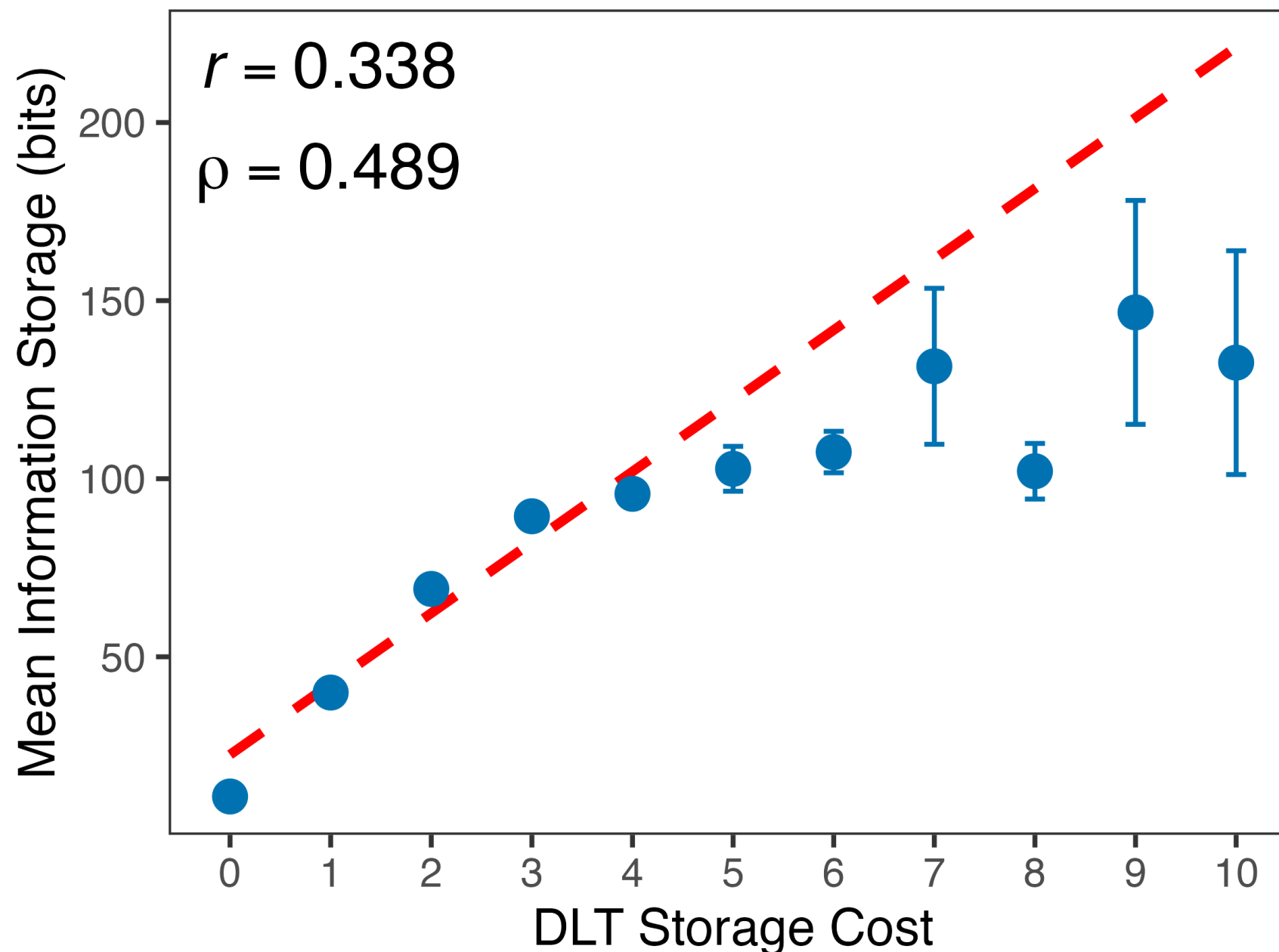
2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GUM [Zeldes 2017] (233,926 words) で確認



2. DLT との相関

- Information Storage は DLT storage と相関するか？
 - UD_English-GUM [Zeldes 2017] (233,926 words) で確認



3. 自然な読み時間での分析

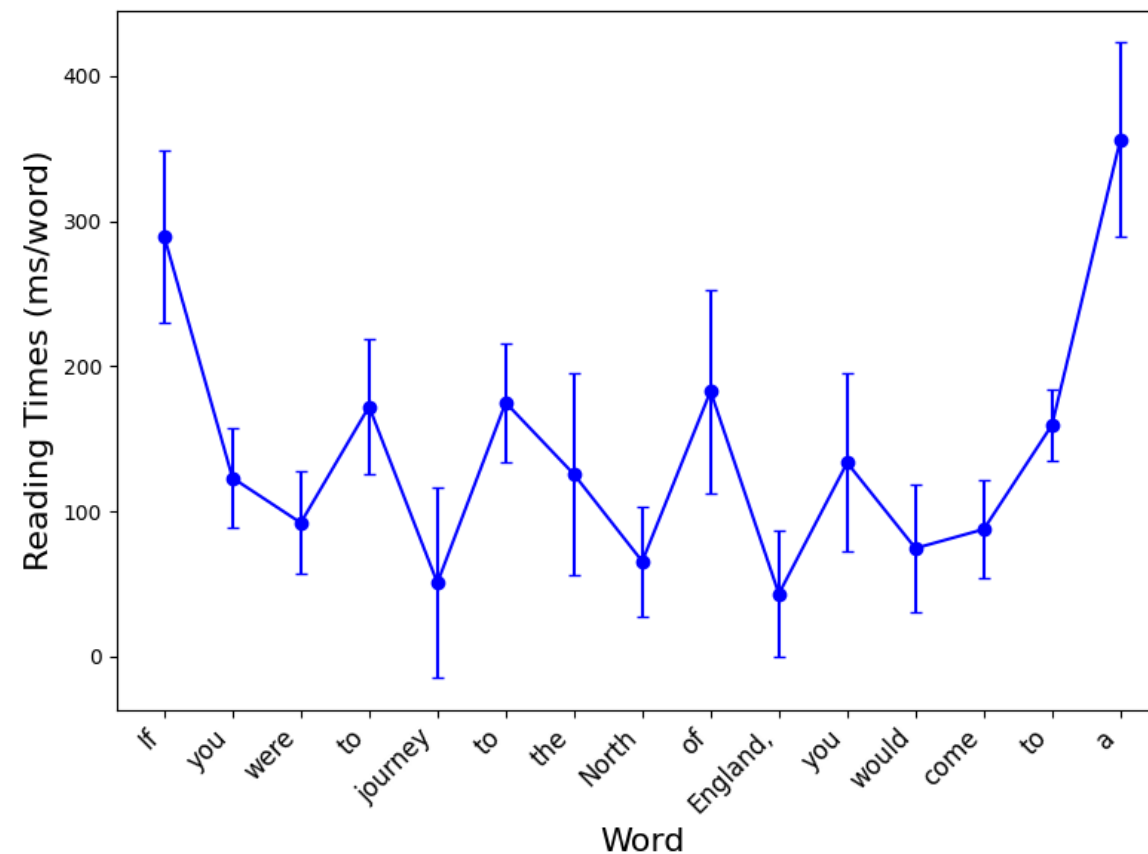
- Information Storage は自然な読み時間を予測できるか？

3. 自然な読み時間での分析

- Information Storage は自然な読み時間を予測できるか？
 - 読み時間の変動を DLT storage より説明できるか？
 - DLT storage は Universal Dependencies で計算

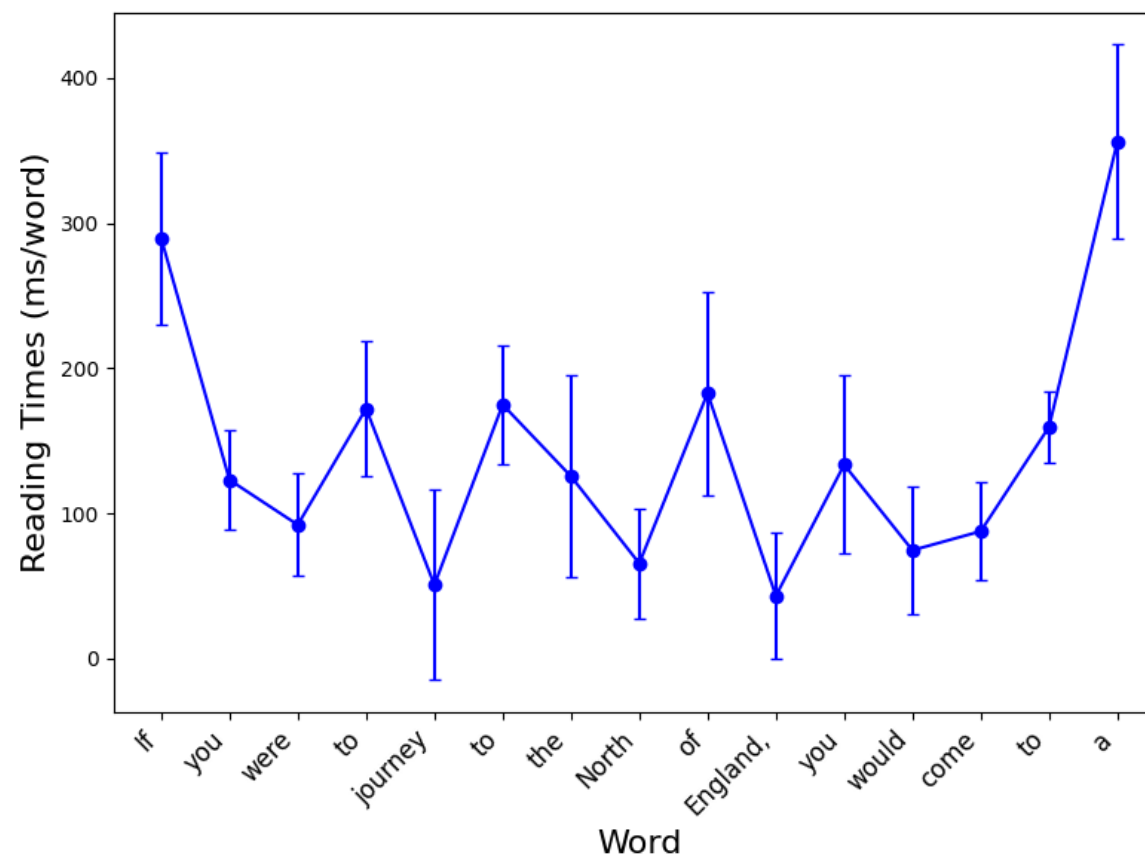
3. 自然な読み時間での分析

- Information Storage は自然な読み時間を予測できるか？
 - 読み時間の変動を DLT storage より説明できるか？
 - DLT storage は Universal Dependencies で計算



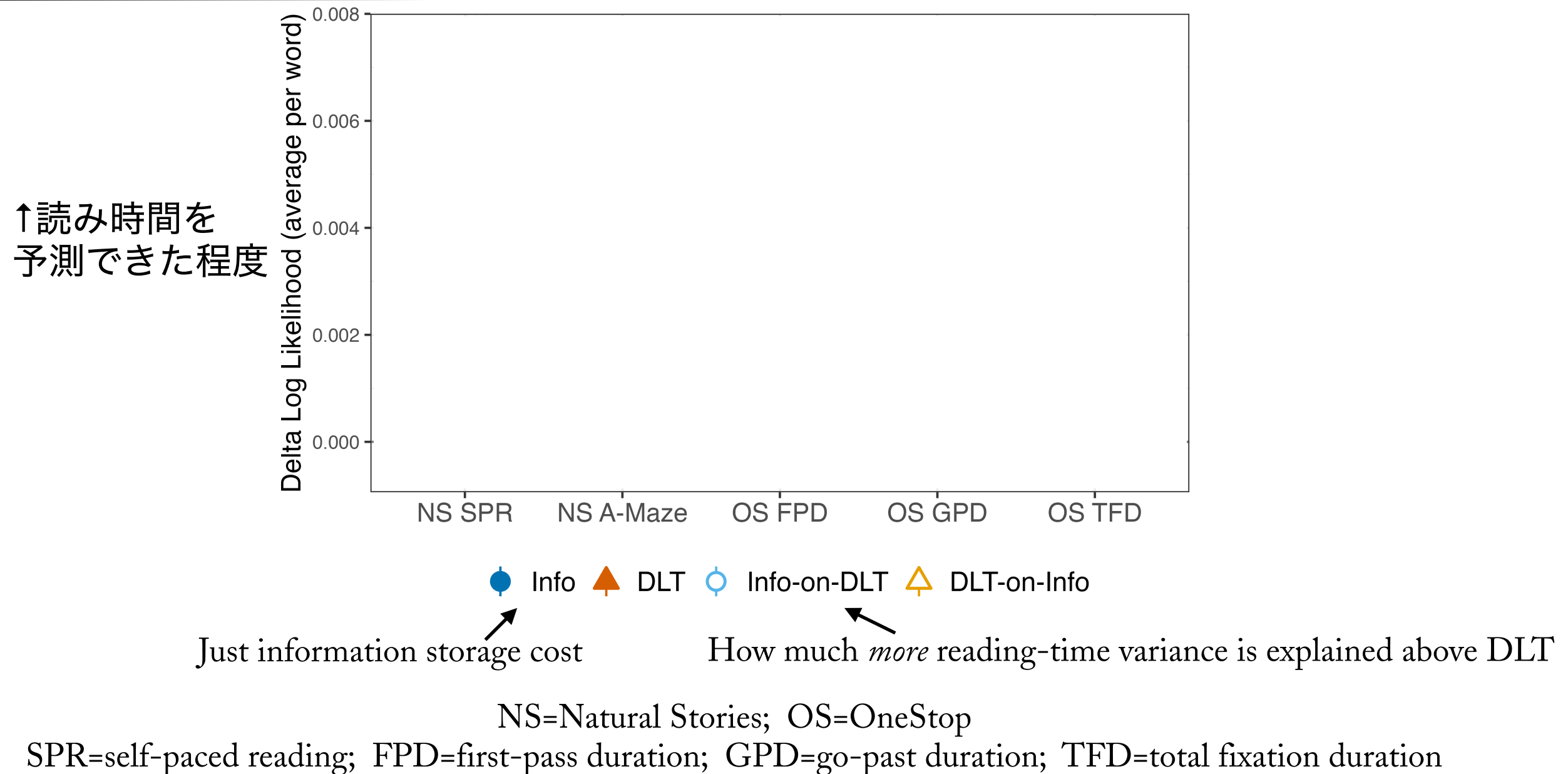
3. 自然な読み時間での分析

- Information Storage は自然な読み時間を予測できるか？
 - 読み時間の変動を DLT storage より説明できるか？
 - DLT storage は Universal Dependencies で計算

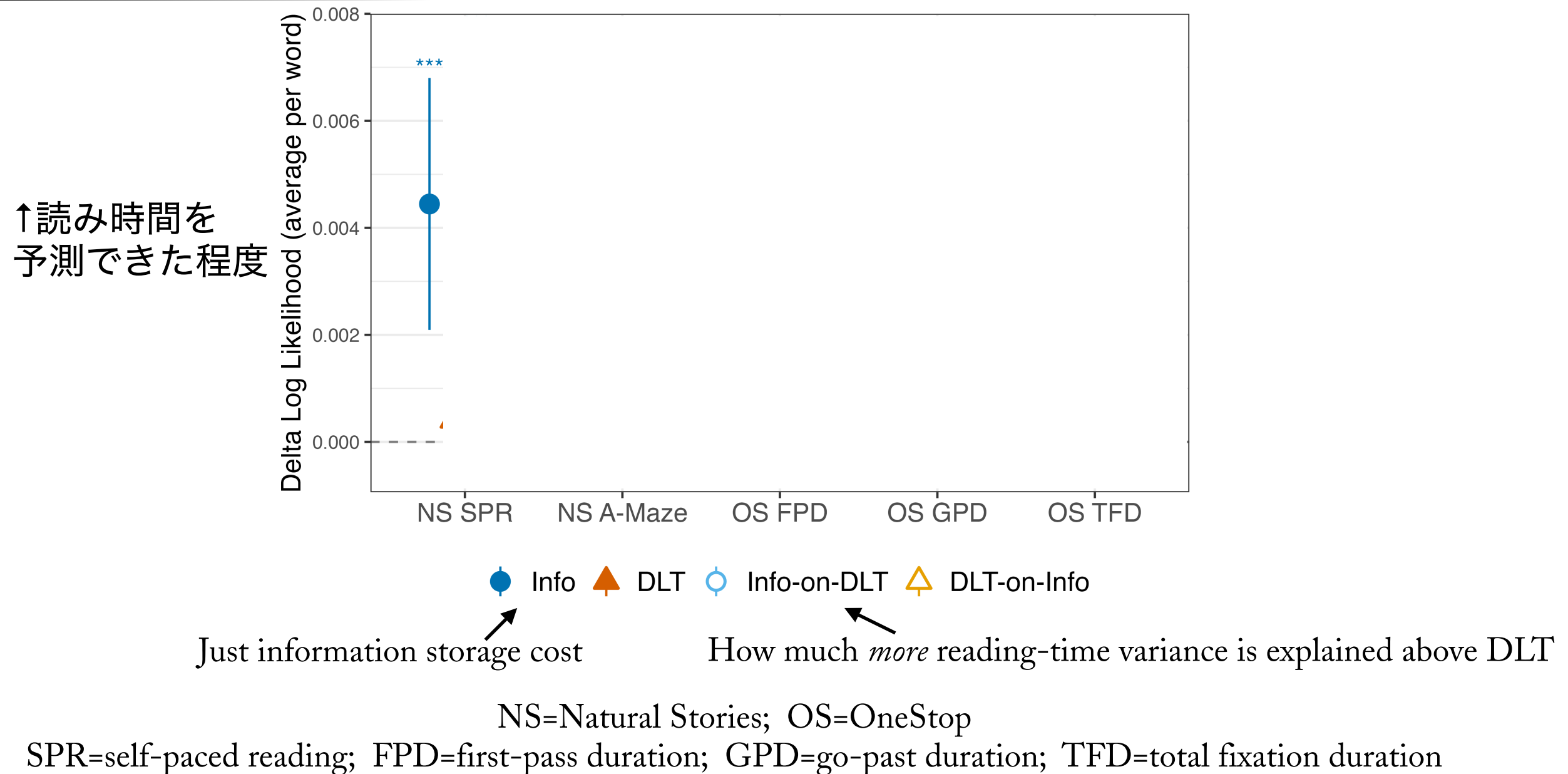


- 大規模な英語のデータセットで検証 (<参加者数>)
 - *Natural Stories* [Futrell+ 2021]
 - *OneStop* [Berzak+ 2025]
 - Self-paced reading (181) [Futrell+ 2021]
 - Eye-tracking (180)
 - A-maze reading (95) [Boyce & Levy 2023]

3. 自然な読み時間での分析

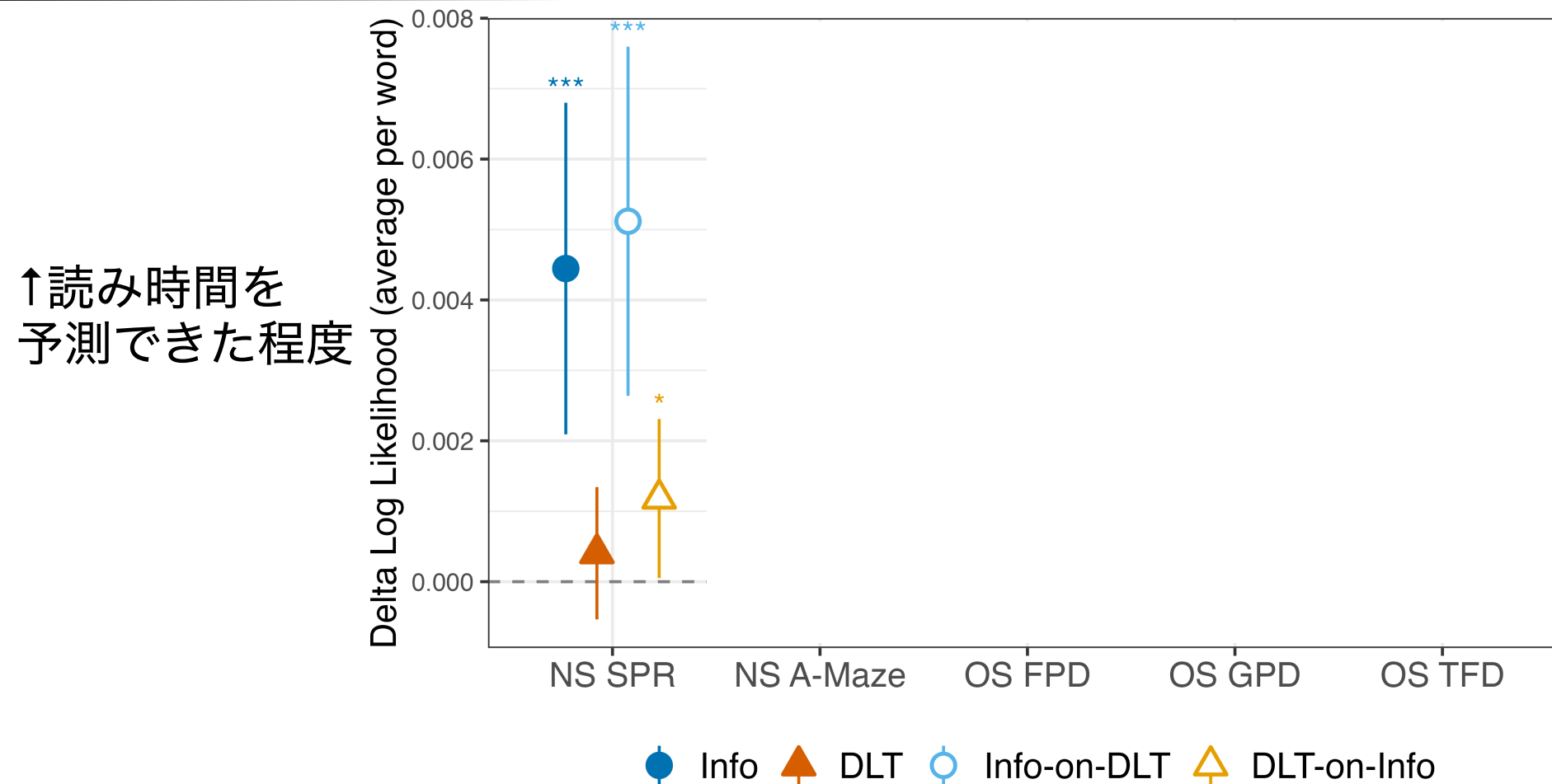


3. 自然な読み時間での分析



- Information storage は有意に読み時間を予測

3. 自然な読み時間での分析



Just information storage cost

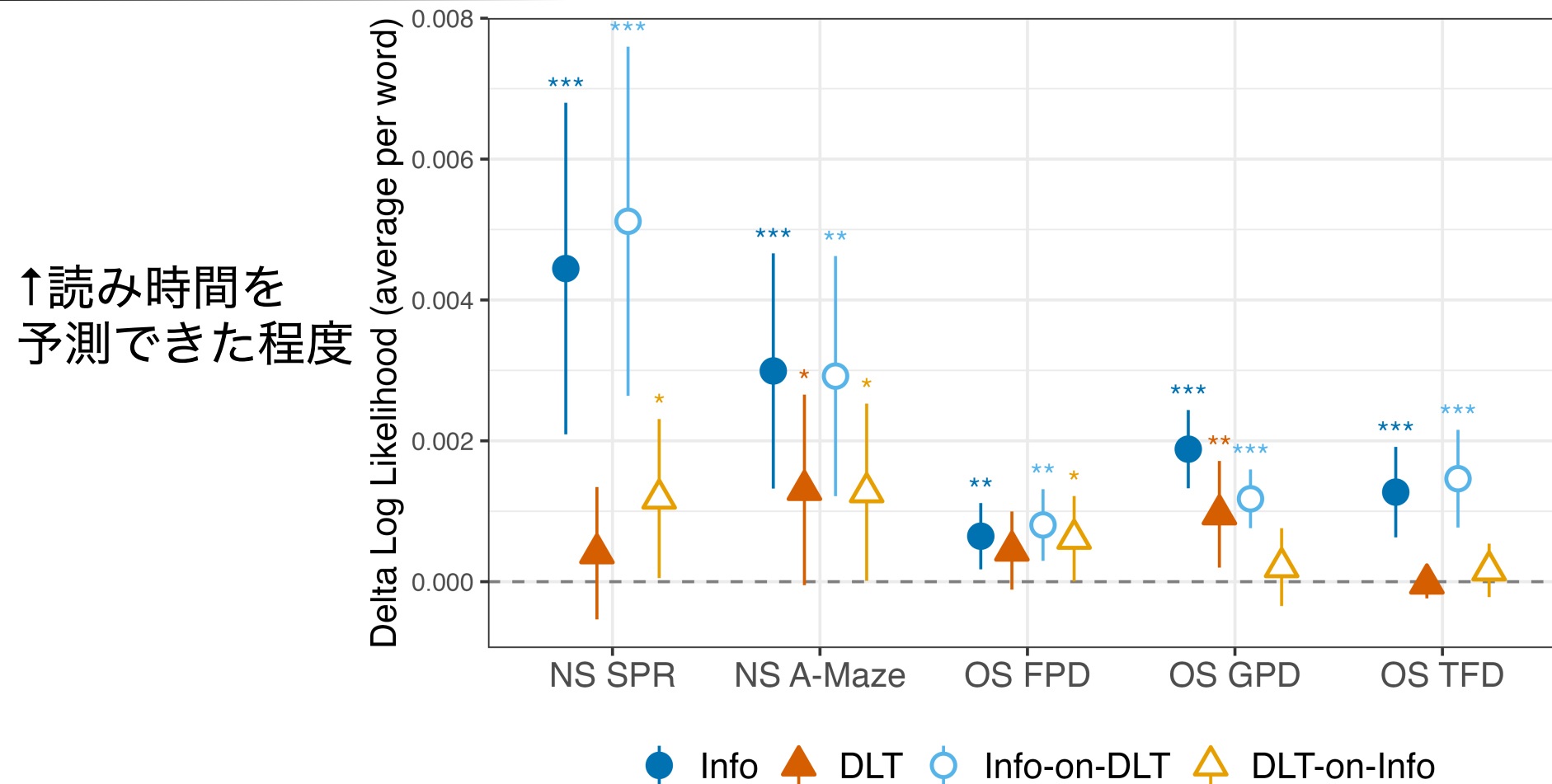
How much *more* reading-time variance is explained above DLT

NS=Natural Stories; OS=OneStop

SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

- Information storage は有意に読み時間を予測
- だが、DLT が捉える全ての変動は説明できていない

3. 自然な読み時間での分析



Just information storage cost

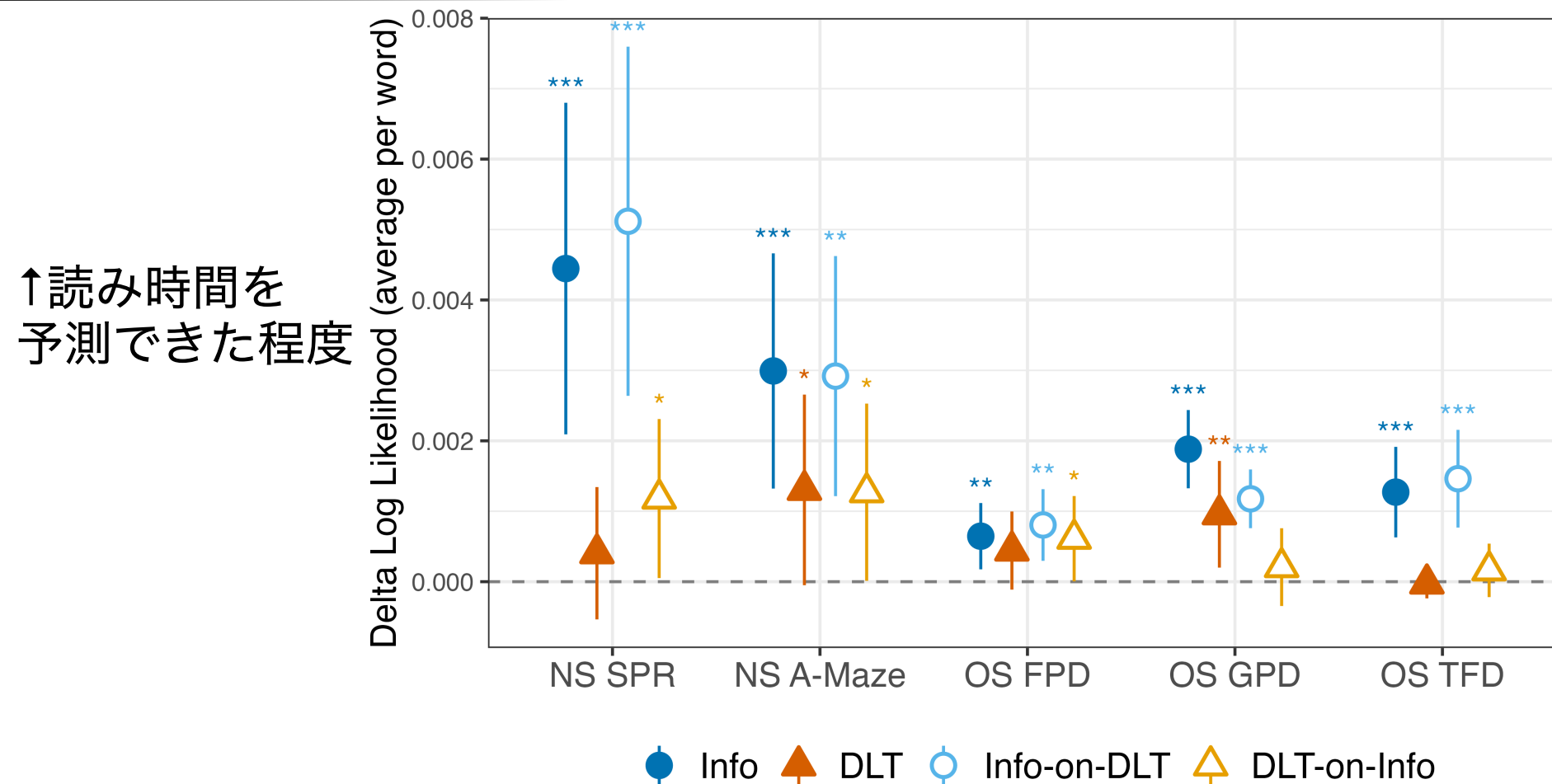
How much *more* reading-time variance is explained above DLT

NS=Natural Stories; OS=OneStop

SPR=self-paced reading; FPD=first-pass duration; GPD=go-past duration; TFD=total fixation duration

- Information storage は有意に読み時間を予測
- だが、DLT が捉える全ての変動は説明できていない

3. 自然な読み時間での分析



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 は有意に読み時間を予測
- だが、DLT が捉える全ての変動は説明できていない

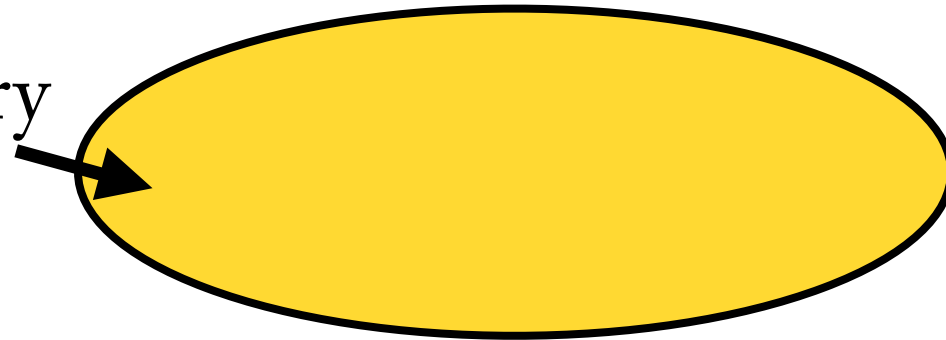
Discussion & Implications

- ストレージコストを過去と未来の (half-pointwise) 相互情報量で定式化

Discussion & Implications

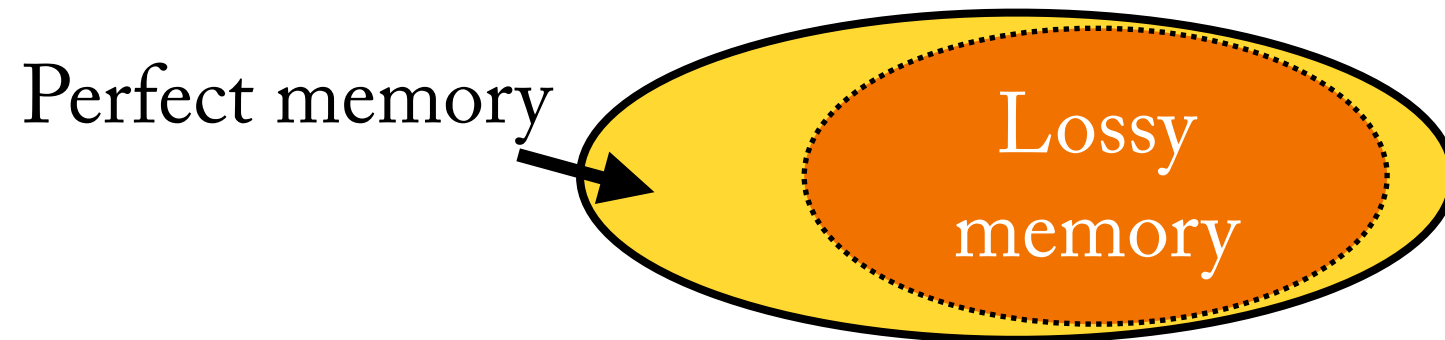
- ストレージコストを過去と未来の (half-pointwise) 相互情報量で定式化
 - 限られた認知資源のもとでの *lossy-context storage* を探る基盤

Perfect memory



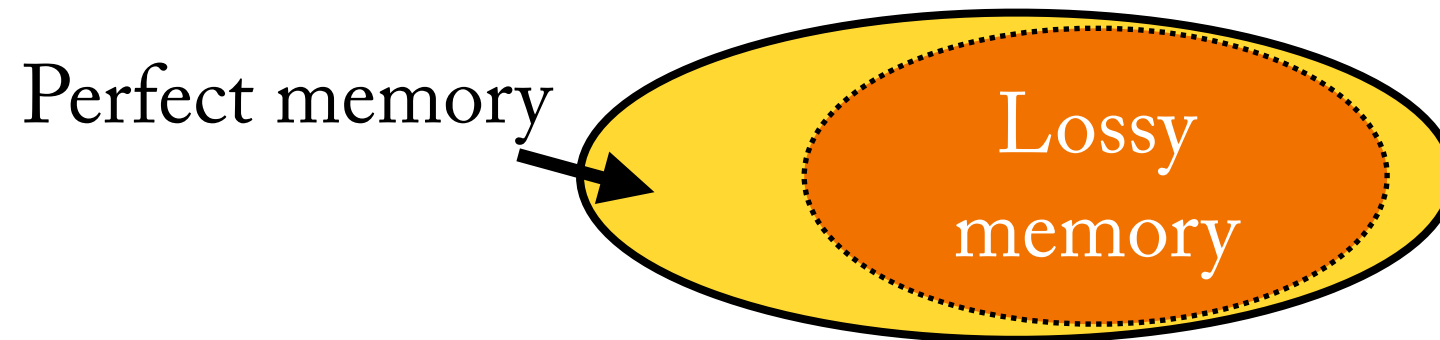
Discussion & Implications

- ストレージコストを過去と未来の (half-pointwise) 相互情報量で定式化
 - 限られた認知資源のもとでの *lossy-context storage* を探る基盤



Discussion & Implications

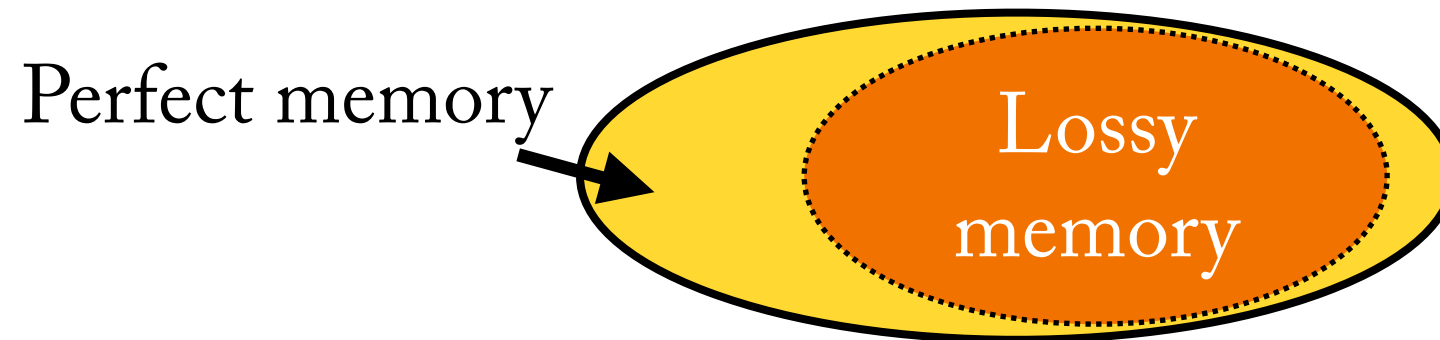
- ストレージコストを過去と未来の (half-pointwise) 相互情報量で定式化
 - 限られた認知資源のもとでの *lossy-context storage* を探る基盤



- Information/DLT storage は読み時間変動の異なるところを捉えている

Discussion & Implications

- ストレージコストを過去と未来の (half-pointwise) 相互情報量で定式化
 - 限られた認知資源のもとでの *lossy-context storage* を探る基盤



- Information/DLT storage は読み時間変動の異なるところを捉えている
 - 統計的な予測と構造的な処理は異なる相補的なメカニズムで動いている [Kajikawa & Isono 2026]

まとめ：情報理論での文理解研究

- 文理解を最も抽象的なレベルの問題として扱う
 - 数式ベースで厳密な議論が可能
 - （今日は紹介しなかったが）合理的分析 (rational analysis)
[Anderson 1990; 2025; LiederGriffiths 2020] との相性が非常に良い

まとめ：情報理論での文理解研究

- 文理解を最も抽象的なレベルの問題として扱う
 - 数式ベースで厳密な議論が可能
 - （今日は紹介しなかったが）合理的分析 (rational analysis)
[Anderson 1990; 2025; LiederGriffiths 2020] との相性が非常に良い
- もちろん、抽象的なのはときに欠点