

Spillover as Rational Processing Delay in Sentence Comprehension



Kohei Kajikawa and Ethan Gotlieb Wilcox

Georgetown University

kk1571@georgetown.edu

Traditional View: *Spillover* is typically treated as a byproduct of modular processing [1,2].

Our Proposal: It reflects a *rational* decision to delay processing under memory constraints.

Prediction: The delay's **benefit**, quantified by *mutual information* (MI) between the current word, W_i , and future words, $W_{>i}$, will be correlated with spillovers. → Confirmed in three reading-time (RT) datasets.

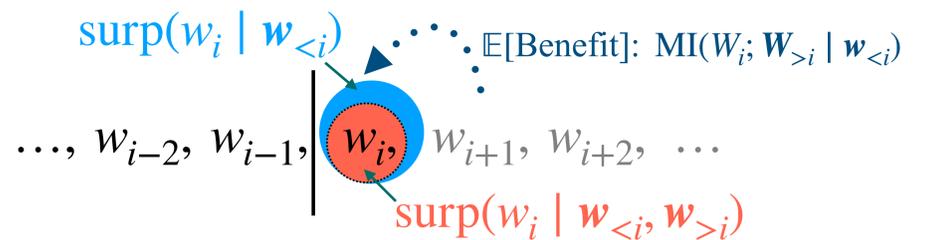
Proposal

- Sentence comprehension is incremental at a *macro* level but not necessarily fully incremental at a *micro* level.
 - Delayed integration: Comprehenders often wait for subsequent input.
 - Reduced surprisal: Integrating words in a richer context lowers processing difficulty.
- Expected benefit of processing delay is quantified by the mutual information between W_i and $W_{>i}$.

$$\Delta \overleftarrow{\text{surp}}_{w_i} := \text{surp}(w_i | \mathbf{w}_{<i}) - \text{surp}(w_i | \mathbf{w}_{<i}, \mathbf{w}_{>i})$$

$$= \text{pmi}(w_i; \mathbf{w}_{>i} | \mathbf{w}_{<i})$$

$$\mathbb{E}_{w_i | \mathbf{w}_{<i}} \left[\mathbb{E}_{\mathbf{w}_{>i} | \mathbf{w}_{<i}, w_i} [\Delta \overleftarrow{\text{surp}}_{w_i}] \right] = \text{MI}(W_i; W_{>i} | \mathbf{w}_{<i})$$



- Prediction: higher benefit at w_i should be associated with greater spillover at w_{i+1} .

Analysis

Estimating word-by-word spillovers

- Two ways of estimating *relative* spillover on Brown [3], Natural Stories [4], and OneStop [5]:

- 1) **Lag-based:** spillover as the contributions of lag predictors

Train: $\text{RT}(w_i) \sim \text{GAM}_\theta(\text{pred}(w_i) + \text{pred}(w_{i-1:i-3}))$

Predict: $\text{Spillover}_{\text{lag}}(w_i) = \sum_w \text{GAM}_\theta(\text{pred}(w_{i-1:i-3}))$

- 2) **Resid-based:** spillover as the residual of RT

Train: $\text{RT}(w_i) \sim \text{GAM}_\theta(\text{pred}(w_i))$

Predict: $\text{Spillover}_{\text{resid}}(w_i) = \text{RT}(w_i) - \text{GAM}_\theta(\text{pred}(w_i))$

- Two types of $\text{pred}(w_i)$ are used:

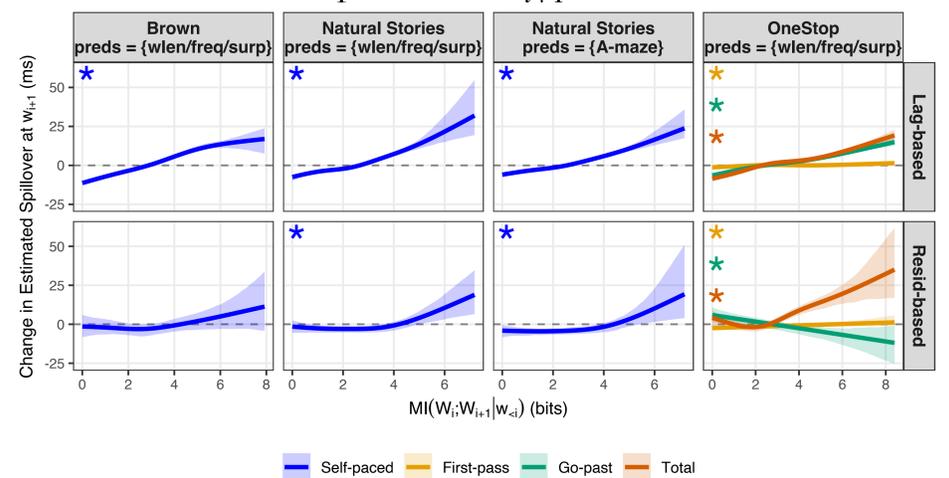
- a) $\text{wlen}(w_i) + \text{freq}(w_i) + \text{surp}(w_i)$

- b) A-Maze(w_i) (for Natural Stories [6])

- We use normalized A-Maze RTs, as this method doesn't suffer from spillover, empirically.

Relationship between MI and spillovers

- Bootstrapped GAM smooths relating MI at w_i to estimated relative spillover at w_{i+1} .

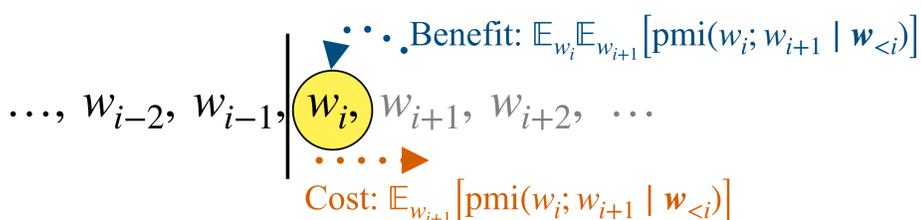


- We find positive relationships supporting the hypothesis.

Next Step

Benefit/cost tradeoff of processing delay

- Developing a *rational* framework to balance the *benefits* and *costs* of processing delays.
 - A temporal extension of surprisal theory.
 - Target phenomena: spillover, deep/shallow processing, word skipping, reanalysis, and relative clause, ...



Information-theoretic storage cost

Information-Theoretic Storage Cost in Sentence Comprehension

Kohei Kajikawa¹, Shinnosuke Isono², Ethan Gotlieb Wilcox¹

¹Georgetown University

²National Institute for Japanese Language and Linguistics (NINJAL)

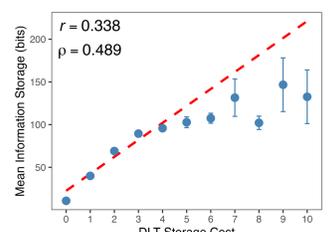
Correspondence: kk1571@georgetown.edu



arXiv preprint

- Maintaining predicted syntactic elements in working memory consumes cognitive resources.
- We redefine syntactic storage cost as a continuous, information-theoretic measure of future predictability.

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



Correlation between DLT storage and information storage in the UD_English-GUM

References:

- [1] Mitchell (1984) [2] Bartek et al. (2011) [3] Smith & Levy (2013)
 [4] Futrell et al. (2021) [5] Berzak et al. (2025) [6] Boyce & Levy (2023)

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