Information Locality in the Processing of English Object Relative Clauses

Hailin Hao (USC), Weijie Xu (UC Irvine),
Michael Hahn (Saarland), Richard Futrell (UC Irvine)

In an effort to provide a unified account for both expectation [1-2] and dependency locality effects [3] in sentence processing, Lossy-Context Surprisal (LCS) [4] proposed that expectations for upcoming linguistic material are not based on a perfectly retained context as in the original Surprisal Theory (Eq. 1) [2], but on imperfect representations of the context, constrained by working memory limitations (Eq. 2). LCS predicts an interaction between expectation and locality effects, termed *information locality*, whereby the processing benefit for two words that highly predict each other decreases when they are separated in linear distance. In this work, we test this prediction using English object relative clauses [5-6].

Experimental Design. The dependency we looked into is between the extracted object and its head verb (e.g., "lawn" and "mowed" in (1a)). To create varying levels of expectations, we manipulated the Pointwise Mutual Information (PMI, as defined in Eq. 3; high vs. low) between the two codependents. PMI was crossed with the Dependency Length (DL; short vs. long) between the two codependents in a 2x2 design, as in (1) (N = 48; 24 target items). The experiment was conducted online using A-Maze [7].

Hypothesis and Predictions. According to Surprisal Theory, the processing difficulty of a word is proportional to its negative logarithm of a word's contextual probability, representing its information content. It can be rewritten as the difference between the word's information content without the context and its PMI with the context, as in Eq. 3, which can be further rewritten as in Eq, 4, whereby PMI(w, c) is replaced with the sum of the PMI between the current word and each past word in the context [4]. Therefore, if the context contains a word that has high PMI with the current word, the surprisal of the current word will be reduced, leading to decreased processing difficulty of the current work. LCS additionally predicts that the facilitation from high PMI will be weakened as DL increases. This is because LCS posits that expectations are formed based on imperfect contexts. That is, a progressive noise term can be added to the PMI between the current word and each past word, as in Eq. 5 [4]. Since the noise term increases as distance increases, the further apart two words are, the less the surprisal will be reduced, therefore causing more processing difficulty.

Results. We fitted a Bayesian linear mixed effects model on the RTs of the right codependent (formula in Eq. 6). As seen in Fig. 1, first, we found a main effect of categorical PMI (β = -88, CrI = [-135, -43]), whereby higher PMI reduces the RT of the critical verb, consistent with expectation-based theories. We also found evidence for locality effects , whereby increased DL leads to longer RTs (β = 151, CrI = [98, 208]). There is also a positive interaction between the two factors (β = 146, CrI = [55, 240]), whereby the effect of PMI is reduced with long DL, and this interaction effect is mostly driven by a neutralization of the PMI effect at long DL. Consistent with our prediction, the benefit of higher PMI in predicting the right co-dependent is compromised when there is more intervening material.

Discussion. By showing a reduced benefit from predictive information for longer dependency length, we demonstrate that predictive processing is subject to WM constraints, such that degraded representations of the linguistic context can lead to less sharp expectations. Our result provides an empirical support for sentence processing models, such as LCS, that incorporate memory-based mechanisms into expectations.

Eq. 1: processing difficulty $\propto$ surprisal(w) = $-\log P(w|c)$
Eq. 2: processing difficulty $\propto$ lossy surprisal(w) = $-\log P(w|c') = -\log \sum_{c} P(w|c)\ P(c|c')$

Eq. 3: PMI = $\log \frac{p(wi, wj)}{p(wi)p(wj)}$

Eq. 3: Surprisal(w) = P(w) – PMI(w, c), PMI(w, c) = $\log \frac{p(w, c)}{p(w)p(c)}$

Eq. 4: Surprisal(w_i) ≈ P(w_i) – $\sum_{j=i}^{i-1}$ PMI(w_i, w_j), PMI(w_i, w_j) = $\log \frac{p(wi, wj)}{p(wi)p(wj)}$

Eq. 5: Surprisal($w_i|c$) ≈ P(w_i) - $\sum_{j=i}^{i-1}(1 - e_{i\text{-}j})$PMI($w_i$, w_j)

Eq. 6: logRT ~ HDMI * DL + (HDMI * DL | subject) + (HDMI * DL | item)

(1) Sample stimuli. The extracted object is bolded. The critical region (i.e., main verb) is underlined.
   a. High PMI; Short Distance
      The **lawn** that the man mowed was green and healthy.
   b. High PMI; Long Distance
      The **lawn** that the man who wore a blue cap mowed was green and healthy.
   c. Low PMI; Short Distance
      The **lawn** that the man watched was green and healthy.
   d. Low PMI; Long Distance
      The **lawn** that the man who wore a blue cap watched was green and healthy

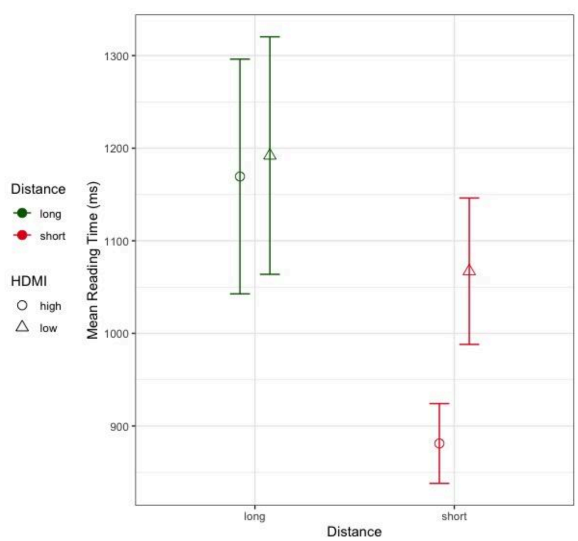

Fig. 1: A-Maze RTs at the critical region

References: [1] Hale (2001) NAACL. [2] Levy (2008) Cognition. [3] Gibson (1998) Cognition. [4] Futrell, Gibson, & Levy (2020) Cogn. Sci. [5] Grodner & Gibson (2005) Cogn. Sci. [6] Bartek et al. (2011) JEP. [7] Boyce, Futrell, & Levy (2020) JML.