Information entropy facilitates lexical processing

Hossein Karimi, Jaden Zinn, & Haomiao Li (Mississippi State University)

Introduction: Past research has shown that comprehenders employ preceding contextual information to generate predictions about upcoming words, and that constraining contexts facilitate the processing of more predictable words (e.g., *banana* in 1c vs. 1a; e.g., Altmann & Kamide, 1999; Brothers et al., 2015; Federmeier & Kutas, 1999). Lexical predictability is traditionally quantified by either cloze probability (Taylor, 1953), or information-theoretic metrics such as surprisal (Levy, 2008) and entropy (Hale, 2003). However, along with sentence constraint, all these measures are calculated per item (i.e., collapsing over participant data) and, as such, may miss nuanced information at the *trial* level (i.e., for each participant, for each item). We examined the potential effect of *trial*-level entropy on lexical processing during natural reading. Although entropy is commonly interpreted as reflecting "uncertainty" about upcoming information, it actually poses an interesting dilemma. On the one hand, if lexical prediction involves pre-activation of full lexical items (e.g., *banana*, *mango* etc.), greater entropy may lead to competition among the lexical candidates, increasing processing cost (e.g., Chen & Mirman, 2012). On the other hand, if prediction involves pre-activation of low-level semantic features (e.g., [*+edible*], [*+tropical*], etc.), greater entropy may lead to stronger pre-activation of more semantic features, thereby facilitating the processing of any upcoming word that shares those features (e.g., Brothers et al., 2023; Luke & Christianson, 2016). Our own prior study (N = 112), showed a facilitatory effect of trial entropy on self-paced reading times on the target word (e.g., *banana*). The current study is an attempt to replicate these findings during *natural* reading.

Method: 48 native speakers of English completed two separate experimental sessions 7-21 days apart. In the "cloze session", they provided up to 8 completions to 216 sentence fragments with varying contextual constraints as illustrated in (1) (the target word *banana* and the following words were removed), along with their estimated probability for each word, as shown Table 1. Trial-level entropy values were calculated using the given responses and normalized probability values (by dividing the original values by the total number of responses). In the "reading session", participants read the same sentences in full while their eye movements were tracked. The cloze session always came first so that the participants were never exposed to the target words.

Results & Discussion: We used mixed-effects regression models including one item-level measure of lexical predictability (either constraint, cloze, or surprisal) as well as the critical trial-level entropy on the target word (*banana*). The correlations between the predictors are reported in Table 2. Because "number of responses per trial" had a stronger effect than trial entropy, we used this predictor instead. Note that this observation suggests that the subjective probability values may be somewhat unreliable. We observed a significant effect of number of responses on regression probability, and total reading times, with greater number of responses, leading to lower regression probability (Figure 1), and faster reading times (Figure 2; see Table 3). The main effects of constraint, cloze probability and surprisal were all significant, with higher constraint, higher cloze probability and lower surprisal leading to lower regression probability and faster total reading times. The effects did not change when cloze was replaced with constraint or surprisal. Trial entropy (which included the probabilities) had a significant facilitatory effect for higher cloze sentences on the two spillover regions (i.e., *at breakfast*).

General Discussion: The results suggest that trial entropy facilitates lexical processing, and that prediction during language processing corresponds to pre-activation of semantic features rather than full lexical items. The results also suggest that rather than capturing uncertainty, entropy may actually capture degrees of *certainty* about the general semantic domain of upcoming information.

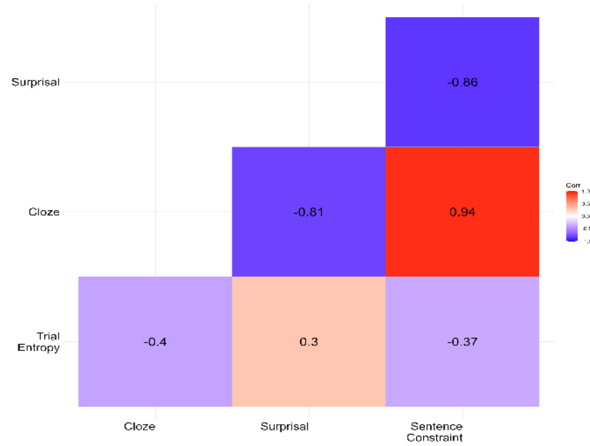
(1) Sample experimental item:

a) Low constraint:	My picky younger brother refused to eat the <u>banana</u> at breakfast before the others.
b) Medium constraint:	Because of fruit flies, he threw away the <u>banana</u> at breakfast before the others.
c) High constraint:	The hungry monkey was trying to peel the <u>banana</u> at breakfast before the others.

Table 1*. Sample responses for(1c)*

	Word	**Probability**
word1	banana	95
word2	orange	60
word3	mango	50
word4	...	...
word5		
word6		
word7		
word8		

Table 2. Correlations between predictors.



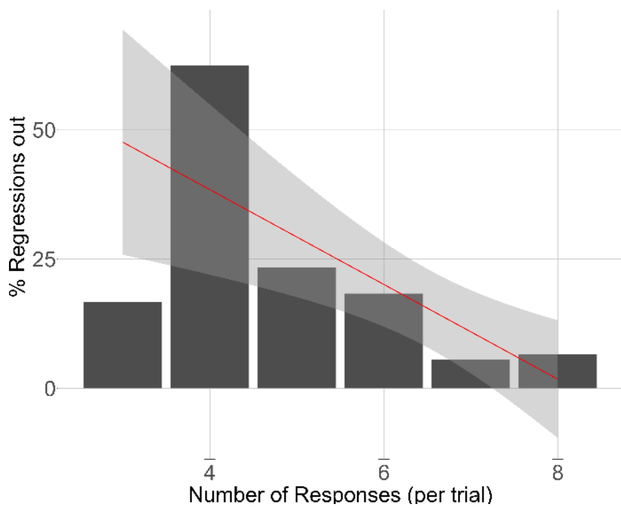


Figure 1. Regression probability on the target word (banana).

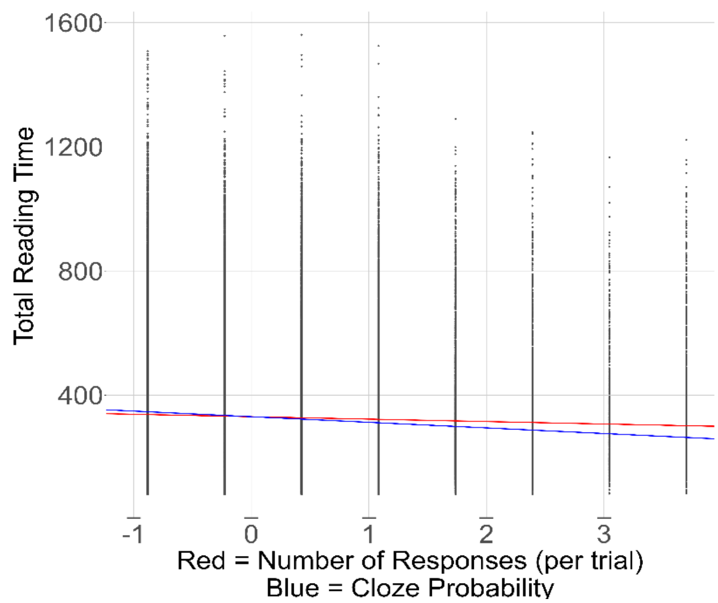


Figure 2. Total reading times on the target word (banana).

Table 3*. Regression results on the target word (banana).*

Region	**Dependent var.**	**Predictor**	***Est.***	***SE.***	***t***	***p***
target (*banana*)	Regression Probability	N Responses	-.13	.06	-2.23	.02
		Cloze	-.14	.04	-3.07	.002
		Interaction	-.07	.05	-1.38	.16
	Total reading time	N Responses	-.02	.008	-2.43	.01
		Cloze	-.04	.006	-6.80	<.001
		Interaction	.008	.008	1.02	.30