Interactive Effects of Frequency and Semantic fitness on Sentence Processing

Introduction.

The frequency of linguistic input is a key factor in predicting sentence processing difficulty, motivating frequency-based probabilistic models of sentence processing (Hale, 2001; Levy, 2008). Another important factor is semantic fitness, which indicates how consistent the linguistic input is with world knowledge. More recently, DeLong et al. (2014) and Kang et al. (2018) provided some evidence suggesting semantic knowledge can be subdivided into predictability and plausibility, which impose separate effects on sentence processing. However, the former studies did not test the effect of semantic fitness across all levels of frequency. Put differently, despite the possibility that semantic knowledge and frequency interrelate to each other, little attention has been paid to their interactive effect. Thus, the present study aims to investigate in a more controlled way what is the interactive effect of world knowledge and frequency in sentence processing, by disentangling thematic-fit ratings and frequency effects. Our results suggest that there is an interaction between thematic-fit and frequency in on-line sentence processing, implying that both world knowledge and frequency of linguistic input affects sentence processing separate.

Method.

We designed a self-paced reading experiment, where the independent variables are: (i) dependency frequency of the verb and its direct object drawn from the Google Syntactic N-gram corpus (Goldberg & Orwant, 2013), and (ii) a thematic-fit (patienthood) rating of the latter word pairs drawn from Experiment 2A of McRae (2010). 12 sets of four items were selected with 2x2 conditions : (1) high vs. low thematic-fit and (2) high vs. low frequency, as illustrated in the items in Table 1. Ratings of patienthood greater than 4 were considered high while less than 4 were considered low, and dependency frequencies greater than 10 were considered high while 0 frequencies considered low. Four lists of sentences were set, in Latin Square Design, having three target sentences from each condition and 24 distractor sentences half of which were followed by comprehension tasks. 10 participants were assigned to each list. The structure of target sentences maintained the same structure having a fifth region as a critical region and sixth region as a spill-over region.

Sentences	Condition	Thematic-fit	Dep Freq
She 1| wanted 2| to 3| accuse 4| the criminal 5| of 6| vandalism. 7|	HighFit/HighFreq(HH)	6.9	62
She 1| wanted 2| to 3| accuse 4| the defendant 5| of 6| vandalism. 7|	HighFit/LowFreq(HL)	6.8	0
She 1| wanted 2| to 3| accuse 4| the prosecutor 5| of 6| vandalism. 7|	LowFit/HighFreq(LH)	2.6	42
She 1| wanted 2| to 3| accuse 4| the witness 5| of 6| vandanlism. 7|	LowFit/LowFreq(LL)	3.2	0

Table1. Item examples with regions, in each of the four conditions

The data from three participants were removed from the analysis for having comprehension question accuracy below 75%, and data points were excluded if the reading time for the corresponding region was less than 100ms or more than 1500 ms. After residual reading time was computed for each participants and according to region length, linear mixed effect models were fit to compare across different conditions at the regions 5 and 6.

Results.

As shown in Figure 1., the HL items were read slower than HH items at region6 ($\beta = 41.50$, $t = 2.72$, $p < 0.01$). Similarly, the LH times were also slower than HH items ($\beta = 29.52$, $t = 2.00$, $p < 0.05$). Subsequent investigation revealed a main effect of frequency ($\beta = 41.61$, $t = 2.68$, $p < 0.01$), near-significant main effect of thematic fit ($\beta = 28.71$, $t = 1.88$, $p = 0.06$), and an interaction between two ($\beta = -45.59$, $t = -2.09$, $p < 0.05$), seen in Figure 2. In other words, frequency only predicts reading time differences when the thematic-fit is high.
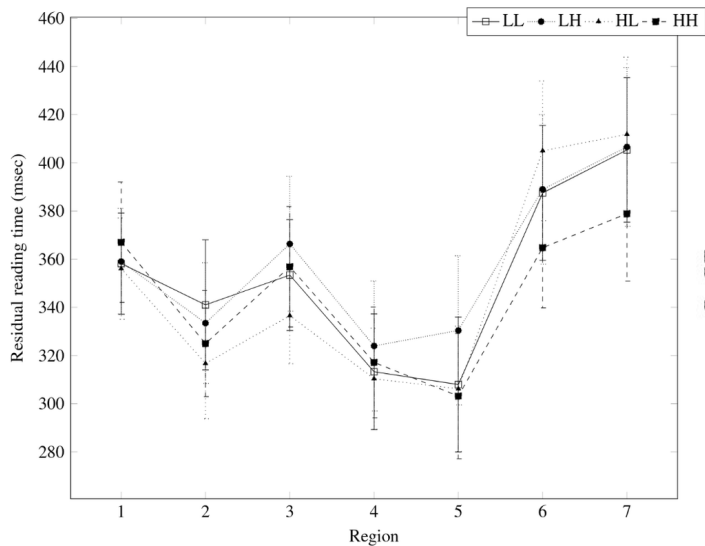

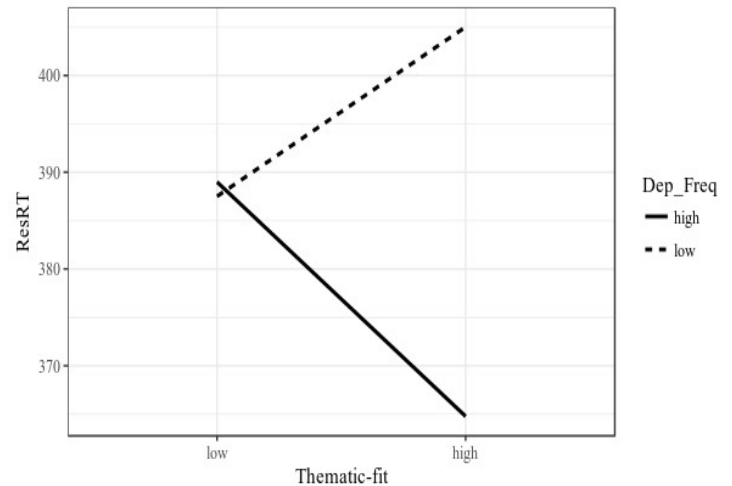Figure 1: Mean residual reading times for all sentence regions

Figure 2: Interaction beween Thematic-fit and Frequency on region 6

Another linear mixed effect regression model was fit on region 6 to determine whether the raw word frequency on the verbs in region 5 predicts the reading time differences in region 6, but no significant effect was found ($\beta = -5.01$, $t = -1.88$, $p = 0.06$). In conclusion, our results suggest that the effect of frequency occurs only when the linguistic input is highly plausible, and that both frequency and semantic fit need to be taken into consideration in models of sentence processing.

References.

Goldberg, Y., & Orwant, J. (2013). A dataset of syntactic-ngrams over time from a very large corpus of english books. In *SEM conference. Proc. of the Main Conference and the Shared Task: Semantic Textual Similarity* (Vol. 1, pp. 241-247).

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. NAACL Conference, pp 159–166.

Kang, Hong Mo & Koenig, Jean-Pierre & Mauner, Gail. (2018, March). *Plausibility is not reducible to predictability*. 31st CUNY Conference, Davis, CA.

Levy, R. (2008a). Expectation-based syntactic comprehension. Cognition, 106:1126–1177.

Mcrae, K. (2010). Thematic Fit Ratings from a number of studies [Data file]. Retrieved from https://sites.google.com/site/kenmcraelab/norms-data.