Introduction

Processing sentences is known to be affected by variety of features. On one hand, frequency of linguistic input is considered as an important factor in predicting sentence processing difficulty, motivating frequency-based probabilistic models for sentence processing (Hale, 2001; Levy, 2008). On the other hand, there have been researches where the way world knowledge affects sentence processing is investigated (Kang, 2018). However, to our knowledge little attention has been paid to the net effect of frequency and world knowledge and their interactive effect despite the fact that they are highly interrelated to each other. To be specific, there is a strong possibility that the more semantically felicitous linguistic input is, the more likely it occurs in the real world, making it obscure to determine whether the ease of frequent or semantically felicitous sentence processing is attributed to the frequent exposure to the input or to their consistency with people's world knowledge. Therefore, this study aims to investigate the interactive effect of frequency and world knowledge in sentence processing by disintegrating the thematic-fit ratings and the frequency. The result shows that there is an interaction between thematic-fit and frequency in processing sentence, implying that both world knowledge and frequency of linguistic input affects sentence processing in a discrete way and that they need to be considered separately for modeling of sentence processing .

Method

We designed a 2x2 within subject self-paced reading experiment, where one independent variable is a frequency of pair of a verb and its corresponding direct object and the other independent variable is a thematic-fit(patienthood) ratings of them. By using Google Syntactic N-gram(Goldberg & Orwant, 2013) and data from for extracting frequency data and thematic-fit ratings respectively, 12 sets of stimuli were selected where four sentences of each condition are included: 1) high thematic-fit/high frequency, 2) high thematic-fit/low frequency, 3) low thematic-fit/high frequency, 4) low thematic-fit/low frequency. Every sentence had a fifth region as a critical region and sixth region as a spill-over region. The structure of sentences maintained the same structure for minimizing possible contamination effects. Four lists were made where three sentences per each condition are included in Latin-square design. 10 participants were assigned to each list. Example of a set of stimuli is given in Table 1.

Before analysis, participants who had accuracy of comprehension tasks less than 70 % were removed and some of reading data were excluded if the reading time for the corresponding region was less than 100ms or more than 1500ms or it is 2 standard deviations away from the mean. After residual reading time was computed by each participants and word lengths. With the residual reading time, each region was analyzed with using linear mixed effect model in R to see if there is any significant difference on residual reading time by conditions.

Sentences	Conditions	Thematic-fit	BigramFreq
They|1 failed|2 to|3 tow|4 the car|5 during|6 the storm.|7	HighFit/HighFreq(HH)	6.1	22
They|1 failed|2 to|3 tow|4 the trailer|5 during|6 the storm.|7	HighFit/LowFreq(HL)	6.5	0
They|1 failed|2 to|3 tow|4 the man|5 during|6 the storm.|7	LowFit/HighFreq(LH)	1.5	10
They|1 failed|2 to|3 tow|4 the family|5 during|6 the storm.|7	LowFit/LowFreq(LL)	1.2	0

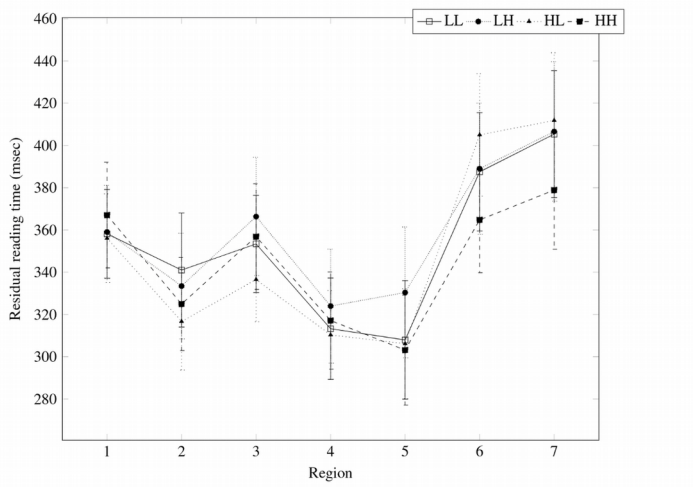
<Table 1. Example set of target sentences>

Results

While no significant differences among conditions is found in any other regions, in region 6, which is the spill-over region, shows a significant difference in processing time by conditions (Figure 1). Closely looking into the relation between thematic-fit ratings and frequency, an interaction between

two factors is found (Figure 2). The interactive relation is revealed to be attributed to the difference of frequency effect dependent on the condition of the thematic-fit ratings. In other words, the predictive power of frequency only exists when the thematic-fit is high, which is the case that the target sentence is harmonious with world knowledge. The noteworthy result here is that the condition where thematic-fit is high and the frequency is low (HL condition) showed the slowest reading time. It is an interesting pattern since we have hypothesized that when the thematic-fit is controlled to be same, the processing time for the pairs for verbs and their direct objects would be reduced when they occur more frequently. Here are two possible interpretations for this unexpected phenomenon: 1) the lexical frequency in region 5 in HL condition would be relatively low compared to other conditions 2) low frequent pairs with high thematic-fit are the the least predicted during on-line sentence processing for it is counterintuitive that semantically natural pairs occur infrequently resulting in the slowest processing time. To check our first hypothesis, additional linear mixed effect regression was carried out on region 6, which is designed to see whether or not the word frequency in region 5 predicts the processing time in region 6, and not a significant difference, was found even though word frequency in region 5 is proven to be a predictor of the processing time in region 5.

In conclusion, by separating thematic-fit from frequency, it is shown that the frequency effect of linguistic input occurs only when the linguistic input is coherent with the world knowledge. This implies that the prediction of sentence processing difficulty can be accomplished neither only with thematic-fit nor with frequency. Rather, both factors should be taken into consideration to make better prediction of sentence processing difficulty.



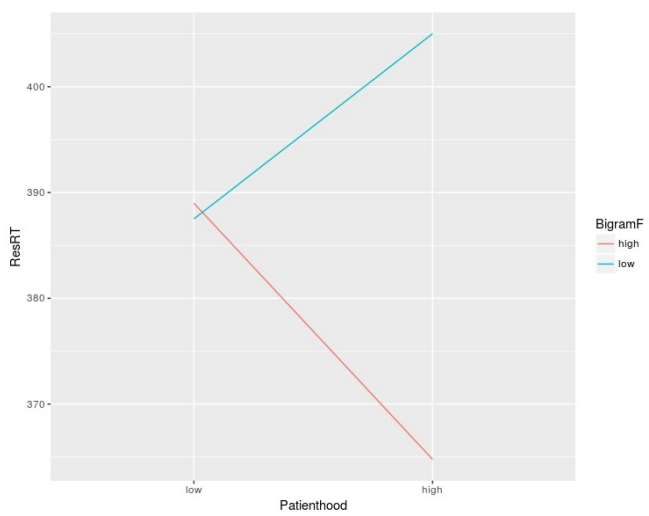


Figure 1: Mean residual reading times for all sentence regions

References.

Goldberg, Y., & Orwant, J. (2013). A dataset of syntactic-ngrams over time from a very large corpus of english books. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 1: Proceedings 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. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, pages 159–166.

Kang, Hong Mo & Koenig, Jean-Pierre & Mauner, Gail. (2018, March). *Plausibility is not reducible to predictability.* Oral presentation at the 31st Annual CUNY Human Sentence Processing Conference, Davis, CA.

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