The Effect of Semantic Similarity is a Function of Contextual Constraint

Hongoak Yun (hyun3@buffalo.edu)
Gail Mauner (mauner@buffalo.edu)
Department of Psychology, 204 Park Hall
Buffalo, NY 14260 USA

Douglas Roland (droland@buffalo.edu)
Jean-Pierre Koenig (jpkoenig@buffalo.edu)
Department of Linguistics, 609 Baldy Hall
Buffalo, NY 14260 USA

Abstract

We investigate how the degree to which a context constrains the words that could occur in a sentence affects the processing of the word that does occur. Roland et al. (2012) found that processing was facilitated when target words were more semantically similar to word alternatives that could have appeared. Because this effect is independent of word predictability, it suggests that comprehenders may have separate expectations for words and more general semantic features. We show that the semantic similarity effect is modulated by the degree of contextual constraint. We found that facilitation due to semantic similarity was greater when contexts were less constraining, and lower when contexts were more constraining, independent of word predictability. We interpret these results as suggesting that in highly constraining contexts, comprehenders may expect specific words, and face difficulties when these expectations are violated, while in less constraining contexts, they may have more general expectations for semantic properties shared between the words that could occur.

Keywords: sentence processing; semantic similarity; predictability; entropy; contextual constraint; expectation-based language comprehension

Introduction

In expectation-based models of sentence comprehension, contextual information has an enormous effect on how words are integrated into sentences. These models predict that the degree of difficulty a reader encounters in integrating a new word into a sentence is either entirely or in large measure a function of how predictable that word is given prior context (e.g., Levy, 2008). Presumably this is because predicted words are activated by context in advance of when they are encountered, making them easier to retrieve from memory or because predictable words are easier to integrate into the representations being constructed during comprehension. The effect of predictability on processing time has been observed in many studies (e.g., Bicknell, Elman, Hare, McRae, & Kutas, 2010; Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; Staub, 2011; DeLong, Urbach, & Kutas, 2005; Federman, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Otten & Van Berkum, 2008; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

The relationship between a word’s predictability and the amount of effort required to process it has been formalized in a number of computational models of language processing known as surprisal models (e.g., Boston, Hale, Patil, Kliegl, & Vasishth, 2008; Hale, 2001; Levy, 2008; Pado, Crocker, & Keller, 2009). In these models, the amount of cognitive effort required to integrate a word into a sentence depends on the negative log probability of that word given its preceding context. Surprisal models have had considerable success in predicting differences in reading times based on a word’s predictability given its preceding context.

As it turns out, the amount of processing effort associated with integrating a word into a sentence cannot be entirely reduced to its predictability given its preceding context. Roland, Yun, Koenig, and Mauner (2012) examined the effects of the semantic cohort of a target word (i.e., the other words that could appear in the same position/context as the target word) on the processing of a target word. They found that words that are more semantically similar to their semantic cohort are easier to process when word predictability and other factors are controlled for. This result is important because it points to a limitation in expectation-based computational accounts of sentence processing that claim that a word’s probability given its context is the sole predictor of processing effort (e.g., Levy, 2008). The results we present further constrain expectation-based accounts of sentence processing by showing that the effect of semantic similarity between a target word and its semantic cohort interacts with the degree of constraint provided by context.

The findings of Schwannfluegel and LaCount (1988) motivate the possibility that the benefits of semantic similarity on word integration might be modulated by contextual constraint. Schwannfluegel and LaCount found that unpredictable words that were semantically related to the most predictable word that could occur in the same position were processed faster than other equally unpredictable words that were also semantically unrelated to that most predictable word. What is crucial for this discussion is that the benefit of shared semantic information was not consistently observed for all unpredictable words. Shared semantic information only facilitated the processing of unpredictable words when contexts were weakly constraining.
To illustrate why the benefit of shared semantic similarity to other words activated by the preceding context would be greatest when a target word is unpredictable and its context is only weakly constraining, consider the sentence contexts in examples (1) and (2), for which we have obtained word completions (this study will be described in greater detail later).

(1) The gladiator jabbed the African tiger with
(2) The aborigine attacked the angry lion with

In both contexts, an instrument noun is most likely to be the next word. However, the types of instruments in the semantic cohort differed across contexts. For context (1), instruments like sword, spear, stick, knife, and spike were mentioned. These instruments share a typical property, i.e., all can be used as “pokers”. In contrast, instruments like fire, net, whip, and rock, which were mentioned for context (2), have few salient characteristics that are common to instruments of attacking. This difference in the degree of shared characteristics suggests that context (1) places greater restrictions on the range of possible instruments than context (2). Using responses obtained from a completion study, we illustrate the distribution pattern of the probabilities of possible instruments for each context. In comparing Figure 1a to Figure 1b, two things become apparent. First, the most probable instruments for jabbing are more likely than the most probable instruments for attacking. Second, the probabilities of the jabbing semantic cohort drop more sharply than do the probabilities of the attacking semantic cohort. One way of quantifying the greater degree of constraint provided by the jabbing context is to note that the top three items have a combined probability of .55. In the attacking context, even the first 6 items do not match that combined probability.

Hypotheses and Prediction

Based on the findings of Schwanenflugel and LaCount (1988), we predict that the semantic similarity effect found by Roland et al. (2012) will be stronger in more weakly constraining contexts and weaker in highly constraining contexts. While Schwanenflugel and LaCount only examined the processing of unpredictable words, we do not expect interactions with word predictability, since Roland et al. found no interaction between similarity and predictability.

Entropy as a measure of contextual constraint

In order to measure the effects of contextual constraint, we need a measure to quantify the degree of contextual constraint. Recall that in a more constraining context, there are larger differences in the probabilities of cohort members, because a small subset of the possible words is more likely, while the others are unexpected. Alternatively, in a less constraining context, there are a larger number of words that are more or less equally likely. The Entropy (H) measure, described in Equation (1), captures this difference in distributions. Entropy is higher when the choices are more similar in probability, as in low constraint contexts, and is lower when choices are less similar in probability, as in more highly constraining contexts.

\[ H = - \sum_{i=1}^{n} P(x_i) \log(P(x_i)) \]  

Equation (1)

Experiment to Generate Reading Times

Participants One hundred thirty native English-speaking undergraduates from the University at Buffalo received partial course credit for participation.

Materials We constructed 3 sets of 60 active declarative sentences with optional prepositional phrases similar to those in Example (3). Sets were differentiated by having an instrument noun that was highly likely (e.g., sword), moderately likely (e.g., spear), or unlikely (e.g., spike). To avoid wrap-up effects on instrument reading times (Just & Carpenter, 1980), all sentences included sentence-final phrases like in the Colosseum. Presentation regions are indicated in example (3) by vertical lines (|).

(3) The gladiator jabbed the African tiger with a sword/spear/spike in the Colosseum.

Selection of target instruments was based on responses from a listing study in which 42 participants produced five instruments for sentence fragments like (1) and (2). Cloze probabilities for highly likely, moderately likely and unlikely instruments were M = .23, S.D. = .06, M = .10, S.D.
An initial fit for the random intercepts and slopes model was performed using the maximal random effect structure, as discussed by Baayen, Davidson, and Bates (2008). We used the entropy of the probability distribution of all possible instruments for a context to measure the degree of constraint provided by the preceding context. Entropy values ranged between 2.55 for the most constraining contexts and 5.02 for the least constraining contexts.

**Dependent Variables** While the primary dependent variable was the reading times for sentences that participants continued to judge acceptable, we examined “No” judgments to ensure that they did not differ as a function of instrument likelihood. Across conditions, percentages of “No” responses adjusted for remaining chances to say “No” (see Boland, Tanenhaus & Garnsey, 1990) were low (under 5% in all conditions) and their variances were small. “Yes” reading times for instrument noun phrases were filtered for outliers such that reading times greater than 4,000 ms or less than 200 ms were omitted. Filtering resulted in the removal of 27 of 3723 (0.7%) reading times.

**Model Results and Discussion**

We provide a summary of the linear mixed-effect regression model in Table 1 and a graphical representation of the interaction between Similarity and Entropy in Figure 2.

**Predictability** Consistent with many previous studies (Ashby, Rayner, & Clifton, 2005; Bicknell et al., 2010; DeLong et al., 2005; Ehrlich & Rayner, 1981; Federmeier et al., 2007; Frisson et al., 2005; Otten & Van Berkum, 2008; Rayner & Well, 1996; Staub, 2011; Van Berkum et al., 2005), more predictable instruments were processed more quickly.

**Similarity** Instruments were read faster when they more similar to the members of their semantic cohort than when
they were less similar. This result replicates Roland et al.’s (2012) results.

Table 1: Summary of fixed factors from the linear mixed-effect regression model, when the effects of random variables were maximized, for predicting reading times of that target noun.

<table>
<thead>
<tr>
<th></th>
<th>Estimated Coefficient</th>
<th>S.E.</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>716.35</td>
<td>16.82</td>
<td>42.60</td>
</tr>
<tr>
<td></td>
<td>(716.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>-50.05</td>
<td>9.17</td>
<td>-5.46</td>
</tr>
<tr>
<td></td>
<td>(-28.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity</td>
<td>-230.36</td>
<td>68.65</td>
<td>-3.36</td>
</tr>
<tr>
<td></td>
<td>(-24.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>24.80</td>
<td>18.52</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(12.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>19.35</td>
<td>2.74</td>
<td>7.06</td>
</tr>
<tr>
<td></td>
<td>(41.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>-32.29</td>
<td>6.47</td>
<td>-4.99</td>
</tr>
<tr>
<td></td>
<td>(-32.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability x Similarity</td>
<td>32.71</td>
<td>90.84</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability x Entropy</td>
<td>-26.48</td>
<td>18.84</td>
<td>-1.41</td>
</tr>
<tr>
<td></td>
<td>(-7.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity x Entropy</td>
<td>-378.59</td>
<td>108.31</td>
<td>-3.50</td>
</tr>
<tr>
<td></td>
<td>(-20.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability x Frequency</td>
<td>18.06</td>
<td>8.86</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(10.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity x Length</td>
<td>-43.63</td>
<td>21.08</td>
<td>-2.07</td>
</tr>
<tr>
<td></td>
<td>(-4.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability x Entropy x Similarity</td>
<td>155.37</td>
<td>176.91</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>(4.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity x Frequency x Similarity</td>
<td>-50.41</td>
<td>21.27</td>
<td>-2.37</td>
</tr>
<tr>
<td></td>
<td>(-11.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy x Frequency x Similarity</td>
<td>-25.34</td>
<td>5.24</td>
<td>-4.83</td>
</tr>
<tr>
<td></td>
<td>(-27.97)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All predictors are centered, frequency predictor is residualized for length and predictability. Parenthetical values below the coefficients are standardized coefficients from an alternate version of the model with standardized predictors. t-values with an absolute value greater than 2 are significant at an alpha level of .05 (Gelman & Hill, 2007).

Entropy There was no main effect of Entropy. Importantly however, Entropy interacted with Semantic Similarity, just as hypothesized. Specifically as shown in Figure 2, there was an effect of Semantic Similarity when Entropy was high (i.e., low constraint contexts) (Estimated coefficient = -392.90, S.E. = 96.62, t-value = -4.07), but no effect of Semantic Similarity when entropy was low (i.e., high constraint contexts) (Estimated coefficient = -69.01, S.E. = 78.92, t-value = -0.87). The fact that Semantic Similarity did not facilitate the integration of instruments in strongly-constraining contexts is consistent with Schwanenflugel and LaCount’s (1988) results.

Figure 2: Interaction of Entropy and Similarity using standardized coefficients.

Length Longer words took longer to read. This is consistent with previous findings showing the effects of length (e.g., Juhasz & Rayner, 2003). While Length interacted with several factors, these interactions are not of theoretical interest in the current discussion.

Frequency Unsurprisingly, more frequent words were read faster than less frequent words. This too is consistent with previous studies (e.g., Ashby et al., 2005; Juhasz & Rayner, 2003; Kliegl, Grabner, Rolfs, & Engbert, 2004; Staub, 2011). Frequency interacted with a number of other predictors because of differences in the sizes of frequency effects under various combinations of conditions.

General Discussion

We found that semantic similarity between a target word and its semantic cohort has a stronger effect on processing when the context provides fewer constraints on what may appear in the target position. Alternatively, the effects of semantic similarity become weaker as the context becomes more constraining. The effect of contextual constraint on the degree to which semantic similarity affects processing has important implications for models of processing. Roland et al. (2012) suggested two possible causes for the semantic similarity effect: spreading activation between the representations for the words that comprehenders were anticipating, and the possibility that expectations for words and expectations for semantic features could have independent effects on comprehension difficulty. Our results suggest that the nature of comprehenders’ expectations may vary with the degree of contextual constraint. In a highly constraining context (i.e., low entropy), there is no effect of semantic similarity, and comprehension difficulty appears to be primarily determined by the predictability of the target word. If the target word was expected, it is easy to process. If the target
word was unexpected, it is difficult to process. On the other hand, in a less constraining context, semantic similarity influences processing. If the target word is expected, it is easy to process. But for an unexpected target word, the degree to which it is semantically similar to the members of its semantic cohort seems to make up for the unexpectedness of the word, thus easing the processing difficulty. Only when a target word is unexpected and semantically distant from its cohort is it truly difficult.

One possible explanation for why contextual constrain modulates the influence of semantic similarity for unpredictable words is that in a highly constraining context, comprehenders may be expecting specific words, and face difficulty when the expectations turn out to be wrong. In a less constraining context, comprehenders may have less specific expectations – anticipating semantic features in common between a set of possible words (in addition to, or as an alternative to anticipating specific words). Thus, they face less difficulty when the target word turns out to be something other than the most likely word – as long as the target word shares some level of semantic similarity with the other likely possible words. Overall, our data suggests that word predictability, semantic similarity, and contextual constraint all have an impact on language comprehension.

References
Bates, D., Maechler, M. & Bolker, B. (2011). Lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4


