Cognition 122 (2012) 267–279

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Semantic similarity, predictability, and models of sentence processing

Douglas Roland [a,*], Hongoak Yun [b], Jean-Pierre Koenig [a,b], Gail Mauner [b]

[a] *Department of Linguistics, University at Buffalo, The State University of New York, United States*
[b] *Department of Psychology, University at Buffalo, The State University of New York, United States*

ARTICLE INFO

Article history:
Received 25 October 2010
Revised 22 November 2011
Accepted 23 November 2011
Available online 23 December 2011

Keywords:
Sentence processing
Prediction
Parsing
Surprisal
Semantic similarity
Spreading activation

ABSTRACT

The effects of word predictability and shared semantic similarity between a target word and other words that could have taken its place in a sentence on language comprehension are investigated using data from a reading time study, a sentence completion study, and linear mixed-effects regression modeling. We find that processing is facilitated if the different possible words that could occur in a given context are semantically similar to each other, meaning that processing is affected not only by the nature of the words that do occur, but also the relationships between the words that do occur and those that could have occurred. We discuss possible causes of the semantic similarity effect and point to possible limitations of using probability as a model of cognitive effort.

1. Introduction

It is a (true) cliché of psycholinguistics that the accuracy of human sentence processing is something of a feat, as words must be processed and integrated very quickly, given the continuous nature of the input stream. A popular, partial explanation for this feat is that, when processing sentences, we use all kinds of information to predict what is coming up next and that preactivation of the upcoming material makes integrating it easier.

Because of the widespread belief in the importance of predictability in sentence comprehension, much work has been done to enumerate the factors that comprehenders use to make their predictions about upcoming linguistic material. Factors that have been proposed to influence comprehension include verb subcategorization biases (e.g., Trueswell, Tanenhaus, & Kello, 1993), thematic fit of noun phrases (e.g., McRae, de Sa, & Seidenberg, 1997; Tanenhaus, Carlson, & Trueswell, 1989), the likelihood of different agents carrying out different actions (e.g., Kamide, Altmann, & Haywood, 2003), and various discourse factors (e.g., Binder, Duffy, & Rayner, 2001; Hare, McRae, & Elman, 2004).

Comprehenders appear to use contextual information to make predictions about upcoming material. DeLong, Urbach, and Kutas (2005) found an N400 response to indefinite determiners in English (*a*, *an*) that did not correspond to the noun that was most likely to occur next given the context. Similarly, Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) found an ERP response when Dutch determiners did not match the anticipated following noun in grammatical gender. Both of these results suggest that comprehenders have formed expectations for specific words to occur in advance of the point at which the words actually occur.

The linking assumption between predictability and cognitive effort is that the cognitive representations for expected words (or phonemes, syntactic structures, etc.) are presumed to be more highly activated than those for less expected ones. Consequently, they are presumed to be easier to retrieve from memory, and require less additional activation to incrementally update the set of representations created during the comprehension of the utterance. In a sentence like *The poor student ate macaroni*

* Corresponding author. Address: Department of Linguistics, University at Buffalo, Buffalo, NY 14260, United States.
E-mail address: droland@buffalo.edu (D. Roland).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.11.011

and cheese, the word *cheese* is highly predictable. As a consequence, processing the word *cheese* results in only minor changes to the overall set of cognitive representations involved in comprehending the sentence. If the word *cheese* were replaced with a less predictable word, such as in *The poor student ate macaroni and caviar*, the processing of the sentence at the word *caviar* would require a larger change to the overall set of activated cognitive representations, and thus more cognitive effort.

The expectations for a particular word also reflect the expectations for structures at levels besides the word level. In a sentence like *The horse raced past the barn fell*, the reduced relative structure is very unexpected, as is the word *fell*. Consequently, processing the word *fell* results in major changes to the set of cognitive representations involved in comprehending the sentence. We will discuss the issue of exactly how expectations at different levels of representation are related to expectations at the word level in the final discussion section of this paper, but informally we assume the inclusion of expectations at all levels when we refer to word predictability throughout the paper.

The relationship between word probability and cognitive effort has been formalized in theories such as the surprisal theory (Hale, 2001; Levy, 2008), which relies on Eq. (1) to make predictions of cognitive effort. This equation indicates that the degree of cognitive effort required to process a word is dependent on the negative log probability of that word, given the preceding context. This measure has been described in several ways that are mathematically equivalent, but which emphasize different aspects of possible cognitive interpretations of the measure. Levy (2008) characterizes the measure in terms of the degree of difference between the probability distributions of the possible interpretations of the message before seeing the word and after seeing the word. Jurafsky (2003) characterizes it in terms of the amount of information conveyed by the word. Hale (2001) characterizes it in terms of the probability mass of the interpretations that are disconfirmed upon hearing a word.

$$\text{difficulty} \propto -\log p(w_i|w_{1\ldots i-1}\text{CONTEXT}) \qquad (1)$$

One commonality across discussions of expectations in comprehension is that the degree of cognitive effort needed to process a particular message tends to be cast in terms of how likely a particular word, structure, or message is, relative to another word, structure, or message. Aside from their relative probabilities, little attention is paid to potential relationships between the various possible words or structures. The other words that could have occurred in that position are only relevant in that, if a particular word is very likely, other possible words must necessarily be unlikely. This indirect relationship arises because the probabilities of all possible words must sum to 1. Importantly, it is assumed that the nature of the other words that could have occurred has no other bearing than this indirect relationship on the level of difficulty faced in processing the target word itself.

We challenge this often implicit assumption that the degree of cognitive effort is determined solely by the properties of the material that actually occurs by providing evidence for our *Semantic Similarity Hypothesis*, which predicts that processing will be facilitated to the degree that the different possible choices that could occur in a given context are semantically similar to each other. One possible cause for the predicted processing facilitation is that activation may spread between the representations of the different possible choices that are being activated during processing (e.g., McRae, Ferretti, & Amyote, 1997). In this view, greater semantic similarity between the possible word choices would result in greater activation of this set of words, and thus greater facilitation in processing. Alternate possible causes of a semantic similarity effect will be addressed in the final discussion section.

To better understand our *Semantic Similarity Hypothesis*, consider the sets of possible instruments that could occur in the sentential contexts shown in (1) and (2). Based on the hypothetical distributions of possible instruments shown in Fig. 1 for these contexts, probabilistic theories of language comprehension would predict that instruments such as *spear* and *sword* would be easier to process than instruments such as *machete* and *rock*, due to their greater degrees of anticipatory activation. This prediction is consistent with a long history of experimental results showing that the degree to which material is predictable from the context affects comprehension processes, as reflected in measures such as reading times (e.g., Rayner & Well, 1996), electrophysiological response (e.g., Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007), and the ability to comprehend degraded input (e.g., Obleser & Kotz, 2010). Probability-based accounts such as the surprisal theory (Hale, 2001; Levy, 2008) and the SynSem Integration Model (Padó, Crocker, & Keller, 2009) have had good success at modeling such differences in relative cognitive loads during language comprehension across a wide variety of psycholinguistic phenomena based solely on knowing how likely a target word is, given its context.

(1) The aboriginal man jabbed the angry lion with a/an —.
(2) The aboriginal man attacked the angry lion with a/an —.

Models which base their predictions only on the probability of target words (e.g., Hale, 2001; Levy, 2008) necessarily also make the following predictions for the contexts shown in (1) and (2), given the distributions of possible instruments shown in Fig. 1. First, because the probability of *spear* is the same in both contexts, *spear* should have the same degree of difficulty in either context. Second, because *machete* has the same probability in context (1) as *rock* has in context (2), *machete* and *rock* should also have the same respective degree of difficulty, once other factors such as length and frequency are taken into account. However, in the examples shown in Fig. 1, there is a difference between the distributions of possible instruments for these two contexts. The set of likely instruments for the *jab* context are typically all sharp, pointy objects. Several of the possible instruments for *attack* also share these properties, but many of the less likely instruments for *attack*, including *rock*, do not. If the representations of the various possible instruments are initially activated based on their respective probabilities, activation may