Cognition 128 (2013) 302–319

Contents lists available at SciVerse ScienceDirect

Cognition

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

The effect of word predictability on reading time is logarithmic ☆

Nathaniel J. Smith [a,*], Roger Levy [b]

[a] Department of Cognitive Science, University of California, San Diego, United States
[b] Department of Linguistics, University of California, San Diego, United States

ARTICLE INFO

Article history:
Received 10 September 2012
Revised 4 February 2013
Accepted 16 February 2013
Available online 6 June 2013

Keywords:
Psycholinguistics
Expectation
Probabilistic models of cognition
Information theory
Reading

ABSTRACT

It is well known that real-time human language processing is highly incremental and context-driven, and that the strength of a comprehender's expectation for each word encountered is a key determinant of the difficulty of integrating that word into the preceding context. In reading, this differential difficulty is largely manifested in the amount of time taken to read each word. While numerous studies over the past thirty years have shown expectation-based effects on reading times driven by lexical, syntactic, semantic, pragmatic, and other information sources, there has been little progress in establishing the quantitative relationship between expectation (or prediction) and reading times. Here, by combining a state-of-the-art computational language model, two large behavioral data-sets, and non-parametric statistical techniques, we establish for the first time the quantitative form of this relationship, finding that it is logarithmic over six orders of magnitude in estimated predictability. This result is problematic for a number of established models of eye movement control in reading, but lends partial support to an optimal perceptual discrimination account of word recognition. We also present a novel model in which language processing is highly incremental well below the level of the individual word, and show that it predicts both the shape and time-course of this effect. At a more general level, this result provides challenges for both anticipatory processing and semantic integration accounts of lexical predictability effects. And finally, this result provides evidence that comprehenders are highly sensitive to relative differences in predictability – even for differences between highly unpredictable words – and thus helps bring theoretical unity to our understanding of the role of prediction at multiple levels of linguistic structure in real-time language comprehension.

1. Introduction

Making probabilistic predictions about the future is a necessary component of essentially every task that the brain performs, to the point that it has been proposed as a fundamental principle underlying its operation (Bar, 2009). One example of this is in language comprehension: As you read this text, you are unconsciously anticipating upcoming words based on the constantly-evolving context. For example, the sentence.

(1) My brother came inside to...

may well continue any number of ways, but native English speakers are in general agreement—and you will likely immediately recognize—that the sentence

* Corresponding author. Address: School of Informatics, University of Edinburgh, Informatics Forum 3.29, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom.
E-mail address: nathaniel.smith@ed.ac.uk (N.J. Smith).

0010-0277/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.02.013

(2) The children went outside to...

is almost certain to continue with the word *play*. Although *play* is perfectly reasonable as a continuation of (1), in (2) it is read more quickly on average (Ehrlich & Rayner, 1981; Kliegl, Nuthmann, & Engbert, 2006; McDonald & Shillcock, 2003a; Rayner & Well, 1996). A wide range of studies have shown that such effects of predictability or expectation for specific words affect not only reading times but also neural responses (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; Kutas & Federmeier, 2011; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003) and interpretation of temporarily ambiguous input (Altmann & Kamide, 1999; Dahan & Tanenhaus, 2004; Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

A second, related strand of research has shown that incremental processing difficulty is also affected by expectations for more abstract levels of linguistic content, including the predictability of different syntactic (Demberg & Keller, 2008; Ferreira et al., 1986; McRae, Spivey-Knowlton, & Tanenhaus, 1998), semantic (Federmeier & Kutas, 1999), and pragmatic (Ni, Crain, & Shankweiler, 1996) structures. However, the relationship between the effects of expectations for specific-words and expectations for more abstract structures remains poorly understood. The most widespread method for assessing expectations for specific words is the cloze task (Taylor, 1953), in which native speakers are asked to write continuations of an incomplete sentence; in the examples above, *play* is the first word in over 90% of continuations of (2) but almost never appears as the first word of continuations of (1). However, the cloze task makes it quite difficult to precisely measure predictabilities <5–10%, and it is commonly assumed that differences in lexical expectation between items in this range do not produce behavioral effects. This contrasts with studies involving more abstract levels of linguistic structure, where expectation-based effects are observed even though the specific word instantiating the structure may rarely or never be produced in a cloze task. To take one recent example, Levy, Fedorenko, Breen, & Gibson, (2012) showed that the word *who* and the immediately following region of the sentence were read more quickly in sentences like (3) than in sentences like (4):

(3) After the show, a performer who had really impressed the audience bowed.
(4) After the show, a performer bowed who had really impressed the audience.

The word *who* never occurs in practice as a cloze continuation in either context (unpublished data), and so this result would conventionally be interpreted as arising from syntactic expectations (Hale, 2001; Ilkin & Sturt, 2011; Levy, 2008; Levy et al., 2012; Lau, Stroud, Plesch, & Phillips, 2006; Staub & Clifton, 2006): In (4) *who* is introduced by a grammatical construction (relative-clause extraposition) that corpus data indicate is both lower frequency and less likely given the grammatical context than the construction in (3) (ordinary postmodification by a relative clause), even though both are infrequent and unlikely in absolute terms.

Probability theory, however, tells us that such differences in syntactic expectation should also produce differences in lexical expectation, even if these latter differences are too small to measure via the cloze task. We can quantify the predictability of a word w in context C as its conditional probability of occurrence in that context, $P(w|C)$. Similarly, we write the predictability of a syntactic construction S as $P(S|C)$. In these particular contexts, $w = who$ can only occur if $S =$ relative clause. The laws of conditional probability then let us decompose the lexical predictability as the product of two terms (Demberg & Keller, 2008; Fossum & Levy, 2012; Roark, Bachrach, Cardenas, & Pallier, 2009):

$$P(who|C) = P(\text{rel. clause}|C) \times P(who|\text{rel. clause}, C)$$

The first term is the syntactic predictability, and the second measures the likelihood that this relative clause will begin with the word *who* (as opposed to, say, *that*). The latter is presumably roughly constant between these two contexts, which means that while the precise lexical predictabilities in (3) and (4) are too small to measure directly, the ratio between them should be similar to the ratio between their syntactic predictabilities.

Motivated by such considerations, Hale has suggested that syntactic and other types of abstract expectations may affect processing difficulty purely by modulating lexical predictability, which under the *surprisal theory* of incremental language processing is measured as log-probability (Hale, 2001; Levy, 2008). Within surprisal theory, lexical predictability forms a "causal bottleneck" through which the many different kinds of more abstract expectation discussed above must act. But as the above example shows, an essential requirement for this theory is that small absolute differences in expectation for low-predictability words must be capable of producing relatively large effects on processing difficulty, and it is not known whether this is the case. In fact, almost nothing is known about the quantitative form of the relationship between word predictability and the measurable correlates of processing difficulty such as reading time. This is in striking contrast to the study of isolated word recognition, where it has been known since the 1950s that recognition time varies almost exactly as a logarithmic function of frequency[1] (Howes & Solomon, 1951), and the need to explain this pattern has motivated a wide range of theories (Adelman, Brown, & Quesada, 2006; Baayen, 2010a; Murray & Forster, 2004; Morrison, Hirsh, & Duggan, 2003; Norris, 2006). But the few extant published studies (Kliegl et al., 2006; Rayner & Well, 1996) that have investigated the quantitive relationship between word predictability and processing time have yielded only limited insights, particularly regarding the shape of this relationship for highly unpredictable words, partly because of the cloze method's

[1] Note that in psycholinguistics, the term *frequency* refers specifically to a word's unconditional probability of occurrence without regard to context, $P(w)$, making it quite distinct from context-dependent predictability.