CHAPTER SIX

Prediction, Information Structure, and Good-Enough Language Processing

Fernanda Ferreira[1] **and Matthew W. Lowder**
University of California, Davis, CA, United States
[1]Corresponding author: E-mail: fferreira@ucdavis.edu

Contents

Abstract

The good-enough language processing approach emphasizes people's tendency to generate superficial and even inaccurate interpretations of sentences. At the same time, a number of researchers have argued that prediction plays a key role in comprehension, allowing people to anticipate features of the input and even specific upcoming words based on sentential constraint. In this chapter, we review evidence from our lab supporting both approaches, even though at least superficially these two perspectives seem incompatible. We then argue that what allows us to link good-enough processing and prediction is the concept of information structure, which states that sentences are organized to convey both given or presupposed information,

Psychology of Learning and Motivation, Volume 65
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2016.04.002

and new or focused information. Our fundamental proposal is that given or presupposed information is processed in a good-enough manner, while new or focused information is the target of the comprehender's prediction efforts. The result is a theory that brings together three different literatures that have been treated almost entirely independently, and which can be evaluated using a combination of behavioral, computational, and neural methods.

1. INTRODUCTION

A critical component of language comprehension is parsing, which refers to the process of generating a structural representation for a sentence and assigning it an interpretation. For several decades, researchers in psycholinguistics have attempted to explain how this process unfolds incrementally, as words are encountered in sequence, and how information from various sources—lexical, syntactic, semantic, and pragmatic—are rapidly integrated online. Two recent perspectives on the problem have greatly enriched our understanding of how these processes unfold. The first is good-enough language processing (Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; Sanford & Sturt, 2002), and the second is the idea of prediction as a key mechanism in comprehension. Good-enough language processing emphasizes the tendency of the comprehension system to perform superficial analyses of linguistic input, which sometimes result in inaccurate interpretations. Prediction approaches assume that efficient comprehension makes use of contextual constraint to anticipate upcoming input, leading to facilitated processing once the input is encountered. An important task for the field is to determine how these new perspectives can be reconciled with the existing literature and with classic phenomena of comprehension. Here we focus on one of these classic ideas, information structure, which assumes that sentences are divided into given and new information. Since at least the mid-1970s, linguists and psycholinguists have argued that sentence content can be divided into that which is given, and which can be linked to prior discourse, and content that is new, which adds information and moves the discourse forward. Up to now, these different ideas about comprehension—good-enough processing, prediction, and information structure—have not been integrated. This is a serious problem because, at first glance, they seem somewhat incompatible, and yet there is good evidence for all of them. In this chapter, we attempt to integrate these ideas and resolve the contradictions by arguing that information structure plays a critical role in good-enough language processing and in prediction. More specifically, we will argue that given information tends to be processed in a good-enough manner, and that new information is what the comprehender tries to predict. To make this case, we will first review the evidence for each component of this story, and then we will present in more detail a model of language processing that provides a role for each during comprehension.

2. THE GOOD-ENOUGH LANGUAGE PROCESSING APPROACH

Prior to our work on good-enough processing, models of parsing were largely based on evidence obtained from online processing measures such as reading time. In a typical experiment, a subject would be presented with garden-path sentences and appropriate controls, and differences in a measure such as reading time were used to motivate sentence comprehension architectures. For example, a sentence such as *While Anna bathed the baby played in the crib* would be shown along with comma-disambiguated controls, and longer reading times in the noncomma version were viewed as reflecting processing difficulty. Theories of comprehension were designed to explain the processing steps involved in arriving at "the interpretation" (presumably, that Anna bathed herself while the baby played in the crib), but little evidence was obtained to show that an accurate meaning was actually computed. One obvious way to assess meaning is to present subjects with comprehension or probe tasks following the sentences, but typically those tasks, when included, were meant simply to make sure participants "paid attention" to the sentences, and often the comprehension data were not systematically reported or analyzed. Interestingly, some of the earliest studies of garden-path processing that did report comprehension data revealed quite low rates of comprehension for garden-path sentences compared with disambiguated controls; for example, Ferreira and Clifton (1986) reported question-answering accuracy rates between 63% and 80%, and MacDonald, Just, and Carpenter (1992) reported that readers with high working memory spans answered questions correctly 70% of the time, and lower span readers were at chance. Despite these striking results, models of parsing largely ignored comprehension data and instead were designed to account for processing time effects as a function of the manipulation of variables such as lexical information, discourse context, and so on.

Our hypothesis that garden-path sentences might often be systematically misunderstood was motivated by a number of findings that were in the literature at that time, although not necessarily in the literature on sentence processing. One was the report of so-called semantic illusions, in which it was noted that people seem to normalize nonsensical sentences such as *This book fills a much-needed gap* (Fillenbaum, 1974) and *More people have been to Russia than I have* (Pullum, 2004) suggesting that they were somehow bypassing normal compositional processing mechanisms to obtain a sensible interpretation. Similarly, the Moses Illusion, which can be seen in people's tendency to answer "2" to the question *how many of each type of animal did Moses take on the ark*, also suggests that processing is superficial—in this case, too superficial to distinguish between key Old Testament figures (Erickson & Mattson, 1981). Otero and Kintsch (1992) also demonstrated that college students were remarkably poor at noticing contradictions in discourse, and O'Brien and colleagues (Albrecht & O'Brien, 1993; Cook, Halleran, & O'Brien, 1998; O'Brien, Rizzella, Albrecht, & Halleran, 1998) showed that a characteristic attributed to a character in a story was hard for readers to inhibit when new and inconsistent information was provided later in the story. All these findings—interestingly, not from the core sentence processing or parsing literature—suggested that processing might be shallow and that an interpretation once computed or retrieved could be tenacious.

2.1 Good-Enough Reanalysis

To examine this phenomenon with respect to parsing, our approach was to examine the meaning that people obtained for garden-path sentences. In our experiments, we presented subjects with sentences such as *While Anna bathed the baby played in the crib*, and then we asked them one of two questions: Did Anna bathe the baby, and did the baby play in the crib (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Christianson, & Hollingworth, 2001). Our prediction was that the second question would be answered correctly but the first would not, based on the following reasoning. The parser incrementally builds syntactic structure and interpretations; given the properties of the verb *bathe* and the semantics of the female name, *bathe*, and baby, the parser would create the interpretation that Anna bathed the baby. Upon encountering the word *played*, the parse built up to that point would break down because this verb requires a subject and there is none in the representation. To repair the structure, the system must reanalyze *the baby* as the subject of *played*, thereby building the correct structure and the right meaning for the main clause: that the baby played in the crib. Now, however, we reasoned that the system might not always go back to its previous decisions and deal with the (mis) interpretation accessed during the initial parse. If not, this would lead to the creation of a semantic representation in which Anna both bathes the baby and the baby plays in the crib. Our experiments confirmed our predictions: We observed almost perfect accuracy for responses to the main clause question, but only about 50–60% accuracy for responses to the subordinate clause question.

One obvious issue that arises is the characteristics of the controls for the garden-path sentences. In Christianson et al. (2001), we used two different controls: one in which a comma separated the two clauses, and another in which an overt object occurred in the sentence (*While Anna bathed the dog the baby played in the crib*), both of which essentially eliminate the temptation to take *the baby* as object of *bathed*. These controls allowed us to address an important issue, which is whether the questions themselves might have led to the misinterpretations. Because accuracy on the main-clause question was the same for garden-path and garden-path controls, but on the subordinate-clause questions accuracy was much lower in the garden-path condition, we reasoned that any reinstatement of the idea that Anna might have been bathing the baby could not explain our entire pattern of results—specifically, the greater tendency to misinterpret only the subordinate clause and only in the garden-path condition. Similarly, any tendency on the part of subjects to believe that Anna might have bathed the baby simply due to world knowledge cannot explain the difference in performance for the garden-path and nongarden-path versions. Another suggestion that has sometimes been made is that the errors would not have occurred if subjects had been allowed to read the sentences for as much time as they liked. However, our study compared the results from experiments in which sentences were presented in rapid serial visual presentation (RSVP) and in which sentences were presented in their entirety, with subjects allowed to read them for as long as they wished. We found that error rates overall were reduced in the full-sentence, self-paced experiment, but the difference in accuracy between the garden-path and garden-path controls was the same in both. Follow-up experiments using a paraphrase/free recall task showed that subjects recalled the sentences as containing the incorrect proposition (that Anna bathed the baby), but only in the garden-path conditions (Patson, Darowski, Moon, & Ferreira, 2009). This result also shows that the misinterpretation effect is not due to the use of question probes.

Finally, a recent eye movement monitoring experiment (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013) demonstrated that the syntactic structure that subjects create for garden-path sentences is well formed, because the structure is sufficient to allow the reader to bind a reflexive pronoun such as *himself* to its proper antecedent (which, according to standard syntactic analyses, must be in a specific structural position relative to the pronoun). This coindexing could not have occurred if the sentence had not been syntactically reanalyzed. This finding led us to conclude that the source of the misinterpretation effect is not a failure to restructure the sentence; given the binding results, the structure appears to be correctly reanalyzed. Instead, we argued that the problem is that the comprehender fails to inhibit the semantic proposition that was generated from the original, incorrect parse. This failure to properly update the propositional representation for the sentence is reminiscent of the findings from the text processing literature mentioned previously, where it has been previously shown that proper updating of information in response to new information contradicting a previously generated idea does not always take place.

2.2 Good-Enough Processing in a Broader Context

As part of our goal in this chapter to integrate these findings into a more general model of language comprehension, we make the following observations about these "good-enough" findings. The first is that, up to now, the structurally demanding sentences have not been presented to subjects in any sort of discourse context. This is potentially important not only because a context might help to reinforce the correct (syntactically mandated) meaning of the sentence, but also because, in the absence of a discourse, the sentences lack any real information structure. That is, these sentences can only be weakly assigned a given-new structure because there is no discourse to link any information as previously given. Second, to the extent that the sentences we have focused on do have a reasonably clear given-new structure, it appears that the part of the sentence that is likely to be misinterpreted is the part that is more likely to be treated as given, even outside a context. For example, in the garden-path examples given above, the misinterpretation involves the subordinate clause of the sentence, which is standardly viewed as a device for communicating given information (Halliday, 1967; Hornby, 1974; Langford & Holmes, 1979). Similarly, in other misinterpreted garden-path sentences, the part that is incorrect is often a relative clause or other type of modifier, and again, relative clauses and other modifiers are standardly used to convey given or presupposed information (for classic analyses of these ideas, see Crain & Steedman, 1985; Altmann & Steedman, 1988). Finally, proponents of prediction-based models of comprehension have suggested that good-enough processing arises from prediction (Kuperberg &

Jaeger, 2016); in particular, their suggestion is that the "strong priors" associated with canonical language structures may overwhelm the analysis of the current input. In other words, the processor predicts the structure based on frequency and experience, and misinterpretations arise for forms that deviate from those expectations.

In sum, research on good-enough language processing indicates that the semantic interpretations built from syntactically challenging sentences do not always reflect their actual structures. This result is not surprising given previous work in other areas of psycholinguistics, including text comprehension. In addition, it appears that good-enough effects may tend to be localized to the given information in a sentence. To flesh these ideas out further, we turn next to a brief review of prediction models of language comprehension.

3. PREDICTION IN COMPREHENSION

A general trend in cognitive science over the last several years is to suggest that the brain should be viewed as a "prediction engine" rather than as a recipient of input to be passively processed (Clark, 2013). The idea is that the brain proactively anticipates what will happen next (Bar, 2009; Den Ouden, Kok, & de Lange, 2012; Friston, 2010; Lupyan & Clark, 2015; Rao & Ballard, 1999), and this is viewed as having two information processing benefits: First, to the extent that some or all features of an upcoming event are preactivated, processing of that event will be facilitated. Second, prediction error is an opportunity for learning to take place: If an outcome fails to conform to expectations, then the brain must update its model of that domain ("update its priors," in Bayesian terms; Doya, Ishii, Pouget, & Rao, 2007), forming a richer and more accurate representation. This approach has been productively applied to the field of language processing, with recent findings indicating that the efficiency of language processing emerges in part from the use of prediction by adults and even children, and during both reading and listening (Drake & Corley, 2015; Kutas, DeLong, & Smith, 2011; Mani & Huettig, 2012; Smith & Levy, 2013). One recent influential idea in psycholinguistics centers on the importance of dialogue and conversation for shaping the nature of comprehension and production processes (Pickering & Garrod, 2004). For dialogue to be successful, the listener must attempt to anticipate what the speaker will say using "forward modeling" to create a model of the speaker's communicative intentions that can then guide comprehension in a predictive manner.

3.1 Prediction of Syntactic Structure

Although the idea that language comprehension might critically rely on prediction is currently very popular, it is not entirely new. To appreciate the importance of the idea of prediction generally, it is useful to remember that standard models of parsing have assumed that structural nodes in a tree are predicted before any lexical evidence for them. For example, many parsers build nodes up from lexical items (for example, from a verb to a verb phrase) but also build nodes branching to the right based on predictions concerning what the constituent likely takes as a complement (for example, building an object for the verb in advance of any lexical evidence for a noun phrase). Indeed, classic garden-path effects arise because the parser predicts a simple syntactic structure for a phrase in advance of evidence from the input about whether that simple analysis will turn out to be correct. In some versions of the garden-path model, the motivation for the parser's simplicity strategy has been described as an attempt to prevent prediction processes from running amok—that is, in the absence of evidence, predict the simplest structure possible to avoid having to eliminate nodes that might later turn out to be unnecessary (Frazier, 1979, 1987). Staub and Clifton (2006) provided evidence for a somewhat different type of syntactic prediction: They provided evidence that readers who read a clause beginning with the word *either* anticipate receiving a later *or*-clause (see also Lau, Stroud, Plesch, & Phillips, 2006 for ERP evidence of syntactic prediction).

3.2 Prediction of Specific Words

Although a number of models of parsing assume that syntactic prediction is essential for successful syntactic processing, currently it is word prediction that is of most interest, possibly because it is assumed that the prediction of a specific word requires the integration of information from a range of sources and seems incompatible with the idea that the language processing system is modular. Some have argued (see, for example, Lupyan & Clark, 2015) that word prediction is an example of radical top-down processing because it supports the generation of linguistic content in the absence of any sensory input whatsoever. By now there are hundreds of published reports on prediction of words in constrained contexts; in the interests of space, we will highlight a subset of results that succinctly illustrate the phenomenon, focusing on three different literatures.

3.2.1 Prediction in Cloze Contexts

Evidence for prediction during language processing has come from three distinct approaches to studying comprehension. The first approach examines the processes that take place when readers or listeners encounter sentences that contain a highly predictable word, with predictability typically operationalized as cloze probability (the percentage of subjects who provide a specific word as a continuation to an initial sentence string, eg, *It was a breezy day so the boy went out to fly his* __; DeLong, Urbach, & Kutas, 2005). A large number of studies using ERPs during reading and listening have shown that the N400 amplitude is modulated by the extent to which a word fits its preceding context (for reviews see: Kutas & Federmeier, 2011; Swaab, Ledoux, Camblin, & Boudewyn, 2012). Work has also shown that ERP effects are indeed anticipatory, as N400 amplitudes measured from a determiner or adjective are reduced when its agreement features are consistent with a predicted noun (Boudewyn, Long, & Swaab, 2015; Luka & Van Petten, 2014; Szewczyk & Schriefers, 2013; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Many of these ERP studies have also found an anterior post-N400 positivity, whose amplitude is larger to incorrectly predicted words, and to unpredicted words in low constraint contexts (eg, Brothers, Swaab, & Traxler, 2015; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Experiments with similar logic have been conducted using eyetracking, where it has been shown that a highly predictable word receives shorter fixation durations and is more likely to be skipped than a less predictable word. Indeed, predictability effects, along with frequency and length effects, make up the "Big three" predictors of fixation time in reading (Clifton et al., 2016). Skipping data are interesting with respect to prediction because it seems reasonable to infer that if a word is not fixated, it is because it was anticipated and so direct perceptual sampling of it is not required. It has also been reported that increased skipping is accompanied by greater probability of a later regression to the skipped word when the prediction is incorrect (eg, Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006).

These results have sometimes been interpreted to suggest that comprehension is successful *because* texts and speech are normally strongly constrained by semantic context, facilitating the task of prediction. It is also assumed that the speed and fluency of normal comprehension is based largely on top-down mechanisms that are assumed to underlie prediction. We would caution against this particular interpretation to the extent that it is based on findings from studies that use the cloze procedure. It is important to appreciate that these stimuli are highly constrained *by design*. They have been created so that a single word at or near the end of the sentence is either easy or difficult to predict to test the notion that predictability facilitates processing. In addition, the cloze method might actually underestimate overall predictability because it standardly focuses on just one or two words in a sentence rather than degrees of predictability across words in a sentence or discourse. In normal language, entire sequences of words might be predictable based on the norms of dialogue or discourse, and, importantly, these predictable items might be more likely to occur at the beginning rather than the end of sentences. We return to this argument after Section 4 in which we discuss given-new structure and link it to good-enough processing.

3.2.2 Prediction as Surprisal and Entropy

A second approach to studying prediction examines continuously varying levels of predictability for all the words of a sentence, which has led to the development of the concept of *surprisal*. The surprisal value of word w_i is defined as the negative log probability of w_i given the words that have come before it in the sentence (ie, w_1, w_2,…w_{i-1}); $Surprisal(w_i) = -log\ P(w_i|w_1 \ldots w_{i-1})$. Surprisal theory (Hale, 2001; Levy, 2008) builds on foundational ideas in information theory and cognitive science (Attneave, 1959; Shannon, 1948), proposing that comprehenders use probabilistic knowledge from past experience to generate expectations concerning the interpretation of the language input so far, as well as predictions about the word likely to come up next. These expectations determine online processing difficulty. Critically, surprisal theory can be computationally implemented (Hale, 2001; Levy, 2008; Roark, Bachrach, Cardenas, & Pallier, 2009) and generates quantitative predictions about online processing difficulty for every word in the text. These predictions have been supported in several reading time studies (Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Roark et al., 2009; Smith & Levy, 2013) and in some of our own recent work (see later discussion). *Entropy* is a related concept and refers to the number of possibilities emerging from a word at a given point in a sentence (Hale, 2001; Roark et al., 2009; Staub, Grant, Astheimer, & Cohen, 2015). For example, entropy is low at *some* in *For the pasta you'll need to boil some* __ because one continuation is highly likely; entropy is higher at *go* in *At the intersection go* __ because there are at least three likely continuations. Following up on a recent study (Henderson, Choi, Lowder, & Ferreira, 2016), we are currently exploring the relationship between surprisal and entropy in the same stimuli and datasets, and it appears that the two measures are only weakly correlated and show different patterns of neural activation, as assessed by coregistration of fMRI and eye movement data during natural reading (see also Henderson & Choi, 2015; Henderson, Choi, Luke, & Desai, 2015).

Let us return to the research discussed in Section 3.2.1, which examines prediction using cloze methods. As we noted, stimuli are designed so that semantic and contextual constraints increase across word position within a sentence, making the word(s) near the end the most predictable. This idea makes some intuitive sense: The more context that has built up, the more constraint there is likely to be. Fortunately, we do not need to rely on intuition to answer this question; it can be addressed empirically by looking at how surprisal values change across a sentence. In recent work, we have undertaken this task by calculating surprisal values for the set of 95 sentences we used in our fMRI work (Henderson et al., 2016) and plotting these values as a function of word position in the sentence. Our findings suggest that there is no systematic relationship between a word's position in a sentence and its surprisal value. These analyses lead us to conclude that natural texts are quite different from cloze sentences. This is perhaps not surprising given that cloze items are designed to differ in predictability at later sentence positions, in part to allow context to build up in the absence of larger discourse context and to facilitate measurement of prediction effects (eg, ERP effects are often measured on sentence-final words so that those effects are not contaminated by responses to later words). This conclusion has some important implications for our analysis of good-enough language processing, prediction, and information structure, which we will consider in Section 5.

3.2.3 Prediction and the Visual World Paradigm

The final approach to studying prediction that we will review here makes use of a technique called the visual world paradigm (VWP) in which subjects' eye movements are monitored as they listen to sentences while at the same time examining visual displays. Eye movements to mentioned or implied objects are then time-locked to word onsets. In the passive version of the VWP, subjects are required to do nothing more than listen to sentences, and they are usually free to examine the pictures any way they like. In addition, as we have noted in our own work on the VWP (Ferreira, Foucart, & Engelhardt, 2013), the visual information typically precedes the start of the sentences by a few seconds, which is essential for expectations to be established. One of the most widely cited demonstrations of word prediction is the Altmann and Kamide (1999) experiment, which used spoken language and the VWP to investigate prediction during comprehension.

Altmann and Kamide showed that in a visual context with a few objects, only one of which is edible, listeners who heard *the boy will eat the cake* made saccades to the cake before its linguistic onset. This result provides some of the clearest evidence for the idea that comprehenders can predict an upcoming word, as there is little doubt that the word was activated prior to the availability of any linguistic input.

We have recently used the VWP to examine the use of prediction in two linguistic contexts: disfluency repairs and focus constructions (Lowder & Ferreira, 2016a, 2016b). Let us consider repairs first. When a person says *I went to the animal shelter and came home with a dog uh I mean*, it seems plausible to assume that the listener will not only anticipate a repair, but will even anticipate a specific word as the repair—in this case, *cat*. Inspired by recent Noisy Channel models of processing (Gibson, Bergen, & Piantadosi, 2013; Gibson, Piantadosi, Brink, Bergen, Lim, & Saxe, 2013), we assume that comprehenders adjust for the possibility of speaker error (as well as other sources of distortion, including perceptual error and environmental noise) and normalize the input. Comprehenders do this in part by modeling the speaker's communicative intention and assessing what the speaker is likely to be trying to say. In the case of our disfluency example, if someone errs in saying *dog* in that sort of sentence context, it seems likely that the speaker intended to say *cat* instead. This proposal, then, uses the independently motivated idea of prediction in comprehension to answer a question that is critical if we are to develop a complete theory of natural language processing, and that is, how do listeners process imperfect input—the kind of stuff that they are likely to encounter in the real world? Our hypothesis was that they predict the repair using constraints from the sentence as well as information from the visual display.

To test this model of repair processing, listeners were presented with utterances such as *I went to the animal shelter and came home with a dog uh I mean a rabbit*, which constituted the repair version, and *I went to the animal shelter and came home with a dog and also a rabbit*, which constituted a coordination control. As the participants heard these sentences, they also examined visual displays while their eye movements were monitored. The visual displays contained four objects presented for 3 s prior to sentence onset and throughout the utterance: a dog, a rabbit, a cat, and a plant. The dog and rabbit are the mentioned objects, and the plant is an unrelated distractor. The key object is the cat, which is never mentioned but is hypothesized to be the predicted word in the repair condition. The coordination structure serves as an appropriate control because it contains the same two critical words *dog* and *rabbit* in the absence of any disfluency. Analyses of eye movements showed that, prior to the onset of *dog*, listeners anticipated mention of both the cat and the dog about equally; after hearing *dog*, looks to the cat dropped, but upon hearing *uh I mean*, looks to the cat returned to a high level. In contrast, in the coordination condition, the sequence *and also* led listeners to anticipate mention of the cat to a much lesser degree. Upon hearing *rabbit*, looks to the cat in the repair condition dropped once again as listeners concentrated their fixations on the mentioned object, the rabbit.

We interpreted these results as showing that listeners predict the repair in these disfluency contexts. Thus, a key function of prediction may be to help comprehenders cope with speaker error, which of course is common in everyday speech (and even in writing, especially of the casual variety). To examine the process further, we conducted a follow-up experiment in which listeners heard the repair and coordination examples already discussed, and also a version that we hypothesized would encourage prediction as well due to the presence of semantic focus. The sentence was *I went to the animal shelter and came home with not a dog but rather a …*, where the sequence *not an X* presumably leads to the creation of an "alternate set" (Rooth, 1992)—a set of candidates likely to contrast with the concept of a dog (eg, cat, gerbil, rabbit, etc.). One way to think of an alternate set is as a set of predictions about the input; the smaller the set, the more specific the prediction. Additionally, it can be assumed that probabilities are assigned to each member of the set, as has been argued for prediction more generally (eg, Kuperberg & Jaeger, 2016). In the example, *cat* would presumably be assigned a high probability, and the other items would be assigned lower probabilities. In the VWP experiment, listeners heard the repair, focus, and coordination versions in the context of the same visual displays. We replicated our finding that predictive looks to the cat occurred more often in the repair than in the coordination condition. In addition, we found that listeners anticipated mention of the cat about as often in the focus as in the repair conditions. This result, we argued, suggests that the prediction mechanism in both cases might be the same: namely, listeners might generate an alternate set of items, weighted by probability, one of which will later be explicitly mentioned. This finding is significant, we believe, not only because it again demonstrates the role of prediction in comprehension, but also because it provides an important piece of the puzzle for our overall argument—namely, that there is a link between focused or new information and prediction.

3.3 Prediction and Top-Down Processing

As mentioned at the start of Section 3.2, part of the reason for the interest in prediction, and word prediction in particular, is that it seems consistent with strongly top-down models of language processing. Not only does it appear that prediction requires information to feed down from higher semantic levels to lower levels of processing, it actually seems as if at times no sensory information might be required at all to generate a representation (see Kutas & Federmeier, 2000, for a strong version of this argument, and Lupyan & Clark, 2015, for a broader perspective). It has been suggested that models of language comprehension have gradually moved in the direction of earlier and earlier semantic effects, from the models of the 1960s and early 1970s which assumed that semantic processing happened only at major syntactic boundaries, to models of the 1980s and 1990s which argued for strong incremental processing (all semantic processing and integration carried out on each word as it is encountered), to models proposed in this century which assume that semantics can get ahead of current perceptual processing (eg, prediction, good-enough processing). The general trend is away from bottom-up, modular (Fodor, 1975) models of language processing and toward models that are highly interactive.

A complete answer to the question whether word prediction is in fact a problem for noninteractive, modular language processing architectures is beyond the scope of this chapter; here, we will simply observe that the answer depends on how and when the relevant sources of information that support the prediction are accessed and integrated. Let us take the example of a sentence containing the words *windy* and *fly* and a person's ability to predict the word *kite*. Although it could be viewed as a top-down effect, an alternative view is that this prediction is attributable to what have previously been termed intralexical associations (Duffy, Henderson, & Morris, 1989; Fodor, 1983; Hess, Foss, & Carroll, 1995; Morris, 1994); that is, within the lexicon itself, the words *windy*, *fly*, and *kite* are associated, and spreading activation within the lexicon is what leads to activation of the concept *kite*. It is striking how many cases of prediction in the literature can be attributed to this kind of passive process; for example, many experiments capitalize on the existence of compound nouns and contrast sentences containing phrases such as *wedding cake* and *wedding pie*. These findings are no more of a challenge to bottom-up, modular architectures of language processing than are results suggesting syntactic prediction, where to our knowledge no researchers have attempted to argue that anticipation of syntactic nodes based on syntactic left context requires any sort of explanation in terms of feedback or interaction (for further discussion on dissociating top-down prediction effects from passive associative priming, see Lau, Holcomb, & Kuperberg, 2013). Surprisal effects on fixation times and even neural activation also do not require a top-down explanation, for the same reasons.

It appears that some of the best evidence for a prediction effect that cannot be explained in terms of more passive processes comes from the VWP and the anticipation of words prior to any input, and in the absence of any obvious lexical associations. Let us recall the basic Altmann and Kamide (1999) finding that *the boy will eat the* __ leads to anticipatory fixations to a depicted cake. Given that *boy, eat*, and *cake* are only weakly associated, the within-the-lexicon account of prediction looks like a nonstarter. However, let us return to our point about the preview of the visual display that is a normal part of every VWP experiment, including our own. We have shown that, without the preview, anticipatory effects do not emerge (Ferreira et al., 2013). We demonstrated that the preview is essential for establishing the concepts that are likely to play a role in the upcoming sentence and even for allowing the listener to anticipate many of the specific utterance properties (eg, whether it will be declarative, what kinds of arguments it is likely to contain, etc.). Returning to the Altmann and Kamide example, the 3-s preview is sufficient to allow the subject to establish the identities and even the phonological forms of each of the depicted concepts. Then, upon hearing *eat*, the listener can easily guess that the next word will be the only item in the visual display that happens to be edible. It is certainly reasonable to treat this as a top-down effect, but an alternative view is that even researchers holding the most modular views of processing have assumed that people are capable of reasoning from linguistic and visual information and using their knowledge to guide their eye movements. Similarly, our VWP experiments providing evidence for prediction in repair disfluency and focus structures can be interpreted as showing that listeners are able to reason from semantic information to enable them to elaborate their semantic representations. An alternate set is a semantic structure that the listener generates based on the semantic content of the utterance, along with some pragmatic knowledge about the speaker's likely intentions. Again, to us, nothing about this process seems to mandate a strongly top-down language processing architecture. Nevertheless, we appreciate that this is an issue that requires a great deal of additional consideration and discussion.

3.4 Summary

We have presented a very brief review of evidence from three different areas of psycholinguistics that demonstrates the role of prediction in language comprehension. We have suggested that prediction in cloze tasks might create a misleading picture regarding the general predictability of everyday texts; moreover, this work might also get the focus backward in the sense of emphasizing high predictability rather than recognizing the inverse relationship between predictability and information. Given that at least one major goal of communication is to acquire information, presumably the system was designed to be able to integrate items that are not highly predictable. Our experiments on repair disfluencies and focus suggest a way of reconciling these perspectives. The idea is that comprehenders process information in sentences by generating a set of candidates based on context as well as semantic knowledge, each of which is assigned a probability. A highly predictable word is one that was assigned a high probability in that candidate set; a word of low predictability is one that was assigned a lower probability; and an entirely unpredictable word is one that is entirely outside the set and not semantically or associatively related to any set member. We will expand on these ideas in the next sections.

4. INFORMATION STRUCTURE: GIVEN BEFORE NEW

Information in sentences is not uniformly distributed; instead, sentences are structured so that some content can be linked to what precedes it in the text, discourse, or conversation, and the remaining content can be treated as new—that is, as *information* in the technical sense of the term—that should be added to the ongoing model of the linguistic material. Perhaps the best-known theory of information structure is Haviland and Clark's (1974) Given–New Strategy, which makes a number of important assumptions about how language is used and processed. First, Haviland and Clark assume that the primary purpose of language is to communicate *new* information. Listeners attempt to identify the given information in any sentence primarily so it can be used to address the contents of long-term memory, and they identify the new information and integrate it with what has already been stored away. Ease of comprehension is determined by how easily the antecedent for the given information can be located in memory, and how easily the new information can be integrated with it. The given-new organization of a sentence is signaled in a variety of ways, the most important of which is the earlier syntactic positioning of given information. In addition, some syntactic forms exist specifically to convey givenness and focus; for example, a cleft such as *It was the sandwich that John ate* conveys the given information that John ate something and the new information that what John ate was a sandwich.

A variety of linguistic devices exist to allow speakers to communicate the given-new structure of an utterance. As already mentioned, the subject of a sentence is the default location for given information, which is reflected in tendencies such as the frequent occurrence of pronouns in subject position. Discourse markers such as *similarly* and *in contrast* also convey information about what is already known and what should be treated as the focus. In general, a variety of tools exist to convey presupposition, including expressions such as *too* (*Mary had lunch too* presupposes that someone other than Mary had lunch), subordinate clauses (*After Mary had lunch John did the dishes* presupposes that Mary had lunch), and restrictive modifiers such as relative clauses. Restrictive modifiers are a particularly interesting case, as they have received careful treatment from Steedman and his colleagues to explain their role in many garden-path phenomena (Altmann & Steedman, 1988; Crain & Steedman, 1985). Steedman and colleagues' argument is that restrictive modifiers presuppose a set of items denoted by the head noun of the phrase; for example, *the soldiers warned about the dangers conducted the midnight raid* presupposes a set of soldiers, a subset of which were warned about some dangers (see also Ferreira & Clifton, 1986, for a slightly different take on the same phenomenon). Thus, a comprehender encountering a subordinate or relative clause will expect those forms to convey information that can be recovered from previous discourse, suggesting that they will trigger a search for matching content in long-term memory. They may also process such content more superficially, because it is redundant and serves mainly to provide a bridge to what has already been communicated.

Another tool for conveying information about the given-new structure of a sentence is prosody. In general, new information tends to be more prominent prosodically than given information, and given information is generally spoken with lower pitch, less intensity, and shorter duration. Consistent with the finding that new information tends to occur later in sentences, linguists have long argued for a default rule of sentence phonology which places main sentence stress at the ends of phrases and clauses (the Nuclear Stress Rule; Chomsky & Halle, 1968; for recent discussion, see Zubizaretta & Vergnaud, 2006). As Haviland and Clark point out, focal stress always falls on new information, and thus the view that emerges from prosody is that, in the default case, new information will be found in later word positions within a sentence. It is perhaps useful to distinguish between two kinds of focus: presentational focus and contrastive focus (for review, see Drubig & Schaffar, 2001). Presentational focus is information that is simply new in the discourse, while contrastive focus is information that is new and requires the exclusion of potentially context-relevant alternatives (Drubig, 1994; Kiss, 1998; see Winkler & Gobbel, 2002, for discussion). Thus, even the default stress that results from application of the Nuclear Stress Rule causes the prosodically emphasized information to be focused, not in the sense that it is contrasted with something else, but simply in the sense that it is not to be found in the previous discourse.

We have already noted that given information is associated with earlier positions within a sentence. This would appear to be helpful to the comprehender on the assumption that the role of given information is to connect new ideas to content already established or known between the communicator and the audience, because then the comprehender knows where to attach the new information in the ongoing representation of the conversation or discourse. Fortunately, speakers tend to order information so that given precedes new (Ferreira & Yoshita, 2003). For example, if I ask the question *What did John eat for lunch?*, you would likely respond by saying *I think he ate a sandwich*, not *A sandwich was eaten by John*. Generally, the subject position is reserved for information that is given. Moreover, the tendency to place given information early is graded, so that the most likely syntactic position for given information is the subject, followed by the direct object, and then the indirect object (Keenan & Comrie, 1977).

What accounts for this tendency on the part of the speaker? An important idea that has emerged in the sentence production literature is that it is a by-product of speakers' preference for starting their sentences with the information that is most accessible to them (Bock, 1982; Bock & Warren, 1985; Tanaka, Branigan, McLean, & Pickering, 2011), a strategy MacDonald has referred to as "Easy First" (MacDonald, 2013). Experimental and corpus investigations of language production have shown that animate, frequent, imageable, and shorter expressions tend to occur earlier in sentences than expressions that are inanimate, less frequent, abstract, and long (eg, Bock, 1982; Bock & Warren, 1985; Ferreira, 1994; Stallings, MacDonald, & O'Seaghdha, 1998; Tanaka et al., 2011). The preference to order given before new (eg, Bock & Irwin, 1980) is thus simply another example of Easy First. Moreover, the given-new ordering preference is observed in languages other than English, including ones whose syntactic system permits fairly free arrangement of grammatical constituents (Christianson & Ferreira, 2005; Ferreira & Yoshita, 2003 for demonstrations in Odawa and Japanese, respectively). The advantage to the speaker of placing easy information in earlier sentence positions is that it provides extra time to work on generating the more difficult content; at the same time that the easy information is being articulated, the more difficult information can get planned.

These findings have motivated what are generally referred to as incremental theories of language production, which assume that speakers do not plan utterances in their entirety prior to speaking, but instead interleave planning and execution processes so as to maximize fluency as well as allocation of resources (Ferreira & Swets, 2002; Levelt, 1989; Meyer, 1990). Incremental theories of language production assume that syntactic flexibility and grammatical options exist in part to allow speakers to maintain this Easy First strategy. For example, if someone asks *What did John eat*, the reply that conforms to given before new is *He ate a sandwich*. But if someone asks *What happened to John at the restaurant*, the speaker can take advantage of the passive form to place given before new, thus saying something like *John was ignored and never got served*. As Ferreira (2000) has argued, this approach treats the decision to place easier concepts early in sentences as the fundamental planning operation; syntactic structures themselves are not selected, but simply emerge as by-products of the planning process. If an entity that is accessible and is placed early in the sentence happens to be a theme or patient, then a speaker of English will be obliged to generate a passive to accommodate that early placement.

5. PUTTING IT ALL TOGETHER: INFORMATION STRUCTURE, GOOD-ENOUGH PROCESSING, AND PREDICTION

As we have seen, biases that are rooted in the architecture of the language production system lead speakers to place given information before new. How does this distribution of information affect the comprehension of language? We assume that the first fundamental step for the reader or listener is to identify the given information in the utterance and link it to ideas that have already been established in the ongoing discourse. Given the reliability of speakers' tendency to order given before new, the comprehender will presumably assume that the given information is to be found in the earlier part of the sentence. Syntactic forms such as clefts, subordinate clauses, and restrictive modifiers will also be treated as redundant and discourse-linked. In addition, this given information presumably only needs to be processed deeply enough to allow a link to long-term memory to be established; as Haviland and Clark (1974) argued, the comprehender's goal is primarily to simply locate an antecedent for the given information so that the new information can be easily integrated. Processing resources should be devoted to the integration of new content, because that is the information that is crucial to the communicative exchange. On this view, the goal of comprehension is to identify and integrate new information, and the given concepts are there primarily to provide hooks onto which the new information can be hung.

One objection that might be raised at this point is that in most investigations of sentence comprehension, subjects are typically shown lists of single, unrelated sentences, or occasionally they might be presented with sentence pairs. It would appear, then, that these sentences lack the discourse context that is presumably required to establish any content as given. Our response is twofold. First, we believe this is an important issue, and moreover, a major advantage of the approach we have advocated in this chapter is that it could serve to renew interest in psycholinguistic studies of sentence comprehension in rich discourse contexts. It would also be helpful to study richer and more varied nonlinguistic contexts, such as more realistic visual worlds (Henderson & Ferreira, 2004). Second, although it is true that it is only in a proper discourse that a piece of information can be genuinely given, it is also true that certain linguistic devices exist to convey the given-new structure of a sentence and that information gets conveyed even without context. For example, a sentence such as *What John ate for lunch was a sandwich* communicates the given information that John ate something for lunch and the new information that that something was a sandwich, and this is true even outside any context establishing that division. Similarly for forms such as preposed subordinate clauses, relative clauses, and certain kinds of prosodic focus—these all communicate givenness and newness by virtue of their linguistic forms, although the effects would likely be reinforced in an appropriate context.

Returning now to good-enough language processing, this approach leads to the prediction that it is not the entire sentence that tends to be processed in a good-enough way, but only the given portion. Is the evidence consistent with this prediction? The answer is that it seems to be, although experiments designed to test the prediction directly would certainly be welcome. But as we mentioned in our previous description of the experiments on the misinterpretation of garden-path sentences, we observed that comprehension errors are quite specific and localized. Following a sentence such as *While Anna bathed the baby played in the crib*, subjects who were asked whether Anna bathed the baby often made the error of saying yes, presumably based on their inability to relinquish the original meaning associated with the incorrect syntactic analysis. But subjects asked whether the baby played in the crib answered the question correctly essentially 100% of the time, with no difference in performance for garden-path and control conditions. Our interpretation of this pattern was that comprehenders succeeded in revising the syntax and semantics of the sentence so as to locate an obligatory subject for the main clause (*the baby played in the crib*), but they did not always inhibit the meaning associated with the syntactic misparse in favor of the one supported by the correct analysis. This description of the data is accurate, but now we can speculate a bit more about why this pattern was observed. According to the analysis we have presented here, it is because comprehenders superficially process the subordinate clause, and that in turn is because it is treated as given information (a bias that the preposing of the subordinate clause would exaggerate). Thus, our suggestion is that good-enough processing takes place for given information, allowing comprehenders to allocate resources to the processing of what is new and potentially informative. The content that is redundant, given, and highly predictable can be attended to minimally in favor of content that is new.

But is there any direct evidence for this idea that information processing resources are allocated differently to given and new information? In fact a great deal of work going back to the earliest days of psycholinguistics supports it. In one study which used the picture-sentence verification task, listeners presented with a sentence inconsistent with the content of the picture tended to overlook the discrepancy when it was about the presupposed content of the sentence (Hornby, 1974). For example, given *The one that the cat is being petted by is the girl*, listeners tended to overlook the presence of a dog in the picture while noticing the inconsistency of a depicted boy. In a cleft structure, the element that is presupposed is in the relative clause (*that the cat is being petted*), and so this finding demonstrates that given information is given less scrutiny than the main assertion of the sentence. Another compelling example of the same tendency is the Moses illusion discussed in Section 1. Recall that given *How many animals of each kind did Moses take on the ark?*, many subjects will answer with "two" rather than challenging the presupposition behind the question (Erickson & Mattson, 1981), in part because presupposed information tends to be "given a pass" during comprehension. Consistent with this idea, Bredart and Modolo

(1998) showed that focusing the critical content (*It was Moses who took two animals of each kind on the ark*) led to higher detection rates (see also Kamas, Reder, & Ayers, 1996). Similar conclusions regarding the superficiality of language processing, particularly for presupposed information, come from work on "change detection" in texts (Sanford, Sanford, Molle, & Emmott, 2006; Sanford & Sturt, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004), which shows that changes to focused information are more likely to be detected than changes to content that is given.

The explanation for this pattern is that people tend to save processing resources for the part of the sentence that is new and in focus. An early demonstration of this tendency to emphasize focused content during processing comes from Cutler and Fodor (1979), who asked participants to monitor for a particular phoneme in a sentence following a question that focused different words and concepts. They observed that phoneme monitoring times were faster for a phoneme that was part of a focused word, indicating that participants devoted more processing resources to information when in focus. More recently, Lowder and Gordon (2015a) systematically varied the degree to which a target word was given versus focused and used eye movement monitoring to measure processing effort. Their findings indicate that increased focus is associated with longer reading times, consistent with the notion that degree of focus correlates with deeper processing (see also Benatar & Clifton, 2014; Birch & Rayner, 1997; Price & Sanford, 2012). Relatedly, Lowder and Gordon (2012, 2013, 2015b, 2015c, in press) have shown in several experiments that propositional content is processed more quickly when in a relative clause or other modifying phrase compared to when it is presented in a main clause, supporting the notion that relative clauses contain given information and are therefore processed quickly and superficially. Overall, as Lowder and Gordon (2015a) point out, focused information is privileged in a number of ways: Compared with given information, it attracts more attention, is remembered better (Birch & Garnsey, 1995; Gernsbacher & Jescheniak, 1995; McKoon, Ratcliff, Ward, & Sproat, 1993; Singer, 1976), and is scrutinized more carefully during performance of validation and verification tasks (Baker & Wagner, 1987; Bredart & Modolo, 1988).

The story so far, then, is that given information is processed more quickly and superficially than new information, leading to good-enough effects. We have also seen that the given information is treated in this way to save processing resources for what is informative. We now turn to the question of what operations those processing resources are being allocated to perform.

And at this point, we return to the idea of prediction in comprehension and propose that the comprehension system allocates resources to the task of implementing the operations that support prediction. In other words, we suggest that the purpose of prediction is to facilitate the integration of new information during language comprehension. This idea may seem counterintuitive because much of the literature on prediction focuses on *successful* prediction and emphasizes factors and findings that enable people to predict right down to the precise word that will come next in the string. As we argued in Section 3.2.1, this emphasis on successful prediction is in part a by-product of the use of cloze methods to investigate the effects of linguistic constraint on processing. But we believe this puts the emphasis in the wrong place: Successful comprehension is not about having predictions precisely confirmed; indeed, if the fundamental purpose of language is to enable the communication of information, then perfect predictability will be rare rather than typical. Instead, successful comprehension is about successfully integrating new information, and prediction mechanisms facilitate that task. To see this point, recall that informativeness and predictability are inversely correlated, so that if a word or event is 100% predictable, it contributes no information. Thus, what the comprehension system presumably evolved to do is to integrate information—less than perfectly predictable content—in an efficient way. The system is designed to spare scarce processing resources by processing given information in a good-enough way, and it devotes those saved resources to the task of integrating the new information.

How does prediction accomplish the task of integrating new information? Our proposal is that the comprehension system uses the prior linguistic context, world knowledge, and other forms of stored knowledge to generate a set of candidates for the new information. Each member of the set is assigned a probability based on those priors. For example, if I ask *What did John eat for lunch*?, your response will begin with the given information; you will say *John ate* __, where the blank indicates the position in which the new information will be provided. (Of course, the speaker might simply respond by saying *a sandwich*; the fact that the given information does not even need to be articulated further highlights the extent to which it is known and recoverable.) Whatever you provide as the object of *ate* is new information, but notice that, as the comprehender, I can get a head start by generating a set of candidates, which would presumably include something like {*a sandwich/a piece of pizza/a salad,...*} weighted by likelihood—eg, *a sandwich* might be assigned the highest probability based on general cultural information, John's own lunch preferences, and so on. Thus, I cannot predict exactly what you will say (if I could, why would I ask the question?), but I do have some idea of the conceptual world from which the answer will come (eg, words such as *snakes* and *rocks* are likely not in the candidate set).

Our proposal, then, is that comprehenders' ability to generate a set of likely candidates for the new content of a speaker's utterance constitutes a set of predictions, and those predictions facilitate the task of integrating new information. On this view, the norm is not for words to be 100% predictable or even to be highly predictable; the norm is for words to be of mild predictability depending on a range of factors. But most importantly, if sentences have a given–new information structure, and if the goal of language interactions is to communicate information, then it seems clear that the comprehender will be on the lookout for new information. Moreover, this idea fits with the general idea in the prediction literature that "prediction failure" is important because it is an opportunity to learn; failures of prediction are an opportunity to update one's priors to include the ideas contained in the unexpected content.

6. CONCLUSIONS

In this chapter, our goal has been to put together three important and influential ideas that have been influential in psycholinguistics for over half a century: information structure, superficial (good-enough) language processing, and prediction. Our argument is that the given–new structure of a sentence guides its processing, as illustrated in Fig. 1. Given information is processed superficially, with the result that the representation can end up

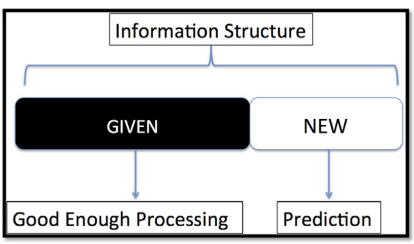

Figure 1 Our framework for linking information structure to language processing. Content identified as given is processed only deeply enough to allow an antecedent to be found in the ongoing discourse representation; new information is integrated by anticipating likely words, structures, or events.

missing key details or even failing to reflect its actual content; this is the effect of good-enough processing on the given portion of the sentence. New information is the target of processing effort, and the mechanism that supports the integration of the new information is the generation of a set of predictions. We characterize those predictions as similar to the alternate sets that have been proposed by linguists in studies of linguistic focus (Jackendoff, 1972; Rooth, 1992). We have updated this idea by hypothesizing that the alternate set contains candidates weighted by probability, with those probabilities determined both by the current linguistic content and by prior knowledge.

We believe this framework has many advantages over current approaches to language comprehension. In contrast to our previous formulation of the good-enough model (Ferreira et al., 2002; Ferreira & Patson, 2007), it puts important constraints on when and where good-enough language processing will take place: This framework states that it is primarily given or presupposed information that will be processed in a good-enough manner. This claim can easily be tested. Moreover, this approach tells us a bit more about what is "good-enough" from the perspective of the comprehender: It is enough simply to locate an antecedent for the given information. Our prediction is that once the antecedent has been identified, processing will proceed to the next proposition in the sentence, creating the potential for superficial and inaccurate analyses. Of course, because that content is redundant, the cost for superficiality or inaccuracy will be minimal, another prediction that we would like to test in future studies. In addition, and in contrast to many current theories of language processing that emphasize the role of prediction, our framework acknowledges the importance of prediction while avoiding the implication that the goal of processing is to discover redundant, predictable content. The approach we advocate assumes that comprehenders seek information in the technical sense of the term (content that is nonredundant and unpredictable; Shannon, 1948), and the purpose of the prediction mechanisms is to ease the integration of that information. The generation of a set of candidates weighted by their likelihood gives the comprehender a head start on integrating the new content.

REFERENCES

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1061–1069.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.
Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Holt, Rinehart and Winston.
Baker, L., & Wagner, J. L. (1987). Evaluating information for truthfulness: the effects of logical subordination. *Memory & Cognition, 15*, 247–255.
Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society: Biological Sciences, 364*, 1181–1182.
Benatar, A., & Clifton, C., Jr. (2014). Newness, givenness and discourse updating: evidence from eye movements. *Journal of Memory and Language, 71*, 1–16.
van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.
Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition, 25*, 653–660.
Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34*, 232–267.
Bock, J. K. (1982). Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.
Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior, 19*, 467–484.
Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.
Boston, M., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam sentence corpus. *Journal of Eye Movement Research, 2*, 1–12.
Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 15*, 607–624.
Bredart, S., & Modolo, K. (1988). Moses strikes again: focalization effect on a semantic illusion. *Acta Psychologica, 67*, 135–144.
Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition, 136*, 135–149.
Chomsky, N., & Halle, M. (1968). *Sound pattern of English*. New York: Harper & Row.
Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition, 98*, 105–135.
Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407.
Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes, 42*, 205–238.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181–204.
Clifton, C., Jr., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language, 86*, 1–19.
Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available? A memory-based view of text processing. *Discourse Processes, 26*, 109–129.
Crain, S., & Steedman, M. J. (1985). On not being led up the garden path: the use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological computational, and theoretical perspectives*. Cambridge: Cambridge University Press.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7*, 49–59.
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.
Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*, 193–210.
Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology, 3*. Article 548.
Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.
Drake, E., & Corley, M. (2015). Effects in production of word pre-activation during listening: are listener-generated predictions specified at a speech-sound level? *Memory & Cognition, 43*, 111–120.
Drubig, H. B. (1994). *Island constraints and the syntactic nature of focus and association with focus. Arbeitsbericht #51*. University of Tubingen.
Drubig, H. B., & Schaffar, W. (2001). Focus constructions. In M. Haspelmath, E. Konig, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (vol. 2, pp. 1079–1104). New York: Walter de Gruyter.
Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 791–801.
Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540–551.
Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84.
Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*, 715–736.
Ferreira, F. (2000). Syntax in language production: an approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press/Taylor & Francis.
Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164–203.
Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.
Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*, 3–20.
Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic parsing. *Journal of Memory and Language, 25*, 348–368.
Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*, 165–182.
Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1*, 71–83.
Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84.
Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research, 32*, 669–692.
Fillenbaum, S. (1974). Pragmatic normalization: further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology, 102*, 574–578.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies* (Ph.D. dissertation, University of Connecticut). West Bend, IN: Indiana University Linguistics Club.
Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.
Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*, 127–138.
Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive Psychology, 29*, 24–58.
Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 8051–8056.
Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24*, 1079–1088.
Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159–166).
Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics, 3*, 199–244.
Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521.
Henderson, J. M., & Choi, W. (2015). Neural correlates of fixation duration during real-world scene viewing: evidence from fixation-related (FIRE) fMRI. *Journal of Cognitive Neuroscience, 27*, 1137–1145.
Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage, 132*, 293–300.
Henderson, J. M., Choi, W., Luke, S. G., & Desai, R. H. (2015). Neural correlates of fixation duration in natural reading: evidence from fixation-related fMRI. *NeuroImage, 119*, 390–397.
Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 1–58). New York, NY: Psychology Press.
Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*, 62–82.
Hornby, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior, 13*, 530–538.
Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: response bias not sensitivity. *Memory & Cognition, 24*, 687–699.
Keenan, E. L., & Comrie, B. (1977). Accessibility and universal grammar. *Linguistic Inquiry, 8*, 63–99.
Kiss, K. E. (1998). Identificational focus versus information focus. *Language, 74*, 245–273.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32–59.
Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.
Langford, J., & Holmes, V. M. (1979). Syntactic presupposition in sentence comprehension. *Cognition, 7*, 363–383.
Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*, 484–502.
Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.
Lowder, M. W., & Ferreira, F. (2016a). Prediction in the processing of repair disfluencies. *Language, Cognition and Neuroscience, 31*, 73–79.
Lowder, M. W., & Ferreira, F. (2016b). Prediction in the processing of repair disfluencies: evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication*. http://dx.doi.org/10.1037/xlm0000256.
Lowder, M. W., & Gordon, P. C. (2012). The pistol that injured the cowboy: difficulty with inanimate subject-verb integration is reduced by structural separation. *Journal of Memory and Language, 66*, 819–832.
Lowder, M. W., & Gordon, P. C. (2013). It's hard to offend the college: effects of sentence structure on figurative-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 993–1011.
Lowder, M. W., & Gordon, P. C. (2015a). Focus takes time: structural effects on reading. *Psychonomic Bulletin & Review, 22*, 1733–1738.
Lowder, M. W., & Gordon, P. C. (2015b). Natural forces as agents: reconceptualizing the animate-inanimate distinction. *Cognition, 136*, 85–90.
Lowder, M. W., & Gordon, P. C. (2015c). The manuscript that we finished: structural separation reduces the cost of complement coercion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 526–540.
Lowder, M. W., & Gordon, P. C. (in press). Eye-tracking and corpus-based analyses of syntax-semantics interactions in complement coercion. *Language, Cognition and Neuroscience*.
Luka, B. J., & Van Petten, C. (2014). Prospective and retrospective semantic processing: prediction, time, and relationship strength in event-related potentials. *Brain and Language, 135*, 115–129.
Lupyan, G., & Clark, A. (2015). Words and the world: predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24*, 279–284.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 226.
MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56–98.
Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 843–847.
McKoon, G., Ratcliff, R., Ward, G., & Sproat, R. (1993). Syntactic prominence effects on discourse processes. *Journal of Memory and Language, 32*, 593–607.
Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables of a word. *Journal of Memory and Language, 29*(5), 524–545.
Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92–103.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: what readers believe versus what they read. *Psychological Science, 3*, 229–235.
O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: a memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1200–1210.
Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 280–285.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–190.
Price, J. M., & Sanford, A. J. (2012). Reading in healthy ageing: the influence of information structuring in sentences. *Psychology and Aging, 27*, 529–540.
Pullum, G. K. (May 6, 2004). *Plausible angloid gibberish [Web log post]*. Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/000860.html.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*, 79–87.
Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21*, 448–465.
Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *In Proceedings of the 2009 Conference for Psycholinguistic Modeling via Incremental Top-Down Parsing* (pp. 324–333).
Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75–116.
Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes, 42*, 109–130.
Sanford, A. J., & Sturt. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences, 6*, 382–386.
Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Journal, 27*, 623–656.
Singer, M. (1976). Thematic structure and the integration of linguistic information. *Journal of Verbal Learning and Verbal Behavior, 15*, 549–558.
Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*, 104–120.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.
Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language, 39*, 392–417.
Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: evidence from *either…or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 425–436.
Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1–17.
Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: an application of the change-detection paradigm. *Psychonomic Bulletin & Review, 11*, 882–888.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck, & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 397–439). New York: Oxford University Press.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: an ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68*, 297–314.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: evidence from Japanese. *Journal of Memory and Language, 65*, 318–330.

Winkler, S., & Gobbel, E. (2002). Focus, p-movement, and the nuclear-stress rule: a view from Germanic and Romance. *Linguistics, 40*(6; Issue 382), 1185–1242.

Zubizaretta, M. L., & Vergnaud, J. R. (2006). Phrasal stress and syntax. In M. Everaert, & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 522–568). Malden, MA: Blackwell.