Brain Research 1770 (2021) 147632

Contents lists available at ScienceDirect

Brain Research

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

Predicting syntactic structure

Fernanda Ferreira *, Zhuang Qiu

University of California, Davis, United States

ARTICLE INFO

Keywords:
Language processing
Prediction
Syntax
Reading
Psycholinguistics

ABSTRACT

Prediction in language processing has been a topic of major interest in psycholinguistics for at least the last two decades, but most investigations focus on semantic rather than syntactic prediction. This review begins with a discussion of some influential models of parsing which assume that comprehenders have the ability to anticipate syntactic nodes, beginning with left-corner parsers and the garden-path model and ending with current information-theoretic approaches that emphasize online probabilistic prediction. We then turn to evidence for the prediction of specific syntactic forms, including coordinate clauses and noun phrases, verb arguments, and individual nouns, as well as studies that use morphosyntactic constraints to assess whether a specific semantic prediction has been made. The last section considers the implications of syntactic prediction for theories of language architecture and describes four avenues for future research.

1. Introduction

This review focuses on language users' ability to predict syntactic categories, a topic that has been somewhat neglected relative to semantic prediction–the ability to anticipate the next word or series of words based on semantic constraints. Just as comprehenders anticipate that they will encounter the word *kite* following *The day was breezy so the boy went outside to fly a* _ (although of course the boy might have chosen to fly his toy airplane instead), comprehenders can also expect that if they hear or read the determiner *those*, they are likely to predict a plural noun phrase. Of course an optional adjective could come between the determiner and noun, delaying confirmation of the prediction, but the same is true in the semantic example as well: A modifier could occur before *kite* (or *airplane*). One reason syntactic prediction is significant is that it is necessarily implied in any account of semantic prediction. The prediction of a specific word based on meaning logically implies simultaneous prediction of a grammatical category as well, since all words belong to syntactic types. In other words, anticipation of the word *kite* based on semantic constraints also implies that a singular noun was anticipated as well. It would appear, then, that any time a semantic prediction is made, a syntactic prediction is generated as well. Note that the opposite does not hold, however: that is, a syntactic prediction may be made that lacks semantic content. For example, if a stranger walked up to you and said "Those…", you would be able to predict a plural noun phrase, but based on the single word and minimal context, not much else can be anticipated. Thus, syntactic prediction may be viewed as the more general phenomenon, reinforcing our contention that the topic deserves more research and more theoretical treatments. This review is an attempt to pull together some of the work that has been done on syntactic prediction to see what sort of picture emerges from the literature on the topic and to suggest productive future directions for research.

2. Prediction in language comprehension

By now a large number of comprehensive reviews of how prediction takes place during language comprehension have been published (Ferreira and Chantavarin, 2018; Huettig and Mani, 2016; Kuperberg and Jaeger, 2016; Nieuwland, 2019), and the current special issue updates those discussions and analyses. We therefore will only briefly summarize the main ideas behind the notion that prediction is essential for successful language processing. This fundamental insight is rooted in the idea that cognitive systems, whether biological or artificial and whether human or nonhuman, are "prediction engines" (Clark, 2013; Nave et al., 2020): Intelligence consists in using information about what has happened and what has worked in the past to anticipate stimuli and events. Organisms and machines that prepare for what is coming in this way are viewed as making faster, more optimal decisions than they would make if they passively waited for information and made decisions only once all relevant sources of information were received and integrated. A core assumption of this approach is that prediction is fundamentally a top-down process: Predictions are based on a cognitive

* Corresponding author.
E-mail addresses: fferreira@ucdavis.edu (F. Ferreira), zkqiu@ucdavis.edu (Z. Qiu).

https://doi.org/10.1016/j.brainres.2021.147632
Received 17 May 2021; Received in revised form 20 August 2021; Accepted 21 August 2021
Available online 25 August 2021

architecture in which higher-level sources of knowledge are used to preactivate sensory information. Consider scene viewing, for example (Henderson, 2017; Henderson and Hollingworth, 1998). The visual system does not merely process visual features and then pass those along to higher-level systems that integrate the information and then generate a conceptual understanding of what the scene is about; instead, each eye movement a viewer makes reflects the viewer's partial understanding of the scene that has been built up to that point, with saccades targeted towards regions of the scene that the visual system predicts will be informative. In other words, we do not merely perceive that a refrigerator is in a kitchen scene because we see it when our eyes land on it; instead, before we saw the refrigerator we predicted it would be there, allowing us to pre-activate some of its features and properties which are then confirmed (or not) when our eyes actually landed on the object. Indeed, evidence suggests that viewers false alarm to highly predictable objects not actually present in the scene based on these strong semantic expectations (Castelhano and Henderson, 2008).

It is not difficult to see how these ideas might be applied to language processing. Based on massive amounts of data, human comprehenders have built up a database of knowledge regarding linguistic patterns. Much of this knowledge is not even linguistic but is based on schemas and scripts that represent our general understanding of objects, events, and situations. If around noon someone says *I'm hungry, I think I'm going to go out and get some* __, based on a number of cultural conventions and your knowledge of English words, you will predict that the speaker will say *lunch* next. So-called "cloze completion" tasks, which are used in psycholinguistic work to quantify these sorts of expectations (Luke and Christianson, 2016; Staub, 2015), would verify that most naive subjects given that sequence will in fact generate the word *lunch* as a sensible continuation. Of course the speaker could say something slightly different–for example, the speaker might provide a specific food item such as *pho*–but recent research suggests this form would also be easily integrated because comprehenders appear to generate a set of semantic features that may be closely linked to a specific word but that will also lead to facilitation if a word consistent with those semantic features is encountered instead (Brothers et al., 2017; Szewczyk and Wodniecka, 2020).

The predictions that the language system makes allow it to get a head start on the input, making processing more efficient when the word or word sequence is actually encountered (Altmann and Kamide, 1999; Federmeier and Kutas, 1999). If the listener anticipates the word *lunch*, then processing of that word in the speech stream will be a simple matter of matching the preactivated features against the sensory input. If the listener anticipates *lunch* but instead hears *pho*, processing is still facilitated because although the preactivated phonological features will turn out to be wrong, many of the semantic features of *pho* overlap with those of *lunch* and so processing should still be faster than if no prediction was made at all (Szewczyk and Wodniecka, 2020). Prediction is also useful for learning. Imagine that someone does not know what *pho* is: Given the preamble and their knowledge of the world, including what people tend to eat around noon, they can infer that *pho* is a kind of food that is commonly eaten for lunch. The next time a sentence about items to eat for lunch is encountered, this person's knowledge representations will have been updated to include the information that *pho* is one possible food candidate. Predictions, then, facilitate processing of current input and learning of what may be encountered in the future.

The evidence for these ideas at this point is fairly massive. What is still less clear are the mechanisms behind semantic prediction. One proposal is that comprehenders may use mental imagery: Preactivation may be based at least in part on activation of images that correspond to linguistic content (e.g., the sound of a word; (Gambi and Pickering, 2016). Evidence also suggests that the tendency to predict during language comprehension varies by age, knowledge, and cognitive ability: Children engage in prediction less than adults do (Mani and Huettig, 2012) (not surprising since they know less about language and about the world), those with more knowledge and stronger literacy skills predict more than those with less knowledgeable and weaker literacy skills, and individuals with greater cognitive capacity such as a larger working memory span are also more likely to engage in prediction and to be successful when they do so. Interestingly, what happens in old age is still under debate. Some evidence suggests older adults predict less than adults in their 20s and 30s because systems that support efficient language processing are in decline (Federmeier et al., 2002; Wlotko et al., 2012; Wlotko and Federmeier, 2012), whereas other studies indicate that older adults predict more than younger adults do because they have a larger knowledge base from which to draw information, and that knowledge helps to compensate for their deteriorating sensory abilities as well as a general reduction in processing speed and fluency (Choi et al., 2017; Rayner et al., 2006). According to current probabilistic, information-theoretical models of language processing, prediction takes place on a word-by-word basis, with even relatively low levels of predictability facilitating the integration of input, following a Bayesian statistical algorithm. Other language models predict over scales other than the single word, ranging from statistical models that predict individual letters (Elman, 1990) to models that predict an entire utterance (Goodman and Frank, 2016).

A challenge for any model of prediction in language processing is to establish clearly that comprehenders have preactivated features of the upcoming input rather than merely integrating information more efficiently when it is encountered. Unfortunately, prediction and integration are difficult to disentangle (Ferreira and Chantavarin, 2018), and in an earlier era of psycholinguistics, it was suggested that if evidence is obtained that is consistent with both, the more parsimonious explanation appeals to integration, as integration clearly must be part of comprehension but prediction may be a special-purpose mechanism. Some have rejected this prediction-integration dichotomy altogether, essentially arguing that if a word is more easily integrated when encountered, some aspects of its representation must have been preactivated–that is, some features of that word must have been predicted (see, for example, Kuperberg and Jaeger, 2016). Others have attempted to distinguish prediction from integration with the use of clever paradigms designed to demonstrate processing difficulty which could only be observed if comprehenders had preactivated some piece of information. For example, in the Visual World Paradigm (VWP), a technique based on the use of eyetracking to monitor the saccades listeners make to depicted objects while listening to spoken sentences (Altmann and Kamide, 1999; Kamide et al., 2003a), studies have shown that an eye movement will target an object well before it is articulated in the speech stream, indicating that it was indeed predicted.

As has been true for the topic of prediction overall, evidence for prediction during language processing has been taken as support for a highly top-down cognitive architecture, one that stands in contrast to feed-forward models in which information is passed along in one direction only, from sensory data to high-level semantic representations. The idea is that comprehenders do not merely process the input; they generate it themselves, internally, based on their knowledge of linguistic constraints and real-world situations. This assumption may be correct for some kinds of prediction but clearly not for all. Indeed, much of the evidence for prediction that has been provided in the psycholinguistic literature does not require a top-down explanation. Consider again the example *The day was breezy so the boy went outside to fly a* _. The process of anticipating the word *kite* may involve simply completing a sequence that is prestored in memory: that is, the associations among *breeze, fly,* and *kite* are strong enough to activate *kite* simply as a completion of the linguistic pattern given *breeze* and *fly* (Jackendoff and Audring, 2020). Intra-lexical priming of this sort can be explained without invoking top-down processing, as the information needed to generate the prediction is in the same database and part of the same processing system as the predicted item.

This issue transitions us to the topic of this review, which is the prediction of syntactic structure. We suspect that one of the reasons syntactic prediction has received far less attention than semantic prediction is that it is less compelling as evidence for top-down processing. Instead, syntactic prediction appears to fall squarely in the category of linguistic pattern completion, or at least some sort of intra-modular processing within the language system (within the syntactic processor or parser). For example, given the word *those*, comprehenders might not know what specific word will come next, but they can predict they will soon encounter a plural noun, and they will show signs of processing distress if a conjunction or preposition follows instead. But to explain this pattern, there is no need to invoke the use of high-level knowledge to constrain low level processing; instead, what is involved appears to be one example of the sort of completion process that Huettig et al. (2021) consider to be true for prediction generally, in which, at multiple levels of representation the language system completes a sequence after encountering the sequence's initial portion. In this review, we focus specifically on the use of syntactic constraints for completion. Based on their knowledge of syntax, speakers of English know that a noun eventually must follow a determiner, and given the plural feature on *those*, they know that the noun must also be plural. This is simply a matter of accessing syntactic constraints and applying them in real time during comprehension; no higher-level semantic knowledge is involved. All the knowledge that is required to generate the prediction of a plural noun is inside the language system and all the relevant representations are given in a linguistic–indeed, in a syntactic–vocabulary. It is important to note that syntactic completions based on the application of structural constraints alone are only one example of how Huettig et al. view the potential completion processes in language comprehension. Other cases of prediction via completion that involve semantic contrasts (e.g., *Are you going to eat those* _ versus *Are you going to paint those* _) will make use of selectional restrictions and more lexically specific co-occurrences (e.g., the objects with which the verb *eat* tends to occur differ from those that co-occur with the verb *paint*).

Despite the inability of syntactic prediction to help us adjudicate between feed-forward and top-down architectures, we still contend that it is an important phenomenon to explore and understand. One reason is that, as we argued earlier, any instance of word prediction implies that syntactic prediction has taken place as well: That is, if you anticipate the word *lunch*, you have necessarily also anticipated a singular noun. Moreover, by exploring how syntactic prediction works, we can see how far we can get with a prediction account that does not invoke any sort of top-down processing mechanisms, which may be useful for theorizing about cognitive architectures. Yet another reason to highlight syntactic prediction is that the prediction versus integration debate is just as salient for syntactic prediction as it is for semantic prediction, but it may be easier to settle due to the more circumscribed nature of syntactic knowledge and grammatical constraints. Finally, we can briefly consider how semantic and syntactic prediction might work together during online processing. Let's return to the example *The day was breezy so the boy went outside to fly a* _. We observed that on the basis of stored lexical knowledge and intra-lexical associations the comprehender will activate the word *kite*, but as a reviewer of this piece noted, other words will be activated as well, including ones of the wrong syntactic category—for example, *soar* and *glide*. Semantic prediction will likely generate a large cohort of possible completions, but the syntactic prediction mechanism has specifically mandated a singular noun for that position in the sentence. The syntactic prediction, then, allows the processing system to winnow down the set of possible semantic completions to only those that are syntactically licensed, potentially making later integration of the input more efficient. Of course, this mechanism would not rule out other singular nouns such as *airplane* or even *bird*, which also may be activated by association and likely are only entirely ruled out once the word is perceptually processed.

3. Parsing models and syntactic prediction

A number of models of parsing make use of syntactic prediction, some more explicitly than others. This section first reviews influential models starting with the general concept of left-corner parsing, followed by the Garden-Path model and constraint-based lexicalist alternatives. We then turn to a discussion of Syntactic Prediction Locality Theory, Good-Enough and Noisy Channel processing, and finally information-theoretic models of surprisal and entropy. The common thread will be the extent to which prediction of syntactic constituents is a core mechanism in these approaches.

3.1. Left-Corner parsing

Some of the earliest arguments for prediction in language processing came from research on models of the language parser and arguments for so-called "left-corner" parsers (for discussion, see Crocker, 1999). A left-corner parser has access to a database of phrase-structure rules or constraints, and when a lexical item is encountered, builds as much structure as the applicable phrase structure rules license. In Crocker's (1999) model, when a noun such as *Cats* is encountered, that noun permits the parser to build its parent Noun Phrase (NP) node, which then licenses a Sentence (S) node for the overall structure. With these steps, the grammatical subject of the sentence is built, and now the parser can also project the Verb Phrase (VP) node before there is any lexical evidence for it–that is, before a main verb, auxiliary verb, or adverb is encountered. This VP node constitutes an expectation made based on the rules of grammar–that is, if we are parsing a sentence and a subject has been built, a verb phrase must follow. Similarly, if a preposition such as *above* is encountered, an NP node for a prepositional object can immediately be predicted based on the phrase structure rule for Prepositional Phrases (PPs): PP → P NP. Syntactic rules, then, specify patterns that are completed based on partial input. In the same way that *The day was breezy so the boy went outside to fly a* _ can be completed with the word *kite* given prestored associations among the content words, so too can the parser assume that the occurrence of an initial NP allows anticipation of a VP node based on the pattern S → NP VP, or that a preposition will be followed by an object based on the pattern PP → P NP.

A more recent version of this approach uses the formalism of Tree Adjoining Grammar (TAG), which is a class of grammars implemented as prestored syntactic treelets anchored by lexical items that get assembled by the parser into a connected structure (Demberg et al., 2013). For example, the word *cats* is the lexical anchor for a treelet consisting of the N and the NP node so that when the word is retrieved from the lexicon, the syntactic nodes it licenses are retrieved as well. That treelet is then available for insertion into a bigger tree. The Demberg et al. model posits a class of what they refer to as Prediction Trees, which are treelets with potential upcoming nodes that can be activated in some circumstances before the occurrence of any sort of lexical anchor. One type of prediction tree is the one formed from S → NP VP; that is, the parser has access to a clausal treelet based on the rule S → NP VP, which constitutes a prediction about the form the overall sequence is likely to take. The model posits several other kinds of prediction trees as well. A separate verification mechanism then checks whether the syntactic prediction is correct and converts predicted to lexically licensed nodes.

3.2. The Garden-Path model and its challengers

In the early 1980s, Lyn Frazier and colleagues proposed what has come to be known as the Garden-Path Model of human parsing (Frazier and Fodor, 1978; Frazier and Rayner, 1982; Rayner et al., 1983), which has been highly influential in the field of psycholinguistics. The Garden-Path model assumes that the parser has access to all the phrase structure rules of the language, and as words are sequentially encountered, they are structured into a connected tree according to those rules. For example, an utterance-initial word such as *cats* would trigger activation of the rules *cats* → N, N → NP, and S → NP VP, yielding a result much like what emerges from a left-corner parser. In the Garden-Path model, then, the parser predicts nodes prior to any lexical evidence for them.

Therefore, based on an initial encounter with a noun, the parser will predict the NP, S, and VP nodes. The challenge for the parser, however, is ambiguity: the possibility of applying more than one rule to the input. For example, given *Mary knows Tom*, the NP *Tom* could be structured via at least two different rules: VP → V NP or VP → V S. The former is the direct object analysis of *Tom* and the latter treats *Tom* as the subject of a complement clause. The solution Frazier proposed is that the parser goes with whichever analysis finishes first, which will be the analysis that requires the retrieval of fewer rules and the creation of fewer nodes. The simpler analysis simply finishes more quickly, and the parser is assumed to move on once it has a viable syntactic analysis. The result will be a preference to build the minimal amount of structure consistent with the input, a strategy that is referred to as the Minimal Attachment strategy.

The relevant point about the Garden-Path model with respect to syntactic prediction is that it assumes a parser that generates syntactic nodes in the tree prior to any encounter with licensing lexical input. The analysis the parser builds constitutes a prediction about how the sentence will continue: At the point where *Tom* is encountered in the *Mary knows Tom* example, the parser does not yet know what role *Tom* will play in the structure, but the parser predicts that it will turn out to be a direct object based on application of the minimal attachment strategy. If the sentence continues with a sequence such as *very well*, the prediction will be confirmed, but if *Tom* is followed by a verb such as *was*, the prediction is disconfirmed and the sentence structure has to be rebuilt. In the Garden-Path model, the parser is also assumed to be a processing module with no access to sources of information other than phrase structure rules and therefore no ability to use lexical constraints, including verb subcategorization information (Ferreira and Henderson, 1990). In addition, the parser was presumed to be insensitive to frequency information, and therefore would be unable to update its parsing strategies based on prediction error (i.e., based on frequency or experience). The combination of these two assumptions meant that the parser would initially misanalyze a sequence such as *Mary realized Tom* essentially every time it was encountered even though *realize* rarely takes a direct object complement, and the parser would also not update its strategy for analyzing the postverbal NP even though the result would almost always yield a prediction error. It is on these points (and several others) that the model was later challenged.

Focusing on the parser's inability to consult sources of information beyond phrase structure rules, follow-up work suggested that at least under some circumstances the parser can use lexical information stored with verbs to pre-acctivate nodes and prevent a syntactic misanalysis (Garnsey et al., 1997; Trueswell et al., 1993). For example, when readers encounter a verb such as *realize*, they also activate its subcategorization frame specifying that a clause is likely to follow (Mitchell and Holmes, 1985). The Garden-Path model's parser could not use this lexical constraint because, according to that model, the structure of a sentence is built up only from phrase structure rules without the benefit of co-occurring lexical information such as verb subcategorization frames. This assumption means that not only is the Garden-Path parser unable to avoid a misanalysis in the case of sentence-complement biased verbs such as *realize*, it also cannot predict structure based on lexical co-occurrence even when no ambiguity is involved. For example, given *Paula put*…, the Garden-Path parser would not be able to predict upcoming the NP and PP nodes even though those are obligatory arguments of the verb. Proponents of more lexicalist models challenged this assumption and developed parsers that operate by using stored lexical information to project the upcoming tree (MacDonald et al., 1994). These lexicalist models are less dependent on phrase structure rules and more reliant on syntactic projection from lexical items. In lexicalist models, prediction of syntactic nodes is an automatic byproduct of encountering a head that takes arguments: At least in a head-initial language like English, processing of a head will lead the parser to predict the head's arguments and insert them into the tree prior to any encounter with the words that will eventually be linked to those syntactic nodes. Interestingly, this trend in psycholinguistics reinforced a shift that was already taking place in generative grammar, where theories of syntax were reducing and even eliminating phrase structure rules and replacing them with collections of grammatical constraints, including those rooted in lexical semantics (Chomsky, 1970; Chomsky, 1981; Jackendoff, 1977).

Parallel, constraint-based models make similar assumptions regarding the use of lexical information during parsing but add the idea that all structures consistent with the current input are activated in parallel, with activation levels proportional to their frequency of use (Spivey et al., 2002). Given a string such as *Mary saw Tom*…, parallel constraint-based models "predict" both the structure in which the postverbal NP is treated as an object and one in which that NP is the subject of a complement clause. Those predictions are weighted by their likelihood of success given previous encounters with the relevant forms in the language. With regard to prediction, the difference between parallel, constraint-based models and the Garden-Path model is twofold, with both differences pointing in the direction of enhanced syntactic prediction: Parallel, constraint-based models can use lexical constraints to predict upcoming nodes based on head-argument dependencies as well as other types of frequent co-occurrences, and they allow the parser to predict multiple structures in parallel, with alternatives weighted by their likelihood of success given past experience. Since parallel, constraint-based parsers do make use of frequency information to determine activation levels of alternative analyses, it clearly follows that these models assume the parser can learn (MacDonald, 2013). More recent versions even allow the parser to adjust its priors based on relatively short exposure times such as in the course of a single psycholinguistic experiment (Farmer et al., 2011) but see (Harrington Stack et al., 2018).

Overall, then, we see that the Garden-Path model and the models that arose in response to it all assume that comprehenders predict syntactic nodes. If a word or phrase is ambiguous between a couple of different syntactic analyses and the parser proceeds anyway, it is essentially predicting how the sentence is likely to turn out. In addition, lexicalist models that build up structure based on frequent co-occurrences among words and syntactic forms can be viewed as parsers that predict structure in advance of the input. Moreover, as these models weight these analyses by frequency and adjust those frequencies based on experience, they also make use of information regarding the extent to which predictions are confirmed.

3.3. *Gibson's (1998) syntactic prediction locality theory (SPLT)*

The core ideas behind the SPLT (Gibson, 1998) are that syntactic processing involves prediction and that maintaining and confirming those predictions is a burden on working memory. As already discussed, many lexical items such as verbs specify the categories with which they like to co-occur, with the parser predicting the presence of those categories at the encountered word. In the SPLT, it is assumed that maintaining predictions is costly: If, for example, a determiner such as *the* is processed, the parser will predict the occurrence of a noun, with the prediction taking up memory units that are proportional to the distance over which the prediction must be maintained. In addition, integration costs are incurred when the parser then associates the current item with the word or syntactic category that predicted its presence, with greater distances between the predicted item and the point where it was predicted leading to greater costs. A good example of how this works is the comprehension of sentences involving wh-movement, which require the parser to resolve long-distance dependencies. If a sentence or clause begins with a wh-word such as *who*, English comprehenders know that that wh-word must be associated with some later position in the sentence for it to receive a thematic role. According to the SPLT, the greater the distance between the wh-phrase and the position in the sentence from which it can be assigned a role, the greater the memory load and subsequent processing complexity. The SPLT uses this principle to explain the well-known filled-gap effect, illustrated in an example such as *My brother wanted to know who Ruth will bring us home to at Christmas*, in which processing difficulty is experienced at the word *us* because the parser predicted that the wh-word *who* could be assigned a thematic role at *bring*, but the presence of *us* signals that *who* gets its thematic role further downstream, from the preposition *to*. In the language of prediction, the parser predicted that *who* would be the object of *bring* but it turned out to be the object of the preposition *to*. Maintaining the prediction over a longer distance and having to reanalyze the interpretation of *who* leads to measurable processing costs. Another phenomenon that provides evidence for the SPLT is the preference for short constituents to precede long ones: *John gave the book he found in London when he was on vacation to Mary* is worse than *John gave to Mary the book he found in London when he was on vacation*. The preference for short phrases to precede longer ones follows from the SPLT as follows: the verb *give* predicts two postverbal phrases, a theme and a recipient. If the first postverbal constituent is long, the prediction associated with the second one must be maintained for more time and integration of that phrase with the verb that licensed it must take place over a greater distance. If, instead, the short constituent precedes the long one, the prediction of the second postverbal phrase is maintained over fewer words and the integration distance is shorter as well, leading to reduced memory cost and easier processing.

This model, then, favors locality as a means of reducing the memory costs that are inherent in syntactic prediction (for additional arguments for locality, see (Hawkins, 2001; Liu, 2019), given the model's core assumptions. Memory costs are associated with the prediction of items based on the current input, and integration costs are tied to the need to link syntactic elements to the categories that license or predict them. As its name suggests, the SPLT model treats prediction as the core mechanism of sentence comprehension, and it is the need to maintain and confirm syntactic predictions that leads to different patterns of processing costs within a sentence and for different sentence types. It should be noted that anti-locality effects have been reported in the literature as well (Levy, 2008; Vasishth and Lewis, 2006), but those findings do not undermine the central idea that syntactic prediction takes place; instead, what is in dispute is whether maintaining a prediction leads to memory load, as assumed by SPLT, or whether holding a prediction strengthens the expectation, thus facilitating processing when the predicted constituent is finally encountered. But, either way, syntactic prediction is posited as a regular part of language processing.

3.4. Good-Enough and Noisy Channel processing

The models of comprehension discussed thus far view language processing as compositional: Any interpretation is a function of the words in a sentence and their syntactic arrangement. This idea is a useful simplification, but it ignores the fact that people often obtain interpretations that are inconsistent with a sentence's content. For example, the first author recently received an email from a colleague with the sentence *I hope you've had a spring break*; it was only when the colleague pointed out in a follow-up message that she'd left out the word *good* before *spring break* that the omission was even noticed. It appears that, in addition to standard mechanisms of processing that build up meaning compositionally from words and their arrangements, there is also a system that normalizes the input, sometimes even distorting the lexical and grammatical content to conform to expectations. This basic insight was the motivation for the Good-Enough model of language processing (Ferreira, 2003; Ferreira et al., 2002) and a later approach that invokes the idea of information processing over a noisy channel (Gibson et al., 2013). Both variants can be related to the idea of prediction, both semantic and syntactic, and their tight inter-connection.

The Good-Enough model treats the comprehenders' priors as a source of information that generates interpretations that exist in parallel with the one derived compositionally. Consider the case of implausible passive sentences such as *the dog was bitten by the man*. The compositional analysis treats the man as the biter and the dog as his unfortunate victim. But a schema-based interpretation based on prestored knowledge and expectations suggests the opposite interpretation. Which one of these two meanings of the sentence "wins" is based on a number of different factors, but the key idea for our purposes is that the system can generate a global sentence meaning based on its conceptual knowledge, and that unlicensed meaning may be founded on syntactic prediction (and consequent distortion). One possible mechanism to explain this tendency is that the parser may sometimes predict an active structure based on the content words *dog* and the lemma *bite* and those prediction mechanisms may in turn lead to the creation of a syntactic structure that is not faithful to the input. A study designed to test this idea showed that implausible passive sentences prime active forms, suggesting that their structure has been normalized to an active form in order to accommodate the schema-based interpretation (Christianson et al., 2001). The Noisy Channel model (Gibson et al., 2013; Gibson et al., 2019) provides some ideas about how this distortion may take place. According to the Noisy Channel approach, the language comprehension system has adapted to the reality of noisy input and flawed perceptual systems by correcting input so that it conforms to expectations, according to principles of rational inference. These corrections include distortions (e.g., treating *bitten* as *bit*), deletions, and additions. In the case of a sentence such as *the dog was bitten by the man*, these normalizations would allow the passive elements in the sentence (the auxiliary verb, the passive morphology on the main verb, and the preposition *by*) to be overlooked so that the sentence's interpretation can be made to conform to expectations (for additional discussion, see Kuperberg and Jaeger, 2016).

More evidence for prediction of syntactic structure based on Good-Enough processing comes from research showing that comprehenders often fail to recover from an initial garden-path misanalysis of a string (Christianson et al., 2001; Slattery et al., 2013). Given the sentence *while Mary bathed the baby played in the crib*, after *bathed* the parser will predict that the following NP *the baby* serves as its object, in part due to lexical expectations (*bathe* is a transitive verb) and due to application of the Minimal Attachment principle. Unfortunately, the main verb *played* needs that same phrase to serve as its subject. Syntactic reanalysis must occur so that *the baby* is made the subject of the main clause and the subordinate clause verb is then reinterpreted as intransitive. However, about half the time comprehenders still end up misunderstanding the sentence's meaning (Christianson et al., 2001) because the interpretation associated with the initial misanalysis on which *the baby* serves as the object of *bathed* lingers in memory (Slattery et al., 2013). We see, then, that an incorrect prediction regarding the role a phrase will play triggers not only the need for syntactic reanalysis but may also cause the comprehension system to misunderstand a key part of the meaning of the sentence.

The development of models that allow for unlicensed interpretations was a major advance for psycholinguistics because these models dramatically expand the range of phenomena that psycholinguistic theories are able to explain. Humans may misinterpret sentences due to well-founded expectations concerning what people are likely to say. These expectations may cause the parser to predict syntactic categories that support the more likely interpretations at the expense of accuracy. But as proponents of Noisy Channel models have demonstrated, this process is calibrated to the probability that a message might have been distorted, reassuring us that the language system is generally rational (in the information-theoretic sense) because it considers the likelihood of the input given a range of relevant factors, including speakers' communicative intentions.

3.5. Surprisal and entropy in Information-Theoretic models of language processing

Syntactic prediction is the key idea behind a set of models of language processing that not only assume word-by-word prediction, but also precisely quantify the strength of those predictions based on statistical analyses of language corpora, which are thought to reflect language users' experience. This enterprise of attempting to explain language processing difficulty in information-theoretic terms obviously assumes not only the reality of prediction, but also that it is a regular process that takes place prior for every word of a sentence or set of sentences. A metric known as surprisal quantifies the expectedness of a word based on its preceding context (Berger et al., 1996; Hale, 2016). The surprisal value of a word is the negative log probability of that word given the words that came before it in the sentence, with higher values indicating greater surprisal. This definition of surprisal collapses across a number of levels of linguistic representation, but some models separate lexical from syntactic surprisal, quantifying the expectation of a particular syntactic category given the left syntactic sentential context. Thus, the surprisal value of any noun following the determiner *the* would be low; in contrast, the surprisal value associated with a verb following the sequence *While Mary bathed the baby* would be high, as the parser predicts after *baby* that another NP will come next based on its assumption that the preceding NP is the object of *bathed*. Roark (Roark, 2001; Roark et al., 2009) showed that reading times are predicted better by a model that separates lexical from syntactic surprisal rather than conflating them into a single measure of expectation. Neuroimaging evidence also indicates that syntactic surprisal values correlate with activation in cortical regions associated with language processing (Henderson et al., 2016). A recent theoretical updating of the concept of surprisal combines the metric with the assumption that memory for the left context is imperfect, thus explaining phenomena such as the locality preferences discovered in the context of the SPLT (Futrell et al., 2020).

A potential criticism of the concept of surprisal as a theory of prediction generally and syntactic prediction in particular is that it might actually reflect integration rather than preactivation of syntactic content. An expectation-based approach naturally leads to the idea that a word will be more difficult to process when it is low probability given the left context, but this could be viewed as the processing system having difficulty integrating that low probability word or phrase at the point at which it is encountered. A complementary metric that deals with prediction more squarely is entropy, another information-theoretic concept (Hale, 2016; Lowder et al., 2017). Entropy refers to the uncertainty of a particular outcome; the greater the number of possible outcomes, the greater the entropy value. For example, there is less entropy at the main verb *realize* in a sentence of the form X verbed Y than there would be if the main verb were *know* because *realize* is almost always followed by a sentence complement, whereas *know* can be followed by a direct object or a sentence complement about equally often. Thus, the parser is better able to predict the upcoming syntactic structure given *realize* than given *know*, and therefore less processing cost should be observed at the verb. Notice that this experimental prediction is different from what would be derived from the Garden-Path model, for example, because in the Garden-Path model, the processing costs should all be concentrated on the disambiguating verb (the main verb of the complement clause); no effect at the main verb of the main clause is expected. This idea of entropy associated with processing difficulty has received some experimental support from a reading time study which showed that the degree of uncertainty about the full upcoming structure correlated with longer reading times (Linzen and Jaeger, 2016). Another eyetracking study showed that increased word-by-word surprisal and entropy reduction led to longer reading times independent of global text difficulty (Lowder et al., 2017). Moreover, entropy reduction increased single fixation and first fixation durations only, whereas surprisal affected all eyetracking measures, providing evidence that the two information-theoretic measures can be dissociated.

Information-theoretic approaches to measuring processing difficulty provide clear-cut evidence for syntactic prediction during online sentence comprehension. The correlations between high surprisal, high entropy, and high entropy reduction, on the one hand, and long reading times as well as greater neural activation, on the other hand, show that encountering a syntactic category that is unexpected as well as experiencing uncertainty about what will come next lead to processing costs. In addition, these metrics indicate that syntactic prediction is not something that happens only in special situations in which a syntactic category is highly likely but instead operates on a word-by-word basis, as a normal part of the parsing process, with some predictions being very low probability and others being more likely to be confirmed.

In this section we have provided a general overview of how various prominent theories of language processing incorporate prediction mechanisms, some more explicitly and directly than others. Whether it is the Garden-Path model, constraint-based lexicalist models, the dependency-locality model, Good Enough and Noisy Channel approaches, or information-theoretic models, all have been shown to treat mechanisms of syntactic prediction as fundamental to language comprehension. In the next section, we review a sample of studies that provide evidence for the prediction of specific syntactic categories in various linguistic contexts–for example, prediction of an *or*-clause following a clause beginning with the word *either*. Due to space constraints, this review will not be exhaustive but will highlight some illustrative experiments that show how the parser predicts specific syntactic forms during online processing.

4. Evidence for prediction of specific syntactic forms

One of the earliest empirical investigations of syntactic prediction examined the *either-or* construction (Staub and Clifton, 2006). A clause that begins with *either* requires a syntactic constituent beginning with the word *or*, as illustrated in *Either Linda bought the red car or her husband leased the green one*. In this eyetracking experiment, subjects read sentences like the example or an identical version without the word *either*. The logic was that if the word *either* was present, then readers could predict an *or*-clause, and this prediction would facilitate processing of that clause relative to the no-*either* condition. The investigators found shorter first-pass reading times on the *or her husband* region of the sentence in the *either*-present condition relative to the no-*either* condition. This facilitation persisted for the remainder of the sentence, and in addition, readers were less likely to make regressive eye movements from this second clause to the first one when *either* was included. This pattern indicates that the word *either* allowed the parser to "pre-build" a structure consisting of a sentence with two coordinated clauses. When the lexical items making up the *or*-clause were then encountered, they could be integrated more quickly because the structure for them had already been constructed. The design of this experiment also allowed the possibility of facilitation in NP-coordination constructions to be tested (*The team took (either) the train or the subway to get to the game*) and the same pattern was observed: the presence of *either* allowed the parser to anticipate coordination, in this case NP-coordination, which then facilitated processing of the second disjunct. This experiment shows, then, that the parser can use syntactic constraints to predict an upcoming syntactic form–in this case, either a second NP or clausal disjunct in a coordination structure. Similar results were reported for processing of coordinate structures in Mandarin Chinese (Qingrong and Yan, 2012).

The cleverness of using *either-or* as a tool for revealing the presence of syntactic prediction is that the construction allows prediction and integration processes to be distinguished. As Staub and Clifton argue based on Sturt and Lombardo (2005), in a phrase such as *the train or the subway*, the first NP *the train* is likely analyzed as a simple direct object and reanalyzed as the first part of a coordinate phrase once the word *or* is encountered. The presence of *either* allows the parser to avoid a garden-path because the parser will predict that that first NP is the first part of a coordinate phrase. Thus, the lack of garden-path effects in this study is the diagnostic for treating the effect of either as one involving prediction rather than integration.

As mentioned in the section on theories of parsing and the assumption of syntactic prediction, a robust source of information for prediction of syntactic constituents are verb subcategorization frames. A large number of studies had already investigated the extent to which these lexical constraints facilitate processing, but few were framed as investigations of syntactic prediction, as the idea of prediction was less trendy at the time. In more recent work, the visual world paradigm has been used to investigate this possibility more directly (Arai and Keller, 2013). In the first experiment, listeners were presented with sentences containing a verb that required a direct object and sentences with verbs that were intransitive: for example, *The nun punished the artist* vs. *The nun disagreed with the artist.* At the same time that listeners heard the sentences they viewed a screen containing three images: a nun, an artist, and an inanimate distractor. The investigators observed that listeners made anticipatory saccades to the artist upon hearing the transitive verb *punished*, indicating that the lexical requirement for the verb to occur with a direct object allowed the parser to predict that constituent, which was then reflected in anticipatory looks to the artist. Note that although listeners anticipated mention of a specific word, namely *artist* in the example, this was not due to the use of semantic constraints from the sentence, as the two sentence versions were identical up to the verb. The anticipation of the artist was simply because only three objects were displayed, one of which had already been mentioned as the sentential subject, leaving only one punishable entity to which to make a saccade. The second experiment showed that a similar anticipatory effect could be obtained based on whether a verb tended to occur in past participle form or not, comparing sentences such as *The song recorded by the nun was about the flower* (*recorded* is frequently used as a past participle) and *The videotape watched by the student was found under the chair* (*watched* rarely occurs as a past participle). The investigators found that, for sentences with verbs such as *watched*, listeners did not make anticipatory looks to any of the depicted objects because they expected something that could be a direct object but were unable to locate anything sensible. But for the sentences with verbs such as *recorded*, which are frequently used as past participles, the parser anticipated the reduced relative clause structure, which supported anticipatory eye movements to the image of the nun. It appears, then, that verb-based constraints can be used to anticipate syntactic structure, specifically a direct object in the first experiment and a reduced relative clause form in the second.

A number of different studies have explored syntactic category violations and their relationship to syntactic prediction. An early study investigated whether left sentential context might permit readers to anticipate an elliptical form that would help to “rescue” what would typically be treated as an ungrammatical sequence (Lau et al., 2006). The dependent measure was ERPs, and specifically, the left anterior negativity (LAN) response on the critical word. For example, subjects were shown *Although Erica kissed Mary’s mother, she did not kiss Dana’s of the bride*. This sentence is ungrammatical, but it could have been licit had it ended at *Dana’s*, due to the availability of an elliptical reading (*Erica kissed Mary’s mother but not Dana’s*). The critical finding was a reduced LAN on the word *of* relative to another ungrammatical version that did not invite an elliptical interpretation (**Although the bridesmaid kissed Mary, she did not kiss Dana’s of the bride*). Both versions are ultimately ungrammatical, but the first one that invites an ellipsis structure was easier to process, even if the prediction of an elliptical form was ultimately contradicted by the input. The parser thus appears to engage in syntactic prediction, and in this case, the prediction of a null form. This finding is reminiscent of the parser’s tendency to predict a “gap” following a highly transitive verb in a sequence such as *Who did Mary see Bill about?*, where processing difficulty is observed on *Bill* because the parser predicted a null element–a gap–following the verb *see* to which it could link the wh-word *who* (i.e., the filled-gap effect). Of course, given that phonetically null elements such as null pronouns, gaps, and elided material are legitimate syntactic constituents, it is not surprising that they would be syntactic forms that the parser might predict.

The expectation of a particular syntactic category has been shown to be strong enough to allow the parser to estimate the form properties of an upcoming word (Dikker et al., 2010). The investigators used magnetoencephalography (MEG) to investigate early processing in the occipital cortex as a function of disconfirmed syntactic category predictions. Subjects in the MEG were presented with sentences that were either grammatical or ungrammatical, implemented by having a noun either follow an adjective or an adverb (*The beautiful princess was painted* vs. *The beautifully princess was painted*). The investigators tested whether the violation would trigger activity in the visual cortex between 100 and 130 ms after onset of the noun, even though previous work had established that the M100 is sensitive only to physical features of the stimulus. They observed that not only was there an M100 response on the noun, but that it was sensitive to the form typicality of the word as a possible noun (based on previous work showing that nouns have somewhat distinct form characteristics). This finding supports the idea that the parser predicted different syntactic categories following an adjective versus an adverb and that the magnitude of the response to the expectation was linked to quality of fit between the presented word and the predicted syntactic category.

A follow-up study was designed to investigate the same core questions but without resorting to a grammatical violation paradigm (Matar et al., 2019). To implement this, the investigators took advantage of some features of standard Arabic, including the fact that verbs and nouns have distinct orthographic patterns. In one version of the sentences, subjects read a subjective-adjective sequence that could be followed by either a noun or a verb, while in the other version, the only grammatical continuation was a following noun. In addition, because in written Arabic nouns and verbs differ from each other orthographically, it was possible to evaluate whether the same M100 component that is known to be sensitive to the form properties of the input is enhanced when there is more uncertainty about the upcoming syntactic category. The results were consistent with the previous study and suggest that the language system makes predictions about the syntactic category of the next word in the absence of relevant semantic information. Moreover, given that the effect was observed in the M100 component, which reflects visual information processing, these results indicate that the prediction involved anticipation of the next word’s form properties. The key innovations of this study, then, are two-fold. The first innovation is the link to the idea of entropy: that is, as syntactic entropy increases (i.e., as the number of licit syntactic continuations goes up), activity in the left visual cortex and the right frontal cortex also increases. The second important innovation is the demonstration that this syntactic prediction effect can be observed in perfectly grammatical sentences, which suggests that predicting the syntactic category of upcoming words is a routine part of language comprehension.

Some evidence concerning the locus of syntactic prediction in the brain comes from an fMRI study that compared linguistic stimuli that were either sentences, two-word determiner-noun sequences, or random word lists, and that were made up of either real content words or jabberwocky (Matchin et al., 2017). An example of one of their sentences with real content is *The poet will recite a verse* and an example of a jabberwocky sentence is *The tevill will sawl a pand.* The logic of this design was that only the sentence condition would tap into syntactic prediction mechanisms because neither the determiner-noun nor the word list items included content that could support syntactic predictions or reflect any processing costs associated with maintaining predictions across intervening material. The investigators observed increased activity in the inferior frontal gyrus and posterior superior temporal sulcus for both the normal and jabberwocky versions of the sentences relative to the word list baselines, but no advantage for two-word sequences over word lists was observed. The fact that the jabberwocky versions of the sentences patterned with the natural sentences suggests the effects are attributable to syntactic prediction separated from semantics, since the jabberwocky versions were essentially meaningless. From these results, the investigators reasoned that syntactic prediction takes place in IFG and in pSTS. At the same time, they acknowledge that the design of their study did not allow them to verify directly that predictions were in fact made in the sentence conditions; instead, they reasoned that syntactic prediction was likely taking place based on the increased activity for jabberwocky sentences and the lack of activity for the two-word licit sequences.

Recall that the Matar et al. (2019) investigation examining the processing of grammatical sentences in Arabic supported the conclusion that syntactic prediction is a routine part of sentence processing, and the Matchin et al. (2017) findings just summarized comparing sentences, determiner-noun sequences, and word lists implicate core brain regions of the language network in syntactic prediction. It might be tempting based on these findings to conclude that syntactic prediction is universal and that the extent of syntactic prediction varies little across individuals, but this inference is inconsistent with the results of a recent study that examined the relationship between literacy and syntactic prediction (Favier et al., 2020). Dutch participants listened to sentences such as *The window is indeed broken by the bull* as they viewed a display consisting of an image of a window, a bull, a newspaper, and a hairbrush. The VWP was used to assess the extent to which listeners anticipated mention of the bull (the only animate entity in the display) by saccading to it prior to its occurrence in the input. This pattern would be evidence for syntactic prediction because the listeners had to determine online that the sentence was in the passive form and that the best object to play the role of the agent would be the one that is animate. The investigators also divided the subjects into a high-literacy and a low-literacy group based on their performance on assessments of experience with written materials (receptive vocabulary, author recognition, reading habits, spelling ability, and word/pseudoword reading). The investigators found that the higher literacy group anticipated mention of the appropriate object sooner than did the lower literacy group, and there were no differences based on individual differences in processing speed, non-verbal IQ, and verbal working memory. The authors argue from these results that the ability to generate syntactic predictions quickly during spoken language processing is affected by reading experience, a result that is likely due in part to the much richer lexical and syntactic content that tends to be conveyed in written texts versus in spoken language (Montag and MacDonald, 2015; Seidenberg and MacDonald, 2018). This study, then, suggests that although basic syntactic prediction may regularly take place, the extent of it seems to differ across individuals depending on the amount of language experience they have accumulated. This finding aligns with previous research showing that the extent of semantic prediction also varies depending on individual difference characteristics such as processing speed and working memory capacity (Huettig and Janse, 2016; Huettig and Mani, 2016).

Another class of studies has used syntactic constraints as a tool to assess whether a semantic prediction has been made. One of the earliest studies using this approach took advantage of Spanish's grammatical gender system according to which a determiner and noun must agree (Wicha et al., 2004). Using ERPs, the investigators demonstrated enhanced positivity between 500 and 700 ms after the onset of a determiner if the gender of that determiner did not agree with that of the expected upcoming noun. A study conducted in Dutch made use of similar grammatical gender constraints between prenominal adjectives and the nouns they modify and yielded similar results (Van Berkum et al., 2005). However, these findings have recently been challenged: In replication studies, investigators have observed an enhanced negativity on an unexpected noun, likely reflecting the difficulty of integrating that noun, but not the prediction effect associated with the gender-mismatched determiner (Ito et al., 2017; Nieuwland et al., 2020). Another team has made use of the more elaborate gender system in Polish and the requirement for agreement throughout the entire NP to establish that the language processing system anticipates not only specific words based on semantic constraints but also broad semantic classes (Szewczyk and Schriefers, 2013; see also Szewczyk and Wodniecka, 2020). This work done with Polish stimuli is particularly interesting because it is used to motivate a model of prediction which assumes that morphosyntactic constraints can be used not just to anticipate upcoming syntactic categories but also to provide information about the type of semantic prediction that the system should make. In other words, according to this approach, the agreement requirement is not only a tool for researchers to take advantage of to test whether a specific lexical prediction has been made, but also a source of information to the language processor to allow expectations to be incrementally updated and to permit the generation of more accurate semantic predictions.

Finally, a recent investigation of the effects of prosodic focus on processing of German sentences showed that a prominent pitch accent on a word supports the generation of both syntactic and semantic predictions, with syntactic predictions having a stronger effect on comprehension (van der Burght et al., 2021). If listeners hear *Yesterday the POLICEMAN arrested the thief, not…*, after the negation they can expect two things to happen next: First, an NP should appear, and second, that NP should stand in semantic contrast to the policeman (e.g., the inspector). Using a paradigm in which comprehenders indicated preferred continuations or explicitly provided completions, the investigators showed not only that both syntactic and semantic predictions were generated based on contrastive focus, but also that only a contradiction of the syntactic prediction undermined overall comprehension of the sentence. These results, then, establish within the same experiment that not only are both types of predictions generated, but they are dissociable and the syntactic predictions seem more critical for successful understanding of the sentence's meaning.

5. Conclusions and future directions

This review has highlighted syntactic prediction, a topic that has received far less attention than has semantic prediction. Our summary of some major models of syntactic processing reveals that most assume the existence of mechanisms that allow the parser to generate syntactic nodes in advance of the input that those nodes will syntactically incorporate. Indeed, there would be no garden-path sentences if the parser simply adopted a "wait-and-see" attitude to syntactic ambiguity rather than trying to anticipate a structure that could turn out to be contradicted later by input that cannot be grammatically integrated. More current models based on information-theoretic concepts explicitly center around the idea of prediction, with word-by-word assessment of the likelihood of various upcoming syntactic categories. We also highlighted a few studies that provide evidence for the prediction of specific syntactic categories during online processing, and taking advantage of a broad range of methodologies: behavioral measures, the VWP, standard reading paradigms, ERPs, MEG, and fMRI. A closely related body of work shows that syntactic predictions support predictions of semantic content based on the application of syntactic constraints to narrow down semantic classes. At the same time, although the evidence for syntactic prediction is compelling, it is also essential to take into account the potential role of individual differences, as almost any time individual differences are assessed in these investigations, it seems that differences are observed in the extent to which both syntactic and semantic prediction tend to occur.

We began this review with a discussion of the implications of prediction for language processing architectures. We pointed out that prediction is often viewed as the application of top-down constraints to facilitate the building of interpretations and as evidence for interactive models of language and cognition. We suggested that this position is based on flawed logic, as many forms of prediction amount to information within a given system affecting later processing in that system, and this is especially true for syntactic prediction. One confusion that needs to be addressed is that the term "top-down" is used in two distinct ways in this literature. In traditional usage, "top-down" refers to the ability of higher-level sources of information such as real-world plausibility to influence lower-level systems such as lexical access or syntactic processing. But another usage of top-down is specific to syntactic parsing and makes reference to the geometry of a phrase structure tree, in which nodes are arranged hierarchically and "top-down" refers to the availability of nodes high up in the tree that allow nodes below them to be inferred and built. For example, an NP node licenses the construction of an N node below it, and this is a "top-down" prediction in this geometric sense. Obviously, though, there is nothing interactive about this type of process, as all the activity is taking place within the syntactic processor and the knowledge that is consulted and the forms that are generated based on that knowledge are also entirely syntactic.

With this point in mind, it is clear that evidence for syntactic prediction does not constitute evidence for interactive language processing architectures. In addition, we also suggest that much of the evidence even for semantic prediction also can be accommodated by a strictly feed-forward architecture because the semantic constraints that are involved are “intra-lexical”–that is, they capitalize on word-to-word associations that are likely stored within the lexical processing system. There is a temptation to interpret any sort of evidence for the sophisticated use of knowledge during language processing as evidence for highly interactive processing systems, but it is important to remember that the arguments concerning modularity were never about *whether* information was used; they were about *when* and *how* sources of information were consulted and integrated during online processing. Even an effect that appears to show an influence of semantic knowledge on syntactic predictions cannot be taken as evidence for interaction if the language system has time to perform a syntactic analysis, evaluate it, and revise it in light of semantic information.

Throughout this article we have pointed out that nearly all theories of human parsing assume some sort of syntactic prediction mechanism as a routine step in building a grammatical analysis. The critical role of syntactic prediction in normal parsing suggests that syntactic prediction is at least as robust as semantic prediction and perhaps even stronger (van der Burght et al., 2021), with each word being integrated into a syntactic tree following the generation of that word’s predicted analysis. For example, the syntactic features linked to the word *cats* in *those hungry cats* would first be generated by prediction based on the features of *those* and then confirmed once the word is encountered in the input. However, there is not yet a great deal of work attempting to contrast the robustness of syntactic versus semantic prediction directly (for one example, see Kamide et al., 2003b as well as the aforementioned van der Burght et al. study), and of course this issue cannot be resolved based on a single comparison between the two types of constraints. What is required is to create an inventory of the types of potential syntactic and semantic prediction effects and then systematically evaluate their relative strengths in both offline and online contexts. This would be a productive line of research for the future.

We will end by presenting four additional issues we believe are worth exploring in future work. Of course, this list is far from exhaustive, and we hope that interested readers will be inspired by this review to come up with additional research directions. First, syntactic predictions may involve the use and integration of different kinds of grammatical knowledge: lexically based syntactic information, phrase structure constraints, constraints on referential dependencies, and so on. How do these different sources of grammatical information interact to yield syntactic predictions? Addressing this question would help illuminate further the extent to which prediction in language processing is based on top-down mechanisms in contrast to the sort of completion process Huettig et al. (2021) have articulated. Second, what is the relationship between linguistic knowledge and prediction? If it is true that those with lower levels of literacy predict less than those who have acquired more reading experience, and if Matchin et al. (2017) are also correct that syntactic prediction is possible but not necessary for successful comprehension, then what exactly are the critical sources of information to support syntactic prediction, and what kinds of benefits does syntactic prediction confer given that it appears to be somewhat optional or at least not universal? Third, are the syntactic predictions made by bilingual comprehenders encapsulated or do the languages interact? For example, do Spanish-English bilinguals expect to encounter an adjective following a determiner only when they communicate in English, or do they have this expectation for Spanish as well, at least to some extent, even though in Spanish adjectives almost always follow the head noun? This research topic would offer a novel avenue for investigating the long-standing issue concerning the extent to which bilinguals’ languages are simultaneously activated (Marian and Spivey, 2003; Spivey and Marian, 1999). And finally, over what distance can syntactic predictions be maintained? The semantic predictions that have been investigated tend to be relatively local in the sense that the semantically predicted word usually occurs right after the biasing context, but syntactic predictions often must be maintained over many intervening items, as in the example of *either-or* and in the case of long-distance dependencies involving fronted wh-phrases. Does the syntactic prediction become stronger as it is held in memory or does it weaken? This research question offers a productive way of examining further the interaction between language and memory systems.

As we can see, although the phenomenon of syntactic prediction is robust, we know little about the details of how this ability emerges, how it interacts with other sources of knowledge, and how it operates in multilinguals. Our hope is that this review will inspire other researchers to view syntactic prediction as a topic as worthy of study as semantic prediction, particularly since, as we noted at the start of this review, syntactic prediction is fundamental to any type of prediction during language processing.

References

Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition 73 (3), 247–264.
Arai, M., Keller, F., 2013. The use of verb-specific information for prediction in sentence processing. Language and Cognitive Processes 28 (4), 525–560.
Berger, A., Della Pietra, S.A., Della Pietra, V.J., 1996. A maximum entropy approach to natural language processing. Computational linguistics 22 (1), 39–71.
Brothers, T., Swaab, T.Y., Traxler, M.J., 2017. Goals and strategies influence lexical prediction during sentence comprehension. J. Mem. Lang. 93, 203–216.
Castelhano, M.S., Henderson, J.M., 2008. The influence of color on the perception of scene gist. J. Exp. Psychol. Hum. Percept. Perform. 34 (3), 660–675.
Choi, W., Lowder, M.W., Ferreira, F., Swaab, T.Y., Henderson, J.M., 2017. Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. Psychol. Aging 32 (3), 232–242. https://doi.org/10.1037/pag0000160.
Chomsky, N., 1970. Remarks on nominalization. In: Jacobs, R., Rosenbaum, P. (Eds.), Readings in English Transformational Grammar. Ginn, Waltham, MA, pp. 184–221.
Chomsky, N., 1981. Lectures on Government and Binding. Foris.
Christianson, K., Hollingworth, A., Halliwell, J.F., Ferreira, F., 2001. Thematic roles assigned along the garden path linger. Cogn. Psychol. 42 (4), 368–407.
Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. The Behavioral and Brain Sciences 36 (3), 181–204.
Crocker, M.W., 1999. Mechanisms for sentence processing. Language Processing. https://books.google.com/books?hl=en&lr=&id=E-l5CwAAQBAJ&oi=fnd&pg=PA191&dq=Mechanisms+Sentence+Processing+Crocker&ots=C219Fn0oBT&sig=EMtDo1bQ_1Bz3ay7RIG2NOwWXVk.
Demberg, V., Keller, F., Koller, A., 2013. Incremental, Predictive Parsing with Psycholinguistically Motivated Tree-Adjoining Grammar. Computational Linguistics 39 (4), 1025–1066.
Dikker, S., Rabagliati, H., Farmer, T.A., Pylkkänen, L., 2010. Early occipital sensitivity to syntactic category is based on form typicality. Psychol. Sci. 21 (5), 629–634.
Elman, J.L., 1990. Finding structure in time. Cognitive Science 14 (2), 179–211.
Farmer, T., Fine, A., Jaeger, T., 2011. Implicit context-specific learning leads to rapid shifts in syntactic expectations. 33rd Annu. Conf. Cogn. Sci. Soc 8, 2055–2060.
Favier, S., Meyer, A.S., Huettig, F., 2020. Literacy can enhance syntactic prediction in spoken language processing. J. Exp. Psychol. Gen.
Federmeier, K.D., Kutas, M., 1999. A rose by any other name: Long-term memory structure and sentence processing. J. Mem. Lang. 41 (4), 469–495.
Federmeier, K.D., McLennan, D.B., Ochoa, E., Kutas, M., 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. Psychophysiology 39 (2), 133–146.
Ferreira, F., 2003. The misinterpretation of noncanonical sentences. Cogn. Psychol. 47 (2), 164–203.
Ferreira, F., Bailey, K.G.D., Ferraro, V., 2002. Good Enough Representations in Language Comprehension. Current Directions in 11 (1), 11–15.
Ferreira, F., Chantavarin, S., 2018. Integration and Prediction in Language Processing: A Synthesis of Old and New. Current Directions in Psychological Science 27 (6), 443–448.
Ferreira, F., Henderson, J.M., 1990. Use of verb information in syntactic parsing: evidence from eye movements and word-by-word self-paced reading. J. Exp. Psychol. Learn. Mem. Cogn. 16 (4), 555–568.
Frazier, L., Fodor, J.D., 1978. The sausage machine: A new two-stage parsing model. Cognition 6 (4), 291–325.
Frazier, L., Rayner, K., 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cogn. Psychol. 14 (2), 178–210.

Futrell, R., Gibson, E., Levy, R.P., 2020. Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. Cognitive Science 44 (3). https://doi.org/10.1111/cogs.v44.310.1111/cogs.12814.
Gambi, C., Pickering, M.J., 2016. Predicting and imagining language. Language, Cognition and Neuroscience 31 (1), 60–72.
Garnsey, S.M., Pearlmutter, N.J., Myers, E., Lotocky, M., 1997. The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. Journal of Memory and Language 37 (1), 58–93.
Gibson, Edward, 1998. Linguistic complexity: locality of syntactic dependencies. Cognition 68 (1), 1–76.
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 (20), 8051–8056.
Gibson, Edward, Futrell, Richard, Piantadosi, Steven T., Dautriche, Isabelle, Mahowald, Kyle, Bergen, Leon, Levy, Roger, 2019. How Efficiency Shapes Human Language. Trends in Cognitive Sciences 23 (12), 1087. https://doi.org/10.1016/j.tics.2019.09.005.
Goodman, Noah D., Frank, Michael C., 2016. Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences 20 (11), 818–829.
Hale, John, 2016. Information-theoretical complexity metrics. Language and Linguistics Compass 10 (9), 397–412.
Harrington Stack, Caoimhe M., James, Ariel N., Watson, Duane G., 2018. A failure to replicate rapid syntactic adaptation in comprehension. Memory & Cognition 46 (6), 864–877.
HAWKINS, JOHN A., 2001. Why are categories adjacent? Journal of Linguistics 37 (1), 1–34.
Henderson, John M., 2017. Gaze Control as Prediction. Trends in Cognitive Sciences 21 (1), 15–23.
Henderson, John M., Choi, Wonil, Lowder, Matthew W., Ferreira, Fernanda, 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. NeuroImage 132, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050.
Henderson, John M., Hollingworth, Andrew, 1998. High-level scene perception. Annu. Rev. Psychol. 50 (1), 243–271.
Huettig, F., Audring, J., Jackendoff, R., 2021. Prediction as pre-activation: A linguistically and psychologically plausible theory of language processing. Unpublished manuscript.
Huettig, Falk, Janse, Esther, 2016. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. Language, Cognition and Neuroscience 31 (1), 80–93.
Huettig, Falk, Mani, Nivedita, 2016. Is prediction necessary to understand language? Probably not. Language, Cognition and Neuroscience 31 (1), 19–31.
Ito, Aine, Martin, Andrea E., Nieuwland, Mante S., 2017. How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. Language, Cognition and Neuroscience 32 (8), 954–965.
Jackendoff, R., 1977. X syntax: A study of phrase structure. Linguistic Inquiry Monographs Cambridge. Mass 2, 1–249.
Jackendoff, R., Audring, J., 2020. Relational Morphology: A Cousin of Construction Grammar. Front. Psychol. 11, 2241.
Kamide, Yuki, Altmann, Gerry T.M, Haywood, Sarah L, 2003a. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. J. Mem. Lang. 49 (1), 133–156.
Kamide, Y., Scheepers, C., Altmann, G.T., 2003b. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. J. Psycholinguist. Res. 32 (1), 37–55.
Kuperberg, Gina R., Jaeger, T. Florian, 2016. What do we mean by prediction in language comprehension ? Language, Cognition and Neuroscience 31 (1), 32–59.
Lau, Ellen, Stroud, Clare, Plesch, Silke, Phillips, Colin, 2006. The role of structural prediction in rapid syntactic analysis. Brain Lang. 98 (1), 74–88.
Levy, Roger, 2008. Expectation-based syntactic comprehension. Cognition 106 (3), 1126–1177.
Linzen, T., Jaeger, T.F., 2016. Uncertainty and expectation in sentence processing: Evidence from subcatergorization distributions. Cognitive Science 40, 1382–1411.
Liu, Z., 2019. A Comparative Corpus Analysis of PP Ordering in English and Chinese. Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 33–45.
Lowder, M.W., Choi, W., Ferreira, F., Henderson, J.M., 2017. Lexical Predictability During Natural Reading: Effects of Surprisal and Entropy Reduction. Cognitive Science. https://doi.org/10.1111/cogs.12597.
Luke, Steven G., Christianson, Kiel, 2016. Limits on lexical prediction during reading. Cogn. Psychol. 88, 22–60.
MacDonald, M.C., 2013. How language production shapes language form and comprehension. Front. Psychol. 4 (April), 226.
MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S., 1994. The lexical nature of syntactic ambiguity resolution. Psychol. Rev. 101 (4), 676–703.
Mani, N., Huettig, F., 2012. Prediction during language processing is a piece of cake–but only for skilled producers. J. Exp. Psychol. Hum. Percept. Perform. 38 (4), 843–847.
Marian, Viorica, Spivey, Michael, 2003. Competing activation in bilingual language processing: Within-and between-language competition. Bilingualism: Language and cognition 6 (2), 97–115.
Matar, Suhail, Pylkkänen, Liina, Marantz, Alec, 2019. Left occipital and right frontal involvement in syntactic category prediction: MEG evidence from Standard Arabic. Neuropsychologia 135, 107230. https://doi.org/10.1016/j.neuropsychologia.2019.107230.
Matchin, William, Hammerly, Christopher, Lau, Ellen, 2017. The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior 88, 106–123.
Mitchell, D.C, Holmes, V.M, 1985. The role of specific information about the verb in parsing sentences with local structural ambiguity. J. Mem. Lang. 24 (5), 542–559.
Montag, J.L., MacDonald, M.C., 2015. Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. J. Exp. Psychol. Gen. 144 (2), 447–468.
Nave, Kathryn, Deane, George, Miller, Mark, Clark, Andy, 2020. Wilding the predictive brain. Wiley Interdiscip. Rev. Cognit. Sci. 11 (6) https://doi.org/10.1002/wcs.v11.610.1002/wcs.1542.
Nieuwland, Mante S., 2019. Do "early" brain responses reveal word form prediction during language comprehension? A critical review. Neurosci. Biobehav. Rev. 96, 367–400.
Nieuwland, M.S., Arkhipova, Y., Rodríguez-Gómez, P., 2020. Anticipating words during spoken discourse comprehension: A large-scale, pre-registered replication study using brain potentials. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior 133, 1–36.
Qingrong, Chen, Yan, Huang, 2012. Processing coordinate structures in Chinese: Evidence from eye movements. PLoS ONE 7 (4), e35517. https://doi.org/10.1371/journal.pone.003551710.1371/journal.pone.0035517.t00110.1371/journal.pone.0035517.s00110.1371/journal.pone.0035517.s002.
Rayner, Keith, Carlson, Marcia, Frazier, Lyn, 1983. The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. Journal of Verbal Learning and Verbal Behavior 22 (3), 358–374.
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. Psychol. Aging 21 (3), 448–465.
Roark, Brian, 2001. Probabilistic top-down parsing and language modeling. Computational Linguistics (Association for Computational Linguistics) 27 (2), 249–276.
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: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 324–333.
Seidenberg, M.S., MacDonald, M.C., 2018. The Impact of Language Experience on Language and Reading. Topics in Language Disorders 38 (1), 66–83.
Slattery, Timothy J., Sturt, Patrick, Christianson, Kiel, Yoshida, Masaya, Ferreira, Fernanda, 2013. Lingering misinterpretations of garden path sentences arise from competing syntactic representations. J. Mem. Lang. 69 (2), 104–120.
Spivey, Michael J., Marian, Viorica, 1999. Cross talk between native and second languages: Partial activation of an irrelevant lexicon. Psychol. Sci. 10 (3), 281–284.
Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., Sedivy, J.C., 2002. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. Cogn. Psychol. 45 (4), 447–481.
Staub, Adrian, 2015. The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. Language and Linguistics Compass 9 (8), 311–327.
Staub, A., Clifton Jr., C., 2006. Syntactic prediction in language comprehension: evidence from either...or. J. Exp. Psychol. Learn. Mem. Cogn. 32 (2), 425–436.
Sturt, P., Lombardo, V., 2005. Processing coordinated structures: Incrementality and connectedness. Cognitive Science 29 (2), 291–305.
Szewczyk, Jakub M., Schriefers, Herbert, 2013. Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. J. Mem. Lang. 68 (4), 297–314.
Szewczyk, J.M., Wodniecka, Z., 2020. The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. J. Exp. Psychol. Learn. Mem. Cogn. 46 (9), 1714–1734.
Trueswell, J.C., Tanenhaus, M.K., Kello, C., 1993. Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. J. Exp. Psychol. Learn. Mem. Cogn. 19 (3), 528–553.
Van Berkum, Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol. Learn. Mem. Cogn. 31 (3), 443–467.
van der Burght, C.L., Friederici, A.D., Goucha, T., Hartwigsen, G., 2021. Pitch accents create dissociable syntactic and semantic expectations during sentence processing. Cognition 212, 104702.
Vasishth, Shravan, Lewis, Richard L., 2006. Argument-Head Distance and Processing Complexity: Explaining Both Locality and Antilocality Effects. Language 82 (4), 767–794.
Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. J. Cognit. Neurosci. 16 (7), 1272–1288.
Wlotko, Edward W., Federmeier, Kara D., 2012. Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. Psychophysiology 49 (6), 770–785.
Wlotko, E.W., Federmeier, K.D., Kutas, M., 2012. To Predict or Not to Predict: Age-Related Differences in the Use of Sentential Context. Psychol. Aging 27 (4), 975–988.