Review

CellPress

Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing

Matthew J. Traxler

Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA

Syntactic parsing processes establish dependencies between words in a sentence. These dependencies affect how comprehenders assign meaning to sentence constituents. Classical approaches to parsing describe it entirely as a bottom-up signal analysis. More recent approaches assign the comprehender a more active role, allowing the comprehender's individual experience, knowledge, and beliefs to influence his or her interpretation. This review describes developments in three related aspects of sentence processing research: anticipatory processing, Bayesian/noisy-channel approaches to sentence processing, and the 'good-enough' parsing hypothesis.

Syntactic parsing: then and now

Syntactic parsing comprises a set of mental processes that bridges the gap between word-level and discourse-level semantic processes. These interface processes serve to build or recover dependencies between words in a string [1–5] (see [6,7] for the role of syntax and grammar in production). Structural dependencies, conceptual information supplied by content words, and principles governing how thematic role assignments are derived from grammatical functions determine the standard, literal interpretation assigned to a sentence. Take, for instance, the content words embarrass, nurse, and doctor. These words are not sufficient, by themselves, to allow a comprehender to say who did what to whom (or how, when, and where). Syntactic cues and syntactic parsing processes supply the information needed to determine who did what to whom. While the meaning assigned to a given utterance depends on multiple factors, no theory of language understanding can be complete without a consideration of syntax (or grammar) and syntactic parsing.

Psycholinguists have long debated the degree to which syntax and syntactic parsing represent an autonomous, modular subsystem within the larger suite of language production and comprehension processes [8–11]. The prevailing view is that, while there are aspects of syntactic parsing that cannot be subsumed by other levels of processing (e.g., lexical or discourse processes), there are strong interactions between syntactic processes and other aspects of linguistic interpretation (e.g., prior linguistic context, concurrent prosody) and between syntactic processes and aspects of cognition beyond strictly linguistic systems [3,12–15]. Examples of aspects of interpretation that require syntactic computations include phrase packaging (inclusion or exclusion of words from phrases), modifier

Corresponding author: Traxler, M.J. (mjtraxler@ucdavis.edu, mjtraxler@gmail.com).
Keywords: syntax; parsing; Bayes' theorem; noisy channels; good-enough parsing.

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2014.08.001

Glossary

Event-related potentials (ERPs): when neurons fire, they generate electrical current that can be detected at the scalp. Neural activity produces systematic oscillations of electrical current. This activity occurs in response to various stimuli. Presentation of a stimulus leads to an ERP – the pattern of electrical activity at the scalp that occurs because of the stimulus.
Grain size: probabilistic accounts, including Bayesian/noisy-channel accounts, suppose that people keep track of patterns in the language. These patterns can occur at different levels of specificity. For example, the most common structure in English is noun–verb–noun (NVN) (subject–verb–object). That is a very large grain size. However, a verb like sneeze almost never takes a direct object. It is most often expressed as noun–verb (NV). So, at a fine grain size, sneeze is most likely to appear in a NV structure. At a larger grain size, the most likely structure is NVN.
Lexical co-occurrence: some words are more likely to appear together than others; police–car is more likely than police–cat.
Lexical item/lexical processing: a lexical item is roughly a single word. Lexical processing refers to the mental operations that retrieve or activate stored knowledge about words as needed during comprehension and production.
N400 response: the N400 is a characteristic of brain waves. The amplitude (size) of the N400 is related to how frequently a word appears in the language, how well the word's meaning fits with its contexts, and other factors that make identifying and integrating the word easier or more difficult.
Reduced relative clauses: relative clauses are expressions that modify preceding nouns. They are often signaled by a relativizer, a word like 'that'. In the expression 'The cat that my sister likes', 'that my sister likes' is a relative clause modifying cat. If the word 'that' were removed, the relative clause would be a reduced relative clause.
Schema: a knowledge structure in long-term memory that reflects an individual's knowledge of a certain kind of event. For example, a restaurant schema encodes the typical participants, objects, and events that occur when one goes to a restaurant.
Sentence complement: sometimes verbs appear with an entire sentence as an argument (as opposed to, for example, just a noun phrase). In the sentence 'John knows the answer is in the book', 'the answer is in the book' is a sentence complement that is governed by the verb 'knows'. (What does John know? That the answer is in the book.)
Syntactic parsing: involves the set of mental operations that detects and uses cues in sentences to determine how words relate to one another.
Syntactic structure: a mental representation that captures dependencies between words in sentences.
Thematic roles: abstract semantic classes that capture common roles that many different entities and objects play in different sentences. For example, in the sentences 'John drank milk' and 'John pokes bears', John is the initiator of the described action. John is a thematic agent. In the two sentences, milk and bears are on the receiving end of the action, so they are thematic patients.
Theta domains: according to Frazier and Clifton [16] some words in sentences assign thematic roles to other words in the sentence. A theta domain is that part of a sentence for which a given word assigns thematic roles.

attachment decisions (in cases where a modifying expression may belong to only one among a set of previous words), and the definition of theta domains and assignment of thematic roles (see Glossary), among others [16–19].

There is general agreement that syntax plays an important role in meaning derivation, but there has been a shift away from strictly bottom-up, serial, encapsulated views of language interpretation and toward more interactive accounts. Three sets of related developments are changing the way that psycholinguists view language interpretation in general and the nature of syntactic parsing processes in particular. These include the relatively recent emphasis on predictive or anticipatory processes, the application of Bayesian probability estimation to language comprehension, and the changing view of language comprehension through the lens of satisficing or good-enough processing [20–25].

The role of anticipation

Language interpretation occurs in a rapid and incremental fashion [26,27]. Comprehenders can identify a word's semantic and syntactic characteristics and the word's relationship to prior context within a few hundred milliseconds of encountering it. Any account that assumes that processing occurs in a strictly bottom-up fashion (signal analysis, followed by word recognition and lexical access, followed by syntactic parsing, followed by integration of new information with prior syntactic and semantic context) is strongly constrained by the speed at which comprehenders can access detailed information about newly encountered words. One way to account for the incredibly rapid and incremental nature of interpretation is to propose that comprehenders anticipate upcoming input rather than waiting passively for the signal to unfold and then reacting to it. Results from various experimental paradigms indicate that comprehenders discriminate between more likely and less likely continuations. In reading, more predictable words are skipped more often than less predictable words [28–30]. Visual world experiments also indicate that comprehenders actively anticipate or predict the imminent arrival of not-yet-encountered information [20]. In these visual world experiments, participants view an array containing pictures of various objects (e.g., a cake, a girl, a tricycle, and a mouse). While viewing the array, participants listen to sentences. If the sentence begins 'The little girl will ride...', participants make eye movements toward the picture of the tricycle even before the offset of the verb 'ride'. This is not simply a reflex based on association between ride and tricycle, however. If the visual array includes a little girl, a man, a tricycle, and a motorcycle, participants make anticipatory eye movements toward the tricycle when listening to 'The little girl will ride...'. However, they make anticipatory movements toward the motorcycle in the same visual array if the subject noun is 'man' (as in 'The man will ride...'). Event-related potential experiments show that prediction-supporting contexts produce smaller N400 responses than less supportive contexts, even when intralexical association is held constant [31].

Results like this indicate that, at least in processing environments where a small number of referents are made visually salient, participants are capable of identifying how a sentence is likely to continue. These predictions may relate to what concept is likely to be mentioned next, but they may be even more specific than that. For example, DeLong and colleagues' event-related potential study [32] produced evidence that comprehenders anticipated the phonological form of an upcoming pair of words (a determiner and a noun). Further, participants appear to act on these predictions before they receive definitive bottom-up evidence confirming or disproving the prediction (i.e., when context makes one continuation more likely than others, the eyes will fixate a picture representing the likely continuation before the comprehender hears a word that refers to that object).

Although researchers agree that anticipation and prediction occur during sentence interpretation, the precise means by which comprehenders derive predictions is currently not well understood. Hence, we need accounts that can tell us how predictions are made, which in turn will tell us why some predictions are made but not others. Hypotheses about how predictions are made include using the production system to emulate the speaker [33,76], relying on intralexical spreading activation [34], or using schematic knowledge of events [35]. A simple word–word association hypothesis is made less plausible by experiments showing that syntactic factors affect the response to a word when lexical association to preceding context is held constant [36]. In addition, words that do not fit a syntactically governed thematic role do not enjoy a processing advantage simply because they are associated with other content words in a sentence [37]. For example, the word axe is strongly associated with the noun lumberjack. Despite this strong association, axe is not processed faster than normal in the sentence frame 'The lumberjack chopped the axe'. Interestingly, the neurophysiological response to a word that is associated with other content words in a sentence changes based on preceding discourse context [31,38–41]. When discourse context activates an event schema that incorporates a particular concept, a word relating to that concept will evoke a smaller N400 response, even when that word is not a good fit given the immediate syntactic context. Hence, event knowledge representations rather than simple lexical co-occurrence appear to provide comprehenders with the basis for deriving predictions.

Bayesian estimation and noisy channels

Comprehenders can anticipate the imminent arrival of specific lexical items relating to activated event representations. Research on syntactic processes suggests that comprehenders may also be able to anticipate structural properties of sentences before bottom-up cues provide definitive evidence for or against a given structural hypothesis. The idea that syntactic parsing processes can be affected by the probability or likelihood of particular syntactic structures has, in fact, been around for a long time [2,3,42]. Trueswell and colleagues were among the first to provide evidence that the conditional probability of a structural analysis in a given context influenced the processing load imposed on the comprehender [10,43]. The precise timing and nature of probabilistic influences are treated differently under different accounts of parsing [19,44–46].