Parsing: Psycholinguistic Aspects

Colin Phillips, University of Maryland

Comprehension of sentences involves a number of different processes: decoding of sounds (or written characters), word recognition, combining of words into well-formed structures, and interpretation of those structures. *Parsing* refers to the mental processes that determine the grammatical structure of sequences of words, and is just one of the processes involved in sentence comprehension. Parsing is normally closed to introspection, and speakers share the intuition that parsing is both immediate and effortless, proceeding incrementally as each new word is heard or seen.

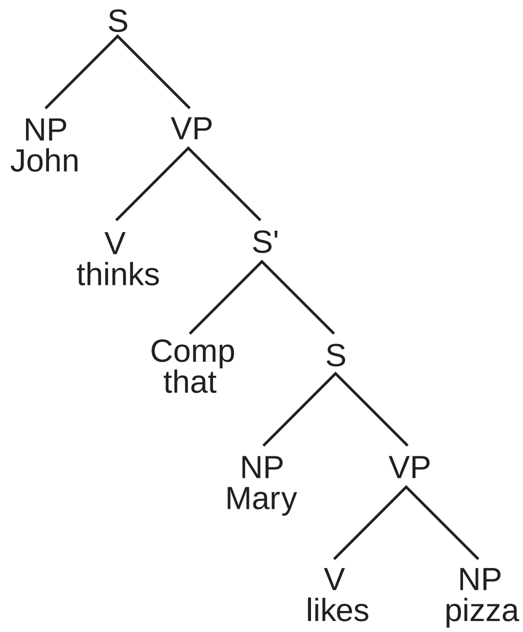
The intuitive ease of parsing belies a set of sophisticated automatic mechanisms that ensure rapid and accurate structure-building, despite the fact that the spoken or written language input contains very few direct cues about sentence structure. The relatively slow progress over the last 40 years in developing computer models of incremental parsing attests to the sophistication of human parsing. If parsing were simple, successful computer models would already have been developed.

Most information about the speed, accuracy and automaticity of human parsing comes from experimental studies. *Speed* of parsing has been demonstrated using eye-movements in reading, event-related brain potential (ERP) responses to the detection of syntactic deformations, and the ‘shadowing’ of speech at delays of only 250-300 msec. These studies show that a number of parsing operations occur within 200-300ms after the presentation of a word. *Accuracy* of parsing has been shown by studies of speakers’ ability to draw on their grammatical knowledge to constrain parsing choices. For example, their knowledge of island constraints on *wh*-movement serves to filter out ungrammatical parses in much the same way that speakers ignore ungrammatical antecedents for pronouns and reflexives, and use subcategorization to constrain choices of complements of verbs.

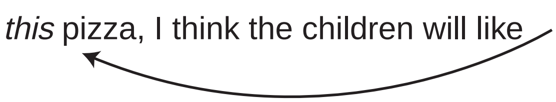
But although some parsing operations occur extremely rapidly, it is unclear how much automatic parsing occurs as quickly as 200-300ms after people encounter a word. The strongest evidence for rapid parsing comes from tasks with spoken language with relatively simple and predictable syntax.. Whether such findings generalize to all of the details of grammatical parsing remains to be seen.

The speed, accuracy and automaticity of successful parsing suggests that speakers have a sophisticated mechanism that allows for rapid and precise grammatical analysis of sequences of words. Is this mechanism part of the mental grammar itself? Most grammatical models provide a recursive characterization of the set of possible sentence structures of a language, but this does not guarantee an effective procedure for incrementally mapping word sequences onto sentence structures. In fact, some researchers have concluded that grammars are *not* conducive to successful parsing, based on the right-branching character of much syntactic structure (1), which hinders the immediate recognition of phrases; based on the existence of movement (2) and deletion processes (3), which remove words from their canonical positions; and based on the head-final character of many languages, such as Japanese (4), in which nouns and verbs occur at the end of noun phrases (NPs) and verb phrases (VPs), respectively.

(1)



(2)



(3) John will drive to London, and Sue ~~will drive~~ to Edinburgh.

(4) Masa-ga $[_{VP}$ Kumiko-ni $[_{NP}$ ringo-o tabeta inu-o] ageta]
M.-NOM K.-DAT apple-ACC ate dog-ACC gave
'Masa $[_{VP}$ gave Kumiko $[_{NP}$ the dog that ate the apple]]'

Attempts to reconcile how successful people are at parsing with the nature of the grammar have taken several forms. One approach compensates for the unhelpful nature of the grammar by introducing a set of 'heuristics and strategies' independent of the grammar, such as a rule that analyzes *noun* + *verb* sequences as subject-verb structures. A second approach revises the model of grammatical knowledge in such a way that it is more conducive to incremental parsing, and so requires relatively few specialized parsing mechanisms. The most common approach to the problem assumes that there is a specialized parsing mechanism that implements a detailed 'covering grammar', i.e., an implementation of the grammar which is specifically adapted for parsing.

Recovering the grammatical structure of a sentence is clearly just one step along the way to understanding a sentence. The structural representations built from parsing may be no more than transient representations, which are disposed of as soon as the semantic interpretation of a structure is complete. Evidence for this view comes from classic studies demonstrating that people preserve in memory the semantic content of sentences better than the syntactic form. But one source of difficulty in parsing are the incomplete structural dependencies that have to be held in memory, thereby delaying immediate interpretation. Examples of such structures are *wh*-dependencies (5), relative clauses (6), and center-embedded structures (7).

(5) Which candidate did the committee think that the voters would like ?

Wh-dependency

(6) The parents finally found food that their daughter was willing to eat

relativization

(7) The articles that the student wanted to read were missing from the library

subject-verb agreement

Rapid and accurate parsing of *unambiguous* word sequences (8) depends on efficient access to grammatical knowledge. In the parsing of *ambiguous* word sequences (9), on the other hand, successful grammatical analysis is not enough, because it may provide two or more alternative structures for a single word sequence, with no immediate way to choose between the competing analyses.

(8) John thinks Mary ... { John thinks [$_S$ Mary loves pizza]

(9) John knows Mary ... { John knows [$_{NP}$ Mary] very well.
John knows [$_{S}$ Mary loves pizza] }

Structural ambiguity resolution has been a major focus of research on parsing. Structural ambiguities, and the 'garden path' effects that arise when speakers select incorrect analyses, were originally investigated to validate the heuristics and strategies proposed for the parsing of unambiguous sentences. Later, the processes involved in ambiguity resolution became a focus in their own right. Structural ambiguities generally seem to be resolved in favor of the simplest alternative, but there has been disagreement over what counts as 'simplicity'. While some investigators have argued that it should be defined in structural terms, others have emphasized the contribution to simplicity of such factors as plausibility, frequency, and referential felicity .

Parsing has often featured prominently in discussions of the 'modularity' of linguistic processes, i.e., whether they are independent of other cognitive processes. The question of whether multiple sources of information (syntactic, semantic, probabilistic) are integrated in resolving syntactic ambiguities has been widely used in arguments for or against the existence of informationally *encapsulated* mental modules. The claim that the parser is informationally encapsulated has come under attack in recent years, mainly because of evidence that 'simplicity' in ambiguity resolution involves non-structural factors. But these arguments are independent of the claim that there is a grammatically *specialized* mental module that accurately and rapidly builds the structure of unambiguous word sequences. Evidence for this remains strong.

Colin Phillips
University of Maryland

Suggestions for Further Reading

Altmann, Gerry. 1998. *The Ascent of Babel*. Oxford, UK: Oxford University Press.

Berwick, Robert, & Amy Weinberg. 1984. *The grammatical basis of linguistic performance*. Cambridge, Mass.: MIT Press.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA.: MIT Press.

Fodor, Jerry A., Thomas G. Bever, & Merrill Garrett. 1974. *The Psychology of Language*. New York: McGraw Hill.

Kempson, Ruth, Wilfried Meyer-Viol, & Dov Gabbay. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Oxford: Blackwell.

MacWhinney, Brian, & Elizabeth Bates. 1989. *The Cross-linguistic Study of Sentence Processing*. Cambridge, UK: Cambridge University Press.

Marcus, Mitchell. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA.: MIT Press.

Miller, George. & Noam Chomsky. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology, volume 2*, edited by R.D. Luce, R.R. Bush & E. Galanter. New York: Wiley.

Pritchett, Bradley. 1992. *Grammatical Competence and Parsing Performance*. Chicago, Ill.: University of Chicago Press.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA.: MIT Press.