

AMBIGUITY RESOLUTION IN THE HUMAN SYNTACTIC PARSER: AN EXPERIMENTAL STUDY

Howard S. Kurtzman
Department of Psychology
Massachusetts Institute of Technology
Cambridge, MA 02139

(This paper presents in summary form some major points of Chapter 3 of Kurtzman, 1984.)

Models of the human syntactic parsing mechanism can be classified according to the ways in which they operate upon ambiguous input. Each mode of operation carries particular requirements concerning such basic computational characteristics of the parser as its storage capacities and the scheduling of its processes, and so specifying which mode is actually embodied in human parsing is a useful approach to determining the functional organization of the human parser. In Section 1, a preliminary taxonomy of parsing models is presented, based upon a consideration of modes of handling ambiguities; and then, in Section 2, psycholinguistic evidence is presented which indicates what type of model best describes the human parser.

1. Parsing Models

Parsing models can be initially classified according to two basic binary features. One feature is whether the model immediately analyzes an ambiguity, i.e., determines structure for the ambiguous portion of the string as soon as that portion begins, or delays the analysis, i.e., determines structure only after further material of the string is received. The other feature is whether the model constructs just a single analysis of the ambiguity at one time, or instead constructs multiple analyses in parallel. The following account develops and complicates this initial classification scheme. Not every type of model described here has actually been proposed in the literature. The purpose here is to outline the space of possibilities so that a freer exploration and clearer evaluation of types can be made.

An Immediate Single Analysis (ISA) model is characterized by two properties: (1) An ambiguity is resolved as soon as it arises, i.e., on its first word (or morpheme); (2) the analysis that serves as the resolution of the ambiguity is adopted without consideration of any of the other possible analyses. Typically, such models lack the capability to store input material in a form which is not completely analyzed. Pure top-down, depth-first models such as classical ATN's (Woods, 1970) are examples of ISA models.

For certain sentences, Frazier & Fodor's (1978) Sausage Machine also behaves like an ISA model. In explaining their Local Association principle, they claim that in the first stage of parsing, structure

can be built for only a small number of words at a time. As a result, in a sentence like "Rose read the note, the memo and the letter to Mary," the PP "to Mary" is immediately attached into a complex NP with "the letter" without any consideration of the other possible attachment directly into the VP, the head of which ("read") is many words back.

A Delayed Single Analysis (DSA) model is also characterized by two properties: (1) When an ambiguity is reached, no analysis is attempted until a certain amount of further input is received; and (2) when an analysis is attempted, then the analysis that serves as the resolution of the ambiguity is adopted without consideration of any other possible analyses (if any others are still possible--i.e., if the string is still ambiguous). A bottom-up parser is an example of a DSA model. Another example is Marcus's (1980) Parsifal. These models must have some sort of storage buffer for holding unanalyzed material.

It is possible for Single Analysis models to combine Immediate and Delayed determination of structure. Ford, Bresnan, & Kaplan's (1982) version of a GSP does so in a limited way. Their Final Arguments principle permits a delay in the determination of the attachment of particular constituents into the overall structure of the sentence that has been determined at certain points. (The GSP's Chart is what stores the unattached constituents.) However, it must be noted that during the period in which that determination is delayed, other attachment possibilities of the constituent into higher-level structures (which are themselves not yet attached into the overall sentence structure) are considered. Therefore, it is not the case in their model that there is a true delay in attempting any analysis. The fundamentally Immediate nature of the GSP requires that some attachment possibility always be tested immediately.

More authentic combinations of D- and ISA could be constructed by modifying bottom-up parsers or Parsifal, which are both inherently Delaying, so that under certain conditions auxiliary procedures are called which implement Immediate Analysis. (There is, though, no real motivation at present for such modifications.) It can be noted that while bottom-up mechanisms are logically capable of only Delayed Analysis, top-down mechanisms are capable of either Immediate or Delayed Analysis.

Another type of model utilizes Delayed Parallel Analysis (DPA). In this type, parallel analysis of an ambiguity is commenced only after some delay

beyond the beginning of the ambiguous portion of the string. Such a model requires a buffer to hold input material during the delay before it is analyzed. Also, any model that allows parallelism requires that the parser's representational/storage medium be capable of supporting and distinguishing between multiple analyses of the same input material, and that the parser contain procedures that eventually oversee a decision of which analysis is to be adopted as resolution of the ambiguity. An example of a DPA parser would be a generally bottom-up parser which was adjusted so that at certain points, perhaps at the ends of sentences or clauses, more than one analysis could be constructed. Another example would be a (serious) modification of Parsifal such that when the pattern of more than one production rule is matched, all of those rules could be activated.

There are actually two sorts of parallelism. One can be called momentary parallelism, in which a choice is made among the possible analyses according to some decision procedure immediately--before the next word is received. The other sort can be called strong parallelism, in which the possible analyses can stay active and be expanded as new input is received. If further input is inconsistent with any of the analyses, then that analysis is dropped. There might also be a limitation on how long parallel analyses can be held, with some decision procedure choosing from the remaining possibilities once the limiting point is reached. (It would seem that some limitation would be required in order to account for garden-pathing.)

In addition, in strong parallelism although multiple analyses are all available, they might still be ranked in a preference order.

A further type of model is characterized by Immediate Parallel Analysis (IPA), in which all of the possible analyses of an ambiguity are built as soon as the ambiguous portion of the string begins. Frazier & Fodor's (1978) parser is partially describable as an IPA model with momentary parallelism. In explaining their Minimal Attachment principle, they propose that an attempt is made to build in parallel all the possible available structures, on the first word of an ambiguity. The particular structure that contains the fewest connecting nodes is the one that is then right away adopted.

Fodor, Bever, & Garrett (1974) proposed an IPA with strong parallelism. As soon as an ambiguity arises, the possible analyses are determined in parallel and can stay active until a clause boundary is reached, at which point a decision among them must be made.

There is another design characteristic that a parser might have which has not been considered so far. Instead of the parser, after making a single or parallel analysis of an ambiguity, maintaining the analysis/es as further input is received, one can imagine it just dropping whatever analysis it had determined. This can be called abandonment. Then analysis would be resumed at some later point, determined by some scheduling principles. Perhaps

the most natural form of a parser which utilizes abandonment would be an IPA model. The construction of more than one analysis for an ambiguity would trigger the parser to throw out the analyses and wait until a later point to attempt analysis anew. Thus, the parser is not forced to make an early decision which might turn out to be incorrect, as in momentary parallelism, nor is it forced to carry the load of multiple analyses, as in strong parallelism. At an implementation level, this abandonment might be realized as mutual inhibition by the several analyses.

Abandonment is also possible in an ISA model. Take, for instance, a generally bottom-up model in which constituents can be held free, not yet attached into the overall sentence structure. A constraint could be placed on such a model which forbade such free constituents, forcing the analyses of the constituents to be abandoned if they cannot immediately be fit into the overall sentence structure. (Such a constraint might be implemented as a limit on storage space for free constituents.) Then, at some later point, a new analysis of the constituents and their attachments would be made.

Abandonment is also possible, though less intuitively satisfying, in delayed models. In these models, there would be a delay in beginning analysis, and then another delay as a result of abandonment.

When analysis is begun again following abandonment, it can proceed according to any of the above models, though of course some would seem to be more natural than others.

2. Experiment

Previous psycholinguistic experiments have often used quite indirect methods for tapping parsing processes (e.g., Frazier & Rayner's (1982) measurements of eye-movements during reading and Chodorow's (1979) measurements of subjects' recall of time-compressed speech) and have yielded conflicting results. The present investigation set out to gather data concerning the determinants and scheduling of ambiguity resolution, through use of an on-line task that provides readily interpretable results.

Subjects sat in front of a CRT screen and on each trial were presented with a series of words comprising a sentence, one word at a time, each word in the center of the screen. Each word remained on the screen for 240 msec and was followed by a 60 msec blank screen. Presentation of the words stopped at some point, either within or at the end of the sentence, and a beep was heard. The subjects' task was to respond, by pressing one of two response keys, whether or not the sentence had been completely grammatical up to that point.

For experimental items, presentation always stopped before the end of the sentence, and the sentence was always grammatical. These experimental sentences contained ambiguities which were shown to be correctly resolved in only one way by the last word that was presented. There were

two versions of each experimental item, which differed only in the last presented word. And these last words of the versions resolved the ambiguity in different ways. An example is shown in (1) (along with possible completions of the sentences in parentheses).

- (1) The intelligent scientist examined with a magnifier [a] our (leaves.)
[b] was (crazy.)

Any individual subject was presented with only one version of an item. If subjects had chosen a particular resolution for the ambiguity before the last word was presented, it was expected that they would make more errors and/or show longer correct response times (RTs) for the version which did not match the resolution that they had chosen than for the version which did match. (Experimental items were embedded among a large number of filler items whose presentation stopped at a wide variety of points. Many of these fillers also contained ungrammaticalities, of various sorts and in various locations in the sentence.)

A wide variety of ambiguities were tested, including those investigated in previous studies. Only a few highlights of the results are presented here, in order simply to illustrate the major findings.

For items like (1b), subjects made a large number of errors--about 75%. This indicates that they were garden-pathed--just as in one's experience in normal reading of such sentences. By contrast, for items like (1a), very few errors were made. Further, the RTs for the correct responses to (1a) were significantly lower than those to (1b). For (1a), RTs fell in the 450-650 msec range, while for (1b) the RTs were 100 to 400 msec higher. Evidently, subjects had resolved the ambiguity in (1) before receiving the last word, and they chose the resolution fitting (1a), in which "examined" is a main-clause past-tense verb, rather than the resolution fitting (1b), in which it is a past participle of a reduced relative clause.

However, quite different results were obtained for items like (2), which differs from (1) only by the replacement of "scientist" by "alien".

- (2) The intelligent alien examined with a magnifier [a] our (leaves.)
[b] was (crazy.)

There was no difference between (2a) and (2b) in either error rate or RT--both measures fell into the same low range as those for (1a). That is, subjects were not garden-pathed on either sentence. They kept open both possibilities for analysis throughout presentation of the sentence.

Several conclusions can be drawn from comparing results of items like (1) and those like (2). First, it is possible to delay the resolution of an analysis. Two classes of parsing models can thus be ruled out as descriptions of the overall operations of the human system: ISA and IPA-with-momentary-parallelism. Second, the duration of this delay is variable, and therefore any model in which the point of resolution

for a particular syntactic structure is invariant is ruled out. Marcus's Parsifal is an example of such a disconfirmed model. By the way, this does not mean that there must always be some delay in resolution. In fact, for items like (1) it does appear the resolution is made immediately upon reception of "examined". This is indicated by subjects' performance for (3) and (4) matching their performance for (1) and (2), respectively.

- (3) The intelligent scientist examined
[a] our (leaves.)
[b] was (crazy.)
- (4) The intelligent alien examined
[a] our (leaves.)
[b] was (crazy.)

It seems then that the delay can vary from zero to evidently a quite substantial number of words (or constituents).

Third, the duration of the delay is apparently due to conceptual, or real-world knowledge, factors. With regard to (1) and (2), one component of our real-world knowledge is that scientists are likely to examine something with a magnifier but unlikely to be examined, but for aliens the likelihoods of examining and being examined with a magnifier are more alike. Thus, it seems that the point at which a resolution is made is the point at which one of the possible meanings of the sentence can be confidently judged to be the more plausible one. So, parsing decisions would be under significant influence of conceptual mechanisms. This fits with work in Kurtzman (1984; Chapter 2), in which a substantial amount of evidence is offered for the strong claim that parsing strategies in the form of preferences for particular structures (e.g., Frazier & Fodor, 1978; Ford et al., 1982; Crain & Steedman, in press) do not exist. It is argued rather that all cases of preference for one resolution of an ambiguity over another can be accounted for by a model in which conceptual mechanisms judge which possible resolution of the ambiguity results in the sentence expressing a meaning which better satisfies expectations for particular conceptual information or for general plausibility. Such a model requires that parallel analyses be presented to the conceptual mechanisms so that it may be judged which analysis better meets the expectations. Therefore, an acceptable parsing model must have some parallel analysis at the time a resolution is made (which is consistent with some previous psycholinguistic evidence: Lackner & Garrett, 1973). This requirement of parallelism then leaves us with the following models as candidates for describing the human parser: DPA with either kind of parallelism, IPA-with-strong-parallelism, or Abandonment-with-parallel-reanalysis. (Abandonment might work in (2) by abandoning analysis upon the attempt at analysis of "examined" and then commencing re-analysis either (a) at a point determined by some internal schedule, or (b) upon a signal from conceptual mechanisms that the conceptual content of the syntactically unanalyzed words was great enough to support a confident resolution decision.)

In contrast to the other remaining models,

IPA-with-strong-parallelism posits that input material is at all times analyzed. A look at results for other stimuli suggests that this might be the case. In a task similar to the present one, Crain & Steedman (in press) have shown that for items such as (5), comprised of more than one sentence, the first sentences (5a or 5b) can bias the perceiver towards one or the other resolution in the last sentence (5c or 5d), which contains an ambiguous "that"-clause (complement vs. relative).

(5a) RELATIVE-BIASING CONTEXT

A psychologist was counseling two married couples. One of the couples was fighting with him but the other one was nice to him.

(5b) COMPLEMENT-BIASING CONTEXT

A psychologist was counseling a married couple. One member of the pair was fighting with him but the other one was nice to him.

(5c) RELATIVE SENTENCE

The psychologist told the wife that he was having trouble with to leave her husband.

(5d) COMPLEMENT SENTENCE

The psychologist told the wife that he was having trouble with her husband.

So, for example, (5c) preceded by (5a) is processed smoothly, while (5c) preceded by (5b) results in garden-pathing at the point of disambiguation (the word "to"). In the present experiment, sentences in which the "that"-clause was disambiguated immediately following the beginning of the clause (5e or 5f) were presented following the contexts of (5a) or (5b).

(5e) RELATIVE SENTENCE

The psychologist told the wife that
was (yelling to shut up.)

(5f) COMPLEMENT SENTENCE

The psychologist told the wife that
to (yell was not constructive.)

It turned out that context had no effect on performance for this type of item. Rather, subjects performed somewhat more poorly when the "that"-clause was disambiguated as a relative (5e), showing about 20% errors and sometimes elevated RTs, as compared with the complement disambiguation in (5f), which showed low RTs and practically no errors. The effect did not differ in strength between the two contexts. These results along with those of Crain & Steedman show that initially the complement resolution is preferred but that later this preference can be overturned in favor of the relative resolution if that is what best fits the context. Now, there is no reason to believe that subjects are actually garden-pated when they end up adopting the relative resolution. Note that there is no conscious experience of garden-pathing, and that the error and RT effects here are much weaker than for classical garden-pathing items like (1). It seems more likely that both possible analyses of "that" have been determined but that one--as a complementizer--has been initially ranked higher and so is initially more accessible. In this speeded task, it would be

expected that the less accessible relative pronoun analysis of "that" would sometimes be missed--resulting in incorrect responses for (5e)--or take longer to achieve. Now, if "that" had simply not been analyzed at all by the time of the presentation of the last word, as in a DPA or Abandonment model, there would be little reason to expect that one analysis of it should cause more errors than the other.

So, we may tentatively conclude that IPA-with-strong-parallelism describes the human parser's operations for at least certain types of structures. Similar results with other sorts of structures are consistent with this claim. This does not rule out the possibility, however, that the human parser is a hybrid, utilizing delay or abandonment in some other circumstances.

Why is the complementizer analysis immediately preferred for "that"? In these items all of the main verbs of the ambiguous sentences had meanings which involved some notion of communication of a message from one party to another (e.g., "told", "taught", "reminded"). In Kurtzman (1984) it is argued that such verbs generate strong expectations for conceptual information about the nature of the message that is communicated. The complement resolution of the "that"-clause permits the clause to directly express this expected information, and so it would be preferred over the relative resolution, which generally would not result in expression of the information. It is also possible that such a conceptually-based preference gets encoded as a higher ranking for the verbs' particular lexical representations which subcategorize for the complement (cf. Ford et al., 1982).

REFERENCES

- Chodorow, M.S. Time-compressed speech and the study of lexical and syntactic processing. In W.E. Cooper & E.C.T. Walker (Eds.), Sentence processing. Hillsdale, NJ: Erlbaum, 1979.
- Crain, S. & Steedman, M. On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language processing. NY: Cambridge University Press, in press.
- Fodor, J.A., Bever, T.G., & Garrett, M.F. The psychology of language. NY: McGraw-Hill, 1974.
- Ford, M., Bresnan, J., & Kaplan, R.M. A competence-based theory of syntactic closure. In J. Bresnan (Ed.), The mental representation of grammatical relations. Cambridge, MA: MIT Press, 1982.
- Frazier, L. & Fodor, J.D. The sausage machine: A new two-stage parsing model. Cognition, 1978, 6, 291-325.
- Frazier, L. & Rayner, K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 1982, 14, 178-210.

Kurtzman, H.S. Studies in syntactic ambiguity resolution. Ph.D. Dissertation, MIT, 1984. (Available from author in autumn, 1984, at School of Social Sciences, Univ. of California, Irvine, CA 92664.)

Lackner, J.R. & Garrett, M.F. Resolving ambiguity: Effects of biasing context in the unattended ear. Cognition, 1973, 1, 359-372.

Marcus, M. A theory of syntactic recognition for natural language. Cambridge, MA: MIT Press, 1980.

Woods, W.A. Transition network grammars for natural language analysis. Communications of ACM, 1970, 13, 591-602.