Center-Embedding Revisited

Michael B. Kac
Dept. of Linguistics
University of Minnesota
Minneapolis, MN 55455

The severe comprehension difficulty associated with certain center-embedding constructions is perhaps the best known of psychosyntactic phenomena. Most attempts at explanation have been variations on a single theme--that the c.e. configuration leads to an overload of short-term memory during processing. That this is not the whole story can be seen from considering the fact, rarely noted, that c.e. constructions exist which are understood quite easily, e.g.

(1) If either the Pope is Catholic or pigs have wings then Napoleon loves Josephine.

To this observation it might be replied that it is not c.e. per se that causes difficulty, but rather MULTIPLE c.e.; thus, (1) would not be expected to pose problems since it is embedded only to a depth of 1. But the same is true of

(2) If if the Pope is Catholic then pigs have wings then Napoleon loves Josephine.

which is, at best, at the outer reaches of comprehensibility.

The facts regarding (1-2) can be accounted for by a few simple assumptions. The first is that constituent recognition is carried out in strict left-right fashion; assume further that in attempting to analyze (1-2), the parser has at some point built a structure of the form

(3) If/either S_1 then/or S_2 then S_3.

and that at this point the following procedures are invoked:

(4) a. When an if is encountered, open an S at that point; the locate the first then to the right of this if and close at the end of the S immediately following.

b. When an either is encountered, open an S at that point; then locate the first or to the right of this either and close at the end of the S immediately following.

Applied to (1), (4) will operate in straightforward fashion; it will close the S beginning with if directly after S_3, and the S beginning with either directly after S_2. Applied to (2), however, it will misparse, closing the S beginning with the first if prematurely after S_2--a garden path effect of a familiar kind. A parallel account can be given of the difference between

(5) a. That for Harry to like Maxine would annoy Fred bothers me.

b. That that Harry likes Maxine annoys Fred bothers me.

where both involve nesting of complementizer-verb dependencies, but where (5b) is considerably more difficult to process than (5a). Assume that complementizers are associated with the verbs that they mark as non-main by the following procedure:

(6) a. Link each for and to to the first infinitive to its right.

b. Link each that to the first finite verb to its right.

In (5a), that is linked to would while the for and to are both linked to like, as desired; in (5b), however, there will be a garden path since the first that will be erroneously linked to likes rather than annoys. Thus (5a-b) and (1-2) are treated analogously in that that errors of prematurity are committed in the cases that pose comprehension difficulty but not in the cases that don't.

The general idea embodied in the foregoing can be further extended. Consider Object relative constructions like

(7) a. The rat that the cat chased squeaked.

b. The rat that the cat that the dog bit chased squeaked.

Examples of this type are perhaps most familiar from discussions of c.e. That (7b) should be more difficult to process than (7a) falls directly out of general design features of a syntactic parser currently under investigation called MULTIGAP ('multiple pass group-analyzing parser') and is, moreover, attributed to a garden path effect much like the ones hypothesized in the earlier cases discussed. In MULTIGAP, although simple NP's (i.e. NP's without clausal modifiers) are identified early in parsing, recognition of complex NP's is forestalled until after an exploratory phase during which the parser builds a structural representation called a PREANALYSIS, in which the sentence being parsed is parcelled up into a sequence of units called BOUNDED GROUPS (b-groups). A key notion here is that of the TRANSITION from one predicate to the next (or from the last predicate to #), i.e. the material that intervenes between the former and the latter. If conditions are satisfied for construing the latter as subordinate to the former (e.g. if there is an overt subordinator in the transition) then the transition is of one type, labelled B'; otherwise, it is of a different type, labelled B. Cross-cutting this dichotomy is a distinction between STRONG and WEAK transitions. If the transition from a predicate of n places to the next predicate, or to #, contains at least n-1 NP's, it is strong, weak otherwise. In (7b), the transitions from bit to chased and from chased to squeaked are both weak, while that from squeaked to # is strong. Abbreviatorily, the four transition types are labelled s (strong B), w (weak B), s' (strong B'), and w' (weak B'). The parser not only types transitions according to this scheme, but also inserts into each a special boundary marker,]], to delineate b-groups. If a transition is of any type other than s,]] is positioned directly after the initial predicate in the sequence under consideration, otherwise directly after the last NP of the transition. (Recall that simple NP recognition has already taken place.) The preanalysis of (7b) is thus

(8) The rat that the cat that the dog
bit]] chased]] squeaked]] #
$\underline{w}_1$ $\underline{w}_2$ $\underline{s}$

Once the preanalysis is set, the parser looks for possible opening points of complex NP's--sequences of the form NP-SUB-V and NP-SUB-NP. (The full MULTIGAP design is capable of dealing with cases where no overt subordinator occurs, but discussion will be limited here to cases where there is such a subordinator.) The opening is labelled as belonging to type 1 (Subject relative) or type 2 (other) depending on which type of opening sequence is found. Accordingly, the parser will find two type 2 openings in (8): the rat that the cat and the cat that the dog. Closure is effected by a procedure which, if the opening is type 1, closes at the first s to the right, but at the first w to the right of a type 2 opening. Thus in (7a), the complex NP will be correctly closed after chased, but in (7b), there will be premature closure of the larger NP at w_1 rather than at w_2, an error of prematurity exactly analogous to those discussed earlier. Note, moreover, that c.e. of a type 1 relative in a type 2 construction poses no problems:

(9) The mouse that the cat who chased the rat
saw squeaked.

While (9) is not absolutely straightforward, it can be understood with a little effort, which is not the case with (7b) even though (9) is of the same depth of embedding. The difference is accounted for in this treatment by the fact that the larger NP is closed off directly after saw (i.e. after the one w) while the smaller one closes directly after rat--at the first s. Because there is only one w, no error of prematurity occurs.

Complex NP's actually come in two varieties, which we might call first and second degree; in a first degree NP. there is only one predicate, while in a second degree NP the main predicate of the construction has a complement. The two types of NP are somewhat different in their behavior in that while type 2 NP's of the first degree always end in w-transitions, a type 2 NP of the second degree may end in either a w or an s; thus compare

(10) a. The boy that Harry believes]] likes
Maxine]] saw Sue]] # w'
$\underline{s}_1$ $\underline{s}_2$

b. The boy that Harry believes]] Maxine
likes]] saw Sue]] s'
w s

In (10a), the Subject NP ends at s_1 while in (10b) it ends at w, corresponding to the fact that in the former case the head NP is the Subject of the complement clause while in the latter it is the Object of that clause. The parser copes with this fact by being equipped with two closure mechanisms, one for first degree complex NP's, and another for second degree NP's. A first degree NP can be identified by checking to see that no B' intervenes between the opening and the first B to the right thereof; if such a B' is found, closure is forestalled. Once all first degree NP's have been closed, the parser closes second degree NP's by simply looking for the first available B to the right of any opening for which no corresponding closure has been made. (Any B that has already been swallowed up into a previously recognized NP is no longer available for consideration.) It is a consequence of this feature of the parser that a second degree complex NP with a c.e. first degree NP should be more easily processed than one with a second degree NP, an expectation that is evidently borne out: compare

(11) The boy who believes that the girl who
kissed Harry likes Maxine saw Sue.

which, like (9), requires some effort, but is within bounds; however,

(12) The boy who believes that the girl who
thinks that Harry likes Maxine saw Sue
kissed Samantha.

evidently is not. In the case of (11), the inner NP is parsed first, then the outer one, and there is no erroneous closure. In the case of (12), by contrast, there are no first degree NP's, and the NP beginning with the boy is closed prematurely, after Maxine.

Honesty compels mention of a perplexing case that this approach cannot handle. If, in a c.e. type 2 relative of the first degree, one of the NP's is replaced by a pronoun, comprehensibility seems to increase:

(13) The rat that the cat I saw chased
squeaked.

I have no explanation to offer for this fact.

A general assumption underlies the treatment advanced here, as follows: in an optimal parser, one seeks the simplest procedures that will apply in error-free fashion to the simplest cases--cases in which a construction of type x contains no other constructions of type x. Effects of the sort observed here are then explained as being due to some of the more complex cases being tractable in terms of maximally simple procedures while others are not. Given the results described above, the principle seems to have a measure of plausibility as at least a working hypothesis.