Syntactic Embedding: What Can People Really Do?

Peter A. Reich and Carson T. Schütze
University of Toronto

Abstract

Recently, several attempts have been made by computational linguists to model various forms of syntactic embedding with reference to psycholinguistic data. Unfortunately, some of these models may be based on an incomplete understanding of precisely what the cited research has to say about purely syntactic aspects of human cognitive processing.

1. Introduction

Computational linguists, as researchers in the general field of artificial intelligence, can be categorized into two types: soft and hard. Hard computational linguists are not concerned with how human beings actually process language—what they care about is what their programs do. Soft computational linguists take as their goal the modelling of language behaviour in human beings. These are, of course, extremes. The hardest computational linguists are not above taking advantage of insights into the way people behave, and the softest are not above making use of computational tricks to get the job done.

One area in which computational linguists find it useful to look at human limitations is the area of clause embedding. Strict limits on embedding allow one to build simpler language processors. The simplest and least powerful processor would be a finite state device. Many linguists continue to believe that because of central embedding more power than this is needed. In some sense, all computational linguistics is done on machines which have a limited number of states. Some of us, especially those of us who call ourselves connectionists, believe that the human brain is one such finite device.

However, the question of how much power is necessary is an important issue, one that refuses to go away. We hope that this paper may offer some clarification to computational linguists about how to interpret papers coming from linguists and psycholinguists.

2. The problem: What linguists say

The issue first arose when Chomsky (1956, 1957) argued that natural language cannot be produced by a finite state device. This argument was based on sentences such as 1, 2, or 3, where the $S_1$ stand for embedded sentences.

1. If $S_1$ then $S_2$.
2. Either $S_1$ or $S_2$.
3. The rat that $S_1$ ate the malt.

The sentences labelled $S_1$ can themselves contain sentences of these types, so that one can have sentences such as 4 or 5.

4. If either $S_1$ or $S_2$, then $S_3$.
5. The rat that the cat chased ate the malt.

Chomsky’s argument was that such sentences are all grammatical because they are “formed by processes of sentence construction so simple that even the most rudimentary
English grammar would contain them” (1957: 23). What Chomsky meant by rudimentary processes was recursion. Within the rewrite rule framework one uses recursion to produce, for example, sentences containing relative clauses modifying the final noun of the previous clause, such as 6.

6. This is the dog that bit the woman who married the vet who inoculated the cat that chased the rat that ate the cheese.

One cannot formulate a rule limiting indefinite expansion to, say, the direct object, because so long as one is expanding the last noun of the clause, the sentences seem fine. For example, in sentence 7 it is the objects of prepositional phrases that are expanded. Furthermore, one must allow central embedding at least to one level, as shown in sentence 8.

7. The woman gave the present to the vet who inoculated the cat with the collar that was made of gold that was mined in Colorado.
8. The rat that the cat chased ate the malt.

In addition to the examples cited by Chomsky, others have discussed additional language structures that cause problems for finite models. One (pointed out by McCawley) is the respectively construction:

9. John, Jim, and Jason took Mary, Marjorie, and Molly to the dance respectively.

Another is nested dependencies in subordinate clauses in German, as in 10 (Bach et al. 1986).

10. Johanna hat die Männer Hans die Pferde füttern lehren helfen.
    Joanna helped the men teach Hans to feed the horses.

A particularly interesting problem is crossed-serial dependencies in Dutch (Bach et al. 1986, Bresnan et al. 1983):

    Jeanine helped the men teach Hans to feed the horses.

Variations of these orders are found in Swiss-German (Shieber 1985), as shown in 12 and 13.

    Jan says that we helped Hans paint the house.

The problem with all of the above examples is that if one accepts that the language can produce sentences of these types with an arbitrary number of central embeddings or crossed-serial dependencies, then a finite system is clearly inadequate.

3. Early studies of embedding

Does natural language contain such sentences? Are the following real sentences in English?

14. The race that the car that the people whom the obviously not very well dressed man called sold won was held last summer. (Miller 1962)
15. The money that the guy whom the black haired girl that we met at Bill’s party introduced to us borrowed from us ought to be paid back soon. (Schaefer 1971)

16. John, Jim, and Jason bought, sold, and gave a pen, a pencil, and a kiss, from, to, and to Mary, Marjorie, and Molly respectively.

There are dozens of papers discussing this issue in the psychological and linguistic literature, from 1956 on. To answer the question properly requires more careful consideration of the issues than is usually given. When is a string of words a sentence in English (or any other language)? There are a number of alternatives.

One possible criterion is that examples of a construction in question can be found to have been published somewhere, with no intention of playing with language (for example, as a letter to the editor of some newspaper, as suggested by Geoffrey Sampson). However, a problem with composed text is that it is not necessarily written linearly—that is, after one writes a paragraph, one decides to modify some sentence. Sometimes such modifications may involve adding an embedded clause. This sort of activity once in a great while leads to multiply centrally embedded sentences, which may or may not be comprehensible. Written language is known to be different from spoken—which is more basic? Most linguists would argue that spoken language (or, in the case of the deaf, sign language) is more basic.

Another criterion—one that is hard to argue with—is that a sentence actually have been uttered, and happened to have been observed by someone interested in these issues. For example, Guy Carden (personal communication) once heard:

17. A lot of the housing that the people that worked in New Haven lived in was back that way.

However, attempts to observe naturally occurring data are not popular. Such complex constructions are, at best, rare, and just because one has not noticed one does not mean that it never occurs. Thus it is that people turn to other approaches. Linguists invent sentences that they think are plausible and then argue that they are a part of the language, for example, sentence 18.

18. The one time that the only car any dealer I could find was willing to guarantee turned out to be a Ford, was when I was hunting around second-hand places in Phoenix. (Bolinger 1971)

In order to discuss such ‘sentences’ as 15 and 18, we feel that it is important to make a distinction between language comprehension and language production. There certainly is a difference. We understand

19. Mowed the lawn John.

even though we would not want to say that that string of words is an English sentence. One can accomplish a significant amount of natural language comprehension without syntax. Schank has shown this in computational linguistics; and more to the point, people with Brocha’s aphasia also show this. People with damage to the area of the brain known as Brocha’s area exhibit what many doctors (but usually not speech-language pathologists) call “production aphasia.” Such patients have trouble putting words together into sentences but exhibit little difficulty understanding what those around them are saying. However, careful tests of such patients have shown that if semantic clues are not adequate to disambiguate what is meant, they have difficulty. Specifically, they have trouble with so called “reversible sentences”—sentences which make equally good sense if two items are interchanged, for example,
20. Sue kissed Bill.

Comprehension, then, may not be such a good test. Yet comprehension is the test most people have proposed. Specifically, some sociolinguists and numerous psycholinguists have invented 'sentences', presented them to experimental subjects, and tested their comprehension of them.

Thus Ellen Schaefer (1971) tested sentences such as sentence 15 by asking questions such as:

21. What ought to be paid back?

She concluded that the ability to handle multiply-centrally embedded sentences declined slowly as the number of embeddings rose. This is a typical finding of such studies. But obviously, one can correctly respond "the money" without parsing the original sentence at all, but rather by simply recalling the noun phrases of the sentence and choosing the one that fits the question best. Could the responses have nothing to do with syntactic parsing?

In order to test this, Reich & Dell (1977) chose eight sentences and 24 questions from Schaefer's study and instead of giving subjects the sentences, they gave them only the noun phrases (or, if appropriate, the verb phrases) to choose from. The performance of their subjects did not differ from Schaefer's subjects, who had heard the entire sentences. Thus they concluded that the extent to which Schaefer's subjects were able to make use of syntax in such sentences was negligible.

If there are problems in testing comprehension, why not ask subjects directly to evaluate grammaticality? But asking subjects to evaluate sentences on the basis of grammaticality is also problematic. Schlesinger (1971) found that judgements were affected not only by amount of embedding, but also by the extent to which semantic cues were found to exist. In other words, in sentence-judging tasks informants cannot keep syntax and semantics completely separate.

We therefore conclude that one must make use of comprehension tasks after all, but one must be more subtle. For example, Labov (1973: 101-102) had his students inject sentences 22 and 23 into natural conversations and observe the reactions of their listeners. Notice that 23 has transposed two of the noun phrases of 22.

22. Say, do you think that the report that the stuff they put in diet soda causes cancer is a hoax?

23. Say, do you think that the stuff that the report they put in diet soda causes cancer is a hoax?

There was no unusual reaction to 22, whereas reactions to 23 indicated that subjects found something strange about it. They could appreciate the difference. Does this mean that subjects can comprehend the syntax of doubly centrally embedded sentences? Even this study has a serious flaw. When a parser is receiving a sentence that exceeds its capacity, it may still be able to handle those parts of the sentence that appear before the point of breakdown. What Labov's example shows is that if there is a parsing problem in these sentences, it hasn't occurred at the point where the substitution was made.

Two studies have given evidence of just where the breakdown occurs. Bruner & Cromer (1967) found that when reading multiply-centrally embedded sentences, subjects' eye movements started retracings at the second verb, indicating that this is the area where the breakdown occurs. Labov's sentences transposed material before this point.
Reich & Dell (1977) ran an experiment to determine the ability of native speakers of English to process the syntax of multiply-embedded sentences. They devised a set of 12 semantically restricted sentences that were centrally embedded four times. Each sentence had a sequence of five consecutive verb phrases; sentence 24 is one such example. From each sentence they generated three semantically anomalous sentences by interchanging the first verb phrase with the second, the second with the third, or the third with the fourth. These are shown in 25-27.

24. The ground that the grass that the hungry deer that the green forest that the fire destroyed sheltered ate covered used to be soft.
25. The ground that the grass that the hungry deer that the green forest that the fire sheltered destroyed ate covered used to be soft.
26. The ground that the grass that the hungry deer that the green forest that the fire destroyed ate sheltered covered used to be soft.
27. The ground that the grass that the hungry deer that the green forest that the fire destroyed sheltered covered ate used to be soft.

The results were that, for the vast majority of the 120 English speakers tested, the syntactic processing failed after hearing the second verb phrase in these multiply-centrally embedded sentences. On the other hand, it is well known that there are no such limits on right branching sentences like 6. Thus, if a computational linguist wishes to model language processing on people, the parser should be able to produce or parse only one level of central embedding, but have no limits on right- or left-branching. It happens that such a parser is a finite state device (Reich 1969).

4. The case of German, Dutch, and Swiss-German

The previous work occurred a generation ago. Thus it is with dismay that we note that history is repeating itself with respect to the Germanic language data. Shieber (1985) used these data to argue that natural language cannot be described by a context-free grammar. He asked four native speakers of the Zürich dialect of Swiss-German to judge the grammaticality of 62 sentences containing up to four clauses. He admitted that the study was not a controlled experiment, stating that “such is...for the most part, unavoidably the case in this area of linguistic research.” He reported that “the vast majority of examples showed unanimity of judgement.” However, he noted that “much beyond triple embedding ... judgements get weaker.” What does this mean? His speakers required more cajoling to say “yes”? They said something like “Well, I guess maybe it’s okay.”? We don’t know. But this point is crucial. Clearly things were getting worse when it came to four clauses. Might not a reasonable line be drawn at three or four? He appears not to have tested five embeddings. His response to the argument that there might be a finite limit is curious: “Down this path lies tyranny. Acceptance of this argument opens the way to proofs of natural languages as regular, nay, finite.” Yes, indeed! Our point exactly!

A much better study methodologically is the recent study by Bach et al. (1986). In this study 72 sentences were given to 30 subjects in Holland and 60 in Germany. Subjects were asked: 1) to rank sentences as to comprehensibility on a nine point scale, and 2) to answer questions testing their comprehension—questions of the form “What did (one of the clause Subjects) do?”

Informally, when Dutch and German speakers saw the sentences, they rejected outright sentences with two or more levels of embedding—a radically different result from that obtained by Shieber for Swiss-German. The results on the formal tasks for both languages showed increasing difficulty with increased levels of embedding, not unlike Schaefer's
findings. An interesting added result was that the comprehension scores differed significantly between the two language groups. The scores for the Dutch were superior to the German at high levels of embedding.

Joshi’s (1990) model, and possibly others, have been influenced by this result. But what does it really mean with respect to syntax? The syntax probably cannot cope with multiple embeddings in either Germanic language much better than in English. But the difference between Dutch and German suggests that slightly more information may be correctly assembled before the point of breakdown in Dutch than in German.

5. Our proposal

We claim that the extent of syntactic processing is both more limited and more sharply defined than many linguists believe. It is simply not true that there are no ways of empirically testing syntactic limitations independent of semantics. Here are four which we hope to explore in the future.

1) Construct sentences with no semantic restrictions guiding the noun phrase–verb phrase pairings, then determine whether the interpretation required by syntax is the one people come up with. Systematically test the first, second, ..., nth clause, to determine where the breakdown occurs.

2) Construct sentences with semantic restrictions, then try using syntax to force a disfavored reading as in Reich & Dell. See whether (i) interpretation changes and/or (ii) processing becomes more difficult and/or (iii) the sentence is considered “less normal.”

3) Compare the comprehension of sentences with that of lists of noun phrases and verb phrases to see whether syntax is contributing to interpretation at all.

4) In languages like Swiss-German, check whether faulty case agreement is noticed at all levels of embedding, as Shieber claims.

If one accepts what we say, the implication is that a finite state device can handle all the problems of embedding raised by linguists. The issue that then arises is what is the structure of this finite model. This we have tried to answer in Schütze & Reich (1990), also being presented at this conference. In that paper we show that a finite system organized as a connectionist network appears to be a viable model.

Acknowledgement

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

References


