PARSING AND SYNTACTIC THEORY

S.G. Pulman
University of East Anglia,
Norwich NR4 7TJ, England.

Summary

It is argued that many constraints on syntactic rules are a consequence of simple assumptions about parsing mechanisms. If generally true, this suggests an interesting new line of research for syntactic theory.

Recent syntactic theory has aimed for a theory of grammar rich enough to be descriptively adequate, but restricted enough to make plausible claims about the acquisition of such systems. Thus in Chomsky's trace theory,[3] sentences, their constituent structures and grammatical relations are determined by a few simple rules which in themselves would generate many non-sentences, but which are constrained by putatively universal principles marking the non-sentences as such. These principles are attributed to the human language faculty, but with no direct implications for models of perception or production, involving as they do many abstract concepts (empty nodes, traces etc.) for which there is no obvious psychological interpretation.

Bresnan[1] presents an interesting alternative: that proposals about the nature of grammar should be responsible to psycholinguistic evidence--to put it crudely, grammars should be 'psychologically real'. By this criterion, many constructions such as passives or infinitive phrases previously regarded as transformational are to be seen as generated directly by phrase-structure rules, with associated conditions on lexical insertion and relationships between lexical entries. Transformations for which there appears to be psycholinguistic evidence, like Wh-movement[11], remain, though under a different formulation than they receive under trace theory.

Taken literally, the requirement of strong psychological reality for grammars makes no sense: I shall take it to mean that such a grammar is one which can be incorporated with little or no adjustment into a model of perception or production, and that the various subcomponents and divisions of labour in that model are both consistent with those implied by the grammar and not disconfirmed by experimental evidence. What sort of parsing model is implied by Bresnan's theory? Bresnan herself envisages realistic grammar being implemented directly as an Augmented Transition Network[1,11] with a HOLD analysis accounting for Wh-phrases. PS rules correspond directly to sections of network, and HOLD is Wh-movement in reverse, so to speak. However, there are at least two deficiencies in such a proposal: firstly, in all ATN models yet proposed, there are many points where exits from a state are tried in a fixed, often arbitrary, order. As a claim about how humans parse this is desperately implausible, for it suggests that at such points we expect an A, look at the next word, find that it is not a possibility for A, revise our expectation to B, find that it is not ... until we have an expectation confirmed or arrive at a default value. There may be some points at which such a situation occurs but in general it does not seem to: more plausible is that we look ahead to the next word without any particular expectation and use the resulting information to decide which exit to take--we go 'bottom-up' in other words. Secondly, in the current generation of ATN proposals there is little or no facility for revising syntactic expectations on the basis of lexical (much less other types of semantic or pragmatic) information, though there is ample intuitive and experimental evidence that humans do this. However, this is probably a criticism of practice rather than of principle.

A 'realistic' parser

A better picture of a parser is one which works at two levels simultaneously: the lowest level mostly bottom up, assembling phrase level constituents NP VP PP etc., possibly incorporating the 'late closure' and 'minimal attachment' principles[6] of Fodor and Frazier. The information given by PS rules and the lexicon would be available to it and guide its operation but would not necessarily correspond directly to portions of a network. The second, more 'top-down' routine would organise these phrases into functionally complete units--roughly, a verb with its obligatory arguments and any optional arguments or modifers which might be present. This second stage uses whatever grammatical information is present in the input (presence of Wh-phrases, complementisers, passive morphology etc.) along with lexical information (subcategorisation restrictions, possibly semantic restrictions on arguments) to determine construction type and assign arguments to their grammatical (and possibly semantic) functions. It is also plausible, and in some cases necessary, to assume that it can guide the operation of the phrase level stage at various points. (A two stage model along somewhat different lines has already been proposed:[6] in support of the claim that functional completeness rather than surface clause segmentation or 'sentoids' guides assembly into units see Carroll et al.[2]).

This is not the place to go into details of implementation; though a preliminary version

of a parser embodying these principles has run with some success, there is still much work to do. The present aim is rather to show how such a model could account for some of the range of phenomena which linguists are currently interested in and of which any theoretical approach m must present a credible account. In particular we want to know whether anything over and above the basic grammatical information represented in an extended lexicalist grammar of the type described by Bresnan will need to be made available to the parser.

To simplify matters, let us ignore most types of complex sentence, and phenomena like ellipsis, anaphora, and conjunction. Most of the simple sentence types of English can then be summarised as follows, ignoring details:

(1) NP { be Adj / V (Prep) (NP) ({NP / S}) } (PP)

Sentences which do not have any of these structures must have undergone grammatical operations of various types, which, using a useful earlier framework[4] we can describe as:

(2) i Root transformations: Subject verb inversion, Topicalisation etc.
ii Wh-movements: Questions, relatives etc.
iii Equi-NP-deletion: infinitive verb phrases etc.
iv Structure-preserving NP movements: Passive, Raising, Tough-movement etc.

or some combination of these.

As well as assigning (phrase level) constituent structure, our parser must of course determine grammatical relations, both overt and covert, in the course of building functionally complete units. On sentences having any of the forms in (1) this is quite straightforward, as surface grammatical relations provide all the information needed. The following definitions will suffice:

(3) Subject (S) = NP/___V tense (V includes Auxiliaries)
Direct Object (DO) = NP/V(Prep)____
Prepositional Object (PO)= NP/P____

Notoriously, though, in sentences resulting from (2)i-iv overt grammatical relations are often different from covert grammatical relations. How would we deal with these? For concreteness, and without any great warrant, let us assume that root transformations are 'unwound', possibly by an extension of the HOLD mechanism for some cases, resulting in structures like (1). (3) can then be used on the result. Similarly, let us assume that something like the existing HOLD treatment of Wh-movements[11] is adequate: a Wh-phrase is put in HOLD until a gap is found, either an obligatory one--a tensed V with no subject, an unfilled subcategorisation slot or stranded preposition; or an optional one--an optional subcategorisation slot or a modifier position. The Wh-phrase is then treated as if it were in the gap and (3) can apply directly. For the sentences in (iii) there is a simple, well-known generalisation:

(4) The subject of an infinitive is the DO of the matrix verb if it has one, otherwise its subject.

The equally well known exceptions to this (promise, appeal, the indirect question sense of ask etc.) can be marked as such in the lexicon, as in all other treatments. For the structures in (iv), surface grammatical relations can be determined directly by (3). Lexical information à la Bresnan is required to determine their underlying grammatical relations, and in the case of passive, grammatical information can be used:

(5) In a structure ... be-V-en/ed ... the (surface) S of be (by rule (3)) is the (underlying) DO of V. The underlying S of V is the PO of by, if present (and a 'possible agent'), or Indefinite (to be pragmatically interpreted) otherwise.

This is essentially Bresnan's lexical redundancy rule.

(4) will account directly for subject-raised sentences:

(6) John seems to like Bill

and (3) will account directly for their putative source:

(7) It seems that John likes Bill.

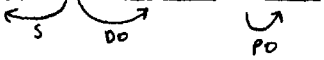
Presumably some semantic statement about the it of such extraposed structures will be needed, as in any account. Other constructions are straightforwardly dealt with by (3) and (4) with the exception of Tough movements:

(8) John is difficult to persuade

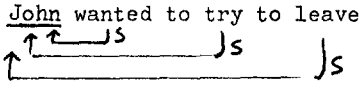
The parser will go doubly wrong here, assigning John as S of persuade (by (4)) and marking the string as ungrammatical because of an unfilled gap after the verb. The obvious solution is to mark Tough movement adjectives lexically so that when they appear in these contexts they cause the parser to put the subject (John) in HOLD, and assign (in this case) Indefinite as the subject of persuade. John will then be correctly inserted in the gap. This treatment captures the many similarities[3] to be found between Tough and Wh-movement.

Some examples will make the operation of (3) - (5) clearer.

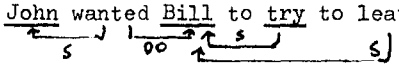
(9) The boy gave the book to Mary (by 3)



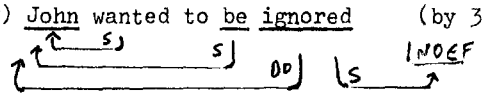
(10) John wanted to try to leave (by 3 & 4)



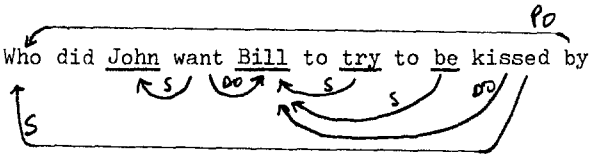
(11) John wanted Bill to try to leave (by 3 & 4)



(12) John wanted to be ignored (by 3,4 & 5)



(13) Who did John want Bill to try to be kissed by



(by 3,4,5 and HOLD)

We can define a 'functionally deviant' sentence as one in which obligatory arguments to verbs have not been found by either stage of the parser (using lexical information, (3) - (5) and HOLD) or where phrases are 'left over', unable to be consistently assigned as arguments to verbs. Such a definition follows naturally from the operation of the parser.

Trace theory

In Chomsky's trace theory[3] all major structures are generated by base rules. Rules inserting lexical material are optional, and some rather simple transformational rules (Move NP, Move Wh-phrase) allow categories to be moved to any other (empty) slot of the same type, leaving behind a 'trace' which is interpreted as identical in reference to the moved category. Traces are used to indicate underlying grammatical relations (of the sort dealt with in 4, 5 and HOLD), to assign aspects of interpretation, and they also function to block the application of syntactic and (some) phonological rules as if they were terminal elements. A sample derivation is the following, where Δ = empty NP node:

(14) i Δ seems $[_S$ Bill to like John$]$

ii Bill seems t to like John

Bill has been moved by Move NP to fill the empty node leaving behind a trace, t which is correctly interpreted as the subject of like, and identical in reference to Bill. If Move NP had not applied (i) would have been marked ungrammatical by a general convertion that no uninterpreted empty node can remain at the end of a derivation. If Move NP had moved John to replace the empty node, the result:

(15) John seems Bill to like t

would have been declared ungrammatical since such an application violates an allegedly universal constraint, either at the stage at which John is moved or when an interpretation rule seeks to relate John to its trace. The constraint in question is the Specified Subject Constraint (SSC), and in the simplest cases it says that no rule can relate or involve positions which are on either side of a full subject:

(16) ... X ... $[_S$ NP ... Y ... $]$

A parallel mechanism, the Tensed Sentence Constraint (TSC), says that no rule can relate a position inside a tensed sentence to one outside that sentence, i.e. in (16), nothing can relate Y or NP to X if S contains Tense. This would correctly rule out the final stage of the derivation:

(17) i we believe $[_S$ the dog is hungry by $\Delta]$

ii t_1 is believed $[$ the dog is hungry by us

iii the dog is believed $[$ t is hungry by us

If the embedded sentence had not contained a Tense, i.e. been ... to be ... the derivation would have been permitted.

Obviously this is a very simplified and incomplete account. In particular it should be noted that Wh-movement, which looks like an exception to SSC and TSC, is not, on Chomsky's account of it.[3] For our purposes, what it is important to focus on is that it is not the rules of grammar themselves which characterise grammaticality, for they will over-generate massively, but the interaction of rules of grammar with constraints like TSC and SSC.

In our framework, however, such constraints are unnecessary. Uninterpreted empty nodes correspond to unfilled gaps, and will be rejected via the functional completeness requirement. Cases like (15) are ungrammatical on two grounds: (3) automatically analyses Bill as DO of seems but this does not match any of its subcategorisation entries and so the sentence is rejected. Furthermore, like is subcategorised for either a DO or an infinitive: neither is present in (15) and so the sentence would

also be rejected on these grounds. All the cases in (17) would be excluded because they contain unfilled gaps (obviously traces and △ are not present in the input string): in (i) there is a stranded preposition demanding a PO, while in (ii) and (iii) is (believed) and is (hungry) respectively cannot be assigned subjects by (3) since there is no overt NP, nor by (4) since they are not infinitives. Thus they too are marked as functionally incomplete.

As for as I can ascertain, all the syntactic effects of TSC and SSC as regards NP movement can be accounted for in a similar fashion, using only that information already required for the recognition of grammatical sentences. If this is true, it suggests a strong and interesting hypothesis: namely that all the syntactic effects of all the constraints proposed at various times would also be automatic consequences of the operation of this type of parser. This has already been suggested for particular instances,[7, 5] if it should prove generally true this would be a fascinating discovery.

More constraints

Another major constraint in Chomsky's framework is known as Subjacency--the requirement that no rules can relate positions separated by two or more cyclic (i.e. NP or S) boundaries. The effects of Subjacency on NP movement result in nothing not accounted for by the earlier discussion but whereas Wh-movement is not affected by SSC and TSC it is not exempt from Subjacency. Subjacency subsumes, as well as other principles, Ross's[10] Complex Noun Phrase Constraint (CNPC), which prohibits movements out of $\angle_{NP}$NP S$\angle$ structures, and some cases of the Co-ordinate Structure Constraint (CSC), which prohibits movements out of conjoined structures. These principles account for the unacceptability of:

(18) (i) *Who did John see the man that Bill hit-
(ii) *Who did John believe the claim that Bill hit -
(iii) *Who did John see - and Bill

Unlike TSC and SSC there is no chance of ruling out the sentences via subcategorisation restrictions or 'functional deviance'--in other words we will have to rely on assumptions about the implementation of the parser. Nevertheless, proposals have been made which will capture the effects of CNPC and CSC and which are in accordance with the logic of our approach, namely, that the most straightforward way of handling grammatical cases will automatically, without further stipulation, rule out the ungrammatical cases. As regards CSC, Winograd[12] has proposed that the simplest syntactic treatment of conjunctions is to interpret them as an instruction to look for a syntactic unit to the right of the conjunction of the same type as one which has been recognised ending immediately to the left of it. This is equivalent to the simple and pervasive generalisation about conjunctions that they conjoin like with like. Thus (18)iii would be ruled out because there is no NP immediately preceding and to match with Bill. This treatment allows for 'across the board' movements:

(19) What does Bill like - and John detest - ?

without qualification. Some elaboration has to be made to allow for anaphorically interpreted deleted elements when dealing with conjoined sentences but otherwise the principle seems essentially correct and is presumably a 'parsing universal' of a substantive kind.

The only workable suggestion concerning the CNPC that I am aware of comes from Ritchie.[9] He proposed an analysis of relative clauses and other Wh-movements similar in spirit to the HOLD analysis. In his programme, the Wh-phrase (presumably a copy of the head NP if no Wh-phrase is present) is stored in a register called WHSLOT which is a local variable. Since it is a local variable, if the relative clause procedure is called recursively, the contents of the register on any earlier call are unavailable to the current call. As he points out, this is a simple and natural way of treating any construction capable of unlimited embedding. The results of this treatment (exactly the same as a HOLD list which is a push-down stack) of relative clauses is that in a sentence like (18)i, the analysis of the relative ... the man that Bill hit ... will invoke a use of the current incarnation of WHSLOT and will therefore put that (or the head NP) into the gap it finds after hit, assigning it DO of hit (assimilating Ritchie's analysis now to ours). The WHSLOT in which the first who was placed is no longer accessible. (Alternatively, the who on the HOLD list is not 'first out'). Thus who could never be placed in the gap and the sentence will be deemed functionally deviant, since at the end of it there is still an unassigned Wh-phrase: all similar cases likewise fall out of this simple and natural (though refutable) principle concerning the treatment of embedded structures involving Wh-movement.

Cases like (18)ii are somewhat more complex. Ritchie points out that at the point at which the parser is aware that it is parsing a complex NP structure, sentences like (18)ii are always potentially ambiguous, between a true relative and a true complex NP reading:

(19) i ... the claim that he made
ii ... the claim that he made a mistake

It seems plausible to assume that the parser postpones a decision until the ambiguity is

resolved. Again, the natural way to do this is to store that (or a copy of the head NP) in WHSLOT or in HOLD, and insert it into a gap if one is found, confirming the expectation that it is a relative clause, or revising the decision about construction type if no gap is found. Either way, no previously encountered Wh-phrase can be inserted into the gap, and so the parser rules out (18)ii as functionally deviant, on two counts in this particular instance: the claim will be marked as DO of hit, and presumably lexical information will show that such an assignment of argument type violates semantic restrictions (you can't hit claims). Secondly, since the gap at the end of the sentence has been filled there is an unassigned Wh-phrase left over.

A further syntactic constraint yet to be accounted for is Ross's Left Branch Constraint (LBC). This prohibits movement of a NP which is the leftmost constituent of a longer NP. Most cases ruled out by this constraint fall under the kind of treatment discussed earlier for TSC and SSC but one class of cases, in which a determiner constituent is fronted, remain:

(20) *Whose did you drink - beer

The acceptability of sentences like:

(21) Whose - did you drink?

where the object of the possessive determiner is understood from context, suggests that when a possessive wh-phrase is encountered, if there is no NP immediately following, a dummy is inserted, to be pragmatically interpreted. In (20) no gap will be found, since beer can be DO of drink. There will thus be an unassigned Wh-phrase whose NP left over and the sentence will be marked functionally deviant.

Clearly, there is much work left to do in exploring the hypothesis that all such constraints can be accounted for by no extra mechanisms than are needed for grammatical sentences. Nevertheless, there seems ample evidence that this is a worthwhile avenue to pursue, and if it is true interesting questions can be raised. What, for example, is the status of TSC and SSC with respect to semantic rules? In Chomsky's framework, rules interpreting reciprocals (each other), reflexives, and pronouns are constrained by TSC and SSC:

(22) i We expect {ourselves / each other} to win

ii *We expect that {ourselves / each other} will win (by TSC)

iii *We expected Bill to warn {each other / ourselves} (by SSC)

In our framework, specified subjects of infinitival sentences are regarded by (3) as direct objects. For sentences like (22), therefore, something similar to a clause-mate account will give the right results: reflexives and reciprocals must refer to a preceding (agreeing) NP to which they are grammatically related via (3), (4) or (5). For pronouns, the reverse is the case: they cannot refer to a preceding NP to which they are grammatically related.

A preliminary attempt to extend such an analysis has been made,[8] though it faces obvious problems with verbs like arrange, which can take a non-finite sentence as complement, and with non-finite relative clauses. If it can be carried through though, then the fact that the functional structures delivered by the parser's second stage provide the right kind of information to state such restrictions on would constitute further evidence that such an approach is on the right lines. Alternatively, it might prove more accurate to retain TSC and SSC as purely semantic constraints.

Another question to be raised concerns the relationship between constraints, grammars and parsers. For Chomsky, constraints are part of the theory of grammar and do not have to be stated for individual grammars (though some language particular parameters might have to be specified). But there is a good sense in which the constraints proposed are arbitrary: it would be just as easy, for instance, to have a Non-Tensed S Constraint, or a Specified Non-Subject Constraint. This fact is what gives plausibility to Chomsky's claim that the constraints are the result of as yet unknown innate mechanisms. But in our framework, if it is anything like correct, a good deal of this arbitrariness is removed. Our grammar uses only the same kind of distributional information as is used in trace theory, and yet does not over-generate to quite the same extent. If we now implement this grammar inside a parser with only such additional information as is required to accept and functionally encode grammatical sentences ((3)-(5), conjunctions, HOLD as a push down stack) it seems to automatically rule out ungrammatical sentences which require constraints in the alternative framework. If this proves to be generally true, the obvious next step is to devise psychological tests to discover whether what seem to be simple and natural assumptions about parsing are in fact an accurate description of the human parsing system. The status of constraints changes radically: instead of being abstract properties of the language faculty to be investigated by some future neurophysiologist they constitute valuable evidence concerning the operation of the human parsing system, and the kind of hypothesis that can be advanced

should be accessible to more direct experimental testing. This, it seems to me, would constitute considerable progress in our understanding of these phenomena.

Footnote

This is a condensed version of a longer paper currently under revision. Some of the ideas are sketched out in an earlier working paper[8] and I would like to thank R.A. Hudson and G. Ritchie for their comments on that. The usual disclaimers apply.

References

(1) J.Bresnan 1978 'A realistic transformational grammar' in M.Halle et al., 'Linguistic Theory and Psychological Reality' M.I.T. Press, Cambridge and London pp.1-59.

(2) J. Carroll, M. Tanenhaus and T. Bever 1978 'The perception of functional relations' in W. Levelt, F. d'Arcais eds., 'Studies in the perception of language' John Wiley and Sons, Chichester and New York, pp.187-218.

(3) N. Chomsky 1977 'On Wh-movement' in P. Culicover et al. 'Formal Syntax' Academic Press, New York and London.

(4) J. Emonds 1976 'A transformational approach to Syntax' Academic Press, New York and London.

(5) J.D.Fodor 1978 'Parsing strategies and constraints on transformations' Linguistic Inquiry Vol. 8 1978 pp.427-473.

(6) J.D.Fodor and L. Frazier 1978 'The sausage machine' Cognition Vol. 6 1978, pp.291-325.

(7) M.Marcus 1977 'A theory of syntactic recognition for natural language' mimeo, M.I.T. A.I.Lab.

(8) S.Pulman 1979 'Grammatical description and sentence processing' UEA Papers in Linguistics Vol. 10, University of East Anglia, Norwich.

(9) G.Ritchie 1977 Edinburgh University Ph.D. Thesis.

(10) J.Ross 1967 'Constraints on variables in syntax' excerpted in 'On Noam Chomsky' ed. G. Harman, Anchor Books 1974.

(11) E. Wanner and M. Maratsos 1978 'The comprehension of relative clauses' in M. Halle et al., op cit. pp.119-161.

(12) T. Winograd 1972 'Understanding natural language' Edinburgh University Press.