North East Linguistics Society

Volume 15
Issue 1 *NELS 15*

Article 8

1985

Rules and Constraints in Sentence Processing

Stephen Crain
University of Connecticut

Janet Dean Fodor
University of Connecticut

RULES AND CONSTRAINTS IN SENTENCE PROCESSING

Stephen Crain and Janet Dean Fodor

University of Connecticut

Some people have hoped that the study of sentence processing will help us choose between competing theories of grammars, where more narrowly linguistic evidence fails to distinguish them. Other people regard this hope as naive. Unfortunately, the pessimists appear to be closer to the truth. Even if a linguistic theory would stand still long enough to be tested experimentally, it very rarely generates any firm predictions about performance. Almost any kind of grammar can be reconciled with almost any experimental results on parsing times, error rates, and so forth, just by changing the way the grammar is implemented in the processing mechanism. It is even possible to implement one kind of grammar and then convert its output into a form characteristic of another kind of grammar; the latter could then be viewed as the competence grammar, while the former, which is actually doing all the work, would be thought of as merely a performance algorithm.

Nevertheless, most people who work on sentence processing have a fairly sturdy sense, even if it's only a matter of intuition at present, of the difference between a direct and an indirect implementation of a grammar. And it is most reasonable to start, surely, by considering the most direct applications of the kinds of grammars we have arrived at by linguistic argumentation. Only if these straightforward models fail do we need to look for more distant and intricate relationships between what people know about language and how they put that knowledge to practical use.

However, even if we accept this methodological rule of the game, it is still remarkably difficult to identify different implications for performance associated with different theories of grammar. For instance, what sort of experimental results would it take to really establish that people mentally derive passive sentences by transformation, rather than by looking them up in the lexicon? What would we expect to observe in an experiment if people were deriving "that-relatives" by deletion rather than movement? Or were applying Subjacency rather than the Complex NP Constraint? Or the ECP rather than Subjacency?

We know of just one difference between formal linguistic devices which does have fairly direct implications for sentence processing, and that is the difference between rules and constraints. We take rules to be positive grammatical statements, which say what CAN occur in a language, and constraints to be negative grammatical statements, which say what CANNOT occur.[1] Some grammars consist solely of rules; others consist of a mix of rules and constraints. The most familiar systems use both. Transformational grammars have contained constraints since the days of the Standard Theory, and the recent emphasis in Government Binding theory has been increasingly on constraints rather than rules to carry the descriptive burden of distinguishing grammatical from ungrammatical sentences.[2]

Grammars without constraints are rarer, for the obvious reason that their rules must be very carefully tailored to the facts; they must not even potentially overgenerate, for without constraints there is nothing to hold the power of the rules in check. Nevertheless there do exist examples of rule-only grammars that are of linguistic interest. In a GPSG system the derived grammar (which is what actually generates the phrase markers for sentences) contains no constraints. (See J.D. Fodor, 1983, for discussion and more detailed examples than we can include here.) A GPSG derived grammar consists solely of a lexicon and a set of context free phrase structure rules (with associated semantic rules, which we will ignore here). Assuming that the grammar is descriptively adequate, these rules generate phrase markers for all and only the well-formed sentences of the language; they do not generate any additional phrase markers which have to be filtered out by constraints.[3] There ARE constraints in a GPSG, but they apply at the level of the metagrammar to limit the set of PS rules constituting a derived grammar; they don't apply in the generation of sentences (phrase markers) <u>per se</u>.

Suppose now that we construct a model of the human sentence parsing mechanism which assumes that the parser consults the competence grammar and uses it to compute for each input sentence a representation of the formal derivation which the grammar assigns to it. There is obviously some leeway with respect to just how these computations are supposed to be organized, but for a GB grammar the natural picture, at least, is that the parser uses its grammatical rules to construct (sequences of) phrase markers for input sentences and uses its grammatical constraints to discard any incorrect phrase markers that it has constructed. This is at least roughly the picture presented by Frazier, Clifton and Randall (1983), and by Freedman and Forster (in press), whose work we will be discussing below. The important point, for these purposes, is that the parser may start to assign a derivation to an ungrammatical sentence, i.e., may assign it a representation at one or more linguistic levels, before this derivation is aborted by the application of a constraint. By contrast, for a parser incorporating a rule-only grammar such as a phrase structure grammar, there could be no such thing as constructing an illicit phrase marker and then applying a constraint which throws it out.[4] For such a grammar, a 'constraint' (e.g., the sentential subject constraint) is just the absence of a rule (e.g., a rule for passing the slash annotation that signifies a gap down onto the node dominating a sentential subject). A parser employing a grammar without constraints will simply apply the rules it has, not apply rules it doesn't have, and the phrase markers it constructs will OBEY the constraints even though there is no sense in which the constraints were applied TO them.

We have argued elsewhere (Crain and Fodor, 1984) for the merits of a GPSG-based parser. Its operations are simple and uniform. Since the grammar contains nothing but phrase structure rules, the parser does nothing but apply phrase structure rules to its input word strings, and there are a variety of effective algorithms for doing that. Since there are no constraints, there is no need for a complex control system to regulate interactions among rules and constraints, establish an order of application, and so forth. Furthermore, the parser never wastes its (presumably limited) resources constructing phrase markers which will subsequently be filtered out.

In our previous work we pointed out that this GPSG-based model of the parser makes some strong predictions, for its performance cannot exhibit any contrast attributable to the difference between rule application and constraint application. (A parser using a grammar with both rules and constraints might or might not exhibit such a contrast, depending on what is assumed about the implementation, the demands of the task, the sensitivity of the experimental paradigm, and so forth.) There have been two challenges to the GPSG prediction in the recent literature. Frazier, Clifton and Randall (op. cit.) have maintained that the parser's application of (some) constraints is delayed relative to its application of rules, so that sentences may temporarily be assigned illegitimate analyses even in cases where there is no real ambiguity, i.e., where the rules and constraints together do uniquely determine the correct analysis. The parser is at first unduly liberal in its assignment of structure to sentences because it is relying solely on the rules of its grammar, which overgenerate; only later does it have access to the information encoded in the constraints. We have argued (Crain and Fodor, op. cit.) against this interpretation of the experimental results presented by Frazier et al., and in favor of an interpretation which is compatible with the strong prediction of the GPSG-based model.

The second challenge to this model, which we will concentrate on here, is by Freedman and Forster (op. cit.), who focus on the issue of whether ungrammatical sentences are assigned partial derivations terminated by the application of a constraint. They have used a sentence matching task which they claim is sensitive to whether or not subjects have access to some well-formed syntactic representation for a sentence, i.e., to whether or not the sentence parsing mechanism is able to construct such a representation in accord with the grammar available to it. The grammar (any grammar) obviously does provide such a representation for a grammatical sentence. For an ungrammatical sentence, as we have seen, the predictions vary. A GPSG provides no well-formed representation at all. Faced with an ungrammatical sentence, the best the parser could do would be to compute a structure for it by analogy with some grammatical sentence which it resembles. By contrast, a GB grammar with constraints discriminates among ungrammatical sentences. None will be assigned complete well-formed derivations, but those which have a well-formed D-structure and only violate post-D-structure constraints will at least have partial derivations.

If successful performance on the sentence matching task requires the mental construction of a complete and well-formed derivation, then both models predict that only fully grammatical sentences will be successfully handled. But if all that is required for sentence matching is the mental construction of some well-formed PART of a derivation, then a GB-based parsing model predicts that performance on some ungrammatical sentences will be poor while for others it will be as good as performance on fully grammatical sentences. Just where the dividing line falls between examples depends on just what one takes the correct derivations to be, and on how much of a derivation is claimed to be necessary for matching. If a D-structure is sufficient, more ungrammatical sentences should be easily matched than if an S-structure is necessary. The predictions of a GPSG-based parsing model remain as before; other things being equal, performance should be equally bad on all ungrammatical sentences since they all equally lack derivations.

The results that F&F have obtained from this task do make a sharp distinction between two kinds of ungrammatical sentence, and it is argued that the nature of this distinction is as predicted by reference to the notion of a partial derivation in a GB grammar. The natural conclusion is that GB theory provides the correct characterization of the grammar employed by the human sentence parsing mechanism, and hence (by the methodological principle of assuming direct implementation if possible) provides the correct characterization of the competence grammar also.

If these experimental findings and their interpretation are valid, they are clearly very important. If they AREN'T valid, it is correspondingly important to ESTABLISH that they aren't. For so much hangs on them. As we have seen, they could genuinely help to choose among competing theories of language. If the sentence matching task really is a divining rod for the existence of partial derivations, it could be used not just to prove that grammars contain constraints, but also to map out the exact division of labor between rules and constraints. Reliable information about that would enormously reduce the field of candidate grammars. Furthermore, an understanding of the rule/constraint distinction could have an impact on broader issues in psychology, because it connects with questions about modularity of mental organization.

A GB grammar is highly modular. Not only is the phonology module separate from the syntax module, which is separate from the semantics module, and so on; but also, WITHIN the syntactic component there are many subcomponents, submodules, each of which may contain as little as one rule, or two or three constraints. GPSG, by contrast, suggests a highly NON-modular organization even WITHIN the syntactic component. Or, to be more precise: the metagrammar could be modular or nonmodular, but all the interactions between its modules will be PRECOMPILED in the derived grammar which actually generates phrase markers. At that level there is nothing but an undifferentiated set of phrase structure rules. (Caveat: see footnote 3 above.)

The efficiency or otherwise of a modular organization for a processing system has recently become a subject of lively debate. We cannot do full justice to it here, and in any case it is still far from a settled conclusion. We will just sketch in a few of the questions, about mental functioning in general and sentence parsing in particular, that have arisen in the context of the modularity thesis (see J.A. Fodor, 1983). For instance: Does modularization of a system always increase its efficiency? If not, what makes the difference? Is it just a matter of grain, i.e., that modules shouldn't be chopped too small? Or does it depend on how distinct the kinds of information are that would be parcelled out between the modules? Or is it perhaps, as suggested by Stabler (1984), at least partly a matter of what the system DOES? If it stores information, which is what a competence grammar does, then modularity may be beneficial, inasmuch as it can unpack interaction effects and thereby eliminate redundancies and compact the representation. But if the system APPLIES information to input, then precomputing the interaction effects might improve efficiency by making it unnecessary (as we noted above) to orchestrate on-line the interactions of many different components.[5]

Exactly the opposite line is taken by Barton (1984), who maintains that the GB goal of reducing a competence grammar to a small set of independent, but interacting, principles should be mirrored in the processing model also. This is far from self-evident, as Stabler's discussion makes clear, and it is not even clear that the modularity of a GB grammar CAN be translated into modularity within a GB parser. There is a difference that is sometimes overlooked between being able to STATE facts independently of each other, and being able to APPLY them independently of each other. Consider, for example, a standard theory transformational grammar with extrinsic rule ordering. One could factor out regularities in a language and encode them in a number of separate transformational rules, but one certainly couldn't construct a parser which applied these rules independently to word strings. The parser proposed by Fodor, Bever and Garrett (1974) employed heuristics which took advantage of surface consequences of complicated interactions among the rules of the grammar. Later, within the framework of the Extended Standard Theory, Marcus' (1980) parser did much the same; its parsing rules do not relate one-to-one to grammar rules, but represent global consequences of the whole grammar. (For example, a great many facts about the grammar of English conspire to make the ending on the verb _take_ a crucial cue to sentence structure in examples like _Have the students taken the exam_? versus _Have the students take the exam_.) Most recently, Berwick and Weinberg's (1984) GB-based parser apparently blends information from the base component and the transformational component to guide its construction of an S-structure. These parsing models were not designed arbitrarily, so the fact that they are less modular than their associated competence grammars points up the moral that an economical representation system does not always make for an economical processing system.

But what, then, about the other desideratum, which is to preserve a close relationship between the competence grammar and the performance routines? One of the features of GPSG theory which makes it interesting in this context is that it provides two distinct levels, the metalevel at which generalizations are expressed and description is compact, and the derived level at which all interactions are precomputed and information is available to the performance routines in a quite direct and 'surfacy' form. A precise algorithm relates the two, so it isn't left to the parser to work out _ad hoc_ what surface consequences the grammar has when uncompacted. The design of the competence system is such as to guarantee provision of that information, and in a systematic format. Thus the goal of tying performance closely to competence can be reconciled with the apparently contrary demands made by each.

There is a common objection to this kind of system, which is that it pays too high a price in terms of the size of the rule system. The number of rules in the derived grammar of a GPSG is far larger than the number of rules in the metagrammar. However this number can be sharply reduced by the collapsing of subsets of rules into schemata by abbreviatory devices. Thompson (1981) shows how rule schemata can be applied directly in parsing.[6] In any case it is not proven that having a large number of rules for parsing is necessarily a disadvantage. Whether or not the need to access more rules will slow down a system depends crucially on the nature of the access mechanism. Some kinds of access mechanism are immune to size of domain. We even have an example of this in language: the lexicon has a huge number of entries but access is rapid, and certainly there is no evidence that people with large vocabularies are slower at word finding than people with modest vocabularies.

What we have here, on both sides of the debate, is a lot of speculation. In the absence of hard facts about how the human brain DOES work, we exchange hunches about how it WOULD work if it were well-designed. Speculation of this kind is not without its value in suggesting worthwhile hypotheses to focus attention on, but of course it can't be decisive. Despite our own prejudices, we admit that there are just too many different ways in which a system could be efficient. Though some formal efficiency proofs have been developed, it is not really known at present how to compare these kinds of efficiency with each other or with the observed efficiency of the human brain. It would be highly satisfactory, therefore, if we had some experimental psycholinguistic results that would allow us to cut through all the pros and cons of the abstract debate. And this is what Freedman and Forster are offering. They appear to have evidence in favor of a distinction between rules and constraints within the syntactic component of the mental grammar, suggesting that division of labor rather than uniformity of operations is what the human language faculty values. We will argue here that F&F's data are open to an alternative interpretation, one which is perfectly compatible with a nonmodular parser which employs phrase structure rules only. We turn now to the details of F&F's experiments.

The sentence matching task that F&F used is as follows. One sentence appears on a screen. Two seconds later another one appears beneath it. The second sentence is either identical to the first, or else differs from it in one word. The subject has to press a button to say whether the second sentence is the same or different from the first, and the response time is recorded. In all cases of interest the sentences do match, so the correct answer is "yes". The mismatch trials are included just to make sure the subject is attending. (Mismatched words are identical in length, so that subjects cannot get by on a superficial visual comparison of the two sentences.)

F&F's experiments 1, 2, and 4 tested four kinds of ungrammatical sentence, and compared their matching times with those of control sentences that were similar but well-formed. The four types of ungrammaticality and their controls are illustrated in (1) - (4) (F&F's examples). One type had incorrect subject - verb agreement, as in (1A); one type had a quantifier misplaced, as in (2A); one type violated the Specified Subject Constraint, as in (3A); and one type violated Subjacency, as in (4A). (For comments on these linguistic characterizations of the examples, see below.)

(1) A. *Mary were writing a letter to her husband.
    B. Mary was writing a letter to her husband.

(2) A. *Lesley's parents are chemical engineers both.
    B. Lesley's parents are both chemical engineers.

(3) A. *Who did the duchess sell Turner's portrait of?
    B. Who did the duchess sell a portrait of?
    C. The duchess sold Turner's portrait of her father.
    D. The duchess sold a portrait of her father.

(4) A. *Who do the police believe the claim that John shot?
    B. Who do the police believe that John shot?
    C. The police believe the claim that John shot someone.
    D. The police believe that John shot someone.

Results from these experiments are shown in Table I (from Freedman and Forster). We have calculated from F&F's figures the portion of the matching time for ungrammatical sentences that can be attributed to the presence of the ungrammaticality itself.

TABLE I
Matching Times in msec by Sentence Type

Number agreement	42 (p<.05)
Quantifier placement	66 (p<.025)
Subjacency	00
Specified Subject Constraint	-31 (n.s.)

For the number agreement and quantifier placement examples, this is derived simply by subtracting from the matching time for the ungrammatical sentence the matching time for its grammatical control. (Note that the A and B versions of a sentence were NOT presented to the subject for comparison against each other; each was matched against itself.) For the Specified Subject Constraint and Subjacency examples, calculation of the matching time increment due to the ungrammaticality is more complex, since differences of lexical content must be controlled for. Comparison of matching times for the A and B versions of these sentences reveal the effect of ungrammaticality plus the effect of the additional words in the A version. The latter can be estimated by subtraction of match times for the C and D versions. To isolate the effect of the ungrammaticality, we therefore calculate ((A-B) - (C-D)).

It is clear from Table I that the matching times show no effect of ungrammaticality due to violation of the Specified Subject Constraint or Subjacency, while there is a significant effect of the ungrammaticality due to improper number agreement and quantifier placement. F&F's interpretation of this difference involves two claims: (i) that examples like (3A) and (4A), but not examples like (1A) and (2A), are "overgenerated", i.e., are assigned certain well-formed syntactic representations by the grammar even though they do not have complete derivations; and (ii) that it is this difference between the examples that is responsible for the difference in matching times. We will argue that the first of these claims is dubious and the second is false.

In an earlier discussion of these results, Freedman (1982) proposed a strong version of the overgeneration explanation, maintaining that a sentence is easy to match as long as it has SOME well-formed syntactic representation. This will account for the experimental results just in case it is also assumed that the number agreement and quantifier placement violations have no syntactic derivations at all, not even a D-structure. But there are problems here, afflicting both kinds of example. First, number agreement cannot be a D-structure phenomenon in a transformational system, such as the GB system that F&F endorse, because agreement is sensitive to the effects of NP-Movement. Therefore agreement must be established either by a late-applying transformation, or by a late-applying filter that discards sentences in which random agreement has led to incorrect results. Is it the case, on either or both of these approaches, that an ungrammatical sentence like *Mary were writing a letter* has a D-structure? The answer is really not clear. Presumably it does have a D-structure on the latter analysis, but perhaps not on the former (since prior to the application of the agreement transformation the verb would have no number feature at all -- but then, does no number feature imply a morphological plural?). The validity of Freedman's hypothesis is therefore highly dependent on the details of the correct linguistic analysis, which we simply don't know at present. In fact, the evidential situation is even worse than this, because of an obscurity in the very idea of a grammar assigning a derivation (even a partial one) to an ungrammatical word string.

Intricacies abound here. We can cut through some of them by narrowing our sights in keeping with the parsing focus of this investigation. Instead of clarification about what a GRAMMAR might assign to an ungrammatical sentence, we need only an account of what the PARSER assigns it. In other words, what matters is whether an ungrammatical sentence is PERCEIVED (consciously or unconsciously) as having a certain incomplete derivation. In many cases this partial derivation would not terminate in a phonetic (orthographic) representation of the ungrammatical sentence.[7] Therefore subjects who detect such derivational possibilities must be engaging in some kind of interpolation, i.e., continuing the derivation through from where it blocked to a surface level representation which does match the input. The interpolated steps will be steps that the grammar does not license. How does the subject compute them? He might continue the incomplete derivation by ignoring a constraint that is in his grammar, or by supplying a rule that is not in his grammar, or by extending the application of one that is.[8] The existence of alternatives like these is very clear in the case of F&F's quantifier placement examples. The ungrammatical sentence (2A) might be assigned a derivation either by ignoring a restriction on the Quantifier Float transformation, or by inventing a phrase structure rule introducing quantifiers in clause-final (or VP-final or NP-final) position. Since Freedman's hypothesis requires (2A) to have no derivation at all, it would have to be claimed that neither of these moves is available to the parser. But if that is the case, what moves are available to the parser for concocting a pseudo-derivation for sentences like (3A) and (4A)?

The only suggestion that seems to work is that (i) the parser faced with an ungrammatical sentence never invents a rule but can set aside constraints, and (ii) a constraint can be set aside only if it is an independent statement in the grammar, i.e., not a contextual restriction embodied within a rule, but something general enough to have been factored out of individual rules and stated separately. On these assumptions, we could account for the pattern of results in F&F's experiments if we could argue that Subjacency and the Specified Subject Constraint are constraints of the latter kind, while number agreement and quantifier placement involve only rules, possibly with some restrictions specifically built into them.

F&F apparently are not confident that all of the premises needed for this earlier explanation could be satisfactorily defended, for they have now suggested a modified hypothesis, which is that successful sentence matching presupposes a well-formed S-structure. (Note that for sentences with well-formed S-structures the syntactic interpolation process is unnecessary; only an LF would have to be concocted.) This modified hypothesis succeeds in excluding the number agreement examples and the quantifier examples on any analysis except one involving a filter over S-structures or an incoherence at LF. However, this revised proposal creates at least as many problems as it solves. In order that sentences like (4A) should have partial derivations sufficient for sentence matching, Subjacency would now have to be a filter over S-structures or a condition on the derivation of LF. Though it may indeed be the latter, it is standardly assumed to apply also to movement operations prior to S-structure. But in that case (4A) will have no more of an S-structure than (1A) and (2A). The status of the Specified Subject Constraint is uncertain in current versions of transformational theory. If it is absorbed into the binding theory (e.g., in the definition of SUBJECT) then it would also qualify as a condition on representations at S-structure, LF or both, but it would not qualify as a condition on the derivation of S-structures.[9]

Our conclusion from this discussion of the linguistic status of F&F's experimental sentences is that a great deal more work will be necessary to establish any natural grouping of them in terms of whether the grammar provides (or the parser can concoct) representations for them at some significant derivational level. Having pointed out the problems, however, we will now adopt the working assumption that some such distinction as F&F have in mind IS linguistically motivated. Our main purpose in this paper is to show that, even on this assumption, F&F's data do not speak to the issue of whether grammars contain constraints.

In what follows we will refer to examples like (1A) and (2A) as _rule violations_ and to examples like (3A) and (4A) as _constraint violations_. These terms are intended as convenient labels in the spirit of F&F's explanation, and we will continue to use them even when we discuss the data in the framework of a phrase structure grammar with no constraints. Our main contention is that this rule-versus-constraint contrast is confounded in F&F's experiments with another factor, which we will call _correctability_, and that it is correctability, not the rule/constraint difference, that is the true source of the observed differences in matching times.

We begin by taking a broader perspective on the matching task. Prior to this latest series of experiments, it had been established that matching times are higher for implausible sentences (e.g., _Susan stroked the yellow bones quickly_) than for sensible ones (e.g., _The young horse galloped away_), and higher still for random word sequences (e.g., _Had heel finite replies Mary the_). These results, due to Murray, are reported by Forster (1979). There is presumably no question of accounting for them in terms of rules versus constraints or the nonexistence of D-structures or S-structures, for implausibility is generally not assumed these days to be a grammatical property. Thus the _yellow bones_ sentence, despite being difficult to match, would have a complete and well-formed derivation.

What these early results suggest instead is the quite obvious idea that a sentence is easy to match to the extent that it is easy to remember, and that it is easy to remember to the extent that it contains predictable word combinations. If we forget the difference between them for a moment, the rule violations and the constraint violations of F&F's experiments fit quite nicely into this general picture. These sentences are certainly ungrammatical (both F&F's subjects and our subjects rated them so), but they do have a considerable amount of coherent structure that could facilitate short term memory, and they do have identifiable meanings which are not distractingly bizarre. So we would expect them to have matching times which differed little or not at all (depending on the sensitivity of the matching task) from the times for fully grammatical sensible sentences. And this is so. F&F's experiment 3 replicated the earlier result for random word strings, with matching times 348 msec longer than for grammatical control sentences. By comparison, the 54 msec mean increment for the rule violations, and of course the null increment for constraint violations, is very small.

What doesn't follow from this very general notion of the syntactic and semantic coherence of a word string is the small but significant difference between the rule violations and the constraint violations, for we probably don't want to claim that the rule violations are significantly less coherent, either syntactically or semantically, than the constraint violations. We had subjects judge sentences of both kinds for grammaticality, and though the constraint violations were rated a little closer to fully acceptable than the rule violations, the difference was minute and nonsignificant.

There are two possible lines of attack on the problem. One line would be that there is some special factor at work in the case of the constraint violations, which makes them easier to match than would be expected given that they are not fully grammatical. This is the line that F&F have taken, with the special factor being the existence of a partial derivation. In exactly the opposite vein, one might claim that there is some special factor at work in the case of the RULE violations, which makes them HARDER to match than would be expected, assuming that they are syntactically and semantically coherent enough to be held in short term memory long enough to be matched. This is the line that we will pursue, with the special factor being the high correctability of the rule violation examples.[10]

Just intuitively, it does seem that the rule violation examples are very easy to correct. In fact, mentally correcting them when one hears them is almost irresistible. Now consider the situation in the experiment. A subject reads <u>Mary</u> <u>were</u> <u>writing</u> <u>a</u> <u>letter</u> <u>to</u> <u>her</u> <u>husband</u>. His parsing routine, designed for normal conversation peppered with speech errors, recognizes what the sentence 'should have been' and constructs a phrase marker for it, throwing away the information about the faulty number agreement -- that kind of information is of no use to anyone except psycholinguists who collect slips of the tongue for a living. But this means that when the second sentence appears on the screen, it will seem NOT to match the corrected first sentence in the subject's head. To give the right response, the subject has to take the trouble to distinguish between what he actually saw and what he mentally corrected it to, and sorting out this confusion could very well cost him 42 msec. Notice that this explanation is exactly comparable to the standard explanation of why, in a lexical decision task, a non-word such as <u>stodent</u> takes longer to reject than a non-word like <u>tordil</u>; <u>stodent</u> is similar to, hence confusable with, a legitimate word, i.e., confusable with an item for which the correct response would be the exact opposite of the one that the subject should actually give (See Forster, 1976).

In contrast with the rule violations, the constraint violations have the curious property that there is NO obvious way to correct what is wrong with them; on the one hand it's perfectly clear what message they express, and on the other hand there is no other, better way of expressing it. So a subject would just have to accept such a sentence as it stands. This means he would have to store an ungrammatical sentence (a defective phrase marker) in short term memory, which might not be ideal, but he would be better off in another respect because he wouldn't have to cope with a confusion about what to match with what. Our hypothesis is that confusion due to the automatic mental correction of ungrammaticalities causes a delay in sentence matching; when there is no such confusion, an ungrammatical sentence is more or less as easy to match as a grammatical one, as long as it is comparably coherent syntactically and semantically. Our own experiments were designed to reveal the effect of correctability, and to show that when this is disentangled from the rule/constraint distinction, the apparent effect of the latter melts away.

We had subjects actually correct rule violations and constraint violations modelled on F&F's examples, and rate how difficult they found it to do so. These are conscious judgements, but subjects seemed able to distinguish between the cases in which 'the right answer' was immediately apparent, and the cases where they had to think and edit, so we think that these ratings do provide at least a rough and ready measure of the spontaneity with which these errors would be corrected when correcting them is not presented as the required task. As we had expected, the rule violations were very much more easily corrected than the constraint violations. The mean rating for the rule violations was 1.20 on a five-point scale in which 1.00 represents the greatest ease of correction. The mean rating for the constraint violations was 3.06. For convenience in what follows we will adopt the terms <u>correctable</u> <u>rule</u> <u>violations</u> and <u>noncorrectable</u> <u>constraint</u> <u>violations</u> for these two types of sentences.

We then crossed the factors of correctability and rule/constraint, so that we would be able to determine which one was affecting matching times. To do this we had to construct two more sets of ungrammatical sentences; we had to add what we will call _correctable constraint violations_ and _noncorrectable rule violations_. The mean correctability rating for the noncorrectable rule violations was 2.74, which is close to that for the noncorrectable constraint violations, as we had intended. The mean correctability rating for the correctable constraint violations was 1.78, which is somewhat higher than for the correctable rule violations but still well below that of either of the noncorrectable types.

As the previous discussion makes clear, it is important to consider the linguistic properties of the experimental sentences. We constructed 24 sentences of each of the four types. With the exception of the noncorrectable rule violations, which we will discuss shortly, there were six subtypes of each type, so that we could be sure that the results were not due to the idiosyncrasies of any one kind of sentence. We attempted to keep our examples safely within the bounds of the intuitive distinction between rule and constraint violations presupposed by F&F. Our noncorrectable constraint violations included Subjacency and SSC violations modelled on theirs, and we added extractions from sentential subjects, from relative clauses, from complex NPs in subject position, and from adverbial clauses. Of course all of these could be argued to be violations of the same constraint, e.g., Subjacency. But at least there is variety in the surface manifestation of the constraint, so that they should not all be open to the same potential garden paths or other parsing problems irrelevant to our present concerns. Our correctable constraint violations included omission of a complementizer on a sentence-initial subordinate clause, subject-aux inversion of a clausal subject, doubly filled COMP, violation of the Fixed Subject Constraint (or for-to filter and that-e filter, or ECP, etc.), and extraction of a conjoined NP where grammaticality could be restored by substituting _with_ for _and_. Our correctable rule violations included agreement and quantifier placement illegalities modelled on F&F's, together with reversals of a determiner with its noun, reversals of a copula and a predicate adjective, substitution of grammatical morphemes such as one preposition for another, and incorrectly case-marked pronouns. The case marking examples are open to the same criticism as F&F's number agreement examples, since case assignment must follow at least NP movement. Apart from this unfortunate choice, our rule violation examples were intended to be perceived by subjects as generated by incorrect base rules, so that F&F would predict high matching times on either version of the overgeneration hypothesis, i.e., regardless of whether a D-structure or an S-structure is assumed to be essential for rapid matching.

There were no subtypes within the noncorrectable rule violations; they were all one-of-a-kind examples. This was because of the extreme difficulty of devising examples of any kind. These sentences had to be matched to the other types in degree of ungrammaticality, they had to violate base rules rather than constraints on transformations or on later derivational representations, they had to be difficult to correct, and on top of all this they had to be comparable to the other sentence types in syntactic and semantic coherence since, as we have noted, this is a powerful factor in determining matching times. The latter requirement is not easy to satisfy, since compositional semantic principles do not apply in any straightforward way to a sentence which has no well-formed D-structure, nor even (since it must be noncorrectable) a D-structure that is a trivial distortion of a well-formed one. Just how a semantic rule should combine the meanings of two constituents depends crucially on the syntactic relationship between them, and this kind of relationship is precisely what is destroyed in a noncorrectable rule violation sentence.

The noncorrectable rule violations, like all of our ungrammatical sentences, also had to be such that we could construct appropriate grammatical control sentences as illustrated above for F&F's examples. Control sets for two of the noncorrectable rule violations are shown in (5) and (6).

(5) A. *The plumber took what the time was to finish the job.
    B. The plumber took the time to finish the job.
    C. The plumber who finished the job asked me what the time was.
    D. The plumber who finished the job asked me the time.

(6) A. *What could possibly show to be part of a bicycle?
    B. What could possibly show that it's part of a bicycle?
    C. It seems very likely to be part of a bicycle.
    D. It seems very likely that it's part of a bicycle.

All 96 ungrammatical sentences were rated by subjects for ungrammaticality (N=24) and for correctability (N=42), and then they and their grammatical controls were used in a sentence matching experiment. The mean matching times for the four types of ungrammatical sentences, relative again to the times for the grammatical controls, are shown in Table II. The results for the correctable rule violations and the noncorrectable constraint violations replicate F&F's findings. There was a significant matching time increment for the former but a nonsignificant decrement for the latter.[11]

TABLE II
Matching Time (msec): Rule vs. Constraint Violations

	Correctable	Noncorrectable	Mean
Rule violations	176 ($p<.001$)	111 ($p<.05$)	144
Constraint violations	254 ($p<.001$)	-17 (n.s.)	119
Mean	215	47	

Our data do not, however, confirm F&F's EXPLANATION for the difference between the correctable rule violations and the noncorrectable constraint violations, for there was no effect at all of the rule-versus-constraint distinction once it was unconfounded, in our four-way design, from correctability. The mean increment for rule violations was 25 msec greater than for constraint violations, but this difference is far from significance. There was, however, a highly significant effect of correctability; the mean increment for correctable sentences was 168 msec greater than for noncorrectable sentences ($p<.001$). This effect is also revealed in the existence of a small but significant negative correlation across all sentences between correctability and matching time ($r = -.31$; $p<.005$). Note that the correlation is negative because highly correctable sentences have low scores on the correctability scale. The data show that highly correctable sentences are the most difficult to match. Most noticeably, the correctable constraint violations were harder to match than any of the other three types, though F&F's explanation (either version) predicts that they should be just as easy as the noncorrectable constraint violations.[12] (A more detailed analysis of these data is presented in Crain & Fodor, in preparation).

Unfortunately, we cannot rest our case here for the artifactuality of F&F's results. F&F have responded to an earlier report on the results of pilot experiments similar to these (Crain & Fodor, 1984) by arguing that correctability couldn't be the source of the matching time differences. Their evidence is the existence of a significant matching effect for sentences that would count, according to our classification, as noncorrectable rule violations. Their experiments 3, 5 and 6 tested three different kinds of such sentences, illustrated in (7) - (9).

(7) *His wife politely the guest introduced.
*What to expect the judge told the thief.

(8) *Tony was granted bail that his plan failed.
*The spy admitted by the judge.

(9) *The girl behind you the subsequent discussion.
*Reminds me of your sister soon got boring.

For convenience we will refer to all three types as phrase structure scrambles (though F&F reserved this term for the first type only). Sentences of the first type were constructed by reversing the first and second half of a grammatical sentence, dividing it either after the verb or before a subordinate clause. (Thus the sources of the examples in (7) were respectively The guest introduced his wife politely and The judge told the thief what to expect.) Sentences of the second type were constructed by exchanging the first half of one sentence with the first half of another, with the split at some phrase boundary (not otherwise specified). (Thus the sources of the examples in (8) were The spy admitted that his plan failed and Tony was granted bail by the judge.) Sentences of the third type were constructed by exchanging the subject NP of one sentence with the matrix VP of another. (Thus the sources of the sentences shown in (9) were The girl behind you reminds me of your sister and The subsequent discussion soon got boring.)

The matching time increments for these ungrammatical sentences were 125 msec, 110 msec and 123 msec respectively (all significant). Note that these effects are larger than F&F obtained for agreement and quantifier placement violations. (Comparisons across experiments probably are legitimate in this case since procedures, subjects, etc., seem to have been comparable.) Recasting this finding in our terms, it appears that noncorrectable rule violations were HARDER to match than correctable rule violations, in contradiction to our hypothesis that correctability was the source of the effect for the correctable rule violations. We believe, however, that this discrepancy can be satisfactorily accounted for.

Note first that our data agree with F&F's to the extent of showing a significant matching time increment for the noncorrectable rule violations, though in this case it is smaller than for the correctable rule violations. Either the so-called noncorrectable rule violations aren't noncorrectable enough, or else correctability is not the ONLY factor that can interfere with the sentence matching task. We think that both are true.

Let's consider first the possibility that correctability is not the only factor affecting matching times. This is a point we have admitted all along, for high correctability is quite patently not the reason why random word strings are so difficult to match. The other factor we have identified is lack of coherence, making retention in short term memory difficult. We suggest that F&F's phrase structure scrambles and our own noncorrectable rule violations were semantically less coherent than the other three sentence types we tested. They were not rated as significantly more ungrammatical; their mean grammaticality rating was 2.77 on a 3-point scale on which 1.00 denotes a perfectly grammatical sentence, with the grammaticality ratings of the other three sentence types ranging from 2.58 to 2.73. Nevertheless, the kind of ungrammaticality involved, as we have already observed, is the kind that interferes with the normal application of semantic composition rules and leaves unclear just what semantic relations are intended to hold among the parts of the sentence. We haven't attempted to quantify this lack of semantic coherence, but it seems quite evident from inspection of the examples presented. Furthermore we would submit that the reason that F&F's phrase structure scrambles showed a larger matching effect (relatively speaking) than our noncorrectable rule violations is that their examples were more severely deficient in this respect than ours were. The F&F examples were constructed by a mechanical procedure without any concern for preserving sense in the outcome. As an extreme case, consider the example <u>Soon</u> <u>became</u> <u>very</u> <u>rusty</u> <u>came</u> <u>third</u> <u>in</u> <u>the</u> <u>race</u> (an instance of the third type characterized above), in which there is no continuity of content at all. In constructing our examples we aimed, at least, for some kind of perceived coherence, though we admit to not always having succeeded as well as we would have liked.[13]

There may also have been some failure to eliminate spontaneous correction of some of these examples. We have noted that the correctability scores for our noncorrectable rule violations were a little lower (i.e., towards the more correctable end of the scale) than for our noncorrectable constraint violations. F&F did not have their experimental sentences rated for correctability at all, and though they claim that these sentences are not correctable, it is hard to agree with their judgement in all cases. It is true that subjects couldn't have been aware of how F&F had derived these sentences, so they could not have corrected them back to their actual sources. But they could certainly have corrected some of them in other ways. For instance, <u>His</u> <u>wife</u> <u>politely</u> <u>the</u> <u>guest</u> <u>introduced</u> could have been perceived as an error for <u>His</u> <u>wife</u> <u>politely</u> <u>introduced</u> <u>the</u> <u>guest</u> (which was not its source), and <u>The</u> <u>spy</u> <u>admitted</u> <u>by</u> <u>the</u> <u>judge</u> could have been perceived as a slip of the tongue for <u>The</u> <u>spy</u> <u>was</u> <u>admitted</u> <u>by</u> <u>the</u> <u>judge</u>. We had subjects rate for correctability the examples of phrase structure scrambles that were cited in F&F's paper (unfortunately only 17 sentences out of a total of 96), and found that they were indeed judged to be slightly more correctable than our noncorrectable rule violations. Correctability, then, could have pushed up the matching times for these sentences.[14]

The conclusion to be drawn from this discussion is that the sentence matching task is far from being a special source of information about grammatical derivations, for it is sensitive to too many factors to provide clear information about any one of them. The earlier results leave no doubt about the importance of syntactic and semantic structure for sentence matching, as would be expected for any task involving short term memory for linguistic material. The data we have reported, particularly the striking contrast between the correctable and noncorrectable constraint violations, strongly indicate the importance of confusability effects, as would be expected on the basis of other tasks involving ill-formed stimuli that closely resemble well-formed models.[15] The mixed results for the phrase structure scrambles and noncorrectable rule violations are then explicable as casualties of the near impossibility of holding one of these factors constant while varying the other. The sentences that violate the 'major' constraints such as Subjacency do emerge from these studies as special, but there is no reason to think that this is because they are 'overgenerated' by a transformational grammar. Quite possibly all that is special about them is the rather unexpected conjunction of properties that they have no obvious grammatical counterparts and yet their meanings are readily recognizable. If we are right, this makes these sentences very special with respect to the matching task, but it does not entail that they have any special status with respect to the grammar the parser employs.

To summarize: What F&F took to be an interesting psychological reflex of a rule-versus-constraint contrast in the competence grammar is apparently not that but rather an artifact of correctability. Correctability is a property of sentences that probably has only a very indirect connection with the competence grammar and certainly does not provide support for grammars containing constraints as opposed to grammars containing only rules. Thus we find ourselves back where we began, with no clear psycholinguistic evidence either way about the shape of the competence grammar, or more generally about the extent of modularity within the human language faculty. That is disappointing in a way, but it is at least better than thinking there is evidence when there is none.

FOOTNOTES

[1]More precisely, we call a statement in a grammar a constraint if deleting it from the grammar would have the effect of increasing the number of derivations licensed by the grammar, and we call a statement a rule if deleting it from the grammar would have the effect of decreasing the number of derivations licensed by the grammar. These definitions capture fairly well, we believe, what linguists have in mind when they talk informally of rules and constraints. Even they have some loose edges, however. For instance, it makes a difference, as will become clear below, what we count as a single statement in a grammar.

[2]Note that we are not distinguishing here between those parts of a grammar contributed by UG and those which are language-specific. For some purposes, such as a theory of language acquisition, this difference is obviously crucial, but from the point of view of sentence processing it is probably not significant (given the unlikelihood, for example, that all and only universal constraints are the consequences of limitations of the parsing routines and hence don't need to be explicitly imposed).

[3]This point is no longer quite so clear as it was when the research reported in this paper was begun. Gazdar, Klein, Pullum and Sag (in press) present a different version of the theory than in earlier works such as Gazdar (1981). In the earlier theory, a metagrammar generated phrase structure rules which generated phrase markers for sentences. In the newer theory, various principles of the metagrammar interact in the generation of phrase markers. There are different ways of formulating the algorithm which defines the set of well-formed phrase markers; on some formulations, some of the metagrammatical principles (e.g., feature co-occurrence restrictions) could be regarded as constraints filtering the effects of others (e.g., feature percolation principles). In the competence grammar this filtering can and should be interleaved with the tree building operations, i.e., it should apply node by node as a phrase marker is being constructed. However, a slightly indirect implementation of the grammar would permit a complete phrase marker to be built before it is checked against a constraint and possibly discarded. This would reduce the differences between the predictions about processing derived from GPSG and those derived from a transformational theory of the competence grammar, such as Government Binding theory. In an attempt to keep the issues sharply defined, we will presuppose here a version of GPSG theory on which there are no explicit constraints at all, even of this minimal variety. We assume that no such construal of GB is possible; the only kind of transformational grammar that could survive without explicit constraints would be one whose transformational rules each contained highly specific contextual specifications.

[4]Such a parser could, of course, assign the wrong phrase marker to an initial fragment of a sentence if it were ambiguous. But that is true of ANY parser that doesn't wait for full disambiguation before computing an analysis, and it is quite different from computing an analysis that is not licensed by the grammar.

[5]Stabler himself appears to find the notion of a competence grammar suspect, given that its properties are likely to differ so radically from those of an optimal performance grammar. We believe there are good grounds for having faith in the psychological reality of a competence grammar, but we will note here only that GPSG theory may resolve the conflict that Stabler perceives between performance and competence. See below.

[6]This is quite different from the approach of Shieber (1984), though also within a GPSG framework. Shieber reduces the size of the rule set by modularizing the grammar. His parser makes direct use of the ID and LP rules of recent versions of GPSG (see Gazdar, Klein, Pullum and Sag, in press), without precomputing the set of linearized phrase structure rules that they jointly entail. Shieber suggests that this may improve efficiency, at least for languages with few LP constraints; however, the advantages may be tied to the Earley-type algorithm that he adopts, which makes grammar size a more significant factor than the amount of on-line computation of rule interactions.

[7]For example, a constraint on representations (i.e., on rule outputs) allows an incorrect word string to be generated and then marks it as incorrect, but a constraint on rule application prevents the formation of the incorrect word string; the only string generated would be the one PRIOR to an illicit movement, deletion, etc. As a slightly more complicated case: a word string appearing at the point at which a derivation is terminated by a constraint on representations might need to undergo further transformations in order to match some given ungrammatical sentence. Assuming that F&F's intuitive notion of "overgeneration" is intended to include cases like these, it cannot be simply defined as actual generation at some non-final stage of derivation. As we will show, it seems that it must involve ignoring or violating constraints.

[8]For a grammar containing a maximally general rule such as <u>Affect alpha</u>, discarding of enough constraints would lead to an infinite number of pseudo-derivations for any word string. It would also result in generation of all strings over the vocabulary, thus permitting no distinctions of the kind that F&F wish to draw. This is why it is useful to limit attention to pseudo-derivations that actually occur to perceivers. We assume that these are relatively minor modifications of legitimate derivations.

[9]In addition to (4A), F&F present another sentence intended to illustrate violation of Subjacency. This is <u>Ann is hard to trust rumors about</u>. The transformation here is Tough Movement rather than WH-Movement (if it is proper to draw such a distinction in a system in which both are instances of <u>Move alpha</u>), but this example, designated as ungrammatical, seems otherwise to be exactly comparable to the sentence which we have given above as (3B) and which F&F employed as a grammatical control sentence for the Specified Subject Constraint. There has been some uncertainty in the linguistics literature as to whether or not Subjacency does or should block extraction from a PP modifier to a noun, but the fact that this confusion was not resolved consistently in one direction or the other in the experimental materials cannot help but make it more difficult to interpret the data.

[10]Note that our explanation does not entail that either the rule violations or the constraint violations can be remembered exactly as well as fully grammatical sentences, or that they will be as easy as grammatical sentences in other experimental tasks. We need to assume only that memory for them is good enough to give rise to no detectable performance decrement in this matching task.

[11]Note that the absolute response times in the two studies cannot legitimately be compared, since subjects, procedures and materials differed in a variety of minor ways. In general, matching times were higher, both absolutely and relative to controls, in our experiment than in F&F's.

[12]Most of the correctable constraint violations involved what are often regarded as filters over S-structure. It is unfortunate that we weren't able, in our attempt to disentangle factors, to separate correctability from the level of application of the constraints. But even if it could be argued that there was some such linguistic difference between the two sentence types, it would be a difference in the wrong direction for the overgeneration hypothesis, since the correctable constraint violations would have more nearly complete partial derivations than the noncorrectable constraint violations and so they should, if anything, have been easier to match. Incidentally, we should also confess that our correctable/noncorrectable distinction appears to be confounded with a contrast between local and global ungrammaticality; the correctable violations can be pinned to a single word or two in the sentence, while the noncorrectable violations typically cannot. But even if this local/global distinction was operative instead of -- or more likely as a contributor to -- what we have called correctability, this does not change the evidence against the importance of the rule-versus-constraint factor.

[13]To some extent the grammatical control sentences in our study permitted the factoring out of effects of semantic or pragmatic unexpectedness, just as in the case of the constraint violation examples, since content is at least roughly constant across all four sentences in a comparison set. By contrast, F&F did not use a four-sentence comparison set for their phrase structure scrambles. Their control sentences were just the source sentences from which the ungrammatical sentences had been constructed, and these were clearly more sensible, as well as more grammatical, than the ungrammatical sentences were. This is probably one of the reasons why their phrase structure scrambles were (comparatively again) more difficult to match than our noncorrectable rule violations.

[14]Interestingly, the F&F sentences that our subjects didn't rate as correctable seem to be the ones on which they couldn't impose a coherent meaning, such as the rusty sentence cited above. This illustrates the difficulty of designing rule violation sentences that simultaneously avoid both threats to efficient sentence matching.

[15]One other aspect of our data for which it would be satisfactory to have an explanation is the fact that the correctable constraint violations were even harder to match than the correctable rule violations, despite being slightly less correctable. We suspect that the correctable constraint violations were more similar to the kinds of errors that occur in normal discourse, and therefore more susceptible to spontaneous correction by whatever mechanisms 'clean up' the input under normal conditions; rule violations such as inversion of determiner and noun seem more artificial and hence, perhaps, are more resistant. We have not attempted to test this hypothesis, but the question of what comes naturally to the error-correction mechanism might prove to be of some interest in its own right.

REFERENCES

Barton, G. E. (1984) Toward a Principle-Based Parser. MIT Artificial Intelligence Laboratory, Memo No. 788.

Berwick, R. C. and A. S. Weinberg. (1984) The Grammatical Basis of Linguistic Performance: Language Use and Acquisition. Cambridge, Mass: The MIT Press.

Crain, S. and J. D. Fodor (in preparation) Overgenerated sentences and the matching task. Submitted to Cognition.

Crain, S. and J. D. Fodor (1984) How can grammars help parsers? In D.R. Dowty, L. Karttunen, and A. Zwicky (eds.), Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives. Cambridge, England: Cambridge University Press.

Fodor, J. A. (1983) The Modularity of Mind: An Essay on Faculty Psychology. Cambridge, Mass: Bradford Books.

Fodor, J. A., T. G. Bever and M. F. Garrett (1974) The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw-Hill Book Company.

Fodor, J. D. (1983) Phrase structure parsing and the island constraints. Linguistics and Philosophy, 6, 163-223.

Forster, K. I. (1976) Accessing the mental lexicon. In R.J. Wales and E. Walker (eds.), New Approaches to Language Mechanisms. Amsterdam: North-Holland Publishing Company.

Forster, K. I. (1979) Levels of processing and the structure of the language processor. In W.E. Cooper and E.C.T. Walker (eds.), Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Frazier, L., C. Clifton and J. Randall (1983) Filling gaps: Decision principles and structure in sentence comprehension. Cognition, 13, 187-222.

Freedman, S. A. (1982) Behavioral Reflexes of Constraints on Transformations. Unpublished doctoral dissertation, Monash University.

Freedman, S. A. and K. I. Forster (in press) The psychological status of overgenerated sentences. To appear in Cognition.

Gazdar, G. (1981) Unbounded dependencies and coordinate structure. Linguistic Inquiry, 12, 155-184.

Gazdar, G., E. Klein, G. K. Pullum and I. A. Sag (in press) Generalized Phrase Structure Grammar. Cambridge, Mass: Harvard University Press.

Marcus, M. P. (1980) A Theory of Syntactic Recognition for Natural Language. Cambridge, Mass: The MIT Press.

Shieber, S. (1984) Direct parsing of ID/LP grammars, Linguistics and Philosophy, 7, 135 - 154.

Stabler, E. (1984) Berwick and Weinberg on linguistics and computational psychology. Cognition, 17, 155-179.

Thompson, H. S. (1981) Chart parsing and rule schemata in GPSG. In Proceedings of the Nineteenth Annual Meeting of the Association for Computational Linguistics, Stanford University.