Cognition, 19 (1985) 193–204 **Discussion**

The psychological relevance of transformational grammar: A reply to Stabler

ROBERT C. BERWICK*
AMY S. WEINBERG
Massachusetts Institute of Technology
University of Maryland

In a recent reply to our article in *Cognition* (1984), E. Stabler criticized us on two main points: our construal of transformational grammar (TG) as a "first level" theory; and our claim that a first level construal of TG is crucial, for its psychological relevance. We'd like to dispute both of these points, and in so doing re-emphasize the psychological relevance of modern transformational grammar.

First, some basic housekeeping: we must clear up the record by explaining our original position, which, contra Stabler's interpretation, does not commit us to viewing a grammar as a first level approximation of a parser. Second, we outline Stabler's reasons for thinking that grammars *should* be first level theories of this kind. We continue by examining Stabler's claim that linguists judge their theories by the inherently *nonpsychological* criterion of relative simplicity. We argue that the psychological notion of learnability is actually the touchstone of comparison. In fact, given this criterion we shall see that one can understand just why linguistic grammars should form the abstract foundation of psychological parsing models.

1. A first level theory?

Stabler takes his demonstration that "it is clear that there are some difficulties in interpreting TG as a first level theory" to be a criticism of our own approach.[1] We are puzzled by this interpretation because our own article makes many of the same points (though it is true that many of these appear in footnotes in the original article). For example, we note that a parsing algorithm could be seen as computing "the extension of the function specified

*Reprint requests should be sent to Robert C. Berwick, NE43814, M.I.T., Cambridge, MA 02139, U.S.A.

[1]Recall that a first level theory merely specifies the function computed by the parser, pairing the same input and output strings (surface and deep structures) as generated by a TG. A first level theory makes no claims whether the grammar and parser use the same algorithm.

0010-0277/85/$4.10

by the grammar." As footnotes 2 and 3 of our original article point out, the possibilities of ambiguity, ability to parse ungrammatical sentences, and human failures on grammatical sentences mean that one cannot say directly that the parser and grammar compute *exactly* the same input–output pairs. That much is obvious, but to carry this observation any deeper is to give too much importance to what is really a quibble.

Let us put aside Stabler's remark about the coextension of grammar and parser then and turn to the main aims of our article. These were two: first, to show, contrary to some popular opinion, that a parsing algorithm mirroring rather directly the rules and representations of transformational grammar could predict DTC results correctly. Since this has sometimes been considered a stumbling block for TG (see, e.g., Fodor, Bever and Garrett, 1974; or Bresnan, 1978, 1982), this was an important demonstration. Note that Stabler's worries about an *exact* grammar–parser input/output correspondence are at right angles to this demonstration. We presented an algorithm mapping from surface to D-structure and another mapping from surface to S-structure. If desired, we could output both levels of representation by running both algorithms simultaneously. The algorithm for computing the mapping function bore a remarkable similarity to the rules of a TG. Second, we tried to show that even if a direct mapping is untenable then researchers interested in constructing a theory of language use should still be interested in the theory of grammar to the degree that one can use that theory to constrain the class of possible parsers (Berwick and Weinberg, 1983, p. 46). Here we drew several possible pictures of the grammar–parser relationship, ranging from the 'covering' relationship to a simple input–output (extensional) equivalence (a first level theory). We were certainly not wedded to any of the models presented, since we provided no evidence in the *Cognition* article to choose among them.[2]

[2]As Stabler mentions, matters are different in our book. There we show that in some respects—particularly with respect to the Subjacency constraint—the parser must mirror the actual form chosen by the grammar for constraints. Stabler seems to suggest that it was some oversight on our part not to realize that this causes problems for our interpretation of TG as a first level theory. This is a peculiar reading of the circumstances because the theory in our book represents a development of the work in the *Cognition* article and thus might be expected to differ from that article. In the article we claimed that *even if the grammar only computed the extension of the function* it would still place constraints on the class of possible parsers; the grammar is therefore still of interest for theorists of language use. In our book we show that there is a much closer relationship than mere extensionality between grammar and parser in the computation of bounding domains. This is an extension of our earlier work, not an oversight.

2. Why a first level theory?

Still, one might ask why Stabler wants to saddle us with the "linguistics as first level" position. He contends that this picture yields a grammar–parser mapping that is principled: if the grammar computes the function then parsing theory will use the easily defined class of input/output representations provided by the grammar in its parsing algorithms. In contrast he also suggests that we allow for an arbitrary grammar–parser relationship. This is something to worry about, since then the class of possible parsers might not be restricted by grammatical theory. If any arbitrary piece or combination of pieces of the grammar constitute a parsing algorithm then notions of grammatical naturalness and consistency can play no role in the algorithm design. Grammatical constraints on parsing algorithms are correspondingly weak. This worry becomes more evident when Stabler discusses the apparent non-use of D-structure by the Marcus parser: "If D-structure can be left aside in the formulation of S-structures, why should we not assume that S-structures might themselves be left aside in the formulation of LF structures?" (Stabler (1985) p. 00).

Before probing into the issue of arbitrariness of grammar–parser pairings more generally, we should first note that Stabler's example is particularly ill chosen because in fact there is no principled reason for the Marcus parser *not* to compute D-structures. In contrast, there are perfectly principled reasons why S-structures *must* be computed.

If the 'constraints' expressed at D-structure prove relevant for parsing, or even if for other, more principled reasons we decide to fix on a first level model, D-structures are easily incorporated in a parsing model. In our *Cognition* article we even present an algorithm based on the Marcus model that maps directly from surface to D-structures. One could simply construct both S- and D-structure representations in parallel.

But to answer Stabler more directly, one must understand what role D- and S-structure play in the grammatical theory. Of particular interest is Chomsky's observation that almost all of the information encoded in the D-structure representation can be recovered from the S-structure representation:

> One variant, call it Ia, assumes as above that Move α forms these S-structures from base generated D-structures. A second ... call it IIb, assumes that it generates S-structures directly ... the virtual interchangeability of theories is clear within the framework of trace theory.
>
> Chomsky, *Lectures on Government and Binding*, pp. 90–91.

In particular, we can recover at S-structure the logical relation that an argument bears to a verb in D-structure. if we use traces.[3] Very little information from D-structure cannot be recovered from the S-structure representation.[4]

In short, the information provided by D-structure relevant to semantic interpretation can be encoded in S-structure. Therefore, it is not surprising that a parser might not have to compute it separately.

In contrast, a transformational derivation or a transformationally-based parser cannot map phonetic form to meaning directly, bypassing S-structure. Lexical arguments (NPs) will not get associated with the correct logical roles if we try to sidestep S-structure. Stabler's charge of arbitrariness falls short of the mark then, because there are perfectly plausible reasons why the parser must compute S-structures but need not compute D-structures.

Still it seems that even this type of argument will leave Stabler unmoved: "we could grant that Berwick and Weinberg have presented a good psychological theory and that this theory is related in interesting ways to current linguistic theory, and still be puzzled by why linguistic theory should have this or any other interesting relation to psychology, ... even a really impressive demonstration ... just makes the puzzle all the more pressing. Is it just some coincidence that there is a psychological theory that is so related to psychological models of language acquisition and use?" (Stabler (1985) p. 13).

[3]Basically this result follows if we note that S-structure differs from D-structure in that the mapping between the two levels involves moving a category from its D-structure to S-structure position. Trace theory gives an encoding of the D-structure position from which a category is moved. So for example, (a) has the corresponding surface (i), S-structure (ii) and D-structure (iii):

(a) (i) The problem was solved
(ii) The problem$_i$ was solved e_i
(iii) Δ was solved the problem

The D-structure says that the NP *the problem* is the 'logical object' of the verb because it appears in the postverbal position. The S-structure encodes this information via the trace.

[4]In fact, the main arguments for an independent D-structure level come from facts about idioms. Idioms appear in either (for example) active and passive forms or just in active forms. We don't find idioms in pure passive form. If one assumes a level of D-structure this can be explained. Since all elements are inserted at D-structure and then possibly moved, D-structure idioms will always be in active form. Since both active and passive forms appear at S-structure, the same explanation won't work here, because then some idioms could appear just in passive form. We could have an idiom such as, *The baker was played by John* (meaning John was fooling around) but not *John played the baker* (with the same meaning) (see Chomsky, 1981, p. 146). The question then becomes whether we have any evidence to think that idiom forming rules are part of the parser. We think that a reasonable approach would store previously heard idioms in a lexicon. The production of new idioms would also not be a part of one-line sentence processing. The point of going through this example is to show that while Stabler is correct that one is owed an argument about why a particular piece of grammatical structure is omitted or recorded in the parser, he is incorrect in suggesting that one cannot provide principled accounts for the parser–grammar discrepancies that he cites.

So, even if we could show that the parser was an *exact* mirror of the grammar, Stabler would not be impressed because he cannot figure out just why the grammar *should* bear this relationship to the theory of parsing. We believe that this puzzlement stems from a serious misunderstanding of the motivation for the rules and representations of linguistic theory.

Stabler claims that the ultimate aim of generative linguistics is to find the *simplest* system capable of describing all natural languages. He then expresses his doubts that this purely *formal* criterion should find its way into a psychological theory of language processing. However, it is Stabler's evaluation of linguistic methodology that is incorrect. The aim of modern transformational grammar is to explain the learnability of natural languages. Simple systems are desirable because they explain language learnability.

In case after case the learnability of a construction–its ability to be learned on the basis of simple data–depends on finding simpler systems of rules and principles. Stabler is just wrong when he claims that "there is no evidence that human representations of their languages are maximally succinct in any natural sense of that term." In fact, the natural sense of "succinct" is just "learnable".

A psychologically valid theory of language learning, then, has to explain how children can acquire their native language under hostile environmental conditions. The drive for maximally simple grammars is thus justified as part of a psychological theory to the extent that such grammars form superior components of plausible theories of language learning.

Rule simplification has proceeded along two paths. First, language-particular rules have been supplemented by universal principles that do not have to be learned. This shift has simultaneously simplified both the number of rules and their form—exactly the drive to maximal succinctness that Stabler cannot accept.[5]

[5]As a concrete example, consider the passive rule. Older versions of this rule included complex contexts that specified the exact environment of the rule's application. Context conditions explained the difference between (i), (ii) and (iii), (iv):

(i) The man was kissed (by the child)
(ii) The man was talked about (by the child)
(iii) *The man kissed (by the child)
(iv) *The man was talked to Mary about by the child

(iii) is out because the passive rule stipulates an association between passive morphology and passive meaning. (iv) is out because the passivized NP must be adjacent to a verb or at most a verb-preposition sequence. Thus (iv) fails to meet the passive rule's application conditions, and passive cannot apply. Both of these stipulations, though descriptively adequate, are problematic from the standpoint of learning. If we cannot use negative evidence in learning (see Brown and Hanlon, 1970), then we cannot explain how the child learns that (iii) and (iv) are ungrammatical if (i) and (ii) are heard. That is, given just (i) and (ii) as evidence, the child ought to hypothesize a rule that would also generate (iii) and (iv). But this never occurs, and there is no adult language

It is important that this inherently psychological mode of explanation be recognized when evaluating Stabler's comparison between Marr and Chomsky. Stabler asserts that Marr is on firm psychological ground because "emphasis on efficiency and the use of efficiency considerations in the research on vision can be motivated by evidence that the human visual system is efficient." (p. 15).

But the same logic applies to transformational grammar. The shift from complicated language-particular rules to simple universal principles is psychologically justified by evidence that grammars are learnable. *Processing* efficiency is not the only relevant psychological consideration. This perspective allows a more adequate reformulation and response to Stabler's question. The question becomes why the representations that linguists arrive at using evidence about language acquisition should be relevant to parsing. Why isn't a completely different knowledge representation used for processing? As Bresnan and Kaplan (1982) observe, this "is the weakest possible hypothesis one could entertain since it postulates multiple stores of linguistic knowledge that have no necessary connection." (p. xix).

That is, the most highly valued theory would be one that could account for all of the functional demands on language (parsing, learning, production) by a single, uniform representational scheme. If someone doesn't like the uniform representation story, then the burden of proof is on them to come up with evidence to counter it.

As a working hypothesis, we would like to establish a uniformity thesis because it helps us to constrain the class of possible parsers, limiting it to those that can process the system of rules specifiable according to the demands of learnability.[6] Now recall Stabler's discussion of a Universal Grammar characterized by maximally simple principles. Stabler claims that we should be wary of this characterization "especially if succinctness is going to exact a price at run time." (p. 18). But given the uniformity principle, then at least as a first hypothesis we should restrict our attention to machine

where all of (i)–(iv) are grammatical. To solve this problem, current theories move the burden of explanation from *ad hoc* stipulations on rules to innate, universal principles. The details will not concern us here (see Berwick and Weinberg, 1984, Ch. 1 for a full discussion). Suffice it to say that once the *ad hoc* stipulations are factored into a universal component, the passive rule can be simply written as Move NP—all cases of overgeneration are ruled out by the universal conditions.

[6]Once we adopt this 'use-based' psychological perspective—once we ask how this uniform representation is adapted for each of the functional demands placed on the language faculty—we see that there will be questions that a theory of parsing must answer that are irrelevant for a theory of language learning. For example, whether rules are accessed in serial or in parallel is plainly relevant for parsing, but not obviously so for learning.

designs that facilitate the efficient use of these representations.[7]

As we noted in our *Cognition* article, there are many alternative mappings between grammar and parser, from isomorphism to mere input–output equivalence (modulo the qualifications discussed in footnotes 2 and 3 of our article). But the mapping between grammar and parser is not arbitrary—as we have discussed in our book and above. By coupling the two, we can gain an account of both learnability and parsability, a psychological account if there ever was one.[8]

In short, we can provide a plausible reason for relating the principles of parsing and linguistic theory if we can show that principles of that linguistic theory are really motivated by the *psychological* concerns of learnability rather than by the formal concerns of simplicity. The history of linguistics for the past twenty years provides just this demonstration.

Stabler tries to counter the view that learnability is a central notion by exhibiting cases where a simpler analysis is chosen even when it doesn't directly contribute to learnability. He argues that if learnability is used to restrict what the language learner must acquire then it is curious that linguists also try to provide succinct representations of universal grammar, since that need not be learned.[9] Only if the drive for succinct representation is seen as an end in itself can we explain why it is pursued in both language universal and language particular forms.

In fact, this distinction is not new. It has been extensively discussed by

[7]Assuming this strategy has already paid off. In Berwick and Weinberg (1984) we were able to show that a certain parsing design and certain constraints that play a role in learnability (Subjacency) guarantee both efficient parsability and easy learnability. This parsing design is thereby supported as a model of language processing.

[8]Stabler's discussion of parsing contains many errors of detail that we will not be able to cover in depth. Let us mention just one here, Stabler's discussion of the connection between grammar size and parsing efficiency.

Stabler seems to imply that a GPSG approach is more "efficient" in part because it expresses all its constraints at a single level of representation. He even suggests (in footnote 10) that Berwick (1982) adopts this view. This is just wrong. Suggesting that a grammar can be viewed as a parser doesn't mean that it can be viewed as an *efficient* parser. In fact, just the reverse is true: by casting all constraints at one level, the ordinary GPSG grammar explodes the size of the grammar enormously, so enormously that the parsing system is not efficient. In practice, GPSG implementors don't 'multiply out' rules until they are needed (see Shieber, 1984). Suppose one 'multiplied out' all possible movement rules and compiled a big table of all these possibilities. This table would correspond to the derived rules used by a GPSG parser. But the parser then has to consider *all* these rules to parse even a simple sentence where no movements have occurred. In other words, the grammar size of the derived system is so big that it slows everything down. This is an exact confirmation of what Berwick and Weinberg (1982) predicted, and a refutation of Stabler's claim.

[9]"but it nevertheless seems possible that a good deal of every attainable grammar might already be determined by the acquisition device. If this is an open possibility, there is no reason to expect any link between the size of grammars ... and their learnability." (Stabler, p. 28).

Chomsky (1981): "The objective of reducing the class of grammars compatible with primary linguistic data has served as a guiding principle since virtually the outset as it should, given the nature of the fundamental empirical problem to be faced—namely, accounting for attainment of knowledge of grammar." Continuing, Chomsky remarks upon a filter that is "too strange to be an appropriate candidate for UG and should be reduced to more natural and more general principles." However, he notes that "the approach I will pursue here can be justified only in terms of its success in unearthing a more elegant system of principles that achieves a measure of explanatory success."[10] However, "it is quite possible to distinguish between these concerns [between the search for elegant characterizations of UG and restricting the class of possible grammars rcb/asw]. For example, a theory of UG with redundancies and inelegant stipulations may be no less restrictive than one that overcomes these conceptual defects." (1981, pp. 13–15).

Chomsky's remarks underscore our point that the drive for succinct representations fits with linguistic methodology only when it helps explain language acquisition.

Stabler's last concern centers on whether the formal analysis of acquisition complexity outlined in Sections 3.1 and 3.2 of Berwick (1982) correctly captures the notion of simplicity as it is used in linguistic theory. At the outset, it should be noted that Stabler's remarks deal with only a small subpart of this much larger work; in fact, Stabler does not discuss any of the applications of the theory to real linguistic cases, as analysis that takes up the bulk of Section 3.2 and the remainder of Chapter 3 of this work.

One aim of Chapter 3 of Berwick (1982) that is explicitly stated in its introduction is to probe the ordinary use of the notion of "simplicity" as used in linguistic theory and *come up with a rationalized account:*

[10]To give the reader an example of what counts as a "measure of explanatory success" we will very briefly discuss the problem of accounting for language variation. English does not allow for either (i) or (ii) below:

(i) What do wonder who saw?
(ii) How many do you think Fred ordered of those books on Tuesday?

These examples are not semantically deviant, as can be seen by (iii) and (iv):

(iii) You wondered who saw what
(iv) How many of those books do you think Fred ordered on Tuesday?

(i) and (ii) are also grammatical in a language like French. In general it seems that if a language does not allow (i) it does not allow (ii). Given earlier conceptions of UG this was surprising, since a special universal principle was invoked for each case (the *wh*-island conditions for (i) and the Left Branch condition for (ii)). Current theories unite these two principles under one condition—Subjacency. This explains the correlation between (i) and (ii), and the language variation.

> The advantages of this formal approach are several. For one thing, it allows us to settle certain questions about the role of notational systems in expressing linguistic generalizations. (p. 219)
> A second advantage of the formal analysis is that it may be used to study the problem of acquisition and to justify several of the "operating principles" used by the acquisition procedure discussed in Chapter 2. Informally, it has long been suggested that the notion of simplicity is intimately connected to acquisition ... one of the major aims of this chapter is to develop and apply a formal characterization of simplicity by pursuing the idea of a notational system as a programming language. The basic approach is to adopt the theory of program size complexity as the right yardstick for simplicity. This model meets several needs. First, it serves as a formal theory of markedness, by equating markedness ... as 'the amount of external information required to fix a grammar.' Second, this theory provides a developmental model of languages acquisition. (pp. 222–223)

Given this, Stabler's exposition is odd, to say the least, because it basically just recapitulates the same critique and search for a rationalization of the notion of simplicity measures as carried out in Sections 3.1 and 3.2 of Berwick (1982).

Stabler's account has two main points:

First, he analyzes a problem with universal TMs and the arbitrary nature of program size measures (pp. 20–23). Stabler's conclusion: notational systems are arbitrary, hence require independent justification.

Second, since the succinctness–explanation connection is to be rejected, ("the first idea that shorter grammars are just more explanatory or more likely to be correct, is untenable." (p. 26)), Stabler discusses whether an alternative account of simplicity based on acquisition is a better metric (p. 27). The obvious measure based on grammar size is wrong, but (p. 28), a measure based on the number of decisions required to learn a grammar seems better ("I think that we are now near Berwick's substantial insight. After noting that 'a strict program size measure is not quite the right one,' he suggests that the right measure is really something like the number of decisions needed to determine the grammar." (pp. 28–29)).

The real irony is that in spite of its surface appearance of disagreement, Stabler hews to precisely the same outline as in Berwick (1982). The aim of Sections 3.1 and 3.2 is to provide a formal account of Chomsky and Halle's often informal remarks on simplicity. Since it is by no means clear just what the right notion of simplicity should be, the chapter adopts an exploratory tone, first presenting the views of Chomsky and Halle (pp. 216–219); then following with a critique of those approaches and an attempt to rationalize them (Berwick, 1982), then presenting the program size measure and some obvious problems with it if interpreted literally, and finally settling on a revised complexity measure based on the number of decisions required to learn a grammar. Note that this is exactly Stabler's outline.[11]

Turning now to particular points of disagreement, let's first consider Stabler's discussion of the arbitrariness of notational systems (pp. 21–23). Part of the difficulty here stems from two different notions of "arbitrary" presented in Berwick (1982) where Stabler sees only one. First of all, Berwick (1982) pointed out that complete computational systems are arbitrary because in a complete system *all* measures do as well as one another. "In thise sense a sufficiently powerful notational system renders the demand to fix notational machinery moot." (Berwick, 1982, p. 221; quoted by Stabler, p. 22). Second, *incomplete* computational systems can provide examples of arbitrary notational systems because one can switch notations and obtain a different complexity result: "if a notational system is not sufficiently rich then there may be some regularity that is inexpressible in that system, but easily captured in another." (Berwick, 1982, p. 219). In this second case, the choice of notation requires independent justification—just as Stabler notes. Now the choice of a notational system becomes an interesting claim open to independent confirmation. This last point is obvious; Berwick (1982) included a full excerpt from Chomsky (1965) showing that if one notational system contains a primitive for cyclic permutations and another does not, then the languages easily describable by these two notations differ.[12]

[11]This correspondence is obscured by Stabler's exposition: he often introduces an idea into the discussion, only to say in the next paragraph that Berwick (1982) says the same thing. For example, Stabler begins his discussion of succinctness measures with a discussion of an "armchair refutation" of their use, based on the arbitrariness of notational systems (p. 21). Only later (p. 22) do we find that this discussion is actually a part of Berwick (1982). Similarly for the discussion beginning on p. 26, and the final proposal of a measure based on decision complexity.

[12]"Let us now consider the possible effect of altering one's notational system. Consider the following example discussed in *Aspects of the Theory of Syntax*...

(15) Aux → Tense (Modal)(Perfect)(Progressive)

Rule (15) is an abbreviation for eight rules that analyze the element Aux into its eight possible forms. Stated in full, these eight rules would involve twenty symbols, whereas rule (15) involves four (not counting Aux, in both cases). The parenthesis notation, in this case, has the following meaning. It asserts that the difference between four and twenty symbols is a measure of the degree of linguistically significant generalization achieved in a language that has the forms given in list (16), for the Auxiliary Phrase, as compared with a language that has, for example, the forms given in list (17) as the representatives of this category:

(16) Tense, Tense–Modal, Tense–*Perfect*, Tense–*Progressive*, Tense–Modal–*Perfect*, Tense–Modal–*Progressive*, Tense–*Perfect*–*Progressive*, Tense–Modal–*Perfect*–*Progressive*

(17) Tense–Modal–*Perfect*–*Progressive*, Modal–*Perfect*–*Progressive*–Tense, *Perfect*–*Progressive*–Tense–Modal, *Progressive*–Tense–Modal–*Perfect*, Tense–*Perfect*, Modal–*Progressive*

In the case of both list (16) and list (17), twenty symbols are involved. List (16) abbreviates to rule (15) by the notational convention; list (17) cannot be abbreviated by this convention.

Chomsky observes that the set described in (17) *can* be abbreviated if a different set of notational conventions is admitted, namely, some notion of cyclic permutation. For example, one could re-interpret parentheses

We are also in basic agreement with much of Stabler's remaining discussion about simplicity metrics. Stabler's ultimate conclusion about simplicity metrics really just reinforces that of Berwick 1982. Thus when he says, on p. 26, "the rough idea we want to flesh out is that if the grammar of a language is 'short,' then the problem of computing it ... should be fairly easy", and then goes on to reject this view, he is, as he says explicitly, simply following the analysis in Berwick (1982). Nothing new is added.

Stabler goes on to briefly consider the implications of simplicity measured as the number of decisions required to learn a grammar. His review is again based on Berwick (1982), but his conclusions are somewhat negative: "the appropriate measures of 'ease of acquisition' and their relations to grammars and parsers are still quite obscure." Certainly, it is true that we are a long way from understanding the connection between acquisition complexity and grammars. Still, what Stabler *omits* from his discussion is quite relevant.

Given that Stabler agrees with us and Berwick (1982) about the need for independent justification of notational systems, the key question is whether one can in fact provide such justification. The remainder of Berwick (1982) (Section 3.2) on simplicity metrics goes on to show that the "decision theoretic" model of simplicity actually can be independently justified. If we adopt this model, it shows that one can explain certain otherwise mysterious facts about possible human segmental systems, facts about the order of acquisition of $\bar{X}$ rules, and other properties of language acquisition.[13] This being so, it becomes the critic's task to show that some other proposed simplicity measure does as well or better as this one; it is certainly not enough to repeat the understood demand for independent justification when this has already been provided.

Stabler's omission of a discussion of the actual application of the formal analysis is, we think, quite symptomatic. In our view, it is only by examining real linguistic cases that one can see whether a proposed formal revision is

surrounding a list of elements as denoting the set of cyclic permutations of those elements—a standard mathematical notation, in fact:

(Tense Modal *Perfect Progressive*) (Berwick, 1982, p. 220).

Plainly, then if we have a computational system that does *not* include the notion of a cyclic permutation, then what we can write down in a compact form changes. This holds if the system is *not rich*. Thus, contrary to what Stabler says is the import of Berwick (1982), independent justifications for a program size measure *is* needed when a notational system is *impoverished*, not when it is complete.

[13]For example, the decision complexity model applied to the case of segmental system learning explains why we don't find sound systems with, say, seven vowels of which four are high. Indeed, given 20 distinctive features there are 2^{20} possible segments and a far vaster number of segmental systems (nearly all the possible subsets of 2^{20} elements—$2^{2^{20}}$). Most of these do not occur. The measure proposed in Berwick (1982) explains why.

doing any real work. In the end, the abstract formal analysis of program size measures is just too abstract unless it is applied to real linguistic examples. For us, the interesting point is not whether there are results connecting, say, the number of nonterminals of a grammar to its acquisition complexity, but whether those results are relevant *in practice*. In the case at hand, the measure based on the number of decisions required to fix a grammar *does* provide an explanation of at least one real, complex example; and there other others given in the remainder of Chapter 3 of Berwick (1982). This is the ultimate test of the analysis in Berwick (1982), and, we think, of any formal analysis. The program of taking TG seriously, as outlined in the *Cognition* article, in our book, and in Berwick (1982), has already led to new insights about the role of grammatical constraints in parsing and language learning, new experimental predictions, and even new practical parsing methods.

References

Berwick, R. (1982) Locality principles and the acquisition of syntactic knowledge. PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Berwick, R. and Weinberg, A. (1984) *The Grammatical Basis of Linguistic Performance*. Cambridge, MA, MIT Press.

Bresnan, J. (1978) A realistic transformational grammar. In J. Bresnan, M. Halle and G. Miller, *Linguistic Theory and Psychological Reality*. Cambridge, MA, MIT Press.

Bresnan, J. (1982) *The Mental Representation of Grammatical Relations*. Cambridge, MA, MIT Press.

Brown, R. and Hanlon, C. (1970) Derivational complexity and the order of acquisition in child speech. In J.R. Hayes (ed.). *Cognition and the Development of Language*. New York, John Wiley.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA, MIT Press.

Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht, Foris Publications.

Fodor, J., Bever, T. and Garrett, M. (1974) *The Psychology of Language*. New York, McGraw-Hill.

Shieber, S. (1984) Direct parsing of ID/LP Grammars. SRI Technical Report.