Cognition, 17 (1984) 155–179 **Discussion**

Berwick and Weinberg on linguistics and computational psychology

EDWARD P. STABLER, JR.*
University of Western Ontario

1. Introduction

According to recent theories in cognitive psychology organisms are to be seen as receiving raw representations of sensory input which are successively transformed into more abstract and more useful representations. Thus, according to one well-known approach to vision, visual input is transformed into a primitive sketch-like representation of two dimensions and then into a sketch with representations of depth and finally into a representation of three-dimensional structure. The main body of research in this approach to vision has aimed to characterize the functions computed in transforming a representation at one 'level' into the corresponding representation at a more 'abstract' level. A similar approach can be taken to language. Here the assumption would be that in speech recognition acoustic input is transformed into a number of more abstract linguistic representations, and similarly for other non-acoustic linguistic input.

David Marr proposes an account of visual processing along these lines, and he suggests that Chomsky is providing the similar account of language processing. He says, for example:

> Vision is an information processing task, and like any other, it needs understanding at two levels. The first, which I call the computational theory of an information processing task, is concerned with what is being computed and why; and the second level, that at which particular algorithms are designed, with how the computation is carried out ... Chomsky calls level 1 theories competence theories, and level 2 theories performance theories. (1979, p. 19)
>
> Chomsky's (1965) notion of a 'competence' theory is precisely what I mean by a computational theory for that problem. Both have the quality of being little concerned with the gory details of algorithms that must be run to express that

*I am indebted to William Demopoulos, Robert Matthews and Zenon Pylyshyn for helpful discussions of this material. I also benefitted from the discussion following a presentation of this paper at the University of Pennsylvania in February 1984. Reprint requests should be sent to Edward P. Stabler, Jr., Department of Computer Science and Philosophy, Middlesex College, University of Western Ontario, London, Ontario, Canada N6A 5B7.

0010-0277/84/$8.00

> competence (i.e., to implement the computation). That is not to say that devising suitable algorithms will be easy, but it is to say that before on can devise them, one has to know exactly what they are supposed to be doing, and this information is captured by the computational theory. (1977, p. 38)

His suggestion is that both theories of vision and linguistic theories specify the functions computed in visual and linguistic processing, respectively. Chomsky agrees. Citing Marr, he says:

> Adopting this framework, we may consider the study of grammar and UG [universal grammar] to be at the level of the theory of computation ... I don't see any useful distinction between 'linguistics' and 'psychology,' unless we choose to use the former term for the study of the theory of the computation in language, and the latter for the theory of the algorithm ... For this reason, I am uneasy with J. Morton's proposal that perhaps '... linguistics is more abstract than psychology, and is better to be considered at a different level ...,' and with a distinction between 'psychological' and 'linguistic' functions. (1980, pp. 48–49)

This is a very appealing view, both for linguists who want to be doing psychology, and for psychologists who want someone to tell them how to do the linguistics. This view has been widely criticized, though. The psychological import of Chomskian linguistics has been debated for many years; neither side seems to have been able to quite come to grips with the concerns of the other. The issues have nevertheless attained a new prominence and importance with the swelling ranks of linguists pursuing research programs that have explicitly sought to make the connection with psychological models more secure.

In recent work, Berwick and Weinberg attempt to dispel the opinion that transformational grammar (henceforth, TG) "has fallen on hard times" in this regard (1984, p. xi). They represent a sophisticated defense of the view that Chomsky's recent "government and binding" approach to TG has an important role to play in the best psychological accounts of human language acquisition and understanding. In fact, they endorse the view of linguistics suggested by Marr and Chomsky in the passages quoted above, arguing that linguistics characterizes *what is computed*, the linguistic 'data structures,' while other psychological investigations explain *how those linguistic structures are computed*. Their defense of this view really does, I think, address the deepest concerns of the various opposing parties. At the very least, the debate is considerably advanced by their work. In this paper I will argue that they do not succeed in making their case, though. Their work fails to support the proposed analogy between the work of Marr and Chomsky; TG is not at all like a first level computational theory. I will first pick on some particular difficulties, but I want to focus on the problem of explaining why theories yielded by anything like the standard methodology of linguistics should have psychological relevance. Berwick develops the standard responses to this problem, and locates the difficulty very precisely in a particular component of linguistic methodology. The problem is precisely located but not resolved. I will consider the possibilities for alternative programs.

2. The parser as a realization of TG

Berwick and Weinberg (1984) begin their new book by arguing that the early attacks on the psychological import of TG are flawed. A wide range of psychological theories are compatible with the account of language provided by TG, and the psychological evidence can hardly be construed as ruling all of them out; in fact, the evidence does not begin to suggest that these theories are not on exactly the right track. Of course, this is a fairly weak point. (As they say, "If we stopped here we would simply be in the same theoretical boat as the early 1960's" (1984, p. xiii).) So what they do is to propose a particular parsing model based on the work of Marcus (1980) and an acquisition theory for that model. They suggest that their parser 'realizes' TG, that it is compatible with available psychological evidence, and that it "explains why certain universal constraints take the form that they do" (1984, p. xiii).

Let us first consider the relation between the linguistic theory and the proposed psychological account. I think that the relationship is a good deal more subtle than the explicit remarks of Berwick and Weinberg indicate. They offer what appears to be a fairly straightforward first-level account according to which the parser simply computes a function specified by the grammar, saying,

> (i) ... it is equivalent to a transformational grammar in the sense that for every surface string (sentence) it associates the same annotated surface structure (annotated by traces) that would be paired with that sentence by a transformational grammar (1984, p. 147).[1]

The literature on Marcus's parser has made it clear that this sort of claim is, at the very least, in need of qualifications that have an important bearing on what we ought to say about the relation between TG and psychological models. These points are given surprisingly little attention by Berwick and Weinberg. Let me go over them quickly.

[1]If these "annotated surfaces structures" are not S-structures, this is a difference between the grammatical theory and the psychological theory which we will want to explain.

Levels of representation

The first thing that should be noted is that the "equivalence" proposed between TG and the parsing model is not the one we might have expected. The theory of TG Berwick and Weinberg have in mind is the recent work of Chomsky, especially his recent (1981, 1982) "theory of government and binding." It associates with each string a set of linguistic structures: a phonetic PF-form, a D-structure, an S-structure, and an LF-structure. Chomsky (1981, p. 33) illustrates the differences between these levels with the following example (in which the relevant structures are only partially indicated),

(English)	it is unclear who to see
(PF)	it is unclear [s′ who [s to see]]
(S-structure)	it is unclear [s′ who_i [s PRO to see t_i]]
(D-structure)	it is unclear [s′ COMP [s PRO to see who]]
(LF)	it is unclear [s′ for which person x [s PRO to see x]]

The primary motivation for the distinct levels of representation is that independent constraints apply at each level; a string is well-formed just in case it has a representation satisfying the constraints at each of the four levels. Ambiguous strings will be associated with more than one such quadruple of structures.

As noted above, Chomsky himself seems to suggest a first level computational interpretation of his theory. So remarks like the following might be taken to be proposals about the domain and range of the functions computed in the exercise of linguistic abilities:

> At the most general level of description, the goal of a grammar is to express the association between representations of form and representations of meaning. (1981, p. 17)
>
> It is reasonable to suppose that the representations of PF and LF stand at the interface of grammatical competence, one mentally represented system, and other systems ... (1981, p. 18)

Certainly the most natural supposition is that humans compute the function from the relatively superficial PF representation to the more abstract LF representation, perhaps by formulating the intermediate levels of D-structure and S-structure.

So the proposal of Berwick and Weinberg is rather peculiar because their parsing model formulates only S-structures. It is plausible that they are assuming that the computation of PF representations from acoustic input would be done by a separate mechanism, and similarly that a separate mechanism may formulate some sort of graphemic representation of visual input. These mechanisms could provide the input for a parser similar to the one proposed.

But it is not plausible that D-structures are formulated by a separate mechanism in this way. This is implausible because the constraints on D-structure must be taken into account in the formulation of S-structures; if they were not, presumably the parser would accept some ungrammatical sentences, *viz.*, those ruled out by specifically D-structure constraints. But the parser is presented as a mechanism that formulates S-structures for all and only the sentences of the language, so it must be implicitly incorporating the D-structure constraints.

I think that there may well be a good explanation of this, but for the moment I just want to make a relatively uncontroversial point:

(1) The psychological model does not suppose that all levels of representation posited by the linguistic theory are formulated.

In particular, I think that D-structures are not assumed to be formulated by Berwick and Weinberg, even though they count their theory as a 'realization' of TG.[2]

So this raises the question: Which levels of grammatical representation *are* formulated by the language processing system? 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? These are important questions for any psychological construal of TG, but again we can rest with a more modest point that follows however these questions are answered:

(2) The motivation for levels of representation in linguistic theory is apparently different from the motivation for levels of representation posited by psychological theories of language processing.

This is obviously a point of some significance if we are interested in the claim that TG is a first level computational theory. We already see that the most natural computational construal, according to which all linguistic representations for each input are formulated, is not what Berwick and Weinberg propose. (Their proposal is, I think, much more reasonable on psychological grounds than the more natural computational construal of TG, but a defense of this claim is beyond the scope of this paper.) Suffice it to say that if TG does have a first level construal at all, it is not one that jumps out at you. It would be nice to have a principled explanation of how one gets the first level theory from the TG given points (1) and (2). Berwick and Weinberg do not consider this problem.

---

[2]If D-structures were to drop out of TG, point (2), below, could still be maintained, though not so easily. See Section 2, below, and especially the defense of point (7).

The language parsed

Another aspect of the Marcus parser which has been extensively discussed is its treatment of garden path sentences. These are sentences like:

(a) The horse raced past the barn fell.
(b) The boat floated down the river sank.

These sentences can be understood as the reduced versions of (a′) and (b′) in the way that (c) is understood as the reduced version of (c′):

(a′) The horse that was raced past the barn fell.
(b′) The boat that was floated down the river sank.
(c) The man taken to the hospital died.
(c′) The man who was taken to the hospital died.

The garden path sentences are grammatical, and they do have well-formed linguistic structures, but they are not understood by all competent English speakers—or at least not readily, without practice with this sort of construction. So Marcus counted among the virtues of his parser the fact that it also fails to parse these sentences. When his parser gets to the last word of (a), it has already constructed the noun phrase "[np the horse]," it has attached that noun phrase under an s, "[s [np the horse]," and it is looking to finish the main verb phrase.[3] In this situation, it cannot handle the last word without undoing some of what it has done, but the parser is deterministic and does not backtrack, so it fails.

The idea that the human parsing mechanism might not accept all grammatical strings is not new. Miller and Chomsky (1963) pointed out many years ago that we should not expect humans or other parsers with finite memory to be able to parse sentences with arbitrary depths of center-embedding. It is not clear how the Marcus parser can account for the special difficulty of center-embedded constructions, but it does, in any case, have an account of the difficulty with garden path sentences. The accounts of deeply center-embedded and garden path sentences are different, though, in a significant respect. The suggestion of Chomsky and Miller was that we may well realize a

---

[3]Marcus *et al.* (1983) note that apparently the same properties of the parser that make it unable to parse garden path sentences or build structures that violate subjacency (see (iii), below) also prevent it from accepting some sentences that it should accept, like:

(d) I drove my aunt from Peoria's car.
(e) Birds eat small worms and frogs eat small flies.

Berwick and Weinberg do not discuss these problems. Nor do they discuss how their psychological model can plausibly account for the "squishiness" of the island constraints in English (discussed in Fodor (1983)) or the more serious constraint violations in Swedish (discussed in Engdahl and Ejerhed (1983)).

procedure which would accept and parse all center-embedded sentences if it were not limited by a finite memory; the difficulty with garden path sentences, on the other hand, is explained by the assumption that we do not realize a procedure which is capable of parsing all grammatical sentences even given unlimited time and memory.

All of this is very well known, of course, but the important thing to notice is that this undermines the attempt to construe TG as a first level computational theory. Garden path sentences are in the set of grammatical sentences but plausibly not in the set of strings for which competent speakers compute more abstract representations. There may also be ungrammatical strings (i.e., strings that a competent speaker recognizes to be ungrammatical and would not be disposed to produce himself) for which more abstract representations are computed, using only linguistic knowledge. Indeed this is most plausible, and of course if it is accepted, it also undermines the simple first-level claim that the parser computes even just the function from inputs to S-structures defined by the grammar. The point is simply:

(3) The linguistic theory does not define the domain of the function computed in language understanding.

Structural relations and parsing functions

One final point to note about the first level claims considered so far is that the relation between grammatical strings and well-formed linguistic structures which is defined by the grammar is not functional; it is one-many. The deterministic parser proposed by Berwick and Weinberg is functional, though; it computes exactly one syntactic representation for each input, even in cases where the input is syntactically ambiguous. In these cases the parser must, in effect, decide which structure to represent on the basis of something other than grammatical knowledge. It is plausible that people understand their language in this way also. It seems unlikely that speakers generally formulate representations of all of the grammatically permissible structures for each input. Marcus used a certain sort of 'semantic' information to guide his parser, but one might plausibly assume that pragmatic influences play a role here. That is, the discourse context of a syntactically ambiguous utterance might well have an influence on the way it is parsed. (This idea obviously runs counter to the new theories of the modularity of linguistic processing coming out of MIT and other places, but I understand that experimental results have not yet settled the matter.) In any case, for present purposes, let me just note the modest point:

(4) The linguistic theory associates ambiguous strings with more than one representation at each level, and most psychological models assume that only one among the specified sets of structures is typically formulated. So in this respect, the linguistic theory again does not specify the function computed in language understanding. Something additional is required to make the ⟨input, abstract structure⟩ relation defined by the grammar functional.

I take it that this point is also relatively uncontroversial. It nevertheless has an obvious bearing on the issues we are concerned with, and it gets surprisingly little attention in Berwick and Weinberg.

Beyond the first level

The main argument offered by Berwick and Weinberg for the psychological importance of TG has an interesting bearing on present concerns. First, it should be noted that they do mention, briefly, the sort of psychological evidence that we might expect. They say (ii):

> There is good evidence that the *data structures*—the units of representation—posited by theories of transformational grammar are actually causally implicated in on-line language processing. (1984, p. 197)[4]

They do not present this evidence themselves, but say, "This, after all, was the burden of Fodor, Bever and Garrett's 1974 argument" (1984, p. 197). The evidence used in the arguments of Fodor *et al.* (1974) is fairly crude (as they recognize); it does not suffice to distinguish the earlier versions of TG from recent versions, and it does not even suffice to distinguish the linguistic representations posited by TG from those posited by 'lexical functional grammars' (LFG) or 'generalized phrase structure grammars' (GPSG). More recent studies of language processing have also failed to indicate with any clarity the best account of linguistic representation.

Berwick and Weinberg hold that other considerations do indicate that their TG-based approach does provide the best psychological theory. They argue that the TG-based parsing model is unique in its ability to explain certain universal features of language (iii):

> By positing that Move α exists and that it is engaged in mental computations, we can actually explain some facts about natural languages—namely, that they will obey subjacency in certain situations and not others—and derivatively some facts about human behavior. In contrast, we do not get the same explanation if we assume other rules and principles, even some that describe the same set of possible sentences (for example, the Augmented Transition Network's 'hold

[4]The emphasis in all the passages quoted in this paper is from the original sources.

> cell' model). And even systems that are written to conform to the same generalizations as transformational grammar—GPSG—but don't use the same mechanisms *fail* to provide an explanation for these constraints. This is then a genuine case where a particular grammatical rule supports a particular claim about what sort of mechanism is responsible for a (mental) computation—just what some have said would be required to show that grammatical rules are 'used' in mental computations. (1984, p. 196)

Now the interesting point for our present concerns is that the virtue claimed for TG here is not something that could be claimed for it if it were merely a first level theory. That is, if TG simply said that a particular function from strings to more abstract representations was computed, then it would logically follow that any first level theory that specified the very same function would also be correct. But GPSG is apparently being faulted here even if it specifies the same structures, or "conforms to the same generalizations." The idea is that TG is better because it not only specifies the right function, but it allows for a natural explanation of why that function is computed. I will return to this idea later, but it is interesting to note that the explanation of subjacency is the mainstay of the argument offered by Berwick and Weinberg for the psychological relevance of TG, and that it apparently sees TG as providing something more than merely a first level account of what is computed.

In sum, it is clear that there are some difficulties in interpreting TG as a first level computational theory. This idea is certainly too simplistic, and Berwick and Weinberg's overall project can be seen as suggesting a more subtle idea. They apparently suppose that there is a more abstract relation between TG and psychology, weaker in some respects than a simple first level computational theory, and stronger in others. The exact nature of this relation is not made entirely clear. My suggestion, though, will be that the problems with the simple first level construal indicate that there is a substantial difference between the kind of work done in linguistics and the kind of work done in psychology. Marr's proposals about what is computed in visual information processing do not have the problems that TG has, and that is no surprise, or so I shall argue. The two projects apparently have very different methodologies, and it is not clear that the methodology of TG is appropriate for psychology. If I can make a case for this view, the problems I have mentioned for psychological construals of TG should make perfect sense.

3. Methodology in psychology and linguistics

Suppose that the difficulties I have raised so far could be avoided by specifying a more abstract sense in which the best psychological theories 'realize'

TG. I think that the status of TG would still be puzzling. Consider Berwick's description of the situation:

> Given that Marr's and the generative grammarian's research strategies are so much alike, it is perhaps surprising that considerable dissatisfaction has been voiced with the grammarian's stance of abstraction, and much less with Marr's. It is hard to distill this uneasiness into a single line or two, but at bottom, it seems to amount to this: these critics of generative grammar (or even of particular theories of grammar) sometimes seem to be saying that a theory that begins and *ends* with grammar cannot be a true account of the language faculty. Marr, it should be noted, does not stop in his account of early visual processing at the abstract theory of the computation; he goes on to probe alternative algorithmic realizations of that theory, and a range of possible machine implementations for those algorithms, ultimately aiming for a full account of the *psychophysical* behavior of human visual perception. Isn't it then possible, these critics go on to say, that by *ignoring* the exigencies of computation—associated algorithms and machine implementations for grammars—that one could arrive at a theory of grammar that, loosely speaking now, literally could not be incorporated into the human nervous system, i.e., a theory of grammar that could not be "psychologically real"? ... another way to view the research reported on here is as an *explicit* demonstration that these fears of "non-realizability" are unfounded ... not only can generative grammars for natural language be "realized" in models of parsing or acquisition, they can be *efficiently* realized ... (1982, p. 12–13)

Now I think that Berwick has got the puzzle exactly right here, and he has correctly indicated the character of his answer. One major source of concern about the psychological import of TG derives from its methodology: there is an apparent disregard for results concerning what humans actually do in language understanding and acquisition. Berwick and Weinberg respond to this worry by providing an actual model that purportedly 'realizes' TG in just the right sense. So far I have urged that that sense is far from clear, but the more important thing to notice is that the worry about methodology will not be assuaged by a demonstration that the currently available product of that methodology can be construed as having some fairly subtle and abstract relation to a good psychological theory. 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 about why linguistic theory should have this or any other interesting relation to psychology. The presence of a really impressive demonstration of the sort offered 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? The way to resolve this puzzle is by considering linguistic methodology with an eye to whether we can see why that method would yield theories with *that* relation to psychological models. So that is what I propose to do.

There is a very striking difference between the work of Marr and Chomsky. Marr does refer frequently to the psychological literature, but he is also very clear about the psychological motivations for his particular proposals. He proposes representations of visual information that are needed to account for our perceptual abilities, and he is concerned that the posited representations can be efficiently formulated, since human visual processing is apparently quite efficient. He makes these points explicitly in remarks like the following:

> To understand vision thus requires that we first have some idea of which representations to use and then we can proceed to analyze the computational problems that arise in obtaining and manipulating each representation. Clearly the choice of representation is crucial in any given instance, for an inappropriate choice can lead to unwieldy and inefficient computations.
> The important thing about a representation is that it makes certain information *explicit* (1979, p. 20).

The 'explicitness' of information I in a representation R is a matter of how complex the computation of I from R is, where the relevant notion of complexity is apparently the standard one. The example Marr uses to illustrate this point is one that is familiar from discussions of the (run-time) resource ('space' and 'time') complexity of computations on particular kinds of machines.[5] He says,

> ... if one chooses the Arabic numeral representation, it is easy to discover whether a number is a power of 10 but difficult to discover whether it is a power of 2. If one chooses the binary representation, the situation is reversed. Thus, there is a trade-off; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover. (1982, p. 21)

This idea of the ease of getting information from a representation, of the 'explicitness' of information in a representation, plays a central role in Marr's work. This is clear both in his general remarks (such as those quoted above) and in his more detailed accounts of the motivation for particular hypotheses.[6] The emphasis on explicitness is an 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. Marr makes this point repeatedly.[7] The real test of the methodology is, of course, the success of its

---

[5]Cf., e.g., Borodin (1973).

[6]See, e.g., the argument for 'terminations' in the primal sketch (1982, p. 80), for 'bit map' representations (1982, p. 81), for 'virtual lines' (1982, p. 149), for aspects of the 2½-D sketch (1982, pp. 276–277).

[7]See Marr (1979, p. 20; 1982, p. 75, 105, 266–267).

theories, but the apparent efficiency of our visual information processing motivates the method.

In Chomsky's work, there is no emphasis on efficiency, and as noted above, the levels of linguistic representation are not motivated by considerations of how any particular functions might be computed in the exercise of linguistic abilities. The role of *efficiency* in Marr's work seems to be filled by *simplicity* in Chomsky's work. He aims to formulate grammars that are simple. Berwick notes that this consideration has played a prominent role in Chomsky's work since his *Morphophonemics of Modern Hebrew* (1951), where he says,

> A grammar of a language must meet two distinct kinds of criteria of adequacy. On the one hand it must correctly describe the 'structure' of the language ... On the other hand it must meet requirements of adequacy imposed by its special purposes ... or, in the case of a linguistic grammar, requirements of simplicity, economy, compactness, etc.
> The criteria of simplicity governing the ordering of statements is as follows: that the shorter grammar is the simpler, and that among equally short grammars, the simplest is that in which the average length of derivation of sentences is the least. (quoted in Berwick (1982, pp. 216, 217))

In *Aspects* (1965), Chomsky is using the same strategy, arguing that we should give up standard rewrite grammars for our account of auxiliary constructions on the grounds that we can reduce the number of symbols (i.e., the number of symbol *occurrences*) from twenty in the standard rules to four in his special parenthesis notation (1965, p. 43). This influence of simplicity considerations is still quite prominent in Chomsky's (1981, 1982) work.

4. Simplicity in linguistics

This aspect of linguistic methodology, the drive to find maximally succinct representations of the language, does not seem to rest on any empirically motivated assumption, in the way that efficiency considerations in vision do seem to be motivated. It should be emphasized again that the first requirement on a theory is that it accounts for the available evidence; the methodologies play the secondary but still important role of guiding theory development along fruitful lines. In accounting for the data we want to make sure we stick to visual processes that are reasonably efficient, but it is not so clear why we should aim for grammars that are maximally succinct. This certainly accounts for some of the worry about the psychological import of Chomsky's grammars.

Let me explain these worries a little more precisely, before considering how they might be defused. In the first place,

> (5) There is no evidence that human representations of their languages are maximally 'succinct,' in any natural sense of that term.[8]

We can make the criticism somewhat stronger, with the following point noted by Berwick:

> (6) More powerful formalisms 'shrink' the size of descriptions for grammars (in the program complexity or information sense) but demand more in the way of computational resources for the languages they generate. (1982, p. 360)[9]

These points can be sharpened, and their application to the present discussion clarified, if we adopt the formal account of modularity and its role in linguistic theory that is developed by Berwick (1982) and Borgida (1983).

Berwick argues that a maximally concise grammar will distinguish independent constraints on well-formedness, rather than simply using a set of rules to generate the right set. He argues as follows. Suppose that a level of representation has to satisfy two constraints. In a simple case, each of these constraints may serve to filter out a regular set. A generator might enforce these constraints, then, by realizing a single machine that filters out the union of the two regular sets of ill-formed elements, or it might filter the representations through a machine defining one constraint and then through a machine defining the other. As Berwick points out, the single machine will need to be very much larger than (i.e., have more states than) the sum of the sizes of the two machines in exactly the case where the constraints are independent. He suggests that Chomsky's (1981, 1982) use of a number of independent theories to constrain particular levels of representation may, in some such formal sense, make the theory more succinct. Berwick calls this "intra-level" modularity.

A similar argument can be used to motivate the definition of a number of levels of representation. In certain cases a considerable gain of conciseness in characterizing a set can be achieved by representing the set as the range of a mapping from another less complex set (and cf. Borgida (1983)). Berwick suggests that this sort of consideration may be used to motivate the use of D-structure in the definition of S-structure, and the use of S-structure in the definition of the mapping from PF to LF structure. Berwick calls this "inter-level" modularity.

---

[8]Some care needs to be exercised here in regard to what we want to count as a mental 'representation' of a language. If some features of the language parsed are determined by features of the parser, for example, it is not clear that it makes sense to say that those features are represented by the parser, or that we can say anything sensible about the size of such a representation (Stabler, 1983).

[9]See, e.g., Constable and Hartmanis (1972); Meyer (1972).

Suppose that Berwick is right about this, that the increasing modularization of Chomskian linguistics is a manifestation of the quest for maximal succinctness in the formal specification of the language. This account of inter-level modularization is of particular interest, since the motivating succinctness considerations are obviously significantly different from considerations of efficient processing. This puts us in a position to see exactly why our earlier worry about the motivations for levels of representation in grammar, (2), might well be exactly on target:

> (7) It would not be surprising if the levels of representation posited by maximally succinct grammars for human languages were not formulated by the human parsing procedure, or by any reasonably efficient parser.

Borgida (1983, p. 30) notes that there are problems which would arise if his two level "stratificational grammars" were used in "practical systems for natural language processing"; *viz.*, some languages definable by these very succinct grammars have no feasible recognition algorithms at all (1983, § 4.4).[10]

Consider the deterministic parser proposed by Marcus. It builds S-structures without first formulating D-structures; it collapses the inter-level modularity. It also achieves a considerable gain in efficiency by formulating a particular linguistic structure in a single left-to-right pass, without backtracking. But to do this, each of its moves must satisfy all of the independent constraints on the level of representation being formulated. Rather than overgenerating and using the independent constraints to filter out unacceptable structures, everything must be brought to bear at once. This deterministic approach is apparently the more resource efficient. Intra-level modularity is not exploited.[11] The relation between levels of representation posited by a maximally succinct grammar and levels of representation formulated by a reasonably efficient or psychologically realized parser remains mysterious. In this framework, though, the problem is at least precisely defined.

---

[10]Berwick and Weinberg (1982, 1984) point out that such resource complexity results need to be treated carefully, since we know that competent speakers themselves are unable to recognize their languages efficiently in the worst cases. Deeply center-embedded and garden path constructions make that clear. The upshot of this is not that resource complexity results are useless, but rather that high complexity in the worst cases is acceptable, so long as reasonable complexity in the more familiar cases is available. So the general point being made is not threatened: the concern is that the neglect of resource complexity and the concern with size complexity will not yield psychologically plausible theories.

[11]Berwick seems to suggest this view also. In his discussion of how GPSG theories assiduously avoid modularization, he notes,

> ... one way to view Gazdar's system is as a kind of "analyzer" or "parser" for sentences, rather than as a theory of grammar. Gazdar's grammar simulates the states of a parser as it moves through a string, left-to-right, checking for well-formedness. (1982, p. 391)

Berwick and Weinberg have discussed some of these issues. One thing they point out is that there will not *always* be a trade-off between succinctness and run-time efficiency. Citing Joshi *et al.*'s (1980) result on the parsing of (context sensitive) 'local constraint grammars' with a polynomial time extension of the Earley algorithm, they say: "… we are claiming that it is *possible* that a more compact grammar, expressed in a more powerful formal system, is more efficiently processed as well." (1982, p. 171; Berwick, 1982, p. 375–376) This point may be granted, but it obviously does not suffice to solve our problem until there are some grounds for believing that the concern with succinctness in grammars for *human languages* is going to yield a notation that is efficiently usable, or at least one that is as efficiently usable as the human notation is.

Chomsky's work suggests two different motivations for the emphasis on simplicity. The first is that formally succinct theories are somehow "more explanatory" than less succinct ones, and the second is that succinct grammars are easy to learn. Berwick (1982) adopts both of these widely accepted views, and takes the daring step of considering how they might be provided with a reasonable, if not rigorous defense. He attempts to defend both in the framework of the theory of program size complexity. He claims that (iv)

> The succinctness of a grammar, its "size" or "descriptive" complexity, is a measure of its explanatory power: "… a description of a set of surface data expresses a *regularity* or constitutes an *explanation* if the description of the data is shorter than a simple list of the data itself" (1982, p. 218)

and (v),

> The succinctness of a grammar is a measure of its ease of acquisition. Since children acquire impressive linguistic abilities with ease, we should aim to find succinct representations of the grammars acquired.

Certainly we want our linguistic theories to explain as much as possible, so succinctness is important if this really is an indicator of explanation. The constraint of feasible acquisition is also well motivated; whatever grammars humans use must be usable *and* learnable. Berwick tries to show that these considerations motivate the emphasis on succinctness.

Succinctness and explanation

It is tempting to think that both of these claims about succinctness can be effectively refuted from one's armchair. The obvious problem is that 'succinctness' or 'size complexity' must be relative to a particular 'size' measure defined over a particular notation. In order to apply these notions at all to different notational systems, we need a very abstract idea of what a grammar is. For example, (interpreting the grammar as a first level theory) we might think of a grammar as a Turing machine that generates n-tuples of linguistic representations. We can enumerate all Turing machines, assigning each a unique 'index' or 'Godel number.' A particular grammar can then be represented by the index of a Turing machine. There will be infinitely many Turing machines generating exactly the same sets of n-tuples of linguistic representations, but we could let the size of the grammar be the smallest index of a machine generating a particular set.[12] We could, but why should we? If our size measure is going to play any significant methodological role, surely its choice should not be arbitrary. But here we have arbitrariness introduced in the indexing of the machines and in the particular definition of the size measure over these indices. It would be astounding if linguists' intuitions about size corresponded to any natural size measure defined over Turing machine generators. We could enforce a correspondence by tailoring our size measure to the linguists' intuitions about succinctness, but then we would apparently still be lacking an empirical justification of the emphasis on this particular notion of succinctness. What we want is some justification of the idea that the linguists' idea of simplicity is tailored to relevant parameters of the acquisition problem (or at least to *something* relevant to human psychology).

Berwick notes the similar point in the context of acquisition theory. We can think of the acquisition device as a universal Turing machine which, on the basis of some limited input, acquires a program which generates a language. (The program itself may be thought of as the index of a Turing machine that generates the language.) The size of the program then just depends on what encoding of programs some particular machine uses. What can be encoded with a short program on one machine may require a longer program on another machine. Berwick locates the difficulty in the fact that we are talking about universal machines:

> With respect to the universal notation system, compact representations of [two different grammars, *i* and *j*] differ only with regard to whether the program for notational system *i* appears before that for system *j*. But this enumeration order is arbitrary—we could just as well have picked an enumeration ... that lists *j* before *i*. In this sense, a sufficiently powerful notation system renders the de-

[12] I follow Berwick in defining size in terms of a minimum size. In the general case, there is no recursive function that determines the minimal size of a program (for *any* standard size measure) (Blum, 1967). Berwick points out that we are not dealing with the general case (1982, p. 240, n 16), but the minimization problem is not even recursive for very weak systems (like context-free grammars, as shown by Taniguchi and Kasami (1970)). Of course there is no problem if the set of grammars is finite (see below), but if not, the proper response to this problem is, I think, to note that while size complexity cannot be effectively computed for many cases, it can be effectively approximated even in the general case to any desired accuracy (cf. Zvonkin and Levin (1970, theorem 1.5)).

> mand to fix notational machinery moot. This result would seem to indicate that a size complexity measure is uninteresting for "complete" computational systems. However, this invariance is *not* necessarily preserved by shifts to "weaker" notations, e.g., those that can specify grammars only for context-sensitive languages, or even only just some restricted subset of all the recursively enumerable languages (as is the case with modern transformational theories that can only specify a finite number of "programs," even if these generate languages that are perhaps not context free or context sensitive). Presumably this is the situation that holds with regard to linguistic theories: a desideratum of current theories of grammar is that they should be tightly constrained, specifying only a finite number of grammars, in some cases. In short, current theories of grammar are *not* universal programming systems; hence the results suggesting that program size measures are useless do not apply. (1982, p. 221)

Let us consider the suggestions here one at a time.

It should be noted first that the 'arbitrariness result' mentioned above does *not* depend on the universality of the programming system. Consider, for example, Turing machines with finite tapes (that is, a class of Turing machines with a particular finite state control—or finite state controls with less than some particular finite number of states—and with tapes less than some finite number of cells long). These are not universal acquisition devices. Obviously, they will only be able to acquire finitely many different languages, since they can only encode finitely many different programs.[13] And yet two such acquisition devices could use different encoding schemes for the same grammars, and so we get exactly the same result: the choice of any particular measure of size complexity is arbitrary.

Berwick claims that this arbitrariness is not necessarily preserved by shifts to weaker representation systems. This is a mistake. No matter how we restrict the class of languages that can be generated, there will be different machines that can generate all and only those languages, and there will be different size measures that can be applied to those machines. We will not be able to compare the sizes of grammars without fixing on a particular notation (or particular machine realization) and a particular measure of size complexity.

If the set of grammars we want to represent is finite, as Berwick suggests, then the problem becomes perfectly obvious. Suppose there are n possible grammars. Then there is obviously a partial recursive function which assigns

[13]Notice that these programs would be able to generate languages that are not regular, since the set generated by the program is the set that it would generate if it were run without limitations on memory or time. This is the situation we finite human beings are presumably in when we know a natural language or when we know how to multiply. In spite of our finitude we can realize procedures which would be able to generate sets that are not regular.

each a distinct numeral; they might be numbered from 0 to n in a base n system, for example, and then each grammar could be represented by a single digit.[14]

Berwick suggests that an explanation just is a relatively succinct description of some set of data. He finds similar ideas in Chomsky:

> We have a generalization when a set of rules about distinct items can be replaced by a single rule (or, more generally, partially identical rules) about the whole set ... (quoted in Berwick (1982, p. 218))

and in Halle:

> N is a natural kind if fewer features are required to designate the class N than to designate any individual sound in N. (quoted in Berwick (1982, p. 218))

Berwick says himself,

> ... the intuition that a generalization exists when a set can be replaced by a rule in fact exhausts the notion of generalization and defines what is or is not a natural class. In sum, what lies behind a generalization or regularity in a set of data is a pattern that permits a "compression" in the description of that set, beyond a simple list of its members (1982, p. 218).

This picture of scientific methodology is not peculiar to linguists, but it is now widely rejected.[15] Berwick and Weinberg use these ideas in criticisms of ap-

---

[14]Unless I am missing something here, it is a mistake for Berwick (1982, p. 240 ff; Berwick and Weinberg, 1982, p. 188 n 33) to think that he will find any support for his approach in the program size complexity theory developed by Kolmogorov and others, since the argument just sketched shows it has no way to provide meaningful comparisons of complexity between arbitrary members of a finite set. Their complexity with respect to any 'optimal' partial recursive function is obviously minimal.

[15]For example, Gregory Chaitin, well known for his work in algorithmic complexity theory, holds a similar view:

> ... the concept of complexity might make it possible to precisely define the situation that a scientist faces when he made observations and wishes to understand them and make predictions ... The simplicity of a theory is inversely proportional to the program that constitutes it. That is to say, the best program for understanding and predicting observations is the shortest one that reproduces what the scientist has observed up to that moment. Also, if the program has the same number of bits as the observations, it is useless, because it is too "ad hoc." ... In summary, the value of a scientific theory is that it enables one to compress many observations into a few theoretical hypotheses. There is a theory only when the string of observations is not random, that is to say, when its complexity is appreciably less than its length in bits (1974, p. 13, cf. also, 1966).

The relativity charge is met by proving the existence of "optimal" programming languages. Certain programming languages can be shown to be less than "optimal," then, in a certain absolute sense, though such comparisons must be rather coarse (cf. also Kolmogorov (1968, p. 298)). The view of science Chaitin suggests is vaguely reminiscent of some of the positivist theories of science which have been pretty well given up as hopeless in the past 20 years or so. I offer a very brief critique below, but for more thoroughgoing and effective criticism see, e.g., Boyd (1979) or Putnam (1978).

proaches in linguistics that propose grammars that have more rules than TG does.

The problem already raised for any such view is that whether something is or is not a regularity becomes a matter of which language it is described in. But fix the language, and there are still problems in saying what counts as a good generalization. To borrow an example from Goodman (1955), suppose that I have only coins in my pocket and we call them c1, c2, ..., cn. It is a fact that c1 is in my pocket and conducts electricity, that c2 is in my pocket and conducts electricity, and so on. More succinctly, everything in my pocket conducts electricity. This is more succinct, but it is not a good empirical generalization that supports counterfactuals (like, "If that emerald were in my pocket, it would conduct electricity"), nor is it an explanation of anything. So being a succinct description of a larger set of data is obviously not a sufficient condition for it to qualify as "explanatory" or as a "generalization." Just as obviously, it is not a necessary condition. A good explanation of why c1 conducts electricity may well be a very long and complicated story. Its virtues as an explanation are not judged by its succinctness but by its veracity as a causal account of the phenomenon in question.

Succinct descriptions of sets of sentences in a language are not *ipso facto* explanatory. If we want to explain human linguistic abilities, we want to find features of the language that have some significance in the causal account of linguistic abilities. It is no surprise that many of the 'regularities' or 'generalizations' about language are plausibly exploited by the human linguistic processor, but it does not motivate a succinctness methodology.

Of course, every science does seek simple, elegant explanations of the phenomena in its domain. There is more to this, though, than seeking succinct descriptions of the data. Chomsky seems to be aware of this when he says,

> The problem is to devise a procedure that will assign a numerical measure of valuation to a grammar in terms of the degree of linguistically significant generalization this grammar achieves. The obvious numerical measure to be applied to a grammar is length, in terms of number of symbols. But if this is to be a meaningful measure, it is necessary to devise notations and to restrict the form of rules in such a way that significant generalizations of complexity and generality are converted into considerations of length, so that real generalizations shorten the grammar and spurious ones do not (1965, p. 42).

The difficulty is to sort out the appropriate notation and size measure for a grammar of psychological relevance. When Chomsky proposes a shorter account of the auxiliary verb system in English, he speculates about the role of the generalization in language learning. This leads us to the second proposal, that 'shorter' grammars, in some specifiable sense of 'shorter,' are easier to learn. The first idea, that shorter grammars are just more explanatory or more likely to be correct, is untenable. Even with arbitrarily weak representational systems, the choice of notation and size measures needs to be motivated. The only way to respond to the worry about linguistic methodology is to show that they are.

Succinctness and learnability

The rough idea we want to flesh out is that if the grammar of a language is 'short,' then the problem of computing it (from certain input) should be fairly 'easy.' But Berwick notes himself that this is not right. He says:

> ... a strict program size measure is not quite the right one. For if acquisition complexity really measures the difficulty of selecting a grammar given input evidence about the language generated by that grammar, then a grammar could be very large and yet acquisition could be extremely easy, even trivial. This might be the case if, for example, there was but a single human grammar. Then literally no external information would be required to identify the grammar; it could be entirely "built in," with no obvious size limitation. (1982, p. 243)

In the terms of the earlier model, this trivial case would presumably be one in which the acquisition device comes all ready with a representation of a particular grammar. Obviously, the human language acquisition system is more flexible, but it nevertheless seems at least 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 (in any natural sense of 'size') and their learnability.

Let us distinguish, then, the part of the grammar which is presumed to be built into the acquisition device, call it 'UG,' from the part that must be acquired. It is clear that Chomskian linguistics has pursued succinctness in its description of *both*. Berwick notes that we might assume that "the space available for specifying UG is limited" (1982, p. 245). Of course it is limited, but there does not seem to be any good reason to suppose that it is *so* limited as to require a maximally succinct UG, especially if the succinctness is going to exact a price at runtime. So with respect to UG, objection (5) still stands.

Perhaps we can still hold something like (v) with respect to what is acquired though. That is, we want to keep the acquisition problem as simple as possible, so in addition to putting as much in UG as is compatible with the evidence, we want to keep the problem of acquiring whatever else is needed as simple as possible. 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. We can capture the Chomskian strategy for simplifying the acquisition problem by minimizing this 'developmental' or 'decision' complexity of grammars. Berwick suggests that this does correspond to a strategy of minimizing a certain program size complexity:

> ... we are to penalize a system that requires a large number of decisions to select the proper grammar, given some initial distribution of possible grammars ... Put another way, the developmental measure attempts to minimize program size, but where "program" is now the developmental program that starts from an initial set of possible grammars and selects a correct final state grammar corresponding to the linguistic competence of the adult community. (1982, p. 243)
>
> If we identify acquisition complexity with the amount of information that must be specified in order to fix the grammar, a specification which is bounded from below by the number of non-terminals in a grammar, then the succinctness gains of a phrase structure representation translates into a significant lessening of the acquisition burden. (1982, p. 361)

This last idea is fairly clear. We can think of the acquisition device as a function LT which maps input into grammars. We want to find a function LT which can effect the appropriate mapping with the smallest amount of input. We can put this into the framework of size complexity theory by thinking of the input as a 'program' which is run by LT to produce a representation of the grammar. We can then define the complexity K of any particular grammar Gi in a set G of grammars in terms of the size of the 'program' needed to generate it from an input. In the standard notation,

$$K_{LT}(Gi) = \left\{ \begin{array}{l} \text{the minimum size input p such that } LT(p) = Gi \\ \text{infinite if there is no such p.} \end{array} \right.$$

In this case, though, we may be able to provide some psychological motivation for relevant aspects of the definition. We can gather empirical evidence about what sorts of input the language learner actually uses, and we can define some size measure over it. Obviously, this size measure will want some psychologically appropriate motivation. The input *might* be, for example, a set of well-formed sentences, and its size might simply be the number of sentences in the set. We can get evidence about what sort of grammar is acquired, then, and project from this evidence in ways that will not lead us to exceed any reasonable 'complexity' in the sense defined. The hope is that this might lead us to grammars that are as easy to learn as human languages are. Now let us consider the idea that more 'succinct' grammars will be less complex in the sense that they will be specifiable with 'smaller' inputs, where the smallness is judged in some psychologically plausible way.

Berwick points out that there are, in fact, some formal results relating input size and grammar size. For example, we can consider functions from subsets of regular languages to finite state automata generating languages which include those subsets. Gold (1967) showed that if G includes all finite state automata, and if the inputs are just finite sets of strings in the language generated by the finite state automaton to be identified, then there is no LT which will always converge on an appropriate automaton. The learning problem is manageable, though, if we know that the finite state automaton to be identified will have less than some finite number, N, of states.[16] Berwick points out that an N-state finite automaton is completely characterized by the set of strings of length 2N − 2 or less that it accepts. Obviously, that is a large number of strings in some cases, but it may nevertheless provide the basis for some inductive methods, some functions, LT, which can correctly characterize a finite automaton on the basis of a subset of the language it accepts. We know that the set of all strings of length less than 2N − 2 is enough, and Angluin (1982) has recently shown that a smaller set will always suffice—a set whose minimum size is a polynomial function of the number of states of the target finite state automaton.

All of these results are rather remote from the acquisition problem for natural languages, but they indicate the kind of connection between grammars and ease of acquisition that would be relevant. Berwick argues that some of the finite state automaton induction methods can be adapted for the induction of the tree automata corresponding to context-free grammars. And he considers how natural language parsers of the sort he proposes (parsers like Marcus's) could be induced using a generalization of one of the finite state automaton methods.

All of this gives us something considerably less than a general defense of the idea that succinct grammars can be learned with 'small' inputs, for any relevant sense of 'small.' We have seen that a relation of this sort holds in some fairly simple formal cases, and Berwick's own acquisition theory may similarly require a number of (distinct) sentences that is some function of the number of nonterminal categories in the parser grammar. There are a number of points to note, though. In the first place, the available results relating the number of nonterminals in a grammar and the number of sentences needed for acquisition are not very robust. If such a relation holds in Berwick's acquisition theory, it is not obvious that it will in any other.[17]

---

[16]Notice that, with a finite upper bound on the number of non-equivalent states, the number of regular languages over a particular finite alphabet is finite, and so with respect to an 'optimal' partial recursive programming language (or 'learning theory'), they are all of equal complexity. Every one can be represented by a one symbol program.

[17]For example, it is far from clear that Berwick's theory can yet make any substantial claim to greater learnability than GPSG can make simply on the grounds that GPSG has more non-terminals. GPSG theorist

In the second place, the use of succinctness arguments in linguistics has certainly not been limited to keeping down the number of non-terminals. Consider Chomsky's argument about the rules appropriate for the auxiliary, mentioned above. So again, the notion of 'succinctness' that is used in linguistics now may not be one that is related to ease of acquisition. Certainly the exact nature of the relation is far from obvious. The appropriate measures of "ease of acquisition," and their relations to grammars and parsers are still quite obscure.

And finally, it should be recalled that large parts of the grammars or parsers humans use, large parts of these mechanisms which define our languages, may not be learned at all, and if they are not learned, we have not been offered *any* grounds for thinking that they are succinct rather than, say, very fast at certain important tasks. The relation between what is learned and what is used in language understanding may be fairly subtle. These are points that Berwick and Weinberg have endorsed, but they threaten the prospects of providing a psychological motivation for the methodology of Chomskian linguistics.

5. Conclusions

In conclusion, I think that it is safe to say that Chomskian linguistic theory is not a first level computational theory. That interpretation of linguistics faces a number of fairly obvious difficulties that do not face a theory like Marr's. The reason that the psychological import of TG has been challenged stems partly from these difficulties, and partly from the difficulty of understanding the methodology of TG as one appropriate to psychology. Berwick does not, I think, succeed in showing that the emphasis on simplicity in linguistics is appropriate for the psychological project of accounting for language acquisition and use, but he does indicate a number of interesting considerations that are worth pursuing. Even if there is a robust relation between succinctness and learnability, as Berwick and Chomsky apparently suspect, some of the challenges to their claims certainly arise because the relation is less than perfectly transparent. It is not as easy to understand as, for example, the emphasis on efficiency in Marr's work. Alternative methodologies in linguistics that try to stay closer to psychological models may well be appropriate. There is no good reason to think that an approach is less explanatory just because its grammars are larger. Of course, every approach will want to

---

will certainly try to formulate a different acquisition theory in which the number of non-terminals in the grammar does not enlarge the required sample size at the rate it would in Berwick's theory.

produce a competitive learning theory, and the work of Berwick and Weinberg makes the competition stiff.

References

Angluin, D. (1982) Inference of reversible languages. *J. Assoc. Comp. Mach.*, *29*, 741–765.

Berwick, R. (1982) Locality Principles and the Acquisition of Syntactic Knowledge. Ph.D. dissertation, MIT Department of Computer Science and Electrical Engineering.

Berwick, R. and Weinberg, A. (1982) Parsing efficiency, computational complexity, and the evaluation of linguistic theories. *Ling. Inq.*, *13*, 165–191.

Berwick, R. and Weinberg, A. (1983) The role of grammars as components of models of language use. *Cog.*, *13*, 1–61.

Berwick, R. and Weinberg, A. (1984) *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA, Bradford Books/MIT Press.

Blum, M. (1967) On the size of machines. *Inf. Contr.*, *11*: 257–265.

Borgida, A.T. (1983) Some formal results about stratificational grammars and their relevance to linguistics. *Math. Syst. Theory*, *16*: 29–56.

Borodin, A. (1873) Computational complexity: theory and practice. In A.V. Aho (ed.) *Currents in the Theory of Computing*. Englewood Cliffs, N.J., Prentice-Hall.

Boyd, R. (1979) Metaphor and theory change. In A. Ortony (ed.) *Metaphor and Thought*. Boston, Routledge.

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

Bresnan, J. and Kaplan, R. (1982) Introduction: grammars as mental representations of language. In J. Bresnan (ed.) *The Mental Representation of Language*. Cambridge, MA: MIT Press.

Chaitin, G.J. (1966) On the length of programs for computing finite binary sequences. *J. Assoc. Comp. Mach.*, *13*, 547-569.

Chaitin, G.J. (1974) Information theoretic computational complexity. *IEEE Trans. Inf. Theory*, *IT-20*: 10–15.

Chomsky, N. (1851) Morphophonemics of Modern Hebrew. M.S. dissertation, Department of Linguistics, University of Pennsylvania.

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

Chomsky, N. (1980) Rules and Representations. *Behav. Br. Sci.*, *3*, 1–61.

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

Chomsky, N. (1982) *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA, MIT Press.

Constable, R.L. and Hartmanis, J. (1972) Complexity of formal translations and speed-up results. *Proceedings of the IEEE Symposium on the Foundatuons of Computer Science*.

Engdahl, E. and Ejerhed, E. (1982) *Readings in Unbounded Dependencies*. Umea, Umea Studies in the Humanities.

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

Fodor, J.D. (1983) Phrase structure parsing and the island constraints. *Ling. Phil.*, *6*, 163–223.

Gazdar, G. (1982) Phrase structure grammar. In P. Jacobson and G.K. Pullum (eds.) *The Nature of Syntactic Representation*. Boston, Reidel.

Gold, E.M. (1982) Language identification in the limit. *Inf. Contr.*, *10*: 447–474.

Goodman, N. (1955) *Fact, Fiction and Forecast*. New York, Bobbs-Merrill.

Joshi, A.K., Levy, L.S. and Yueh, K. (1980) Local constraints on programming languages. *Theor. Comp. Sci.*, *12*, 265–290.

Kolmogorov, A.N. (1968) Three approaches to the definition of the concept of the "amount of information". *Selected Translations in Mathematical Statistics and Probability*, 7, 293–302.

Marcus, M. (1980) *A Theory of Syntactic Recognition for Natural Language.* Cambridge, MA, MIT Press.

Marcus, M., Hindle, D. and Fleck, M. (1983) D-theory: talking about trees. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics.*

Marr, D. (1976) Representing visual information. In W.E. Reichardt and T. Poggio (eds.) *Theoretical Approaches in Neurobiology*. Cambridge, MA, MIT Press.

Marr, D. (1977) Artificial intelligence—a personal view. *Artific. Int.*, *9*, 37–48. (This also appeared as MIT AI Memo 355.)

Marr, D. (1979) Representing and computing visual information. In P.H. Winston and R.H. Brown (eds.) *Artificial Intelligence: An MIT Perspective*, Vol. 2. Cambridge, MA, MIT Press.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, Freeman.

Meyer, A.R. (1982) Program size in restricted programming languages. *Inf. Contr.*, *21*, 382–394.

Putnam, H. (1978) Meaning and the Moral Sciences. Boston, Routledge.

Stabler, E.P. (1983) How are grammars represented? *Behav. Br. Sci.*, *6*, 391–421.

Taniguchi, K. and Kasami, T. (1970) Reduction of context free grammars. *Inf. Contr.*, *17*, 92–108.

Zvonkin, A.K. and Levin, L.A. (1970) The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russ. Math. Surv.*, *25*, 82–123.