North East Linguistics Society

Volume 24
Issue 1 *NELS 24: Volume One*

Article 9

1994

Nominative Case Assignment and Parameter-Setting in V2

Edward Gibson
Massachusetts Institute of Technology

Kenneth Wexler
Massachusetts Institute of Technology

NOMINATIVE CASE ASSIGNMENT AND PARAMETER-SETTING IN V2[1]

Edward Gibson
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
gibson@psyche.mit.edu

Kenneth Wexler
Department of Brain and Cognitive Sciences
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
wexler@psyche.mit.edu

1. Introduction

In recent work, we have formalized the idea that grammatical acquisition can take place by the use of *triggering sentences*: sentences which can allow the learner to correctly set a parameter (Gibson and Wexler, 1992; 1994; henceforth GW). We have shown that although the simple algorithm that re-

[1]We would like to thank the following people for their comments on presentations of this work: Masha Babyonyshev, Kevin Broihier, Ana Cardinaletti, Dominique Estival, Danny Fox, Dawn MacLaughlin, Oren Percus, Colin Phillips, Paul Smolensky, Michal Starke, and audiences at the University of Geneva computational linguistics seminar and NELS 24. We would especially like to thank Alessandra Tomaselli, who suggested the idea which led to the exploration in this paper.

North East Linguistics Society, Vol. 24 [1994], Iss. 1, Art. 9

sults from this formalization is guaranteed to converge on the target grammar if triggers are always present, it need not converge if triggers are not always available. A linguistically natural parameter space involving X-bar parameters together with a verb-second (V2) parameter was demonstrated to pose exactly this problem: there exist source grammars in this space for which there are no triggers for incorrectly-set parameters, so that the learner will never converge on the target from such a state using this algorithm. In this paper we present a new prospective solution to this problem, to be compared with the solutions that were proposed earlier. The new solution hinges on the observation that two of the grammars in the original parameter space are grammars for which there may be very few, if any, grammatical sentences. If these grammars can therefore be ruled out by UG principles, the learner may not need to consider them as prospective targets. The learner may thus be able to "see through" these grammars, to grammars one additional step away, and thus trigger out of grammars that had previously been impossible to escape.

This paper is organized as follows. Background information is given in Section 2. Section 3 presents the new solution, and some conclusions are given in Section 4.

2. Background

Within the principles and parameters framework (Chomsky, 1981; 1986), it is assumed that Universal Grammar (UG) consists of a set of universal principles together with a finite number of parameters, each with a finite number of possible values. Linguistic variation is therefore explained by the existence of multiple parameter values. Furthermore, grammatical acquisition is greatly simplified in this framework, relative to an unconstrained framework, because so little has to be acquired: just the target values of the parameters.

Although much work has gone into the exploration of the universals and parametric variations that exist across languages, comparatively little work has been done in trying to discover what the algorithm for learning is within such a framework. Two general constraints on possible algorithms are as follows:

- The learning algorithm must be able to converge on the target grammar given severe resource limitations. Thus algorithms that require large quantities of memory and/or computation at each learning step are ruled out. As a first approximation, we hypothesize that the learner has no memory of previous parameter settings. Furthermore, we hypothesize that a minimal amount of search takes place when the learner attempts to make changes in her hypotheses.

- Second, because there is no evidence that children make large scale reorganizations in their grammars in short periods of time, any model of the child's learning algorithm should have the property of *conservatism*, under which the learner tries small grammatical changes before she attempts larger changes.

Although not adequately formalized until recently, it has often been tacitly assumed in the (psycho)linguistic literature that grammatical acquisition of non-subset parameter values[2] takes place by the use of linguistic *triggers*: sentences from the target grammar that are grammatical if and only if a certain parameter P is set to its target value.[3] In fact, this assumption is part of the conceptual foundations of the Principles and Parameters approach because it seems to solve the problem of explanatory adequacy, of learnability, so neatly. The idea is that a child would hear such a triggering sentence, and if she had not yet arrived at a grammatical representation for this sentence, then she could change the value associated with parameter P to its target value and thereby get closer to the target grammar. Of course, she might also consider changing the value of other parameters, but these alternative changes wouldn't do any good, because the triggering sentence demands that P be set correctly. The rough idea is that by encountering enough of these triggering sentences, and by eventually choosing the right parameter to change for each, the learner would eventually acquire all of the target parameter settings. A formalization of this idea was provided in GW, in the form of the Triggering Learning Algorithm (TLA):

1. Input sentence S.
2. If S can be syntactically processed, then do nothing.
   Else do the following:
   (a) **Single Value Constraint**: Change one parameter value and reprocess S.
   (b) **Greediness Constraint**: If S can now be syntactically processed, then adopt the new parameter value.
   Else retain the old value.

[2]For discussion of the acquisition of subset parameter values, see Angluin 1978, Williams 1981, Berwick 1985, Manzini and Wexler 1987, Wexler and Manzini 1987, and Clark 1992.

[3]Triggers may be thought of as either *local* or *global*, depending on whether the sentence is considered with respect to a particular source grammar (local), or with respect to all possible source grammars (global). However, we demonstrated in GW that global triggers probably don't exist for most parameter values, because of the interdependence of parameters in a system. Local triggers, on the other hand, seem to exist for most parameter settings in linguistically natural parameter spaces. (See below.)

Because the TLA relies on the notion of trigger from the linguistic literature, it is linguistically natural. Furthermore, the TLA satisfies the psychological constraints given above. In particular, (1) no memory of previous parameter settings is required; (2) no search is necessary when an unparsable sentence is encountered: a single parameter change is tried at random; and (3) a newly proposed grammar is only one parameter change away from its source grammar, so that changes are as small as possible.

In GW, we showed that if triggers exist with sufficient frequency for all parameter values, no matter what the source and target grammars are, then the TLA is guaranteed to converge on the target grammar in the sense of Gold (1967).[4] However, if triggers are not present for all parameter values, then the learner using the TLA might not be able to converge on the target. In GW we provided an example of a linguistically-natural parameter space which contains parameter settings for which there are no triggers for any of the incorrectly-set parameters. This parameter space contains three parameters:

- Two base word order X-bar parameters (cf. Koopman 1983, Hoekstra 1984, Travis 1984):
  - The specifier-head parameter with values spec-first and spec-final;
  - The complement-head parameter with values complement-first and complement-final.
- A verb-second (V2) parameter, with values +V2 and -V2, which indicates that a finite verb moves from its base position to the second position in root declarative clauses (see Bach 1962, Bierwisch 1963, Thiersch 1978, den Besten 1983, Travis 1984, and Haider and Prinzhorn 1986 among many others).

The sentence patterns to be considered within this parameter space — which are presented in Figure 1 — consist of the possible root declarative orderings of main verbs, auxiliary verbs, subjects, two objects, and sentence-initial adverbs.[5] This collection of constituents was selected because it allowed each grammar to be distinguished from every other, with no grammar the subset of any other.

This parameter space contains six source/target grammar pairs such that for each of these pairs, there is no path via triggering data from the source to the target. We refer to the source grammars in these pairs as *local maximum*

[4]The result follows for either a global or local definition of trigger.

[5]See Gibson and Wexler (1994) for more details on how particular sentence patterns are derived in each grammar.

Parameter Settings	Data in Defined Grammar
Spec-Final Complement-final −V2 (VOS)	V S, V O S, V O1 O2 S Aux V S, Aux V O S, Aux V O1 O2 S, Adv V S Adv V O S, Adv V O1 O2 S, Adv Aux V S Adv Aux V O S, Adv Aux V O1 O2 S
Spec-Final Complement-Final +V2 (VOS+V2)	S V, S V O, O V S, S V O1 O2, O1 V O2 S, O2 V O1 S S Aux V, S Aux V O, O Aux V S S Aux V O1 O2, O1 Aux V O2 S, O2 Aux V O1 S Adv V S, Adv V O S, Adv V O1 O2 S Adv Aux V S, Adv Aux V O S, Adv Aux V O1 O2 S
Spec-Final Complement-First −V2 (OVS)	V S, O V S, O2 O1 V S V Aux S, O V Aux S, O2 O1 V Aux S, Adv V S Adv O V S, Adv O2 O1 V S, Adv V Aux S Adv O V Aux S, Adv O2 O1 V Aux S
Spec-Final Complement-First +V2 (OVS+V2)	S V, O V S, S V O, S V O2 O1, O1 V O2 S, O2 V O1 S S Aux V, S Aux O V, O Aux V S S Aux O2 O1 V, O1 Aux O2 V S, O2 Aux O1 V S Adv V S, Adv V O S, Adv V O2 O1 S Adv Aux V S, Adv Aux O V S, Adv Aux O2 O1 V S
Spec-First Complement-Final −V2 (SVO)	S V, S V O, S V O1 O2 S Aux V, S Aux V O, S Aux V O1 O2, Adv S V Adv S V O, Adv S V O1 O2, Adv S Aux V Adv S Aux V O, Adv S Aux V O1 O2
Spec-First Complement-Final +V2 (SVO+V2)	S V, S V O, O V S, S V O1 O2, O1 V S O2, O2 V S O1 S Aux V, S Aux V O, O Aux S V S Aux V O1 O2, O1 Aux S V O2, O2 Aux S V O1, Adv V S Adv V S O, Adv V S O1 O2, Adv Aux S V Adv Aux S V O, Adv Aux S V O1 O2
Spec-First Complement-First −V2 (SOV)	S V, S O V, S O2 O1 V S V Aux, S O V Aux, S O2 O1 V Aux, Adv S V Adv S O V, Adv S O2 O1 V, Adv S V Aux Adv S O V Aux, Adv S O2 O1 V Aux
Spec-First Complement-First +V2 (SOV+V2)	S V, S V O, O V S, S V O2 O1, O1 V S O2, O2 V S O1 S Aux V, S Aux O V, O Aux S V S Aux O2 O1 V, O1 Aux S O2 V, O2 Aux S O1 V Adv V S, Adv V S O, Adv V S O2 O1 Adv Aux S V, Adv Aux S O V, Adv Aux S O2 O1 V

Figure 1: Data in the V2/X-bar parameter space.

Source Grammar	Target Grammar
VOS+V2	SVO
VOS+V2	SOV
OVS+V2	SVO
OVS+V2	SOV
SVO+V2	OVS
SOV+V2	OVS

Figure 2: Local maxima for the V2/X-bar parameter space.

grammars. If the learner somehow ends up a local maximum, she will never converge on the target via the TLA. The six local maximum/target grammar pairs are given in Figure 2.

For example, the grammar VOS+V2 ([spec-final, complement-final, +V2]) is a local maximum relative to the grammar SVO ([spec-first, complement-final, −V2]). This means that if the learner happens to hypothesize the grammar VOS+V2, then she will never converge on the target SVO. In fact, she will never change her hypothesis from VOS+V2. To see this, consider the data from the target SVO that the learner might encounter (see Figure 1). A number of these sentence patterns are grammatical in the source VOS+V2: *S V*, *S V O*, *S V O1 O2*, *S Aux V*, *S Aux V O*, and *S Aux V O1 O2*. If the learner encounters one of these sentence patterns, she will not change her hypothesis from VOS+V2, because she can already build a grammatical structure for each given her current grammar hypothesis. Hence these sentences will not cause the learner to change her hypothesis.

If presented with any of the other sentence patterns from the grammar SVO, the learner abiding by the TLA will attempt a parameter value change to see if the new hypothesis allows a grammatical analysis of the sentence. But each sentence from SVO which is ungrammatical in VOS+V2 is also ungrammatical in each of the grammars that differ from VOS+V2 by one parameter value: VOS, SVO+V2, or OVS+V2. For example, the pattern *Adv S V O* from grammar SVO — which is not grammatical in VOS+V2 — is also not grammatical in any of the grammars VOS, SVO+V2, or OVS+V2. Thus the learner who adheres to the TLA (in particular, the Greediness component of the TLA) and who ends up in VOS+V2 will never converge on the target SVO.

A number of prospective solutions to this problem are proposed in GW. Perhaps the most interesting of these solutions is a markedness/parameter ordering solution. In this solution the learner starts with her V2 parameter set to −V2 and with no values for either of her X-bar parameters (or equivalently, with both X-bar parameters set to to the special value *unset*), and proceeds according to the TLA. It turns out that the learner can never enter one of the local maxima from this initial state no matter what the target grammar is, so that the problem is circumvented. This solution works in large part because of an interesting property of the local maxima: they are all +V2 grammars relative to -V2 targets. Because of this property, it is possible to start in a general -V2 grammar and only hypothesize a +V2 grammar if the grammar is truly +V2.[6]

3. A new solution

Although the solution in which the lack of V2 is the unmarked setting seems natural and has a number of advantages, the study of whether parameter-setting works is new enough that we think it is useful to consider a variety of approaches to the problem of local maxima. Here we will explore an approach which relies on the following observation; some of the grammars in the V2/X-bar parameter space seem in fact not to be possible grammars. If the grammar space is smaller, perhaps the local maxima will cease to exist, even without a markedness solution.

In particular, Tomaselli (1989) adopts the idea from the V2 literature that in a V2 grammar, V (in Comp) assigns Nominative Case to Spec-IP under government. Tomaselli assumes that V must be *adjacent* to Spec-IP in order to assign Case to it (cf. Travis (1984), den Besten (1985), Platzack and Holmberg (1989)). Because Spec-IP is not adjacent to V (in Comp) in a spec-final grammar (see Figure 3), nominative Case can't be assigned to Spec-IP in a V2 grammar.

Thus, in a spec-final/+V2 grammar, the Case filter will rule out sentences with a Spec-IP that needs Case. Since this is all (or almost all) sentences in the grammar, we can thus assume that such grammars don't exist.[7] In particular, two grammars from the GW parameter space (see Figure 1) are ruled out, as in (1), leaving 6 grammars in the space:

(1)
a. * OVS+V2
b. * VOS+V2

[6]A possible problem with the simple unmarked state solution is that it is susceptible to noise: if certain ungrammatical sentence patterns occur in the input stream, they might shift the learner to a local maximum. A more robust solution is one in which, in addition to an unmarked state, there is a latency period associated with the possibility of altering the value of the V2 parameter.

[7]We thank Alessandra Tomasselli for suggesting this argument to us. We return at the end of this paper to a discussion of the reasonableness of assuming that a grammar is ruled out because the representations it generates are ruled out.

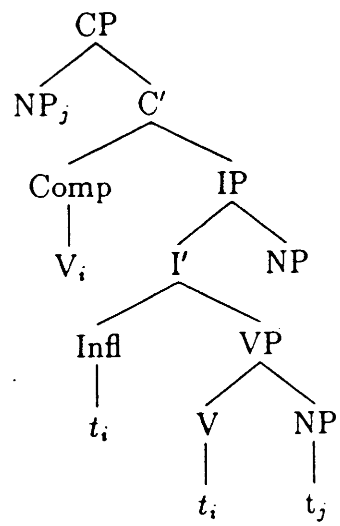

Figure 3: The phrase structure for a spec-final +V2 root clause.

The knowledge that the two grammars listed in (1) are ruled out might be given initially, in the statement of the grammatical space, or it might be calculated by the child (as a linguist would) from the initial knowledge state; we do not discuss the issues surrounding the choice of which of these two possibilities is correct.

Let us refer to the new parameter space, without the two grammars in (1), as GW′. The two grammars in (1) are the local maximum source grammars for 4 of the 6 local maxima listed in Figure 2. Thus only two of those local maxima remain:

(2)
a. (SVO+V2, SVO)
b. (SOV+V2, OVS)

Deleting grammars in the space has eliminated some local maxima, as anticipated. However, it turns out that deleting these grammars has also *added* two new local maxima:

(3)
a. (SOV+V2, VOS)
b. (SOV+V2, OVS)

For example, consider the source grammar SVO+V2 relative to the target VOS. There are three possible parameters that the learner could change, resulting in three possible new grammars: SVO, SOV+V2 and VOS+V2. However, there are no sentences (triggers) to take the learner to either SVO or SOV+V2. That is, there are no sentences from the target VOS that are in

SVO (SOV+V2) but are not in the source SVO+V2. There would be triggers for VOS+V2, but, by hypothesis, VOS+V2 is not an allowable parameter setting for the learner.

In other words, we have demonstrated that deleting grammars from the space has actually added some local maxima. This is because the grammars in (1) acted as "stepping stones" to other grammars, allowing triggers to work. Without the deleted grammars in (1), the Greediness and Single-Valued Constraints of the TLA cannot be satisfied with respect to these two additional cases. The learner is therefore stuck in grammars that previously (in GW) she could have moved on from. This is therefore a further illustration of the fact that the relation between size of grammar space and local maxima is not a simple one: making a search space smaller can have the result of making some learning problems more difficult. This is the result of natural constraints on the learning process as envisioned in the theory of parameter-setting assumed in the Principles and Parameters framework.

Hence simply using the smaller GW′ space doesn't eliminate the problem of local maxima. However, let us see if there is a way for the learning system to both take advantage of the reduction in some local maxima that the move to a smaller space allows while still retaining the possibility that the grammars that have been removed from the space can act as "stepping stones" for the learning system. The idea is to allow a new way to change a parameter value: a way which is licensed by the structure of the space of possible grammars rather than by the nature of the input. In other terms, we will consider the possibility of a *UG-induced* parameter change (in contrast to the *data-induced* parameter changes already assumed in the TLA). In particular, we seek to find a way to allow the learning algorithm to use the removed grammars as stepping stones, thus allowing the algorithm to take advantage of the reduction in grammar space size while not losing the power of the grammars as stepping stones.

Suppose that the knowledge of unallowable parameter combinations is represented as a filter on those combinations in the learner's hypothesis space, as in (4):

(4) * [spec-final, +V2]

We now add to the TLA the possibility of the following *UG-induced parameter change*:

(5) If a newly proposed grammar G is blocked by a UG filter, then in G randomly change the value of one of the parameters involved in the filter.

The idea of (5) is that the newly proposed grammar G, which is blocked by the UG filter, acts as a stepping stone on the way to a grammar in the space.

The change mechanism in (5) attempts to find the closest possible grammar to G which doesn't violate the UG filter. Note that the new grammar proposed by the learning step in (5) is guaranteed *not* to violate the filter, because it must differ from the filter in one of its parameters.[8] We consider the part of the learning algorithm described in (5) to represent a *UG-induced* change because it is a change driven not by the introduction of a new datum, but rather by the UG filter and the nature of the parameter space.

Let us add (5) to the TLA, and call the new learning algorithm TLAF (for TLA + Filter). Calculations show that for the new TLAF learning algorithm, there are no local maxima in the GW space with the filter in (4).

For example, consider the source SOV+V2 with respect to the target OVS. Suppose that the sentence *O V Aux S* is input, which is not grammatical in the current grammar. If the learner (randomly) selects the spec-head parameter for changing, then the resulting grammar will be OVS+V2, a non-UG grammar by (4). She then randomly chooses between the V2 and spec-head parameters (the two parameters in (4)), and changes one of these. If she happens to choose the spec-head parameter (again), then the resulting grammar is SOV+V2 (the initial grammar), and the learner tests to see whether the sentence is grammatical here (Greediness). A failure to parse results, and no changes are made. On the other hand, if she chooses to change the V2 parameter, then the resulting grammar is OVS (the target), and the sentence is now parsable, and the learner has converged. Thus a local maximum under the TLA, namely (SOV+V2, OVS) is no longer a local maximum under TLAF with filter (4). All other states that were local maxima before under the TLA are similarly no longer local maxima under the TLAF with filter (4).

Strictly speaking, the addition of (5) to the TLA allows for violations of the Single Value Constraint, since two parameter values are changed on presentation of an input: one change induced by the input under TLA, and a second change induced by (4). But recall that the original motivation for the Single Value Constraint within the Principles and Parameters framework was to satisfy the two cognitive constraints of resource limitations and conservatism, along with linguistic naturalness. First, let us consider the TLAF with respect to the conservatism constraint, which dictates that the algorithm for selecting parameter changes opt for the smallest changes. Allowing the UG-induced change in (4) still satisfies conservatism: we generalize from "Try the smallest changes" to "Try the smallest *useful* changes."

Now consider the constraint of resource limitations. No additional memory is required by the TLAF, and the computation in the TLAF is only slightly more complex than in the TLA. The only addition is an extra parameter change

[8]We assume that a filter is simply a filter on a conjunction of parameter values.

when the learner ends up in a UG-filtered grammar. However, this parameter change is automatic, involving no additional search or grammaticality tests compared to the TLA that lacks UG-induced changes.

We can contrast the TLAF with the following idea, which is guaranteed to converge:

(6) A search intensive extension of the TLA:
If all single-value changes don't work (i.e. they don't satisfy the Single-Value and Greediness Constraints), try random two-value changes. If the two-value changes don't work, try three-value changes, etc.

(6) satisfies conservatism but potentially requires a large amount of computational resources, because all grammars one parameter value away must be searched (testing for grammaticality) before a two-value change is made and so on. Under the TLAF, no search of this kind is involved.

4. Conclusions

Supplementing the TLA with a mechanism to allw UG-induced parameter changes solves the local maxima problem in the GW search space without the need to assume unmarked values, while still satisfying the resource limitation and conservatism constraints on the learning algorithm. More generally, we have demonstrated in this paper that removing grammars from the hypothesis space may in fact create new local maxima. Furthermore, impossible grammars in the hypothesis space might in principle be considered useful as "stepping stones" in setting parameters.

Which of the two solutions discussed in this paper is to be preferred, the original "non-V2 is unmarked" solution or the new "UG-filter" parameter-change solution? (Or perhaps one of the other solutions discussed in GW.) The UG-filter solution has the disadvantage that an additional step in the TLA is necessary. This may be a relatively minor disadvantage since, as we have pointed out, the computation of TLAF is still simple.

The "non-V2 is unmarked" solution has the advantage, as we have argued in Gibson and Wexler (1994), that it fits in well with current attempts to understand economy of derivations in linguistic theory (Chomsky 1991, 1993). There is another reason to prefer the "non-V2 is unmarked solution." Note that the assumption is that in a V2 language, the adjacency requirement on case assignment under government only rules out representations, as in Figure 3, not grammars. This is the ordinary mode of action of UG principles. A [+V2, spec-final] grammar is actually not ruled out by UG; rather it is simply that all or almost all derivations are filtered out.[9] Thus there is no direct way to calculate from UG that [+V2, spec-final] grammars are impossible, and hence there seems to be no UG motivation to create the filter in (4). Rather (4) is motivated on *functional* grounds. Namely, many, probably most of the sentences that would be useful are ungrammatical in a [+V2, spec-final] grammar. So it is clear that it is not very likely that a [+V2, spec-final] grammar would appear, because too many useful sentences would not be possible in such a grammar. But we would not expect the lack of *functional* usefulness to be represented as a filter (like (4)) in the space of possible grammars.

Thus we conclude that the original "non-V2 is unmarked" solution seems more likely to be correct than the TLAF with the filter in (4). Nevertheless, we have learned much from the attempt to formulate this new approach to the problem of local maxima. For example, we know that reduction in the parameter space may add new local maxima, and that UG-filters together with a certain natural expansion of the learning algorithm may remove local maxima. And it might turn out that in fact UG filters and the TLAF may play a useful role in the theory of parameter-setting, especially if filters can be derived that act on grammars rather than on representations.

5. References

Angluin, Dana. 1978. Inductive inference of formal languages from positive data. *Information and Control*, 45:117-135.

Bach, Emmon. 1962. The order of elements in a transformational grammar of German. *Language*, 38:263-269.

Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge, Mass.: MIT Press.

den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In *On the formal syntax of the Westgermania*, Papers from the 3rd Groningen Grammar Talks, Groningen, 1981, ed. Werner Abraham, 47-131. Amsterdam: J. Benjamins.

den Besten, Hans. 1985. The ergative hypothesis and free word order in Dutch and German. In *Studies in German grammar*, ed. J. Toman. Dordrecht: Foris.

[9] If, for example, it turns out that some expletive pronouns do not require case then a derivation such as in Figure 3 will be allowed with an expletive pronoun in Spec-IP.

Bierwisch, Manfred. 1963. *Grammatik des deutschen Verbs*, Berlin: Akademie-Verlag.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin and use.* New York: Praeger.

Chomsky, Noam. 1991. Some notes on the economy of derivation and representation. In *Principle and parameters in comparative grammar*, ed. Robert Freidin, 417-454. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 1-52. Cambridge, Mass.: MIT Press.

Clark, Robin. 1992. The selection of syntactic knowledge. *Language Acquisition*, 2:83-149.

Gibson, Edward and Kenneth Wexler. 1992. Parameter Setting, Triggers and V2. GLOW. Lisbon, Portugal.

Gibson, Edward and Kenneth Wexler. To appear, 1994. Triggers. *Linguistic Inquiry.*

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control*, 10, 447-474.

Haider, Hubert, and Martin Prinzhorn, eds. 1986. *Verb second phenomena in Germanic languages*, Dordrecht: Foris.

Hoekstra, Teun. 1984. *Transitivity.* Dordrecht: Foris.

Koopman, Hilda. 1983. *The syntax of verbs.* Dordrecht: Foris.

Manzini, M. Rita, and Kenneth Wexler. 1987. Parameters, binding theory and learnability. *Linguistic Inquiry* 18: 413-444.

Platzack, Christer and Anders Holmberg. 1989. The Role of AGR and Finiteness. *Working Papers in Scandinavian Syntax*, 43: 51-76.

Thiersch, Craig. 1978. *Topics in Germanic syntax.* Doctoral dissertation, MIT, Cambridge, Mass.

Tomaselli, Alessandra. 1989. La sintassi del verbo finito nelle lingue Germaniche. Doctoral dissertation, University of Pavia, Italy.

Travis, Lisa. 1984. *Parameters and effects of word order variation.* Doctoral dissertation, MIT, Cambridge, Mass.

Wexler, Kenneth, and M. Rita Manzini. 1987. Parameters and learnability in binding theory. In *Parameter setting*, ed. Thomas Roeper and Edwin Williams, 41-76. Dordrecht: Reidel.

Williams, Edwin. 1981. Language acquisition, markedness and phrase structure. In *Language acquisition and linguistic theory*, ed. Susan L. Tavakolian, 8-34. Cambridge, Mass.: MIT Press.