Edward Gibson
Kenneth Wexler

Triggers

1 Introduction

The principles-and-parameters framework has become a crucial part of the effort to understand the fundamental question of linguistic theory: namely, how any child can learn any natural language. Under this framework it is assumed that linguistic theory is to be characterized by a set of principles called Universal Grammar (UG) and that there exists a finite number of parameters, each of which has a finite number of values (Chomsky 1980, 1981, 1986b). A setting of these parameters instantiates a grammar for a natural language. In this way, linguistic variation can be explained. Furthermore, by restricting variation possibilities to a finite number of parameters, each with a finite number of values, the learning problem is simplified. Searching over a finite space is preferable to searching over a possibly infinite one.

In this article we investigate the properties of parameter setting within this framework. In particular, we examine a tacit assumption made by linguists and first-language acquisition researchers: namely, that the algorithm the learner uses to set parameters relies on the existence of *triggering data*. Triggering data are data from the target language that can be analyzed if and only if one of the target language parameters is set correctly. This assumption implies that triggering data always exist. We demonstrate, however, that triggering data need not exist in the general case. Furthermore, we give an example of a linguistically natural parameter space involving X-bar parameters and

a verb-second parameter that contains grammars for which no triggering data exist. As a result of the lack of triggering data, any learning algorithm that relies on such data cannot succeed in all cases. Nevertheless, there are a number of possible solutions to this problem, which we discuss in detail.

Before we can discuss these issues, it is necessary to formalize the notion of trigger. An implicit assumption underlying this notion is that the values of parameters can be set on the basis of the primary data, simply by "fitting" the parameter values to the data. That is, the learning algorithm simply says to choose a set of parameter values such that the data are grammatical given these parameter settings. This "fitting" hypothesis is what Wexler (1990) has called the General Motor of Learning (GML). A survey of almost any article in syntactic theory that allows for variation and attempts to state parameters shows that the GML is implicitly, if not explicitly, assumed.[1]

Given the GML hypothesis, it is reasonable to explore exactly how parameters can be set by a learner, given input data. Although the learning problem faced by the child is simplified by the assumption that the parameter space to be searched is finite, it is still not trivial. Given a relatively small number of parameters—say, 40—along with at least two possible values for each parameter, a space of $2^{40} \approx 10^{12}$ possible grammars results. This is an enormous search problem if every parameter-setting combination must be considered in turn.

The enormous search space is a problem not only for performance theory, but also for linguistic theory. As Chomsky (1965) points out, besides the criterion of explanatory adequacy, linguistic theory must meet the even more stringent requirement of feasibility, which dictates that grammar must be learnable given the actual restrictions of limited memory and computational resources that exist in a human learner. In an attempt to comply with these feasibility restrictions (see, e.g., Wexler and Hamburger 1973, Wexler and Culicover 1980, Pinker 1984), researchers investigating the theory of parameter setting have assumed that there are sentences in the child's experience that point directly to the correct settings of the parameters. In particular, it has been assumed that for any setting of a nonsubset parameter, there is a sentence that is grammatical under that setting but not under any other. Hearing this sentence allows the learner to determine that the appropriate parameter setting is the one that allows for the grammaticality of that sentence. This sentence—which determines that a parameter is set to a certain value—has often been called a *trigger* for that value of that parameter.[2] Because a trigger

---

[1] Any theory in which the learner does not test to see whether a given input is generated by a proposed grammar would not satisfy the GML. A (cognitively implausible) example of such a theory would be the following: If the first input sentence presented to the learner has n words, then set the grammar to grammar G_n.

[2] Sometimes *trigger* is used in other ways. In these other usages, it is supposed to mean something like an experience that has nothing theoretically to do with a parameter setting, but nevertheless determines the setting of a parameter. See footnote 1 for an example, and Fodor 1978 and Atkinson 1990 for discussion of this issue.

for the value of a parameter may or may not depend on the values associated with other parameters, two classes of triggers can be defined: *global* and *local*.[3]

(1) A *global trigger* for value v of parameter P_i, $P_i(v)$, is a sentence S from the target grammar L such that S is grammatical if and only if the value for P_i is v, no matter what the values for parameters other than P_i are.

(2) Given values for all parameters but one, parameter P_i, a *local trigger* for value v of parameter P_i, $P_i(v)$, is a sentence S from the target grammar L such that S is grammatical if and only if the value for P_i is v.

As should be clear from the definitions, a global trigger is a special case of a local trigger: one that happens to work for all the settings of the other parameters. Thus, if a global trigger exists for value v of parameter P, then this global trigger is also a local trigger for each of the possible settings of all the other parameters.

Note that there can be no trigger (global or local) for a parameter value that yields a grammar that is a subset of a grammar resulting from another setting of the same parameter. Parameters whose different values result in grammars that are subsets of one another are referred to as *subset* parameters (Angluin 1978, Williams 1981, Berwick 1985, Manzini and Wexler 1987, Wexler and Manzini 1987). There can be no trigger for the subset value of a subset parameter, since, by hypothesis, all data that are acceptable in the subset parameter setting are also acceptable in the superset parameter setting. (On the other hand, there are, of course, triggers for the superset grammars.) Thus, only nonsubset parameters have been thought to have triggers for all of their possible values. Our discussion of triggers will therefore be restricted to nonsubset parameters.

Although not explicitly stated, it has been assumed in the (psycho)linguistic literature that the learner uses an algorithm that relies on the existence of triggers in order to succeed. Perhaps the most natural such algorithm is the *Triggering Learning Algorithm* (TLA), given in (3). Under this algorithm, the learner changes her hypothesis about the current grammar depending on whether she can analyze the current input sentence, where a sentence is said to be analyzable for the learner if the learner's current grammar generates the sentence.

(3) *The Triggering Learning Algorithm (TLA)*

Given an initial set of values for n binary-valued parameters,[4] the learner attempts to syntactically analyze an incoming sentence S. If S can be successfully analyzed, then the learner's hypothesis regarding the target grammar is left

---

[3] As far as we know, the distinction between global and local triggers has not been made in the previous literature. It seems to us that in the informal discussions of triggers that linguists give they mostly have global triggers in mind, although there seems to be no necessary reason for this. As we will show, however, global triggers very often do not exist. We define both kinds of triggers for completeness.

[4] For simplicity, we restrict ourselves here to binary-valued parameters. The algorithm and proof of convergence can be generalized in a straightforward way to parameters with more than two values. See footnote 10.

unchanged. If, however, the learner cannot analyze S, then the learner uniformly selects a parameter P (with probability $1/n$ for each parameter), changes the value associated with P, and tries to reprocess S using the new parameter value. If analysis is now possible, then the parameter value change is adopted. Otherwise, the original parameter value is retained.[5]

Following Wexler and Culicover (1980), the TLA is error-driven, in that the learner does not attempt to change her hypothesis as long as the current input sentence can be syntactically analyzed.[6] Furthermore, since there is strong empirical evidence that children use only positive instances of acceptable sentences in linguistic acquisition (Brown and Hanlon 1970, Wexler and Hamburger 1973, Baker 1979, Marcus 1993), the TLA also uses only positive evidence. In addition, it requires no memory of either earlier data or previous parameter settings, so as to limit reliance on the child's memory. Finally, the TLA is conservative, in that its proposals for new grammars are very close to the previous proposals, differing by only one parameter value. This conservatism is desirable from a psychological point of view to the extent that there is no evidence that children make large-scale reorganizations of their grammars in a single learning step.

Note that under the TLA, the learner has no knowledge of which parameters are relevant to a given input, so after a particular input sentence, the learner may actually end up farther from the target grammar than before. This can happen if the input sentence is not analyzable in the current grammar, but is analyzable in the grammar that results by changing the value of a parameter that is currently set correctly. Thus, although there exist triggers for all parameters, the learner does not know which sentences these triggers are or which parameter a triggering sentence is relevant to. Even so, it turns out that convergence is guaranteed in the limit under the assumption that there is always at least one local trigger for some incorrectly set parameter. (The proof is given below.)

In order to make use of the existence of triggers, the TLA considers only grammars that differ from the current grammar by the value of one parameter, thereby behaving conservatively. Furthermore, a proposed grammar is adopted only if it can be used to obtain an analysis for the current sentence. These two constraints on the child's learning algorithm will be referred to as the *Single Value Constraint* (Clark 1989, 1990)[7] and the *Greediness Constraint*. Clark's (1990) statement of the Single Value Constraint is given in (4).

---

[5] The TLA as it is presented here ignores the effects of noise in the input data. A variant of the TLA might consider changing parameters based on only highly frequent kinds of input data. Such an extension of the TLA would avoid changing parameters based on noisy data, if such inputs were of too low a frequency to be considered possible triggering data.

[6] Following Chomsky (1965), we assume that a sentence may still be interpretable even if it does not receive a full syntactic structure. Thus, a child may be able to interpret a sentence that is ungrammatical, while knowing (at some level) that the sentence is ungrammatical. The TLA concerns the acquisition of syntactic parameter values and hence is concerned only with the child's ability to arrive at a syntactic structure for an input sentence. Whether or not the sentence is still semantically interpretable in the context is not at issue.

[7] For a related assumption in a nonparametric framework, see Hamburger and Wexler 1975 and Wexler and Culicover 1980. We should point out that Clark (1990, 1992) ends up rejecting this assumption.

(4) *The Single Value Constraint*

Assume that the sequence $\{h_0, h_1, \ldots, h_n\}$ is the successive series of hypotheses proposed by the learner, where h_0 is the initial hypothesis and h_n is the target grammar. Then h_i differs from h_{i-1} by the value of at most one parameter for $i > 0$.

The Greediness Constraint is given in (5).

(5) *The Greediness Constraint*

Upon encountering an input sentence that cannot be analyzed with the current parameter settings (i.e., is ungrammatical), the language learner will adopt a new set of parameter settings only if they allow the unanalyzable input to be syntactically analyzed.

The Greediness Constraint implies that the child will not replace a hypothesis that does not account for the current sentence S with another hypothesis that still does not account for S. The learner only changes her hypothesis if that change does some local good. This behavior is characterized as greedy behavior in the algorithms literature (see, e.g., Cormen, Leiserson, and Rivest 1990 and the references there).

The proof of the convergence of the TLA relies crucially on the assumption that local triggers exist for every possible grammatical hypothesis that the learner might (temporarily) entertain.[8] This assumption is stated more formally in (6).[9]

(6) *The existence of triggers*

For each possible pair of source and target grammars (G_s, G_t), $G_s \neq G_t$, there exists at least one incorrectly set parameter P in G_s for which there is a trigger S for the target value of P, that is, the value for P in G_t.

Note that the probability that the learner encounters the trigger S is greater than a fixed lower bound b ($b > 0$). This follows immediately from the observation that there are only a finite number of relevant sentence patterns that the learner might encounter, namely, those that are degree-0 (or perhaps degree-1), and the assumption that each sentence pattern has a probability greater than 0 of occurring (see Wexler and Culicover 1980).

Recall that under the TLA, a parameter is selected for a value change with probability $1/n$, assuming n parameters. If all of the parameters are binary-valued, the assumption that triggers are always available to the learner with a probability greater than a fixed lower bound b implies that on a particular learning trial, the probability that the TLA will trigger a parameter to its target value is bounded below by a constant $q > 0$,

---

[8] Thanks to Partha Niyogi and Bob Berwick for pointing out an error in an earlier version of this proof.

[9] Although we prove that (6) is sufficient to guarantee convergence of the TLA, it is not actually necessary. Slightly weaker conditions will allow convergence. However, we believe that (6) is the most natural assumption from the standpoint of linguistic theory and language acquisition.

where $q > b/n$.[10] Therefore, if G_s is the current grammar at time τ, and G_t is the target grammar, which differs from G_s by m parameter values, then the probability that the TLA will hypothesize G_t at time $\tau + m$ is greater than the lower bound $q''' > 0$. Since there are n parameters overall, two grammars may differ by at most n parameters. Thus, there is a constant $r = q'' > 0$ such that for any two grammars, G_s and G_t, G_s the current grammar at time τ and G_t the correct grammar, the probability that the TLA will hypothesize G_t at time $\tau + n$ is greater than r. Furthermore, the probability that the learner starting at G_s will hypothesize G_t at time $\tau + 2n$ is $\geq r + (1 - r)r$. In general, the probability that the TLA will hypothesize G_t at time $(\tau + i * n) \geq r * \sum_{j=0}^{i-1}(1 - r)^j$. Because $\sum_{j=0}^{i-1}(1 - r)^j = 1/r$ as i goes to infinity (for $0 < r < 1$), the probability that the TLA will hypothesize G_t in the limit is therefore 1. Q.E.D.[11]

One component of the TLA that has not yet been discussed is the method by which a parameter is selected to have its value changed at the point when a learner cannot analyze an input sentence. The algorithm given in (3) assumes that when the learner encounters a sentence that she cannot analyze, she randomly selects a parameter (with a uniform distribution over parameters) to have its value changed. In fact, this may be an oversimplification of the learner's parameter selection algorithm in at least the following two general ways:

- First, it is possible that the learner has some knowledge (either built in or computed) of what parameter values should be altered when certain unanalyzable input sentences are encountered. For example, if an unanalyzable sentence contains only a subject and a verb, the learner might know not to bother trying to change values of parameters that have nothing to do with subjects (specifiers). In such an example, the learner might therefore know not to consider a parameter that affects only internal complements of a head.

  However, although there may be instances in which the learner can decide which parameter should be altered based on linguistic knowledge, in many cases it will be impossible for her to tell which parameters are responsible for the inability to analyze the input sentence, without trying alternatives. For example, suppose that a learner acquiring English hears a sentence pattern *Subject Verb Object,* but cannot successfully analyze this sentence, because the parameters that restrict word order are set incorrectly. How is the learner to know whether

---

[10] Given binary parameters, if the correct parameter is selected, changing it must result in the target value, with probability 1. For nonbinary parameters, if we assume that the TLA randomly chooses one of the alternative values for a parameter, then $q > b/n * 1/(m - 1)$, where m is the largest number of values that any parameter has in the system.

[11] The theorem that the target grammar can always be learned can also be proved using simple Markov chain theory. Assuming a random-walk problem in a finite space with a single absorbing state (the target), where the probability in every nonabsorbing state is greater than 0 of moving toward the absorbing state, it follows immediately that the absorbing state is always reached in the limit; see, for example, Lindley 1965: 215–218. See also Niyogi and Berwick 1993, where, following up on our observation, this method is applied to derive some numerical results regarding the parameter space discussed in this article.

the grammar has an underlying word order of SVO with no constituent movement, or whether the surface word order is derived from a different underlying order—say, by verb-second movement (see section 2)? The grammars resulting from these alternative parameter settings are not compatible with one another. Similarly, multiple incompatible grammars are possible for the analysis of many other sentence patterns.[12]

- Second, it is possible that the probability distribution that the learner uses to decide which parameter to change is not uniform: for some reason(s) there may be a greater likelihood of selecting some parameters for value change rather than others. This distribution may or may not depend on the (recent) context, consisting of the sentence patterns that the learner has just encountered.

  As long as the probability is greater than a lower bound $b > 0$ of selecting parameter P when presented with a sentence pattern that serves as a local trigger for the target value of P, the TLA is guaranteed to succeed (assuming also, of course, that triggering data are always going to be presented to the learner). For the sake of simplicity in this article, we assume that the learner uses a uniform probability distribution to decide which parameter to change. This is probably an oversimplification, but it suffices for current purposes. None of the results in this article depend upon assuming uniform distribution: all that is required is that the probability of selecting a parameter to be changed is greater than some lower bound, which is greater than 0.[13]

Although the assumption that triggering data exist for all parameter values is appealing because of the natural learning algorithm that results, it turns out that even rather simple parameter spaces need not obey this restriction. In this article we show how traditional assumptions about the parameterization of word order and verb-second phenomena lead to a parameter space that has no global triggers whatsoever, in the sense defined above. That is, there exists no input that is grammatical if and only if a parameter has a certain value. Thus, although the idea of global trigger seems to capture an intuition found in the linguistic literature, our results show that such triggers might not exist even in simple parameter spaces.

Furthermore, we show that this parameter space has grammars for which there are no local triggers for any incorrectly set parameters with respect to certain target grammars. That is, there exist target grammars in this parameter space such that there are

---

[12] Although a given sentence pattern may be analyzable in more than one grammar, most nontarget grammars will probably not allow a successful analysis of an arbitrary input sentence. Thus, it is not feasible for the learner to keep trying different parameters until she finds one whose new value allows an analysis. Such an algorithm potentially requires n parameter changes and reprocessing attempts for a given sentence: too much computational effort to be considered feasible.

[13] In follow-up work to this article, Frank and Kapur (1993) discuss a variant of the TLA that assumes no lower bound on selecting parameter P: that is, an algorithm in which the learner never follows particular paths from source to target grammars. In order to guarantee convergence, they are then forced to assume the existence of a stronger kind of trigger.

certain grammars (different from the target grammar) for which there are no local triggers at all. Thus, if the learner somehow sets her parameters so that she ends up in one of these states, there will be no way for her to escape from this grammar, given only examples from the target grammar. We thus refer to states having this property as *local maxima*.

Although initially it might seem that these results imply that triggering theory must be abandoned, this is not the case. It turns out that there exists a default initial grammar such that the learner can never get to the local maximum grammars from this initial parameter setting using the TLA. Furthermore, there also exists a parameter-ordering solution to the local maximum problem identified here. In particular, the local maxima can be avoided by restricting the parameters that the learner initially considers to parameters for which there are no local maxima. The existence of these solutions within the framework of triggering theory means that triggering theory may still be essentially correct.

We also consider four other possible solutions to the local maximum problem. One possibility is that there are some grammars—the targets for which there are local maximum sources—that are unlearnable. However, it turns out that all such target parameter settings in this parameter space are present in many of the world's grammars.

A second possibility is that the parameter space associated with UG is not like the parameter space investigated here, in that UG's parameter space contains no local maxima. If so, it may still be possible to retain a learning algorithm based on triggering data as discussed above. However, for such an explanation to be correct, the word order effects discussed in this article must be explained in terms of parameters other than those assumed here. Thus, it would be necessary to explore alternatives to the verb-second and base word order parameters assumed here.

A third possibility is that there exist word order data not considered here that allow the learner to trigger out of the local maximum grammars. Thus, the word order data considered here may be insufficient. Note, however, that this possibility will always exist in any parameter space, no matter how many parameters or data are considered. We can only observe that we have not yet been able to identify any such data for the word order parameter space presented here, in spite of an extensive search.

The final possibility that we will consider here is that the learner does not need triggers as currently conceived in order to acquire the target grammar. Hence, in terms of the TLA, it may be that the learner does not adhere to one or both of the Greediness and Single Value Constraints. In violation of the Greediness Constraint, the learner might be able to change hypotheses without requiring successful analysis of the current input. Or the learner might be able to explore possible grammars that differ from the current one by values of more than one parameter, in contradiction to the Single Value Constraint. Thus, it may be that notions of trigger other than the one described here may allow the parameter space to be void of unwanted local maxima.

We will discuss each of these possibilities. Although we cannot conclusively decide among them, the most promising solution currently appears to be the default state

parameter-ordering solution, because it fits well with current (psycho)linguistic theory, and also because it makes clear predictions. Thus, the computational study of learnability in parameter setting appears to serve the useful purpose of suggesting ways in which the theory of parameter setting may be improved, while maintaining its essence. The spirit of this enterprise, then, is very similar to the spirit of linguistic theory, in which arguments are brought to bear that modify aspects of the theory at the same time that they provide further evidence for it.

The remainder of this article breaks down as follows. Section 2 gives an example parameter space containing two parameters based on simplified cross-linguistic word order assumptions. It is shown that this space contains a global trigger for one of its parameters and local triggers for the other. Section 3 gives a more complex parameter space—one extra parameter to account for verb-second effects—that is shown to have neither global nor local triggers for all of its grammars. Section 4 discusses the ramifications of the existence of such a parameter space. Section 5 presents conclusions.

2 Word Order Acquisition: A Simplified Case

Consider the case of cross-linguistic word order variation. The following X-bar schemata (Jackendoff 1977, Chomsky 1986a) may be parameterized according to head-satellite order in order to explain some facts about cross-linguistic word order variation (see Kayne, to appear; also footnote 21):

(7) a. $X' \rightarrow X$, Complement
    b. $XP \rightarrow X'$, Specifier

The X-bar rule in (7a) says that complements appear as sisters to that head (X), where the location of that sisterhood (left or right) is unspecified. Rule (7b) dictates that specifiers appear as sisters to the single-bar level, where once again the direction of sisterhood is unspecified. Cross-linguistic word order variation may be partially explained by assuming the existence of a two-valued parameter associated with each of the above X-bar rules, that is, with values comp-first and comp-final for the complement-head parameter (7a), and with values spec-first and spec-final for the specifier-head parameter (7b) (see Koopman 1983, Hoekstra 1984, Travis 1984). Given this parameterization, the possible surface word orders for declarative sentences are as shown in table 1 for verbal heads with one specifier (a subject) and one complement (an object).

Surface word orders are assumed to be the result of base generation of word order together with other processes, for example, phrase structure movement. Thus, the fact that the surface word orders VSO and OSV do not appear in table 1 does not mean that these surface word orders cannot occur under this theory of word order. It only means that these word orders are not base-generated.[14]

---

[14] Of course, it may be that word order differences are better explained in terms of the direction of government (Hoekstra 1984), or direction of Case assignment (Koopman 1983, Travis 1984), or perhaps some other difference. What is important in the context of this article is the parameter space that results. It would be interesting to investigate whether other ways of parameterizing word order phenomena have the same properties as the parameters assumed here.

Table 1

Simplified word order parameter space

Specifier-head parameter	Complement-head parameter	Word order
Spec-final	Comp-final	VOS
Spec-final	Comp-first	OVS
Spec-first	Comp-final	SVO
Spec-first	Comp-first	SOV

Given the above parameters, consider the following simplified acquisition problem. The child hears sentences from a single grammar that contain a single verb along with one or two noun phrases, where the noun phrases are either subjects or objects for the verbs. Furthermore, these sentences are assumed to reflect the base-generated word order of the input grammar directly without any phrasal movement. Thus, if interrogatives involve some constituent movement, suppose that the input consists of declaratives only.[15] Additionally, it is assumed that there are no nonlexical elements, so that an object may not appear without a subject also being present. (We are therefore considering only non–pro drop grammars here.) Finally, it is assumed that the child knows (or can compute from the context using linking rules, knowledge of θ-grids, etc.) which grammatical function—subject or object—is appropriate for each noun phrase in some cases.

The child's goal is to acquire the correct settings of the parameters in table 1 given only positive instances of the target grammar. The data associated with all possible parameter settings given the constraints defined above are listed in table 2.

It turns out that acquisition of the target grammar is possible for all grammars using the TLA because of the existence of triggering data in all possible parameter states. In particular, the pattern $S\ V$ is a global trigger for the value spec-first of the specifier-head ordering parameter. Similarly, the pattern $V\ S$ is a global trigger for the spec-final value of this parameter.

The complement-head parameter has no global triggers in this simple parameter space, because every word order pattern that includes an ordering of a complement and a head (a verb and an object here) also includes an ordering of a specifier and a head (a verb and a subject). Thus, the value of the complement-head parameter depends on the value of the specifier-head parameter, and no global triggers exist. However, although the complement-head parameter has no global triggers in this parameter space, there are local triggers for its values in every parameter state.

---

[15] This is not to say that the input stream for the child does not include interrogatives: the input stream will, of course, include interrogatives. However, it is assumed that the child can distinguish some declaratives from interrogatives by using cues other than word order (e.g., situational or intonational cues).

Table 2
Simplified word order parameter space

Specifier-head	Complement-head	Data in defined grammar
Spec-final	Comp-final	VS, VOS
Spec-final	Comp-first	VS, OVS
Spec-first	Comp-final	SV, SVO
Spec-first	Comp-first	SV, SOV

The result that there are no global triggers for some parameters in this parameter space is perhaps not unexpected. To the extent that the structures generated by a parameter setting depend on processes that interact, we will not find global triggers. This poses no problem for the theory of triggers and parameter setting, as long as there are local triggers. If there are global triggers, fewer triggers are needed, because only one trigger is necessary for each parameter value, whereas for local triggers, a different trigger is potentially needed (in the worst case) for each parameter value and combination of other parameter values.

To see how triggers work in the space defined here, suppose that the target grammar has the specifier-head parameter set to spec-first and the complement-head parameter set to comp-final, resulting in an SVO grammar like English. No matter what the child's initial hypothesis for parameter values associated with the target grammar, the pattern $S\ V$ will trigger the spec-first setting of the specifier-head word order parameter. For example, suppose that the child initially hypothesizes OVS word order, the worst hypothesis that she could make given the target grammar. The pattern $S\ V$ is not syntactically analyzable under her current hypothesis, so a change is forced. If the learner decides to consider changing the value of the specifier-head parameter as a result of this pattern, then she will be able to analyze the input, so that the change will be adopted. Thus, upon hearing the input $S\ V$, the learner will eventually end up hypothesizing an SOV grammar. The local triggering pattern $S\ V\ O$ will then eventually force her to move to the target grammar. Thus, under the TLA, convergence to the target grammar is possible because of the presence of the triggering data.

3 Word Order Acquisition: A More Complex Case

Although triggering data are present for all parameter values at every possible grammar in the simplified X-bar parameter space given in section 2, it turns out that parameter spaces need not always have this property. We call such parameter settings *local maximum* parameter settings:

(8) A grammar G_s is said to be a *local maximum* with respect to the target grammar G_t if there is no path via triggering data from G_s to G_t.

The simplest example of a local maximum parameter setting is one in which there exist no triggers for any parameters, so that the learner cannot move from G_s. Such a state is called an *absorbing state* in the theory of Markov chains. In particular, suppose that the learner's current linguistic hypothesis G_s allows the successful syntactic analysis of the set of sentence patterns $\{a_1, a_2, \ldots, a_i, \ldots\}$. Suppose also that the target grammar G_t allows perhaps some, but not all, of $\{a_1, a_2, \ldots, a_i, \ldots\}$ along with the sentence patterns $\{b_1, b_2, \ldots, b_j, \ldots\}$. The existence of at least one pattern b_j that is in the language generated by G_t (notated $L(G_t)$), but not in $L(G_s)$, means that $L(G_s)$ is not a superset of $L(G_t)$, so that the subset problem is avoided. Finally, suppose that for every parameter P, changing the value of P in G_s gives a grammar that allows the analysis of some but not all of $\{a_1, a_2, \ldots, a_i, \ldots\}$ along with some other sentence patterns, but where none of the additional sentence patterns overlap with any of the sentence patterns from $L(G_t)$. Thus, after changing a single parameter from G_s, the learner can analyze no sentences from G_t beyond the ones that she could already analyze with the hypothesis G_s, so that the Single Value and Greediness Constraints force her to stay in G_s. Hence, the TLA cannot move the learner to any other grammar. The grammar G_s is therefore an example of a local maximum grammar relative to the target G_t, because there is no path via triggering data from G_s to G_t. Intuitively, if only one parameter is changed in G_s, then there is no previously unanalyzable sentence that now becomes analyzable.

More generally, a source grammar may be a local maximum grammar even if it is possible to move from this grammar to another via triggering data, as long as there is no path from the source to the target via triggering data. For example, suppose that the learner finds herself in grammar G_{s1} relative to target G_t. Suppose also that there is only one parameter that can be triggered given G_{s1} as the source grammar, and that G_s is the resultant grammar when this parameter is changed. Finally, suppose that there is no trigger from $L(G_t)$ to take the learner out of $L(G_s)$ (i.e., G_s is an absorbing state). In such a case both S_{s1} and S_s are local maxima relative to S_t.

3.1 Verb-Second Effects

Now consider a concrete example of a parameter space with states like the source state in (8): a parameter space consisting of the X-bar parameters presented in section 2 together with a verb-second (V2) parameter. A V2 parameter, when set, 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; see Haegeman 1991 for an overview).[16] Many Germanic and Scandinavian languages have this property, including German, Dutch, Swedish, Danish, Norwegian, Icelandic, and Yiddish. For example, consider Dutch and German, languages that allow the following declarative word orders:

---

[16] In fact, in some languages (e.g., Yiddish; Diesing 1990), the V2 property applies to all tensed clauses, not just root clauses. This is a further parametric difference among languages, which will not be investigated here.

(9) a. S V O

   b. O V S

   c. S Aux O V

   d. O Aux S V

   e. Adv V S O

   f. Adv Aux S O V

   g. . . . C S O V

   h. . . . C S (Adv) O V Aux

Furthermore, languages like Dutch and German disallow the following word orders:

(10) a. *S O V

   b. *S Aux V O

   c. *O Aux V S

   d. *Adv S V O

   e. *Adv O V S

   f. *Adv Aux V S O

   g. *Adv V S O Aux

   h. *. . . C S V O

Examples of such patterns from both Dutch and German are given in (11) and (12).[17]

(11) a. Dutch   Hans koopt het boek.

      German   Hans kauft das Buch.

                Hans buys the book

                'Hans buys the book.'

   b. Dutch   Het boek kocht Hans gisteren.

      German   Das Buch kaufte Hans gestern.

                the book bought Hans yesterday

                'The book Hans bought yesterday.'

   c. Dutch   Hans heeft het boek gekocht.

      German   Hans hat das Buch gekauft.

                Hans has the book bought

                'Hans has bought the book.'

   d. Dutch   Het boek heeft Hans gekocht.

      German   Das Buch hat Hans gekauft.

                the book has Hans bought

                'The book Hans has bought.'

   e. Dutch   Gisteren kocht Hans het boek.

      German   Gestern kaufte Hans das Buch.

                yesterday bought Hans the book

                'Yesterday Hans bought the book.'

---

[17] The examples either are taken directly from Haegeman 1991 or derive from Haegeman's examples.

  f.   Dutch   Gisteren heeft Hans het boek gekocht.
       German   Gestern   hat   Hans das Buch gekauft.
                 yesterday has   Hans the book bought
                 'Yesterday Hans has bought the book.'

  g.   Dutch   . . .dat Hans het boek koopt.
       German   . . .daß Hans das Buch kauft.
                 . . .that Hans the book buys
                 '. . .that Hans buys the book.'

  h.   Dutch   . . .dat Hans (gisteren)   het boek gekocht heeft.
       German   . . .daß Hans (gestern)   das Buch gekauft hat.
                 . . .that Hans (yesterday) the book bought has
                 '. . .that (yesterday) Hans bought the book.'

(12) a.   Dutch   *Hans het boek koopt.
       German   *Hans das Buch kauft.
                 Hans the book buys
                 'Hans buys the book.'

  b.   Dutch   *Hans heeft gekocht het boek.
       German   *Hans hat   gekauft das Buch.
                 Hans has   bought the book
                 'Hans has bought the book.'

  c.   Dutch   *Het boek heeft gekocht Hans.
       German   *Das Buch hat   gekauft Hans.
                 the book has   bought Hans
                 'The book Hans has bought.'

  d.   Dutch   *Gisteren Hans kocht het boek.
       German   *Gestern   Hans kaufte das Buch.
                 yesterday Hans bought the book
                 'Yesterday Hans bought the book.'

  e.   Dutch   *Gisteren het boek kocht Hans.
       German   *Gestern   das Buch kaufte Hans.
                 yesterday the book bought Hans
                 'Yesterday Hans bought the book.'

  f.   Dutch   *Gisteren heeft gekocht Hans het boek.
       German   *Gestern   hat   gekauft Hans das Buch.
                 yesterday has   bought Hans the book
                 'Yesterday Hans has bought the book.'

  g.   Dutch   *Gisteren gekocht Hans het boek heeft.
       German   *Gestern   gekauft Hans das Buch hat.
                 yesterday bought Hans the book has
                 'Yesterday Hans has bought the book.'

h. Dutch *. . .dat Hans koopt het boek.
    German *. . .daß Hans kauft das Buch.
    . . .that Hans buys the book
    '. . .that Hans buys the book.'

As a result of data like these, German and Dutch are assumed to be languages whose underlying word order is SOV (spec-first, comp-final), with V2 effects in main clauses. Thus, although the word order *S O V* is an appropriate base structure for these languages, this word order is not an appropriate surface form in matrix clauses. Constraints on the surface form necessitate (a) the presence of the matrix tensed verb in the C position (to the left of the IP sentential structure) and (b) the presence of a constituent in the specifier position of the complementizer phrase (to the left of the C position). Hence, although *S O V* is not a possible matrix clause word order, both *S V O* and *O V S* are acceptable matrix word orders in Dutch and German.

There are many theories about why this V2 movement takes place, most of which rely on a property of the zero-level complementizer C^0. Although the results we obtain here do not rely on a particular theory of V2, for concreteness let us assume, following Holmberg and Platzack (1991), that in $+$V2 grammars Comp has the finiteness feature $[+F]$, which is only licensed if the X^0 that bears this feature governs an element with nominative Case.[18] Thus, a $+$V2 grammar is one in which C must have the feature $[+F]$; a $-$V2 grammar has no such requirement. Adding this parameterization to the X-bar parameters presented in section 2, we now have three parameters, each with two values, giving a total of eight possible grammars.

As in section 2, suppose that a child faces a simplified acquisition problem. In particular, suppose that the child hears sentences from a single grammar in which an object may not appear without a subject also being present. Furthermore, assume that the child knows (or can compute from the context) which grammatical function—subject or object—is appropriate for each noun phrase. Since all $+$V2 grammars are identical in the set of surface orders that they admit when only a subject, a verb, and an optional object are considered—namely, all and only the surface declarative orders SVO, SV, and OVS—additional constituents must be considered in order to tell the eight possible grammars apart.

The additional patterns that we will consider here derive from including (a) a second object; (b) a single finite (tensed) auxiliary verb, notated *Aux*, which is assumed to subcategorize for the verb, following the grammar's standard complement-head word order; and (c) a sentence-initial adverbial adjunct, notated *Adv*.

Furthermore, we will consider only cases in which C^0 is on the left of its complement IP, so that verb movement to C^0 in $+$V2 grammars is leftward movement. Of course, the C^0-IP word order is not universal (e.g., C^0 is assumed to be on the right in Japanese

---

[18] See Vikner 1991 for a summary of other approaches to V2 phenomena.

and Korean), but note that the C^0-IP word order varies independently of the verb-complement word order. For example, in German and Dutch the verb appears to the right of its complement, but C^0 appears to the left of its complement IP. Thus, a single head-complement parameter will not suffice to account for the possible head-complement word orders across languages. In order to simplify the parameter space under consideration, we therefore fix the value of the head-complement parameter for the category C to be comp-final, so that the verb always moves to the left in $+$V2 grammars.

This simplification will not affect any of the nonlearnability results to follow, because the learnability problem is at least as difficult (and possibly more difficult) when further parameters are added to a parameter space, when considering the same sentence patterns. Thus, any problems that are identified in the smaller parameter space will also be problems for the larger space.[19]

Similarly, we assume a fixed spec-first word order for the category C, so that movement of the root-clause-initial phrase is leftward in $+$V2 grammars. Again, we make this assumption to simplify the parameter space, because the value for C's specifier-head parameter varies independently of the value of I's specifier-head parameter. Even grammars whose clausal subjects appear on the right do not have rightward *wh*-movement or rightward topicalization. For example, in the Amazonian language Pacaas Novos, whose underlying word order is VOS, topicalization and *wh*-movement still involve leftward movement (Dan Everett, personal communication).[20]

For similar reasons of simplicity, we have also restricted our attention to root CPs and IPs. Because the ordering of specifiers and complements may depend upon the head category in question, consideration of variation within other categories (e.g., nouns, prepositions) may necessitate consideration of additional parameters. In the worst case, two additional parameters (specifier-head and complement-head) may be required for each category.

The input data to be considered with respect to the TLA therefore consist of 12 possible constituent patterns for a $-$V2 grammar and 18 possible such patterns for a $+$V2 grammar. These data, which will allow us to differentiate every grammar from every other grammar, are summarized as follows:

(13) a. Some order of *Verb* and *Subject*
     b. Some order of *Verb, Subject,* and *Object:* one sentence pattern for a $-$V2 grammar, two patterns for a $+$V2 grammar
     c. Some order of *Verb, Subject, Object1,* and *Object2:* one sentence pattern for a $-$V2 grammar, three patterns for a $+$V2 grammar

---

[19] If additional patterns of data are considered as well as further parameters, then there is no simple relationship between the learnability problems in the smaller and larger spaces. The learning problem might be easier, harder, or of the same difficulty in the larger space. See section 4.3.

[20] In fact, none of the linguists that we consulted knew of a language in which the specifier of CP appears on the right: movement processes like *wh*-movement and topicalization seem to be universally leftward movement processes, if they are present at all (i.e., they are never rightward). Thus, it might be a linguistic universal that the specifier of CP appears on the left.

d. Some order of *Verb*, *Auxiliary*, and *Subject*

e. Some order of *Verb*, *Auxiliary*, *Subject*, and *Object:* one sentence pattern for a −V2 grammar, two patterns for a +V2 grammar

f. Some order of *Verb*, *Auxiliary*, *Subject*, *Object1*, and *Object2:* one sentence pattern for a −V2 grammar, three patterns for a +V2 grammar

g. The pattern *Adverb* followed by some order of *Verb* and *Subject*

h. The pattern *Adverb* followed by some order of *Verb*, *Subject*, and *Object*

i. The pattern *Adverb* followed by some order of *Verb*, *Subject*, *Object1*, and *Object2*

j. The pattern *Adverb* followed by some order of *Verb*, *Auxiliary*, and *Subject*

k. The pattern *Adverb* followed by some order of *Verb*, *Auxiliary*, *Subject*, and *Object*

l. The pattern *Adverb* followed by some order of *Verb*, *Auxiliary*, *Subject*, *Object1*, and *Object2*

All of the data for each of the possible grammars are summarized in table 3. For notational purposes, grammars in the given parameter space that have no V2 movement will be referred to by their base orders of subjects, verbs, and objects. Thus, a grammar like that of English will be represented as SVO. Grammars that do have V2 movement will be represented by their base word orders appended with the string +*V2*. Thus, the grammars of Dutch and German in the defined parameter space will be notated SOV + V2.[21]

Having set up the grammar space to be considered, we can now determine whether all of the possible grammars are learnable under the TLA. In particular, we can ask whether every value of every parameter has a global or local trigger.

Note that in this set of grammars, no grammar is the subset of any other, so that subset issues do not play a role. That is, if it turns out that a particular parameter value

---

[21] Kayne (to appear) proposes that there is a deterministic relation between c-command and linear spell-out (before movement applies); namely, if X c-commands Y, then X precedes Y. If this Linear Correspondence Algorithm (LCA) is correct, then of course a number of the grammars that we consider will not be directly generable, before movement. For example, there can be no underlying spec-final grammars. More generally, there can be no parameters that specify underlying order.

The LCA will not necessarily change the nature of the problem that we investigate here, however. Many, perhaps all, of the surface patterns that we specify in table 3 will still exist. For example, we observed earlier that there appear to be languages with a subject-final surface pattern. If the LCA is correct, these languages will have to be derivable via further operations, presumably movement operations. Exactly how the sentence patterns distribute over grammars will be determined by the parametric variation that defines the surface variation in these languages. Thus, once a theory is determined that allows for the varied surface patterns, and the relevant parameters are specified, it would be appropriate to investigate how the TLA fares with respect to that theory. There is no a priori reason to believe that the results will not be similar to those obtained under the current theory, which allows for the possibility of specifying order parameters. The lesson is the familiar one: a change in linguistic theory that eliminates certain assumptions of variation (e.g., certain parameters) does not necessarily make the learning problem easier if the theory must specify other possibilities for variation. As Chomsky (1965) points out, the problem of learning (of explanatory adequacy and of feasibility) is the problem of how the relevant grammars "scatter" with respect to the inputs. The properties of the inputs do not necessarily change in a crucial way even if the theory of grammar changes. Of course, actual results will depend on a clearly specified theory of grammar with respect to possible grammars.

Table 3
Simplified X-bar and V2 parameter space

Parameter settings	Data in defined grammar
Spec-final Comp-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 Comp-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 Comp-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 Comp-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 Comp-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 Comp-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 Comp-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 Comp-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

does not have a trigger, it is not because it defines a grammar that is a superset of another grammar, since there are no grammar pairs in the defined parameter space such that one is the subset of another. Thus, the acquisition problem noted by Berwick (1985), Manzini and Wexler (1987), Wexler and Manzini (1987), and Clark (1989, 1990) is not at issue here.

Let us first consider global triggers. It turns out that, unlike the simplified X-bar parameter space examined in section 2, the parameter space defined above contains *no* global triggers for any parameter values. That is, for every parameter value in the three-parameter system defined here, there is no sentence pattern that is grammatical if and only if that parameter is set to the requisite value. In order to get an intuitive grasp of why this is so, let us consider the specifier-head ordering parameter with respect to data consisting of a subject and a verb. If this parameter is set to spec-final, then the order VS is possible, as long as the V2 parameter is not set; otherwise, only the order SV is possible. That is, the specifier-head parameter is not the only parameter involved in determining the surface order of the specifier and the head: the value of the V2 parameter is also crucial. Similar observations can be made for all parameter values, so that there are no global triggers whatsoever. The absence of global triggers results from the interdependence of the parameter values upon one another, especially the interaction of the V2 parameter with the other two parameters.

Let us now investigate the possible existence of local triggering data for the V2 and specifier-head parameters. It turns out that given a target grammar as defined by a set of parameter settings, most of the possible source grammars have local triggers that enable the learner to reach the target by use of the computationally simple TLA. However, there exist pairs of source and target grammars from the parameter space given in table 3, such that no data from the target grammar will ever shift the learner to the target grammar. Thus, if the learner ends up in one of these local maxima, there is no way that she can ever converge on the target grammar, given the TLA. Out of the 56 possible source-target grammar pairs, there are 6 such pairs of source local maximum and target grammars.[22] These local maxima, which were discovered by means of a computer search of the parameter space, are given in table 4.[23] The algorithm by which these local maxima were discovered is given in appendix A. This algorithm consists of an exhaustive search to see which grammars can be arrived at from a source grammar given a target grammar, using the TLA. Once this search is complete, the local maximum grammars consist of those source grammars for which there is no path to the specified target. Because the search for all local maxima involves a large number of calculations,

---

[22] 56 source-target pairs = 8 possible source grammars * 7 possible target grammars (different from the source).

[23] The Common Lisp code that implements this search can be obtained by anonymous ftp to the machine psyche.mit.edu (Internet address 18.88.0.85). The code is found in the file called loc-max.lisp in the directory pub/gibson. The code may also be obtained by sending a request (either by regular mail or by electronic mail) to the first author.

Table 4
Local maxima for the word order parameter space defined in table 3

Source (local maximum) grammar	Target grammar
Spec-final, Comp-final, V2 (VOS + V2)	Spec-first, Comp-final, − V2 (SVO)
Spec-final, Comp-first, V2 (OVS + V2)	Spec-first, Comp-final, − V2 (SVO)
Spec-final, Comp-final, V2 (VOS + V2)	Spec-first, Comp-first, − V2 (SOV)
Spec-final, Comp-first, V2 (OVS + V2)	Spec-first, Comp-first, − V2 (SOV)
Spec-first, Comp-final, V2 (SVO + V2)	Spec-final, Comp-first, − V2 (OVS)
Spec-first, Comp-first, V2 (SOV + V2)	Spec-final, Comp-first, − V2 (OVS)

it was conducted by computer. However, because the search is exhaustive, the results presented here are exact, not approximate. An example of how the algorithm works is given in appendix B.

Despite the size of the search problem in obtaining all the local maxima in a parameter space, it is easy to verify that a particular grammar is a local maximum relative to another. This verification involves seeing that for every sentence S that is in the target and is not in the source, S is not in any of the grammars that differ by one parameter value from the source. For example, consider the first local maximum in table 4, the grammar VOS + V2, which is a local maximum with respect to the grammar SVO. Suppose that the learner currently hypothesizes the grammar VOS + V2. Furthermore, suppose that she hears only data from the target grammar, grammar SVO. Under the current assumptions there are 12 patterns that the learner might hear, as given in table 3:

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

Of these, the patterns *S V, S V O, S V O1 O2, S Aux V, S Aux V O*, and *S Aux V O1 O2* are all grammatical in grammar VOS + V2, the source grammar, so that the learner would not change her hypothesis based on them. On the other hand, the other six patterns from grammar SVO are all ungrammatical based on the learner's current hypothesis. Thus, upon hearing one of these unanalyzable patterns, the learner would attempt to change her hypothesis in order to account for it. However, it turns out that changing the value of exactly one parameter yields grammars in which all of these patterns are still ungrammatical. For example, consider the pattern *Adv S V O*. This pattern is ungrammatical in grammar VOS + V2, but it is also ungrammatical in grammars VOS, OVS + V2, and SVO + V2, the grammars that differ from grammar VOS + V2 by the value of one parameter. Thus, when presented with any unanalyzable pattern, the learner does not have a viable alternative to her initial hypothesis, given her inherent computational constraints. Hence, she will remain in the local maximum grammar forever and will therefore not converge on the target grammar.

The second local maximum given in table 4 is OVS + V2 relative to the target SVO. This is an example of a local maximum from which it is possible to move via triggering data, but from which it is not possible to traverse all the way to the target SVO. In particular, the pattern *S Aux V O* is a local trigger for the complement-head parameter, so that hearing a sentence pattern of this form can cause the learner to correctly set this parameter to comp-final. However, there are no triggers for either of the other parameters. Moreover, the resulting grammar is VOS + V2, which is a local maximum for the target SVO. Thus, OVS + V2 is also a local maximum for SVO. The remaining four source-target grammar pairs given in table 4 are cases of the simpler form of local maximum, such that there are no triggering sentence patterns whatsoever, for any of the incorrectly set parameters in the source grammars.

4 Implications for a Parametric Theory of Language Acquisition

The fact that no triggering data exist for certain grammars in the parameter space defined here has a number of possible implications for language acquisition theory or, more generally, linguistic theory. We will explore the following possibilities:[24]

1. There is a default setting for each parameter resulting in a grammar from which it is impossible to get to a local maximum, no matter what the target grammar is. Furthermore, there may also be a latency period associated with this initial setting, such that the learner can only change values associated with the parameters that are initially unset. This generalization of the default-setting solution is a *parameter-ordering* solution (possibly maturational).

   Because these solutions do not require any changes to the learning constraints implicit in the TLA, they are consistent with triggering theory as currently envisioned.

2. The target grammars for which there are local maxima are unlearnable grammars. This possibility can be dispensed with right away, because these target parameter

---

[24] A logical possibility not mentioned in this list is that negative evidence in the environment might alleviate the local maximum problem in the parameter space under consideration. In fact, it turns out that if the learner can make use of parental correction, all local maxima disappear. (We would like to thank Bonnie Schwartz for making this observation.)

In order to see this, we need to examine each of the local maxima in turn. Suppose that the learner currently proposes one of these local maxima, the grammar VOS + V2 relative to the target SVO. Furthermore, suppose that she utters a sentence of the form *O V S* and that she receives correction (negative evidence), since no sentence of this form is in the target. Based on this evidence, the learner may try to change any of the three parameters. Assuming that the learner will only shift to a grammar in which the input item is ungrammatical (following the spirit of the Greediness Constraint), she may change the V2 parameter, since the resulting grammar VOS is a grammar that does not allow sentences of the form *O V S*. Thus, the learner can move from the grammar VOS + V2 to the grammar VOS by use of negative evidence. Furthermore, the presence of such negative evidence would allow the learner to escape all of the grammars listed as local maxima in table 4. However, although negative evidence will solve the problem under consideration, negative evidence is not always available in a useful way to all children (see, e.g., Brown and Hanlon 1970, Wexler and Hamburger 1973, Baker 1979, Marcus 1993). Hence, we do not give the same weight to this possible solution as we do to the others.

settings—SVO, SOV, and OVS—are all present in the world's languages. Although OVS word order is rare, the word orders SVO and SOV are very common and are therefore not unlearnable.[25]

3. UG does not, in fact, have any such local maxima, so that the problem we are examining here is epiphenomenal. This situation may come about in one of two different ways:

   a. The structure of UG may not resemble the structure assumed here, so that the parameters that are being considered are not relevant to the real UG.

   b. There may exist as-yet-unconsidered data from certain grammars that allow the learner to trigger out of the grammars that are currently thought to be local maxima.

4. The definition of trigger as given here is overly restrictive. In particular, relaxing one or both of the Greediness and Single Value Constraints causes the local maximum problem to disappear.

4.1 An Unmarked Initial Set of Parameter Values

One possible answer to the puzzle posed here is that there exists an initial set of parameter values—the unmarked values—such that the learner could never get into the local maxima from this grammar given only positive evidence from the target grammar. For example, to account for cross-linguistic empirical data, Hyams (1986, 1987) proposes such an initial parameter setting for the pro drop or null subject parameter. There are at least three distinct kinds of initial parameter-setting states: (a) it may be that all parameters receive an initial value from the set of possible values associated with each; (b) it may be that all parameters are initially unset; and (c) it may be that some parameters receive a default initial value from the set of their possible values, and some are initially unset.

4.1.1 Default Values for All Parameters   In the first kind of default grammar outlined above, each parameter receives a default value from the set of possible values for that parameter. It turns out that, in the parameter space defined here, there is no set of initial values for all three parameters such that the learner can always avoid the local maxima when starting from this grammar. That is, no matter which grammar is hypothesized as the unmarked initial grammar, for some target grammar, the learner can always get into one of the local maxima using the TLA while hearing only positive evidence from the target.

In order to see this, let us consider each possible grammar in turn. First of all, if every grammar is learnable, then it cannot be the case that any of the local maximum source grammars is the unmarked grammar. If one of these grammars were the unmarked

---

[25] Wexler (1991) notes that linguists usually consider the lack of evidence for a particular parameter setting as evidence against that parametric theory. Thus, linguists implicitly assume that all logically possible grammars, corresponding to all possible parameter settings allowed by grammatical theory, are attainable by the learner.

initial state, then clearly it would be impossible to arrive at the target, since these are local maxima with respect to their targets. Thus, the initial grammar is not any of VOS + V2, OVS + V2, SVO + V2, or SOV + V2, the four grammars represented in the source local maximum grammar column in table 4.

Furthermore, if the unmarked initial grammar is VOS and the target grammar is grammar SVO, then the sentence *S V* can lead the learner to turn on the V2 parameter, giving grammar VOS + V2. Since grammar VOS + V2 is a local maximum with respect to grammar SVO, the learner will never escape this hypothesis. Hence, the grammar VOS is not the unmarked state.

Similarly, the unmarked state cannot be grammar OVS since the learner can also get from this state to grammar VOS + V2 using only positive instances from grammar SVO. The relevant input is *S V*, which can cause the learner to switch from grammar OVS to grammar OVS + V2, followed by *S Aux V O*, which can cause the learner to switch from grammar OVS + V2 to grammar VOS + V2.

To see that the unmarked state cannot be grammar SOV, suppose that the target grammar is grammar OVS. Hearing the input *O V S* can lead the learner to change her parameter state hypothesis to grammar SOV + V2, which happens to be a local maximum with respect to the target grammar.

Finally, grammar SVO also cannot be the initial state. To see this, suppose that grammar SVO is the initial state and that the target is grammar OVS. The input *O V S* can cause the learner to alter her hypothesis to grammar SVO + V2, which is a local maximum with respect to the target.

Since we have now exhausted all the possible grammars as an initial state, there can be no unmarked set of parameter settings from which the learner starts.

4.1.2 Unset Initial Values for All Parameters   In the second kind of default grammar that we will consider here, the initial value for each of the parameters is the special value *unset*. However, it turns out that such an initial state does not help the learner to avoid all local maxima, under plausible learning assumptions.

Suppose that the learner is given three unset parameters for the parameter space being explored here. Now suppose that the learner is presented with a sentence from the target grammar. Since no parameters are currently set, this sentence will not be analyzable. If the learner sets all three parameters (perhaps arbitrarily) in order to attempt to analyze the input sentence, then the acquisition problem reduces to the acquisition problem with default values for all parameters, which, as just demonstrated, does not avoid the local maximum problem.

In contrast, suppose that the learner can set at most one parameter at a time, following the Single Value Constraint. Furthermore, suppose that the learner obeys the Greediness Constraint so that she sets a parameter value only if setting that value allows her to syntactically analyze a previously unanalyzable sentence. But since all of the data in the space being considered require values for at least two parameters in order to be analyzable, the learner will never be able to set any parameters. Consider, for example,

the sentence pattern $S\ V$. The learner cannot set exactly one of the specifier-head and V2 parameters and be able to analyze this pattern. If she sets the specifier-head parameter to spec-first, then she still cannot build a structure for the input pattern, since she does not know whether or not there is V2 movement. In particular, if there is no V2 movement, then the resulting structure will have the NP specifier in IP. If, on the other hand, there is V2 movement, the resulting structure will be a CP with an NP specifier. Similarly, if the learner sets the V2 parameter to +V2, she still cannot build a complete structure for the input sentence, since she does not know whether the specifier of IP is to the right or the left of the head verb. Since all sentence patterns that depend on one of these parameters also depend on the other, the learner cannot set one parameter at a time under the proposed learning constraints. In order to retain the Single Value and Greediness Constraints, one of these parameters must have a default value.

4.1.3 Unset Initial Values for Some Parameters and Default Values for Other Parameters   In the third kind of default grammar to be considered here, one or two parameters are initially unset, and the other(s) receive default values. It turns out that there exists such a default set of parameter values such that the learner can never get from this initial state to any of the local maximum grammars, no matter what the target.

Consider once again the list of local maxima for the parameter space defined here, given in table 4. Note that each of the local maxima has the V2 parameter set to +V2 and that the target grammar in each case has the V2 parameter set to −V2. As a result of this pattern, it turns out that local maxima can be avoided altogether if the V2 parameter has the default value −V2, and the values for the other two parameters are initially unset.[26]

In order to see how the initial state accounts for the acquisition of all of the possible grammars in the space considered here, let us first consider target grammars that are +V2. Since there are no local maxima with respect to +V2 targets, +V2 grammars can be acquired no matter what the initial state. Thus, the learner can always acquire these targets after sufficient presentation of data using the TLA.

Now consider −V2 target grammars. By hypothesis, the learner will initially propose a −V2 grammar, as desired. She then attempts to set the surface word order parameters to account for constituent patterns that she cannot analyze. Since (a) the specifier-head and complement-head parameters initially have no values and (b) all constituent patterns being considered here involve either specifiers or complements, no constituent pattern will be analyzable at this point. Furthermore, because the learner can set only one parameter at a time (because of the Single Value Constraint), only constituent patterns that require values for either the specifier-head parameter or the complement-head parameter (but not both) will be possible triggers. In particular, con-

---

[26] It may be that the markedness values associated with a given parameter can be derived from the economy of the derivations associated with such values (see Chomsky 1991, 1993). That is, a grammar that allows V2 movement is in some sense more complex than one that does not allow such movement. We leave it to future work to develop this idea further. See also Roberts's (1993:156) Least Effort Strategy for related ideas.

stituent patterns that include no subjects and patterns that include no internal comple-
ments will be the only possible triggers. The only such pattern in the space under con-
sideration is an order of *S* and *V*, and this pattern will correctly trigger the specifier-
head parameter value. Once the specifier-head and V2 parameters have been successfully
set (as will now be the case), any data from the target that cannot be successfully analyzed
will eventually trigger the complement-head parameter.

For example, suppose that the target grammar is SVO. Initially, the learner hypothe-
sizes a −V2 grammar, with values for the specifier-head and complement-head parame-
ters unset. At this point the constituent pattern *S V* is the only pattern that can cause
a parameter change. When this pattern appears, it forces a value of spec-first for the
specifier-head parameter. Once the learner hypothesizes a −V2, spec-first grammar,
any sentence that requires a value for the complement-head parameter forces a comp-
final setting for this parameter. For example, suppose that the sentence *S Aux V O* is
presented to the learner at this point. Only the value comp-final allows a successful
analysis of the input sentence pattern. The same is true of all other sentence patterns
in the space under consideration.

Hence, in spite of the presence of local maxima in the parameter space, the learner
can never arrive at any of these grammars as long as she starts in the initial state of
−V2 with the other two parameters unset. The existence of this initial unmarked state
suggests that the three parameters should not have the same status in linguistic theory.
The two base word order parameters have a simple binary structure: each has two values,
but there is no hierarchy among these values. The V2 parameter differs from the base
word order parameters, since the −V2 value is unmarked with respect to the +V2
value. However, the standard ways of stating the V2 parameter do not involve such
hierarchical structure. For example, Holmberg and Platzack (1991) assume that the val-
ues associated with the V2 parameter are given by the distinction between the finiteness
feature's being in I and its being in C. Such a statement does not directly reflect any
kind of hierarchy; from a formal point of view the V2 parameter is stated in the same
way as the other two parameters. If the ordering view suggested here is correct, then
perhaps this view of the V2 parameter is wrong; that is, perhaps there *should* be a
hierarchical relation among the values. Given these considerations, we would expect
the correct linguistic theory of V2 to predict the ordering; perhaps the notion of economy
of derivation suggested in footnote 26 would be an appropriate path to follow.

Note that if it had turned out that +V2 rather than −V2 was the default value for
a successful solution (one in which no local maxima exist), we would have been unable
to suggest that economy of derivations plays a role in setting initial values of parameters.
Thus, we have evidence for such a role. As a first attempt, we might say that if linguistic
theory finds one option (e.g., verb movement) more costly than another (e.g., no verb
movement), then the parameter value associated with the less costly operation represents
the initial state of the parameter. It is intriguing that economy considerations, from
syntactic theory, are compatible with learnability considerations concerning parameter
setting. Such a convergence suggests that we may be on the right track.

4.2 Parameter Ordering

Although the default initial parameter setting of − V2 allows the learner to avoid all the local maxima in the parameter space, this solution to the observed problem is not robust if the learner somehow sets her parameters incorrectly early in the acquisition process. For example, suppose that the learner sets her parameters arbitrarily after a few unanalyzable inputs, so that at least some analysis can be attempted. That is, if the learner cannot get any analysis at all for an input because of unset parameters, it may be that she sets these values—perhaps arbitrarily, or by some default order—so that some analyses can be calculated. However, such an initial parameter-setting scheme will allow the learner to arrive at a local maximum, no matter what the default state. For example, suppose that the default state is − V2, as proposed in section 4.1.3. Suppose also that the learner initially hears a string of utterances that she cannot analyze after changing a single parameter value. Suppose then that she arbitrarily chooses values for the other two parameters. No matter what these values are, she will be able to arrive at a local maximum grammar for some target grammar, as was demonstrated in section 4.1.1.

Furthermore, the presence of noisy data will allow some learners to arrive at a local maximum in certain circumstances. Recall the variant of the TLA that was proposed in footnote 5 in order to deal with the presence of noise in the data: namely, that only highly frequent data should be permitted to act as triggering data, causing a change in parameter value. Although such a policy will allow most learners to successfully ignore noisy data (since such data are assumed to be infrequent), in a fraction of cases a learner will encounter noisy data in sufficient frequency to cause a parameter change. In these cases a learner may end up in a local maximum grammar and therefore may never arrive at the target.[27]

Thus, the default state solution by itself is not robust with respect to alternative parameter-setting strategies and/or noise in the input data. A more robust way to avoid the local maximum problem is by means of *parameter ordering*. A parameter-ordering solution is a refinement of an initial state solution. In a parameter-ordering solution, parameters are associated with default initial values as in an initial state solution. However, a parameter-ordering solution is additionally restrictive on the learner because such a solution allows the learner to change only a subset of all the parameters for an initial period of time. Only after a sufficient time has elapsed are the later parameters considered, at which time they are also set on the basis of triggering data.

The existence of an initial state solution implies the existence of a parameter-ordering solution. Thus, the parameter space defined here has a parameter-ordering solution. One such ordering is as follows:

- There is a default value for the V2 parameter: the value − V2, so that V2 movement

---

[27] Of course, noise in the data may also allow learners to escape local maxima, but by hypothesis, this noise is infrequent enough so that it rarely causes parameters to trigger. Hence, the likelihood of triggering out of a local maximum via noise is also very low.

is blocked. There need not be such default values for the base word order parameters.

- The learner initially considers changing only the specifier-head and complement-head parameters with respect to data that cannot be analyzed.
- After enough data have been processed so that the values for the base word order parameters can be set with confidence for those grammars without V2 movement, the learner then considers the alternative value for the V2 parameter.

In order to see how this ordering solution accounts for the acquisition of all of the possible grammars in the space considered here, let us first consider the $-$V2 grammars. The learner initially proposes a $-$V2 grammar. She then attempts to set the other two parameters in order to be able to analyze those sentences that she cannot currently analyze. Since there are no local maxima when the V2 parameter is set to $-$V2, the learner cannot get into a local maximum state. Thus, after a sufficient quantity of examples from the target grammar, the learner will eventually hypothesize the target grammar.[28] Additional examples from the target grammar then serve to strengthen this hypothesis. At some point the V2 parameter becomes active, so that the learner can consider the alternative value. However, the learner can already analyze everything from the target grammar, so she has no reason to alter her hypothesis. Thus, the learner always converges on the target grammar when the grammar is a $+$V2 grammar.

Now consider the case in which the target grammar is a $+$V2 grammar. As above, the learner initially hypothesizes that the grammar is not a $+$V2 grammar. She then tries to find a $-$V2 grammar that best fits the data from the target. No matter what her hypothesis, there will be a number of sentences that she will not be able to analyze. Thus, even after the same number of sentences that caused the base word order parameter values to be strengthened in the $-$V2 grammar case, the learner will still not know what the base word order parameter values are. At some point the alternative V2 parameter value becomes active, finally allowing the learner the chance to find the target grammar. Since no target $+$V2 grammar has a local maximum associated with it, the learner can now successfully converge using the TLA.

Since we have exhausted the two possibilities, the learner will converge on the target grammar given the ordering of parameters proposed here. Furthermore, the above ordering solution is more robust than the initial state solution proposed in the previous section. Recall that the initial state solution will not necessarily work if the learner ever accidentally changes the value of the V2 parameter to $+$V2 early on in grammar acqui-

---

[28] We make the crucial assumption that if the grammar is $-$V2, the learner will hypothesize the correct grammar before the point at which $+$V2 becomes available. Of course, given our other assumptions, this can only be guaranteed with some probability approaching 1. Thus, there remains an extremely small probability that the learner will in fact wind up in a local maximum. (We can make this probability as small as we like by extending the period of the initial state.) In this regard, the (maturational) learning theory of this section is slightly at variance with the technical details of the learnability-in-the-limit assumptions (Gold 1967). In our opinion, though, it is in the spirit of those assumptions.

sition when the target is $-$V2. Such a possibility is ruled out by fiat in the ordering solution proposed here.

There are a number of ways of thinking about the period during which the child does not allow the $+$V2 value. At one level, one might just say that there is a period of time during which this value is not allowed, and that the possibility simply arises at a certain age. Thus, the consideration of an alternative V2 parameter value may *mature*.[29] Alternatively, we can note that a child learning a $+$V2 grammar, who has adopted the $-$V2 value (the only one available at first), will not be able to arrive at a grammar that analyzes exactly the sentences in her experience. It may be that the system is set such that after the learner has encountered a particular quantity of data that she cannot analyze (i.e., after a particular amount of "failure"), the possibility of $+$V2 emerges.[30] At the moment, no empirical data that we know of distinguish these two possibilities, although in principle they may be distinguished. For example, a child learning a grammar at a somewhat later age might arrive in a local maximum with respect to a $-$V2 target grammar if the age of the child is the relevant factor in determining whether or not both values of the V2 parameter are to be considered. If it is not the age of the child that is relevant, but a quantity of failure in attempting to analyze input data, then the older learner should go through the same stages as the younger learner and thus have no trouble with the $+$V2 grammar.

Given that this parameter-ordering solution allows a triggering learner to converge on any of the possible target grammars, it is interesting to see what the empirical consequences of such an ordering are. In general, ordering the parameters predicts that the learner should go through stages of development corresponding to the number of distinct parameter partitions. This contrasts with the initial state hypothesis, which predicts no such stages of development. The partitioning proposed here consists of the base word order parameters in the first stage, followed by inclusion of the V2 parameter in the second stage. Thus, this partitioning predicts the following:

- No child learning a $-$V2 grammar should ever hypothesize that she is in a $+$V2 grammar.
- All children learning $+$V2 grammars should go through a stage in which they consider only base word order parameters and do not consider the V2 parameter.

As far as we know, the first prediction is correct. Evaluating the second prediction is more complicated. At early ages, children presented with $+$V2 languages produce both finite and nonfinite forms in matrix sentences. However, the word orders are correct; when the forms are nonfinite, they are in base position, and when they are finite, they are in V2 position (see, e.g., Weissenborn 1990, Whitman, Lee, and Lust 1991, and

---

[29] See Borer and Wexler 1987, 1992 for evidence in favor of maturation. See also Weinberg 1987, Atkinson 1990, and Hoekstra 1990 for comments on the work done by Borer and Wexler.

[30] Notice that there is no UG-determined relation between $+$V2 and "failure to analyze." Thus, if failure does determine the development of the possibility of V2, we may say that failure is a *trigger* for this development in the alternative sense of trigger mentioned in footnote 2.

Poeppel and Wexler 1993 for relevant German data and analyses; see Wexler, to appear, for a summary of the data in additional languages along with an analysis). Hence, these children know the processes that cause V2. However, there is a general decrease in the number of nonfinite, non-V2 sentences as children grow older. What is not known is whether there is an earlier stage in which there are no finite forms (and thus no V2). At the moment there does not seem to be much evidence for such a stage, but the data are too sparse at these very early ages for us to be certain. If such a stage exists, we could say that the +V2 value was not available at this stage, consistent with the hypothesis that +V2 develops late. If such a stage does not exist, the hypothesis that +V2 develops late would have to say that the effects of the stage cannot be seen because the development of +V2 takes place before children begin producing two-word utterances. At the moment we take no position on this issue, awaiting further empirical research.

4.2.1 Alternative Parameter Orderings for the V2–X-Bar Parameter Space   It turns out that the parameter ordering we have just given is not the only one that allows a learner to converge successfully on all possible target grammars in the V2–X-bar parameter space. The ordering just given can be represented as follows:

Initial state: −V2
Partial order of parameters: ((Specifier-Head, Complement-Head) (V2))

Since this ordering is successful, an ordering that forces a strict order on the specifier-head and complement-head parameters will also be successful. Thus, the parameter orders ((Specifier-Head) (Complement-Head) (V2)) and ((Complement-Head) (Specifier-Head) (V2)) are also possible solutions to the local maximum problem posed in this article.

Deciding which of these parameter-ordering hypotheses is the most plausible will of course ultimately depend on empirical data, but until such data are available, the hypotheses with the fewest partitions are to be preferred over those with more partitions, since those with fewer partitions involve fewer stipulations, independent of the empirical consequences. That is, hypothesizing additional partitions implies hypothesizing additional stages of development. If we have no independent evidence of these additional stages of development, the most parsimonious theory will not hypothesize such states.

However, if there is more than one successful parameter ordering with the same number of partitions, then parsimony cannot decide between them. Although empirical data will ultimately make the decision here, as well, it may be possible to choose between the alternatives on linguistic grounds. It turns out that the V2–X-bar parameter space given here has a second possible two-partition parameter ordering that allows a learner to avoid local maxima while learning any of the possible grammars. This ordering is as follows:[31,32]

Initial state: $-$V2

Partial order of parameters: ((Specifier-Head) (Complement-Head, V2))

Thus, there are two successful length-2 partitions for the V2–X-bar parameter space. However, the first partition seems preferable to the second for linguistic reasons— namely, that the base word order parameters are separated in the second, whereas they are grouped as a set in the first.

4.2.2 Parameter Spaces with No Useful Parameter Orderings   Not all parameter spaces contain parameter orderings that allow a triggering learner to successfully converge on all target grammars. The fact that all of the local maxima in the given space consist of $+$V2 grammars relative to $-$V2 target grammars makes the ordering hypotheses possible. If there existed a single $-$V2 local maximum relative to a $+$V2 target grammar, then the orderings proposed here would not work.

For concreteness, suppose that there is an additional local maximum, the grammar SVO relative to the target grammar VOS$+$V2. In the proposed space, then, the grammar SVO is a local maximum for the grammar VOS$+$V2 and, symmetrically, the grammar VOS $+$ V2 is a local maximum with respect to the grammar SVO. An example of such a space is provided in table 5. This parameter space is identical to the V2–X-bar parameter space given in table 3 with the exception that some data have been removed from the first, second, and sixth grammars in order to make the grammar SVO a local maximum for grammar VOS $+$ V2.[33] Given this parameter space, there is no initial state such that a learner can converge on all possible target grammars using the parameter orderings proposed for the original V2–X-bar parameter space. The reason that the parameter orderings proposed above do not work on the new parameter space derives from the fact that it is no longer true that the only local maxima are $+$V2 grammars relative to $-$V2 grammars; there is now a $-$V2 local maximum relative to a $+$V2 grammar. Thus, if the initial state is $-$V2, then the learner can now accidentally end up in a local maximum before the V2 parameter is activated. This possibility did not exist before. Furthermore, if the initial state is a $+$V2 state, then the learner can still end up in a local maximum before the V2 parameter is activated (i.e., before the possibility of $-$V2 exists), as was true in the unmodified parameter space. Since there is no possible initial state, the parameter orderings proposed above will not work.

---

che.mit.edu (Internet address 18.88.0.85). The code is found in the file called ordering.lisp in the directory pub/gibson. The code may also be obtained by sending a request (either by regular mail or by electronic mail) to the first author.

[32] Although the complement-head parameter is not in the first set of parameters to be considered, it makes no difference which value for this parameter is selected initially. Thus, no value for the complement-head parameter is given in the initial state.

[33] Since some data have been removed from the original linguistically motivated parameter space, the resulting space has less linguistic significance. Thus, the values of the relevant parameters in table 5 are represented as 0s and 1s rather than more linguistically grounded values.

Table 5

A parameter space with no parameter orderings that allow the triggering learner to converge on all grammars

Parameter settings	Data in defined grammar
000	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 Aux V S, Adv Aux V O S
001	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 O1 O2 S, Adv Aux V O1 O2 S
010	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
011	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
100	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
101	S V, 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
110	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
111	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

Although these arguments for the inadequacy of the proposed ordering hypothesis with respect to the new parameter space are intuitively plausible, they do not constitute a proof. The proof of this inadequacy is detailed and complex: it involves checking that there is no initial state such that every target state can be attained using a parameter-ordering hypothesis from that initial state without ever possibly ending up in a local maximum. This proof involves searching a large number of cases, and hence the search was performed by computer.[34] The result was that there is no initial state such that the parameter orderings given above would always allow convergence on all grammars in the parameter space.

Moreover, it turns out that the parameter space in table 5 is such that there is no possible parameter ordering at all from which a learner can always converge on all possible grammars. It is therefore an example of a parameter space for which parameter ordering will not help a triggering learner. Thus, the fact that a parameter-ordering solution is available for the original V2–X-bar parameter space is not simply due to the fact that ordering solutions are available for all parameter spaces: ordering solutions are not in general available for arbitrary spaces. The fact that one exists for the V2–X-bar parameter space is an interesting feature of that space. Whether ordering solutions are available for all linguistically possible parameter spaces is an important open question.

4.3 Universal Grammar: No Local Maxima

Although the problem for linguistic parameter-setting acquisition is real in the parameter space examined here, it is also possible that the problem is epiphenomenal, resulting from an inadequate selection of parameter space. In particular, it may be that the parameters identified here are sufficiently different from the parameters in UG that the problem observed here never occurs. However, the parameters that we have assumed here are fairly widely accepted, although with many variations. So until a better analysis of the range of data to be handled by these parameters is proposed, we will assume that these parameters are correct.

Another possibility is that the parameter space is close to that of the relevant part of UG, but that there exist data in addition to those considered here that enable the learner to escape the current local maxima, so that there are no local maxima. The data considered here are clearly a small subset of those that need to be considered in order to conclude definitively whether or not the V2–X-bar parameter space truly contains local maxima. For example, the following kinds of data are relevant to a V2–X-bar parameter space, although such data have not been considered here. Consideration of some or all of these kinds of data could cause the local maximum problem observed here to disappear (see Bach 1962, Bierwisch 1963, Den Besten 1983, Travis 1984, and Webelhuth 1989 among many others for discussion of these and other kinds of data):

---

[34] See footnote 31 for directions for obtaining the code that performs this search.

1. Degree-1 and higher-order data. In the present study we consider only degree-0 sentences (in the sense of Wexler and Culicover 1980), so that no embedded clauses are considered with respect to the V2–X-bar parameter space. Although Lightfoot (1989, 1991) claims that linguistic acquisition takes place using only degree-0 data, it still might be the case that appropriate triggers exist among sentences that are degree-1 or higher.

2. Imperatives, which may indicate the base ordering of the head verb and its complement directly (as in English).

3. Sentence patterns including separable prefix verbs. In many + V2 languages, a finite verb can be fronted, leaving behind a verbal prefix in its base position, whereas infinitival forms of the same verbs never split apart in this way.

4. Stranding effects. In a + V2 language it is often possible to front constituents that are subparts of other constituents.

5. Scrambling in non-V2 positions. In + V2 languages, alternative word orders are frequently possible in non-V2 positions (the third and higher positions).

6. V1 orders. Some + V2 languages allow V1 orders under certain circumstances. For example, conditionals in German are V1 rather than V2.

7. Fronting of constituents other than NPs and adverbs. V2 topicalization can front VPs, V's, and even subject + Verb combinations under certain conditions (Webelhuth 1989).

8. Intonation. Information regarding which item(s) in a sentence pattern are stressed may allow the existence of triggers where none existed before. For example, topicalized nonsubject pronouns must be stressed in + V2 languages.

Our further investigations indicate that consideration of data sets 1 and 2 does not alter the results reported here. In particular, the addition of degree-1 data does not change the character of the parameter space in any helpful way for the learner, partly because an extra parameter is required: one that indicates whether or not embedded sentences are V2. Nor does the addition of data that directly indicate the underlying complement-head word order (e.g., imperatives) change the parameter space in any interesting way. Exactly the same local maxima exist, and the same solutions are again possible.[35]

The data patterns 3–8 have not yet been fully explored. As discussed above, it might be the case that the addition of (some of) these data could cause all of the local maxima in the V2 parameter space to disappear. However, it should be noted that, no matter what parameter space we are considering, it is always possible to examine more data whose addition might cause the disappearance of local maxima. We must keep in mind

---

[35] Addition of the base complement-head word order data does allow the learner to globally trigger the complement-head parameter (which was not possible before), but these global triggers do not alleviate any of the learning problems caused by the existence of local maxima. That is, although the learner may now be able to successfully trigger the complement-head parameter, the other two parameters are still interdependent for all input data.

that it is also possible that the addition of these data might not change the properties of the V2 parameter space in any interesting way. That is, it is also possible that none of the extra data are local triggers for the current local maximum grammars. These data may instead serve to strengthen currently existing local maxima, making the local maxima more like their corresponding targets, without adding the needed triggers.

Furthermore, we must be careful to ensure that the additional data do not also necessitate additional parameters. For example, there will be a parameter distinguishing grammars that are V2 in matrix clauses only (e.g., German, Dutch, and Swedish) from those that are V2 in embedded as well as matrix clauses (e.g., Icelandic and Yiddish). In order to properly characterize degree-1 and higher-order data in a V2 parameter space, it is necessary to consider this additional parametric difference. As soon as further parameters are added, more local maxima may appear. Thus, although the addition of data may cause the disappearance of the local maxima observed here, these additions might only cause such disappearances if no further parameters are considered.

Finally, and perhaps most importantly, it should be noted that although adding new data may cause the removal of local maxima, it may also *create* local maxima. For example, suppose that pattern x_1 is a local trigger for parameter value $P(v)$ from initial grammar G_i. Consideration of additional kinds of data may necessitate adding x_1 to grammar G_i, so that its local triggering status disappears. This observation leads us to distinguish two kinds of data that can be added to a parameter space: (a) data that include (new) orders of previously considered constituents; and (b) data all of which include new kinds of information, so that no orders of any of the constituent patterns to be added are in the current parameter space. The first kind of data can cause new local maxima; the second cannot. (Both kinds can also cause the disappearance of local maxima.)

The additional data sets 1–3 and 8 are of the second kind, because the sentence patterns from each would contain information not present in the current parameter space. For example, consideration of separable prefix verbs (item 3) cannot possibly cause the addition of a pattern that consists only of a new order of current constituents, because each new pattern added here will contain a constituent that was not previously considered: a separable-prefix verb (and possibly its separable prefix). On the other hand, the additional data listed in 4–7 might cause additional local maxima, because these additional data sets may include new orders of previously considered constituents. Constituent scrambling among the non-V2 positions of a V2 grammar is one such case. This scrambling may involve the addition of data that used to act as triggers, so that local maxima are created where they did not previously exist.

Thus, although it is certainly true that the space we consider here is a great simplification of the actual V2 parameter space, it is not clear that this is problematic. The more general space may well have the same properties as the simplified one. Until we discover evidence indicating that the simplified space is categorically different from the more general one, it is reasonable to tentatively extrapolate the properties of the simplified space to be those of the more general one as well.

4.4 Alternative Conceptions of What Constitutes a Linguistic Trigger

The existence of local maxima in a particular parameter space crucially depends on the existence of the Single Value and Greediness Constraints as they are defined here:

- First, if a learner does not obey the Single Value Constraint, so that her parameter-setting hypotheses can differ by arbitrarily many parameter values, then the concept of local maximum vanishes, because every set of parameter settings is attainable from every other set of parameter settings in one step.[36]
- Second, if a learner does not need to be able to analyze the current sentence, contrary to the Greediness Constraint, then local maxima can no longer exist, since nothing forces the learner to remain in a particular nontarget state. One possible alternative to the Greediness Constraint would allow the learner to change a parameter value by deducing that this parameter could not be set correctly when presented with the current constituent pattern. As a result of such deductive capabilities, the current pattern need not be analyzable under the new parameter setting, contrary to the Greediness Constraint.

  Another possible alternative to the Greediness Constraint follows from the observation that the constraint as defined here allows a change to a new grammar only when that grammar allows a complete analysis for the currently unanalyzable sentence. It is possible to relax this constraint so that the new parameter setting need only be better than the current one with respect to the input data (see Clark 1990, 1992).[37] We explore several possible metrics for ranking grammars along these lines.

When these constraints are relaxed, alternative conceptions of grammatical triggers arise. These alternatives are discussed in the following subsections.

4.4.1 Removing the Single Value Constraint

The Single Value Constraint restricts the learner from considering grammars that differ from the current grammar by values of more than one parameter. If this constraint is removed, it turns out that all local maxima disappear. However, before we simply remove the Single Value Constraint, we must make sure that a language acquisition system that lacks the constraint retains the original motivations for it. These motivations include the following:

1. Conservatism in grammar hypotheses. Children do not seem to vary wildly from day to day with respect to their grammatical hypotheses. The Single Value Con-

---

[36] Relaxing the Single Value Constraint to an n-value constraint, where n is a constant larger than 1, causes the disappearance of all the local maxima in the V2–X-bar parameter space examined here. However, under such a relaxed constraint, local maxima may still exist in parameter spaces with more parameters.

[37] This possibility was pointed out to us by David Pesetsky.

straint (along with other constraints) implements this conservatism by ensuring that new grammars are close to old grammars.[38]

2. Resource limitations. The child's resources for language acquisition are limited. The Single Value Constraint, coupled with the constraint that allows only one new grammatical hypothesis to be formed and tested per input item, ensures that only limited resources are necessary at each step in the parameter-setting process.

3. Linguistic naturalness. The theory of parameter setting has assumed that a parameter setting is tied to an input in a natural way.

If we remove the Single Value Constraint, so that the child's new grammatical hypothesis can be arbitrarily distant from the current hypothesis, we end up with a parameter-setting algorithm that is no longer conservative. The need to account for an unanalyzable input item can lead the learner to shift to a grammar that is nothing like the current one, thus allowing the possibility of a massive immediate change in her linguistic behavior. Thus, abandoning the Single Value Constraint without an alternative constraint that retains conservatism is undesirable.

One way to retain conservatism while circumventing the Single Value Constraint is to reduce the likelihood of a large change in the grammar hypothesis by trying more conservative changes first. That is, the learner might try to change only one parameter first, and if that fails to give a useful new grammar, she could then try grammars that differ by two parameter values, and then three, and so on. Although such a scheme maintains conservatism as much as possible (since no change is made unless it is really necessary), the algorithm that results requires too much processing to be realistic. For example, suppose that the learner is actively considering values for 10 parameters. Before she can try combinations of 2 parameter changes, she must verify that changes to any of the 10 current parameter values do not give a successful new grammar. And before combinations of 3 parameters are tried, all 45 possible grammars resulting from changing exactly 2 parameters must be checked. Because of the limitations on the resources available to the child acquiring language, such a scheme is not psychologically plausible.

4.4.2 Removing the Greediness Constraint   The existence of local maxima also depends on the Greediness Constraint. A learner who does not need to be able to analyze the current input under a new set of parameter values will always be able to change her hypothesis grammar as long as there are data in the target that are not in the source (see Clark 1989, 1990, 1992 and Nyberg 1991 for models lacking the Greediness Constraint). However, removing the Greediness Constraint without replacing it by another constraint

---

[38] Note that the definition of conservatism given here does not imply that the child can only propose grammars that are consistent with some subset of the utterances presented to her by adults. Contrary to such a constraint, children regularly produce utterances that are not acceptable in the adult grammar. The conservatism discussed here restricts only the quantity of grammar change that the child proposes over small amounts of time.

has unwanted consequences. Like the Single Value Constraint, the Greediness Constraint is motivated by (a) the desire to keep the learner's behavior conservative, (b) the requirement to limit the resources consumed by the learner when attempting a grammar change, and (c) linguistic naturalness.

Without the Greediness Constraint, the learner can radically alter her grammatical hypothesis very quickly, no matter how close to the target she has come. For example, suppose that the learner has set all but one of her parameters correctly, and that she is presented with a sentence pattern from the target grammar that is not in the current grammar. A nongreedy learner might change any of the parameters because of the unanalyzable data. If the learner changes the value of the wrong parameter (a likely occurrence if the grammar has many parameters), then she proposes a new grammar that is farther from the target than the original. The next time an unanalyzable sentence appears (now a more likely event, since the current grammar is worse than the original), the same process is likely to repeat itself, and the learner gets farther and farther from the target. Lacking the Greediness Constraint makes learning the grammar very difficult: a learner might be next to the target many times over, each time with a different parameter set incorrectly, before she finally accidentally achieves the target.

Although the Greediness Constraint does not guarantee that children will not go through stages in which they cannot analyze sentences that they could analyze earlier, this constraint intuitively minimizes such behavior. The parameter space is assumed to be such that most sentences are grammatical in only a fraction of all possible grammars. Hence, a greedy learner will only change her hypothesis based on an unanalyzable sentence a fraction of the times that a nongreedy learner will. Consider the case described above in which the learner has correctly set all but one of the parameters in the grammar and is now faced with a sentence that she cannot analyze. Since this sentence is assumed to be grammatical in very few possible grammars, there are many fewer possible alternative grammars for a greedy learner to adopt than for a nongreedy learner. If it is the case that, of the grammars that differ from the source grammar by one parameter value, the only grammar in which the current sentence is grammatical is the target grammar, then the learner will either (a) select the target grammar if the appropriate parameter is considered or (b) stay in the current state. Thus, in this case the greedy learner will not select a grammar that is worse than the original, whereas the nongreedy learner will likely select a worse grammar. The greedy learner will remain in the current state until she runs across more unanalyzable sentences, and she will eventually change the appropriate parameter. This learner will then converge on the target. No such easy convergence is possible without a constraint on what new grammars can be adopted. A learner who lacks such a constraint might therefore change parameter values arbitrarily and thus move randomly through the parameter space. The Greediness Constraint is one of the simplest possible such constraints.

4.4.2.1 Changing Parameter Values by Logical Deductions However, if we replace the Greediness Constraint with another constraint on grammar selection, it is possible that

we can avoid the problem of local maxima, while also avoiding the random-walk prop-
erties described above. One such constraint allows a parameter change when the learner
can logically deduce that the current setting is impossible given the input, independent
of whether or not the new hypothesized grammar would allow the successful analysis
of that input. That is, if the learner can deduce that a particular parameter must have a
particular value given the input, independent of whether or not she can now analyze the
input under the proposed grammar, then she should make this parameter change. Al-
lowing such deductions in place of the Greediness Constraint results in an alternative
definition of trigger, one in which the triggering sentence is grammatical *only if* the
requisite parameter is set correctly, rather than *if and only if* as in the definitions given
in section 1. We will call such triggers *deductive triggers* (abstracting local and global
triggers together):

> (14)   A *deductive trigger* for value v of parameter P_i, $P_i(v)$, is a sentence S from
> the target grammar L such that S is grammatical only if the value for P_i is v.

Although the use of deductive triggers allows the learner to avoid the problem of
local maxima (since the Greediness Constraint has been removed), we still need to
determine how the learner could have access to the information that a particular parame-
ter setting is incorrect. One possibility is that the learner makes this determination based
on the input sentence pattern together with calculations on different possible grammars.
That is, the learner might try all possible grammars in which the parameter in question
is set to the requisite value. If none of these grammars allow successful analysis of the
input sentence, then the learner could logically deduce that the parameter value in ques-
tion is incorrect. However, although such an algorithm is logically possible, it requires
that the child possess a large processing capacity, much more than is cognitively plau-
sible. In particular, the child needs to evaluate whether a given sentence pattern is
analyzable in any of a very large number of grammars: all those grammars that contain
a certain value for one parameter. Given just 10 active parameters, this calculation entails
examining 512 grammars in a binary-valued parameter system. This seems like far too
much processing ability to assume the child has.

If, however, the child can make the appropriate deductions without consuming vast
resources, or alternatively, if the required deductions are simply built in as part of the
learner's knowledge, then deductive triggering may yet be of use. For example, the
learner might know or somehow deduce that a + V2 value for the V2 parameter forces
the existence of a tensed verbal element in the second position of the matrix clause.
Given such knowledge, if the learner hears a sentence pattern in which the tensed verbal
element is not in second position, then she can conclude that the target grammar is not
a + V2 grammar. Note that such reasoning does not require that the learner be able to
analyze the given sentence: the learner might not be able to analyze the sentence and
still safely conclude that the target grammar is − V2. Furthermore, note that such a
conclusion does not require a large amount of computation: the knowledge that + V2

grammars always have a tensed element in second position is assumed to be either built in or easily deduced.

If deductive triggers of this form are used in conjunction with greedy triggers, all of the grammars in the V2–X-bar parameter space presented here can be learned no matter which initial parameter settings the learner has, because the only local maxima in this space are +V2 grammars relative to −V2 grammars. In particular, when the learner ends up in a +V2 grammar relative to a −V2 target, she will eventually come across a sentence in which the tensed verbal element is not in second position. When such a sentence (a deductive trigger) appears, the learner will eventually toggle the V2 parameter, thus escaping the state that used to be a local maximum.

Because deductive triggers alleviate the local maximum learning problem when used together with greedy triggers, it is reasonable to explore whether deductive triggers can replace greedy triggers altogether. In order for this to be possible, it is necessary for the learner to have or deduce knowledge of each parameter in the grammar, so that for every value of every parameter, certain sentence patterns rule out that parameter value. Possible candidates for acquiring the values of the specifier-head and complement-head parameters in the V2–X-bar parameter space defined here might be the knowledge that the presence of a tensed auxiliary verb means that the order of the main verb with respect to nonfronted subjects and objects reflects their base relative orders. Given this knowledge, the learner who is presented with the sentence pattern $Adv\ Aux\ S\ O\ V$ could correctly deduce that the value for the specifier-head parameter is spec-first and that the value for the complement-head parameter is comp-first.

Although the presence of such knowledge allows the learner to acquire the target grammar by the use of deductive triggers alone, the learner who uses deductive triggers needs either (a) a great deal of redundant knowledge of her language to be built in or (b) certain as-yet-unspecified deductive capabilities that consume small amounts of resources, but allow the learner to make appropriate deductions about which parameter values are ruled out by certain data patterns. The possibility that the learner has these deductions built in is not attractive, since this extra knowledge is completely redundant with the rest of the grammar. Thus, we may rule out this possibility.

The alternative that the learner can somehow logically deduce necessary information from her knowledge of grammar is more plausible than having these deductions built in. However, this alternative requires that the learner make a number of deductions regarding the grammar being acquired before she acquires values for the parameters. No such deductions are necessary if the learner relies only on greedy triggers. Furthermore, and perhaps most importantly, the general form of the deductive processes that the learner can perform is not clear, nor is it clear how to constrain these processes so that they do not require too many resources. In fact, the difficulties associated with quantifying which kinds of grammatical deductions are possible and which are not provided a primary motivation for researchers in linguistic theory to (implicitly or explicitly) postulate a greedy trigger theory of parameter setting. As a result, the assumption that the

learner reasons about sentences she cannot grammatically analyze is at odds with the parameter-setting theory as it is usually laid out. In this literature, the existence of greedy triggers is assumed to be sufficient for learning. The assumption that the kind of reasoning outlined above is possible would be more compatible with a pre-parametric "hypothesis-testing" theory of language acquisition (Chomsky 1965). Thus, although the use of deductive triggers may solve the learning problem presented here, the required additional knowledge and/or deductive power on the part of the learner makes this possible solution less attractive than some others that have been suggested.

4.4.2.2 Changing Parameter Values Based on Partial Analyses   Another alternative to the Greediness Constraint as stated is a relaxation of this constraint that requires the existence of only a *better* analysis for an unanalyzable sentence pattern, according to some predefined metric, rather than a *perfect* analysis. Under such an extension of the Greediness Constraint, a parameter value would still be altered only when a data pattern is unanalyzable, but changing a parameter value would now not require a complete analysis of the current pattern.

For example, suppose that the learner currently proposes the grammar SOV + V2 relative to the target SVO. Suppose also that the current data pattern is *Adv S V O.* Since this pattern does not have an analysis in the grammar SOV + V2, the learner considers changing a parameter. Suppose that the parameter the learner chooses is the V2 parameter, so that the prospective new grammar is the grammar SOV. Although the sentence pattern *Adv S V O* is not grammatical in the grammar SOV, the learner may move to this grammar anyway if the current sentence pattern can be analyzed better in the grammar SOV than it could in the source grammar SOV + V2. In particular, the learner's function for evaluating relative goodness of a grammar for a particular sentence pattern may rate the grammar SOV more highly than SOV + V2, so that the change will be made. For example, the learner may rate the grammar SOV as better than SVO + V2 because there is an analysis for the first three components of the input pattern *(Adv S V)* in the grammar SOV, whereas there is no such possible analysis in the grammar SOV + V2. One metric that has this effect rates one grammar as better than another if there exists an analysis for some prefix of the input pattern, ignoring the rest of the items. This metric allows a complete analysis of the input sentence to trigger a parameter (as under the strict definition of Greediness), while also allowing patterns like *Adv S V O* to trigger a parameter value without requiring complete analysis under the new parameter setting.

Although the existence of some metric that might circumvent the local maximum problem is plausible in the abstract, it turns out to be difficult to provide such a function that is (a) effective, (b) psychologically plausible (not requiring too much computational effort or memory on the part of the child), and (c) independently motivated (i.e., not stipulative). In order to see some of the problems associated with providing such a metric, let us consider the grammar VOS + V2 relative to the target SVO, a local maximum grammar when the complete analysis of a data pattern is required in order to change

parameter values. Suppose that the learner encounters the pattern *Adv S V O*. Since this pattern is not grammatical in the source grammar, it is a prospective triggering sentence. Of course, because this is a local maximum grammar under the strict Greediness Constraint, none of the alternative grammars one parameter value away (call these *local* grammars) allow the complete analysis of this sentence. Let us therefore consider the possible existence of a metric that ranks this sentence higher in a local grammar than in the current grammar. Let us first consider the metric discussed above that rates one grammar as better than another if there exists an analysis for some prefix of the input pattern. It turns out that this metric does not rank any of the grammars one parameter value distant from the grammar VOS + V2 as better than VOS + V2. For example, consider the grammar VOS. None of the patterns *Adv, Adv S, Adv S V*, or *Adv S V O* is analyzable in either the grammar VOS + V2 or the grammar VOS. The same is true for the grammars OVS + V2 and SVO + V2. Hence, this metric is of no use here.

A second possible metric—an extension of the first—rates one grammar G_1 as better than another G_0 if (a) there exists a prefix P of a sentence pattern S (in the form of constituents or words) that is accepted by G_1 and (b) there is no sentence pattern accepted by G_0 for which P is a prefix. In such a case S is a trigger for G_1. For *Adv S V O* to be a trigger from VOS + V2 under such a metric, one of the following sentence patterns must be (a) unanalyzable under VOS + V2 and (b) analyzable under a local grammar: *Adv xxx, Adv S xxx, Adv S V xxx*, or *Adv S V O xxx*, where *xxx* covers any input string. The first of these patterns, *Adv xxx*, does not distinguish any of the grammars, since all have sentences satisfying this pattern. The second pattern is equally unhelpful, because although there is no sentence pattern of this form that is analyzable in the source VOS + V2, this pattern is also not analyzable in any of the grammars one parameter value away. Similarly for the other two patterns. Thus, this metric does not allow the sentence pattern *Adv S V O* to act as a trigger, and it does not seem to work for other sentence patterns either.

Another way of rating grammars relative to a given input sentence pattern is in terms of the number of positional matches each grammar allows for alternative orderings of the same constituents. For example, the pattern *Adv S V O* from grammar SVO is unanalyzable in grammar VOS + V2. The only ordering of these constituents that is analyzable under the grammar VOS + V2 is *Adv V O S*. Only one of the four items in this pattern is in the same position as in the input sentence pattern: *Adv*. By the above metric, the grammar SVO + V2 contains a better match for the pattern *Adv S V O*, namely, the pattern *Adv V S O*, since two of the positions in this pattern match those of the input sentence *(Adv* and *O)*. Thus, a learner using this metric can be triggered by the sentence pattern *Adv S V O* out of the grammar VOS + V2 toward the target SVO, a grammar transition that was impossible when complete analysis was required. (The pattern *Adv S Aux V* provides a similar triggering sentence under the proposed metric.) Furthermore, this metric allows the learner to escape all of the other previously problematic states, so that the local maximum problem disappears.

Although the proposed metric would allow the learner to shift parameter settings as desired, it must also be demonstrated that this metric is psychologically plausible, not requiring excessive memory or computation on the part of the learner. As stated, the metric requires that the learner find an analysis for the constituents in the given input pattern that has the most positional overlap with the original order. The learner may arrive at this analysis in one of two ways: either she finds all analyses of all possible orders of the words in the input, or she has some built-in knowledge that enables her to rule out many of these computations, so that she needs to construct only a small number of such analyses. The first possibility is clearly too expensive for a learner with small finite resources. Even the second possibility seems too expensive, since it requires the learner to construct analyses of alternative orders of the input string, a constraint that requires a great deal of memory. Thus, although it seemed promising, this metric might require too much processing capacity on the part of the child to be psychologically plausible.

A related metric that does not seem to require as much memory is one that rates a grammar highly if it allows an analysis of a pattern that has the same first and last items as the input pattern. Thus, given the pattern *Adv S V O*, this metric accords a high rank to grammars that allow some pattern of the form *Adv xxx O*, where *xxx* is a (possibly empty) list of intervening patterns. This metric would distinguish the grammar SVO + V2 from the grammar VOS + V2, because although the grammar VOS + V2 does not allow a pattern of the requisite form, the grammar SVO + V2 does. Furthermore, this metric requires less memory on the part of the learner than its predecessor, because there is now no constraint on what kinds of constituents may come between the first and final items in this pattern.[39]

Although the new metric requires fewer resources than the previous one, it is very stipulative. In particular, there exist simpler metrics of similar forms that do not give the desired result. Given the existence of the simpler metrics, the question arises why the learner would use a more complex metric when simpler ones are available. For example, the first metric described above ranked grammars by their ability to analyze some prefixed subcomponent of the input. This metric is simpler than the one just defined because it examines only the prefix of the input string; but it turns out that such a metric is not useful in the problem under consideration. Unless independent motivation is given for a complex metric, we rule out such a metric as a stipulative solution to the local maximum problem.[40]

---

[39] This lack of constraint makes it easier for the learner to come up with analyses that fit the pattern. However, if no such patterns exist, then the computation to verify that none exist becomes more difficult. Thus, we must also assume that this computation is bounded above by some maximal level of effort that could be spent on any one input pattern. Of course, given this upper bound, the learner could erroneously decide that there is no analysis matching a given pattern, when in fact there is such a pattern. Such mistakes are inevitable in a resource-limited system.

[40] There exist many stipulative metrics that would allow the learner to escape the local maxima in the space examined here. One such comparison function that enables the learner to escape the source grammar VOS + V2 on the way to the grammar SVO considers only the second position of the source and target gram-

Thus, although there may exist a metric that the learner could use to avoid the local maximum problem observed here, we have not been able to identify such a metric that is simple, effective, and psychologically plausible. An important open question is whether there is any such metric that avoids local maxima for all possible linguistic parameter systems. Because the existence of such a metric would allow the Greediness Constraint to be relaxed so as not to necessitate a complete analysis of the input string in a new grammar, we will continue to search for such metrics in future work. An important open question is whether there is any such metric that avoids local maxima for all possible linguistic parameter spaces.

5 Conclusions

In this article we have formalized the notion of linguistic trigger and explored whether a grammar learner uses triggers when deciding among possible parameter settings in a linguistic parameter space. A learner who relies exclusively on triggers can end up in a state, a local maximum, from which she will never converge on the target grammar. In addition, we have presented an example of a linguistically motivated parameter space that contains such local maxima, and we have explored a number of possible implications of the existence of local maxima. Those that we did not logically rule out are as follows:

1. There is a default state, − V2. The learner who obeys the Single Value and Greediness Constraints can never arrive at a local maximum grammar.
2. The learner orders her consideration of the parameters so that no local maxima are ever entered.
3. The local maxima found in this parameter space do not exist in UG, either because the parameters explored here are not found in UG, or because additional data allow the learner to escape the local maxima located here.
4. The learner does not obey either one or both of the Single Value or Greediness Constraints, instead relying on an alternative notion of trigger.

In order to choose among these alternatives, more research will need to be conducted. Specifically, additional parameters and further data may help to rule out one or more of these choices. Empirical work in which the order of acquisition of the relevant parameters is examined may also help to tease apart the possibilities.

---

mars: if the constituent in the second position in each is the same, then this function returns *success*, otherwise it returns *failure*. Thus, one grammar is ranked over another based on an input sentence if there exists an analyzable pattern whose second element is the same as in the pattern in question. For example, the learner who relies on such a metric could be triggered from VOS + V2 to VOS after being presented with a sentence whose second item is S (e.g., Adv S V O) because (a) there is no pattern in VOS + V2 whose second element is S, and (b) there exists such a pattern in the grammar VOS (e.g., V S).

Although using this metric may allow the learner to escape the local maximum VOS + V2, it must be kept in mind that the only pattern that allows the triggering to take place is the pattern V S. Because the patterns V S and Adv S V O do not seem like close matches in any place except their second positions, the metric seems rather implausible as it stands.

The major purpose of this article has been the systematic analysis of the notion of trigger. Our results on computational learnability studies of parameter setting show that the theory of triggers can provide the basis for the theory of parameter setting, with only slight modifications of the original theory. From a more general point of view, we believe that we have demonstrated once again that the attempt to formally show that a linguistic theory of variation has the property of feasibility can shed light on the nature of grammar and its development.

Appendix A: The Local Maximum Search Algorithm

Function GET-LOCAL-MAXES (parameter-space)

1. For each target grammar, consisting of a vector of parameter settings G_{target}, do the following:
   a. Compute as follows all the single-move transitions that are possible from one grammar G_i to another G_j given data from G_{target}:
      i. TRANSITIONS (G_{target}) = { }
      ii. For each initial grammar G_i:
         For each final grammar G_j differing from G_i by the value of exactly one parameter:
         Add the pair (G_i, G_j) to TRANSITIONS (G_{target}) if there exists a sentence pattern from G_{target} (a local trigger) that is in G_j but not in G_i.
   b. For each possible initial grammar G_{source}, compute all the grammars that can be arrived at from G_{source} given data from G_{target}. This is accomplished by applying the recursive function CALCULATE-CONNECTED-GRAMS to the grammars G_{target} and G_{source}. The result of this function call is stored in the matrix CONNECTED-GRAMMARS (G_{target}, G_{source}).
2. LOCAL-MAXES = { }
3. For each target grammar G_{target}:
   For each source grammar G_{source}:
   If the G_{target} is not in CONNECTED-GRAMMARS (G_{target}, G_{source}), then add the pair (G_{source}, G_{target}) to LOCAL-MAXES.
4. Return LOCAL-MAXES.

Function CALCULATE-CONNECTED-GRAMS (G_{target}, G_{source})

If there is a value in CONNECTED-GRAMMARS (G_{target}, G_{source}), then return this value. (This value could have already been calculated by this function in an early call to it.)

Otherwise:

1. CONNECTED-GRAMMARS (G_{target}, G_{source}) = $\{G_{source}\}$
2. For each transition pair from TRANSITIONS (G_{target}) = (G_{source}, G_{new}):
   If G_{new} is not already in CONNECTED-GRAMMARS (G_{target}, G_{source})
   Then set CONNECTED-GRAMMARS (G_{target}, G_{source}) to be the union of

CONNECTED-GRAMMARS (G_{target}, G_{source}) and CALCULATE-CONNECTED-GRAMS (G_{target}, G_{new}).

3. Return CONNECTED-GRAMMARS (G_{target}, G_{source}).

Appendix B: A Partial Trace of the Local Maximum Search Algorithm for the Target Grammar SVO

First, we compute the single-move transitions that are possible from a grammar G_s, given data from a target grammar G_t. For example, let us consider the source grammar OVS relative to the target SVO. We want to compute whether the TLA, given input from the grammar SVO, will allow a shift to any of the grammars OVS + V2, VOS, and SOV: the grammars that differ from OVS by one parameter value. Consider the grammar OVS + V2. The intersection of the sentence patterns in the target grammar SVO and the new grammar OVS + V2 is nonempty:

```
(intersection
 {(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)}
 {(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)}
)
= {(S V) (S V O) (S Aux V)}
```

Furthermore, there are sentence patterns in this intersection that are not in the source grammar OVS:

```
(set-difference
 {(S V) (S V O) (S Aux V)}
 {(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)}
)
= {(S V) (S V O) (S Aux V)}
```

Each of the resulting sentence patterns from SVO can therefore act as a trigger to take the learner from OVS to OVS + V2. Thus, we add the pair (OVS, OVS + V2) to the set of single-move transitions. The remaining single-move transitions are constructed similarly.

Next, for each grammar we calculate which other grammars the learner can travel to, using the single-move transitions. (The function CALCULATE-CONNECTED-GRAMS per-

forms this step.) We calculate that the learner, given input from SVO, can travel from OVS to SVO. (In fact, there are three possible cycle-less paths that the learner could take: 1. OVS → VOS → SVO; 2. OVS → SOV → SVO; and 3. OVS → SOV → SOV + V2 → SVO + V2 → SVO.)

Given the matrix of connected grammars, the local maxima fall out as simply the grammars that are not connected to their respective target grammars. For example, it is calculated by CALCULATE-CONNECTED-GRAMS that it is possible to arrive at the target SVO from all grammars except VOS + V2 and OVS + V2. Thus, VOS + V2 and OVS + V2 are local maxima with respect to the target SVO.

References

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

Atkinson, Martin. 1990. The logical problem of language acquisition: Representational and procedural issues. In *Logical issues in language acquisition*, ed. Iggy M. Roca, 1–31. Dordrecht: Foris.

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

Baker, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10:533–581.

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

Besten, Hans den. 1983. On the interaction of root transformations and lexical deletive rules. In *On the formal syntax of the Westgermania*, ed. Werner Abraham, 47–131. Amsterdam: John Benjamins.

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

Borer, Hagit, and Kenneth Wexler. 1987. The maturation of syntax. In *Parameter setting*, ed. Thomas Roeper and Edwin Williams, 123–172. Dordrecht: Reidel.

Borer, Hagit, and Kenneth Wexler. 1992. Biunique relations and the maturation of grammatical principles. *Natural Language & Linguistic Theory* 10:147–189.

Brown, Roger, and Camille Hanlon. 1970. Derivational complexity and the order of acquisition of child speech. In *Cognition and the development of language*, ed. John R. Hayes, 11–53. New York: John Wiley.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1980. Principles and parameters in syntactic theory. In *Explanation in linguistics: The logical problem of language acquisition*, ed. Norbert Hornstein and David Lightfoot, 32–75. London: Longman.

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

Chomsky, Noam. 1986a. *Barriers*. Cambridge, Mass.: MIT Press.

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

Chomsky, Noam. 1991. Some notes on the economy of derivation and representation. In *Principles 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. 1989. On the relationship between the input data and parameter setting. In *Proceedings of NELS 19*, 48–62. GLSA, University of Massachusetts, Amherst.