The Development of Language

Acquisition, Change, and Evolution

David Lightfoot
University of Maryland

BLACKWELL
Publishers

6

Cue-Based Acquisition and Change in Grammars

Grammars, in our perspective, are mental entities which arise in the mind/brain of individuals when they are exposed as children to some triggering and shaping experience. In that case, the central mystery for historical linguists is why they have anything to study: why do languages have histories? Why do changes take place, and why are languages not generally stable? In particular, why do changes sometimes take place abruptly and catastrophically, as we have argued they must? If people produce utterances corresponding fairly closely to the capacity of their grammars, then children exposed to that production would be expected to converge on the same grammar. This is what one would expect if grammars have structural stability, as we have claimed, and children are not "trigger happy," developing different grammars whenever their trigger experiences differ just a little. In that case, diachronic change would be expected only if there were some major disruption due to population movement.

6.1 Models of Learnability

Not only is this what one would expect naively and pre-theoretically; it is also what many learnability models would lead one to expect.[1] For example, Chomsky's *Aspects of the Theory of Syntax* (1965), now a classic, viewed children as endowed with a metric evaluating grammars which can generate the primary data to which they are exposed, along with appropriate structural descriptions for those data. The evaluation metric picks the grammar which conforms to the invariant principles of UG and is most successful in generating those data and those structural descriptions. The child selects a grammar which matches his or her input as closely as possible. Again, if the data and the associated structural descriptions to which the child is exposed correspond fairly closely to the grammatical capacity

of some older individual, one would expect the child's evaluation metric to select the same grammar as that older individual's. This expectation is reinforced if the space of available grammars has scattered peaks, as in the *Aspects* view, and if many aspects of the input have no effect on the mature system – for example, the order in which the input is presented to the child.

The same point holds for more recent models. Gibson and Wexler (1994) posit a Triggering Learning Algorithm (TLA), under which the child-learner uses grammars to analyze incoming sentences and eventually converges on the correct grammar. They distinguish global and local triggers, but both are sentence types (p. 409). If the child-learner cannot analyze a given sentence with the current grammar, then he or she follows a certain procedure to change one of the current parameter settings and then tries to reprocess the sentence using the new set of parameter values. If analysis is now possible, then the new parameter value is adopted, at least for a while. So the TLA is error-driven and permits the child to pinpoint which parameter setting is incorrect when the learner's grammar does not give the right results. There is much to be said about the way in which this model works, and Dresher (1998) has some illuminating discussion; but what is crucial here is that the model has the child seeking grammars which permit analysis of incoming data, where the data consist of more or less unanalyzed sentences. Gibson and Wexler's table 3 (1994: 424), here reproduced as table 6.1, correlates three parameter settings (specifier-final/initial, complement-final/initial, +/– verb-second (V2)) and sets of data (listed here in terms of primitives like subject, verb, first object, second object). When exposed to some data set (right-hand column), the child selects the appropriate grammar (left-hand column), although it is not easy for the child to know which data set she is exposed to.

Clark (1992) offers a similar kind of model, but one which differs from that of Gibson and Wexler in that the child cannot pinpoint the source of a grammar's failure and revise particular parameter settings. Clark posits a Darwinian competition between grammars needed to parse sets of sentences. All grammars allowed by UG are available to each child, and some grammars are used more than others in parsing what the child hears. A "genetic algorithm" picks those grammars whose elements are activated most often. A Fitness Measure compares how well each grammar fares, and the fittest grammars go on to reproduce in the next generation, while the least fit die out. Eventually the candidate grammars are narrowed to the most fit, and the child converges on the correct grammar.[2] Clark and Roberts (1993) used this model to give an account of changes affecting the

Table 6.1 Gibson and Wexler's grammar–language pairs

Parameter settings	Data in defined grammar
Spec-final	V S, V O S, V O1 O2 S
Comp-final	Aux V S, Aux V O S, Aux V O1 O2 S, Adv V S
–V2	Adv V O S, Adv V O1 O2 S, Adv Aux V S
(VOS)	Adv Aux V O S, Adv Aux V O1 O2 S
Spec-final	S V, S V O, O V S, S V O1 O2, O1 V O2 S, O2 V O1 S
Comp-final	S Aux V, S Aux V O, O Aux V S
+V2	S Aux V O1 O2, O1 Aux V O2 S, O2 Aux V O1 S
(VOS + V2)	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	V S, O V S, O2 O1 V S
Comp-first	V Aux S, O V Aux S, O2 O1 V Aux S, Adv V S
–V2	Adv O V S, Adv O2 O1 V S, Adv V Aux S
(OVS)	Adv O V Aux S, Adv O2 O1 V, Aux S
Spec-final	S V, O V S, S V O, S V O2 O1, O1 V O2 S, O2 V O1 S
Comp-first	S Aux V, S Aux O V, O Aux V S
+V2	S Aux O2 O1 V, O1 Aux O2 V S, O2 Aux O1 V S
(OVS + V2)	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	S V, S V O, S V O1 O2
Comp-final	S Aux V, S Aux V O, S Aux V O1 O2, Adv S V
–V2	Adv S V O, Adv S V O1 O2, Adv S Aux V
(SVO)	Adv S Aux V O, Adv S Aux V O1 O2
Spec-first	S V, S V O, O V S, S V O1 O2, O1 V S O2, O2 V S O1
Comp-final	S Aux V, S Aux V O, O Aux S V
+V2	S Aux V O1 O2, O1 Aux S V O2, O2 Aux S V O1, Adv V S
(SVO + V2)	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	S V, S O V, S O2 O1 V
Comp-first	S V Aux, S O V Aux, S O2 O1 V Aux, Adv S V
–V2	Adv S O V, Adv S O2 O1 V, Adv S V Aux
(SOV)	Adv S O V Aux, Adv S O2 O1 V Aux
Spec-first	S V, S V O, O V S, S V O2 O1, O1 V S O2, O2 V S O1
Comp-first	S Aux V, S Aux O V, O Aux S V
+V2	S Aux O2 O1 V, O1 Aux S O2 V, O2 Aux S O1 V
(SOV + V2)	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

verb-second properties of early French, by allowing an arbitrary degree of misconvergence by children.

What these models have in common is that learners eventually match their input, in the sense that they select grammars which generate the sentences of the input. It is only accurate grammars of this type which are submitted to Chomsky's (1965) evaluation metric, and Gibson and Wexler's error-driven children react to inaccurate grammars by seeking new parameter settings until a sufficient degree of accuracy is achieved. Models of this type can characterize instances of language stability straightforwardly. The child converges on a grammar which analyzes the input successfully, where the input consists of sets of sentences. In that case, the grammar will resemble closely the grammar or grammars which generate that input. Such models can also handle cases of mixed input under conditions of population movement. There again the child is presented with a set of data, in this case data yielded by diverse grammars; he or she converges on a grammar which is most successful in generating that data set, sometimes a grammar quite different from any of those in the previous generation. This would be a case of grammar change, and the new grammar might yield structural descriptions and some sentences which differ from those of the input; but the new grammar would result from the child's effort to match the input sentences as closely as possible.

These are not *pure* input-matching models, of course, of the type advocated by MacWhinney and Bates (1990), in which it is mysterious why children should ever produce non-adult forms in any systematic way. Clark, Gibson, and Wexler's children are not dependent only on the input; they operate in a space defined by UG. Consequently, each intermediate stage for the developing child is represented by some set of UG-defined parameter settings, and that set may generate non-adult forms.

However, it is a fact that sometimes children do not match their input at any stage, including the final stage. One instance would be abrupt, catastrophic change, discussed in the last two chapters. Consequently, it will not be productive to insist that successful language acquisition always has children converging on grammars which generate the sentences that they hear.

It is hard to see how these input-matching models can succeed when children are exposed to unusual amounts of artificial or degenerate data, which in fact are not matched. In particular, it is hard to see how they could account for the early development of creole languages, as described by Bickerton (1984, 1998) and others. In these descriptions, early creole speakers are not matching their input, which typically consists to a large degree of pidgin data. Pidgins are primitive communication systems, cobbled

together from fragments of two or more languages. They are not themselves natural languages, and they tend not to last long, before giving way to a creole with all the hallmarks of a natural grammar. The first speakers of creoles go way beyond their input in some ways, and in other ways fail to reproduce what they heard from their models, arriving at grammars which generate sentences and structural descriptions quite different from those of the input.

Let us call this the "abrupt" view of creolization (following Thomason and Kaufman 1988). There is a dramatic discrepancy between what early creole speakers hear in childhood and what their mature grammars eventually characterize as well-formed, much greater than in noncreole contexts. Bickerton deals with plantation creoles, where new languages appear to be "formed in the space of a single generation" (1998). He argues, surely correctly, that situations in which "the normal transmission of well-formed language data from one generation to the next is most drastically disrupted" tell us something about the innate component and how it determines acquisition. They certainly show that children do not always proceed by converging on a grammar which matches the input.

The abrupt view of creolization is more controversial than it should be. It offends a commitment to the proposition that languages generally change only gradually, which we discussed in chapter 4. This commitment is linked to a highly data-driven view of language acquisition, and it is widely and deeply held, including by creolists. Creolists committed to gradualism – for example, Carden and Stewart (1988) – insist that creoles emerge gradually as a result of changes introduced primarily by adults, as they re-lexify their own languages. However, if this is generally true, if this is most of the story, and if creolization for the most part mirrors adult second language learning and is not abrupt and instantiated by children, then there is little reason for theoreticians to be interested in the phenomenon. Our data about the early stages of creole languages are generally not very rich, and if one is interested in adult second language learning, one is probably better off refining theories in the light of better data sources.

Here I want to argue the following: that existing models of learnability and existing models of language change commit us to insisting that languages are basically stable. This conforms to the views of some historians that change is inherently piecemeal and gradual. But a better model of learnability enables us to better understand historical change. On this model we shift our focus from language change to change in grammars, and we expect grammar change to be abrupt, sudden, and what in chapter 4 I called "catastrophic." Our learnability model, in turn, will allow us to

capture the contingent nature of historical change and to avoid the excessively principled accounts of change offered by some historians.

I shall make some more detailed claims about how grammars are acquired and offer a different model of acquisition. I shall discuss the nature of the experience which triggers the development of grammars, arguing that children scan their environment for designated structures or "cues," and that they are not influenced by the set of sentences generated by their grammars. Indeed, there are no independent "parameters"; rather, some cues are found in all grammars, and some are found only in certain grammars, the latter constituting points of variation.

Ironically, the best-worked-out model of parameter setting comes from phonology and the work of Dresher and Kaye (1990). The notion of binary parameters has not played an extensive role in the phonological literature, but Dresher and Kaye identified parameters for stress systems, a rather well-studied area of phonology. Furthermore, they developed a "cue-based" theory of acquisition, now clarified, elaborated, and generalized by Dresher (1998). On this view, UG specifies not only a set of parameters, but for each parameter a cue. I amend this view slightly and say that cues which are realized only in certain grammars constitute the parameters, the points of variation between grammars. A cue is some kind of structure, an element of grammar, which is derived from the input. The cues are to be found in the mental representations which result from hearing, understanding, and "parsing" utterances. As a child understands an utterance, even partially, he or she has some kind of mental representation of the utterance. These are partial parses, which may differ from the full parses that an adult has. The learner scans those representations, derived from the input, and seeks the designated cues.

The child scans the linguistic environment for cues only in simple syntactic domains; this is the "degree-0 learnability" of some earlier work (Lightfoot 1991, 1994). Learners do not try to match the input; rather, they seek certain abstract structures derived from the input, looking only at structurally simple domains, and they act on this without regard to the final result. That is, a child seeks cues and may or may not find them, regardless of what the emerging grammar can generate; the output of the grammar is entirely a by-product of the cues that the child finds, and the success of the grammar is in no way based on the set of sentences that it generates, unlike in input-matching models. The child's triggering experience, then, is best viewed as a set of abstract structures manifested in the mental representations which result from parsing utterances; some of those representations constitute partial parses, which lack some of the information in mature, adult parses.

Dresher (1998) illustrates the cue-based model of acquisition with some phonological parameters. The essential feature is that a cue-based learner does not try to match target input forms, but uses them as sources of cues. The trigger consists not of sets of sentences, but rather of partially analyzed syntactic structures, elements of the internal grammar, "I-language;" these are the mental representations resulting from parsing utterances. So cues are intensional elements, grammar fragments. A cue-based learner sets a specifier-head parameter (specifier precedes/follows its head) on the basis of exposure to data which must be analyzed with a specifier preceding its head – e.g. [$_{Spec}$[John's] $_{N}$[hat]]. This parameter can only be set, of course, when the child has a partial analysis which treats *John's* and *hat* as separate words, the latter a head noun, etc. In this way, the order in which parameters appear to be set, the "learning path" (Lightfoot 1989), reflects dependencies among cues and follows from their internal architecture. To take another example, exposure to a phrase *student of linguistics* may trigger an analysis which generates complements to the right of their head noun, but this can happen only when the child already knows that *student* is a noun which assigns a thematic role to the phrasal element *linguistics*.

Some version of this cue-based approach is implicitly assumed in some work on acquisition – for example, in the work of Nina Hyams (1986, 1996) and in my own work (Lightfoot 1989, 1991). More recently, Janet Fodor has developed similar ideas, developing the parsing devices needed for a child to identify "structural triggers" – what I have called "cues."

The idea that language acquisition is cue-based and does not proceed directly by input matching results to some extent from work on abrupt language change, where children arrive at grammars which generate data quite different from grammars of an earlier generation. In the next sections I shall suggest that the model gives us a way of understanding two syntactic changes in the history of English. Then I shall turn to the analysis of creoles and signed languages.

The next two sections deal with two well-studied catastrophic changes affecting the history of English, and I shall give a cue-based account of the changes. I will discuss the changes as theory-neutrally as possible, but the change involving V to I raises questions about the relationship between morphological and syntactic properties and needs to be treated in the context of models which postulate a substantive connection between morphology and syntax.

These two changes are partially understood, and they illustrate how the study of a change is intimately connected, on this approach, with work

on grammatical theory and with work on cue-based acquisition. They also illustrate what further work is needed for a fuller understanding.

6.2 Cue-Based Acquisition and Loss of Verb-Second

Verb-second languages, like Dutch, characteristically show the finite verb in second position in simple clauses. It doesn't matter how the utterance begins; a finite verb always occurs in second position. The verb is preceded by any phrasal category: a subject NP/DP (1a), a direct object NP/DP (1b), a PP (1c), or an adjective phrase (AP) (1d, e). The verb may not occur in any position other than second (1f).

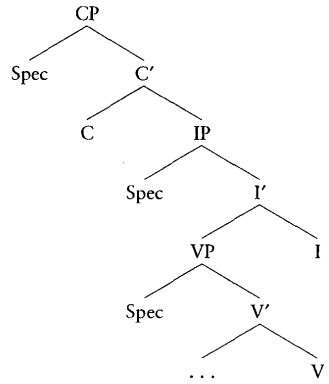
1 (a) $_{DP}$[Wij] $_{V}$zagen vele studenten in Amsterdam.
"We saw many students in Amsterdam."
(b) $_{DP}$[Vele studenten] $_{V}$zagen wij in Amsterdam.
(c) $_{PP}$[In Amsterdam] $_{V}$zagen wij vele studenten.
(d) $_{AP}$[Boos over de regering] $_{V}$zijn de studenten.
"The students are angry about the government."
(e) $_{AP}$[Vaak] $_{V}$zagen wij vele studenten in Amsterdam.
"Often we saw many students in Amsterdam."
(f) *$_{PP}$[In Amsterdam] wij $_{V}$zagen vele studenten.
*$_{AP}$[Vaak] wij vele studenten in Amsterdam $_{V}$zagen.

The most familiar verb-second languages are asymmetric and show verb-second order in simple, matrix clauses only, while embedded clauses usually have the verb in some other position. In Dutch, the embedded verb occurs in clause-final position (2), but other verb-second languages like Swedish have the verb in medial position – Dutch grammars have object–verb order, and Swedish have verb–object.

2 (a) . . . dat wij vele studenten in Amsterdam $_{V}$zagen
. . . that we many students in Amsterdam saw
(b) . . . dat de studenten boos over de regering $_{V}$zijn
. . . that the students angry about the government are

In chapter 3 (section 3.2) I outlined the basic structure of a clause for grammars of English speakers, but in Dutch and German it would be slightly different, because these languages are I-final and V-final (3).

3



Hans den Besten (1983) assumed this kind of structure when he proposed what became the standard analysis for verb-second languages (4).

4 $_{CP}$[Spec C $_{IP}$[Spec $_{VP}$[. . . V . . .] I]]

The finite verb moves to I and then to C, and some phrasal category moves to the specifier of CP; that may be NP, DP, PP, or AP (hence "XP").

But this standard analysis provides (a) no relation between XP-to-Spec movement and V-to-I-to-C movement and (b) no explanation for the obligatory character of the verb-second phenomenon (positing an attractive feature in C merely restates the problem). Linguists know that there is a relationship between the two movements, that an inflected verb *must* move to C if a XP moves to the specifier of CP, and they know this because of negative data, i.e. *un*grammaticality judgments: if some XP moves to the specifier of CP and no finite verb moves to C, the resulting structure is ungrammatical (1f). Therefore, the explanation for the relationship cannot be data-driven; it must come from UG. Similarly for the obligatory character of the verb-second phenomenon, as we saw in chapter 4. The question I want to raise here is: What does the verb-second child learn, and how does he or she learn it?

In Lightfoot 1991 I argued that verb-second children learn that utterances begin with an arbitrary XP with no particular grammatical or thematic role. It is irrelevant whether the XP is a subject or a direct object (1a, b), or whether it is an Agent (1a), a Patient/Theme (1b), a Location (1c), etc. If the specifier of IP is associated with subjecthood, then the arbitrary XP must be in the specifier of some functional category above I, i.e. Spec of CP in the system of (3). Furthermore, as noted, we need a UG condition that lexical material in the specifier of CP needs to be licensed by a lexically filled C; one possibility, due to Uriagereka (1988), is that IP is a barrier to extraction unless it is governed by a lexical C. The only head which can move to C without violating the usual conditions on head movement (Aoun et al. 1987; Rizzi 1990) is a verb, moving through I and therefore picking up finiteness features. This gives rise to the structure (5), and we account for why the finite verb *must* move to C.

5 $_{Spec_{CP}}$[XP] $_{C}$[Vf] $_{IP}$[. . .

If this analysis is correct, then what the verb-second child learns is that utterances begin with an arbitrary phrasal category. The rest of the information comes from UG. Therefore, the *cue* is (6), an abstract structure and an element of I-language:

6 $_{Spec_{CP}}$[XP]

Dutch and German children know that the initial element is in the specifier of CP, because it is invariably followed by a finite verb, unlike, say, topics in English, French, and other non-verb-second languages; the topics in these languages occur in a fronted position but are not followed by a finite verb, and therefore are not in the specifier of CP (7), presumably reflecting some kind of adjoined structure.[3]

7 (a) Peter, I like (him).
(b) Pierre, je l'aime.

Although the initial XP is in principle of arbitrary grammatical function, statistical counts for Dutch, German, Norwegian, and Swedish show that it is a subject about 70 percent of the time in conversational speech (Lightfoot 1993 has the details). Presumably it is those 30 percent nonsubjects which are a crucial trigger for inducing children to postulate that the XP is of arbitrary grammatical and thematic function, hence in the specifier of

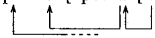
CP and not in the specifier of IP (or whatever position is associated with subjecthood). That is to say, the cue (6) must be attested robustly in the primary linguistic data. What we are interested in is the "expression" of the cue (adapting some terminology from Clark 1992), i.e. those utterances which can *only* be analyzed as $_{SpecCP}$[XP]. In this case, this means initial nonsubjects, because a sentence like (1a) might be analyzed with the subject NP/DP in the specifier of IP. Here we can quantify the robustness of the cue: at least 30 percent of matrix clauses express the cue. There is no reason, of course, to believe that there is anything magical about the 30 percent figure, and no reason to believe that there should be a general, cross-cue definition of robustness. We shall return to the matter of the robustness of the cue later in the chapter.

There are many questions glossed over here, but this seems to be a plausible account of how verb-second properties are acquired, which addresses the problems of the standard analysis. Now I shall argue that this cue-based account provides a good basis for understanding how the verb-second phenomenon was lost in the history of English.

Old English/Middle English texts sometimes show verb-second order and sometimes other orders. They *appear* to have optional verb-second, which became more frequent during the Old English/Middle English period. If the appearance were real, this would be highly problematic, because one could now not invoke UG to explain the obligatoriness of the movement in Dutch and German, as we discussed in section 4.5.

Happily, Kroch and Taylor (1997) have provided arguments for the existence of two dialects in Middle English; if they are right, then there is no homogeneous system with optional verb-second. First, there is a northern, Scandinavian-based verb-second grammar (8), which works like Dutch except that it was I-initial and V-initial – in other words a verb-second version of modern English.

8 [Spec C [Spec I [V DP]]]



This would yield (modernized) sentences like (9).

9 (a) [We] $_{V}$saw many students.
(b) [Many students] $_{V}$saw we in London.
(c) [In London] $_{V}$saw we many students.
(d) [Angry about the government] $_{V}$are many students.
(e) [Often] $_{V}$saw we many students in London.

Second, there was a southern, indigenous grammar which lacked V + I-to-C and was not verb-second.[4] Early English shows three major alternations (10) (not necessarily three distinct parameters, of course): VPs may show verb–object or object–verb order (10a), I may precede or follow the VP (10b), and there may or may not be an operation moving an inflected verb to C (10c). This leads us to expect the initial structures of (11) with or without V-to-I raising. This is enough to generate what one finds in the texts. Therefore, at least (12) exists alongside (8).[5]

10 (a) VO/OV
(b) I-medial/I-final
(c) V + I to C

11 (a) $_{CP}$[Spec C $_{IP}$[Spec I $_{VP}$[V DP]]]
(b) $_{CP}$[Spec C $_{IP}$[Spec I $_{VP}$[DP V]]]
(c) $_{CP}$[Spec C $_{IP}$[Spec $_{VP}$[DP V] I]]
(d) $_{CP}$[Spec C $_{IP}$[Spec $_{VP}$[V DP] I]]

12 . . . $_{IP}$[Spec I $_{VP}$[. . . V . . .]]

(12), assuming verb–object order and V-to-I raising, generates structures like (9a), but not (9b–e) (recall that sentences like (9a) constitute 70 percent of what verb-second children hear). With an adjunction operation, one would get equivalent expressions like (13), which could not occur in a verb-second system like (8).

13 (b) [Many students] we saw in London (equivalent to 9b)
(c) [In London] we saw many students (equivalent to 9c)

If there were multiple grammars along these lines, then the problem concerning the loss of an optional verb-second system is reconstrued: there were (at least) two coexisting grammars, one obligatory verb-second (8) and one with no V + I-to-C movement, i.e. (12), not verb-second in the usual sense. The first of these ceased to be attained, and was lost. The problem now is to find out why this grammar died out.

On our cue-based model of acquisition and assuming the diglossic analysis of Middle English, we can identify what is likely to have militated against the survival of the verb-second grammar. First, children in Lincolnshire and

Yorkshire, as they mingled with southerners, would have heard sentences whose initial elements were nonsubjects followed by a finite verb less frequently than the required threshold; if we take seriously the statistics from the modern verb-second languages and take the threshold to be about 30 percent of matrix clauses with initial nonsubject in the specifier of CP, then southern XP-Vf forms, where the finite verb (Vf) is not I-final and where the initial element is not a wh- item or negative, are too consistently subject-initial to trigger a verb-second grammar. Southerners overwhelmingly produced only subjects preceding a finite verb, and, as a result, a mixture of southern and northern speech would have less than 30 percent of nonsubject followed by Vf. Ans van Kemenade has provided me with some statistics from *Sawles Warde*, an early-thirteenth-century, southwest Midlands text. She counted 152 matrix clauses, excluding coordinate clauses with missing subjects, dislocated structures (i.e. with resumptive pronouns), and initial *þa*. She found non subject-Vf (where the initial element is not a wh- element, of course) in 26 cases, i.e. 17 percent, well below the required threshold.[6] So the evidence suggests that 17 percent of initial nonsubjects does not suffice to trigger a verb-second grammar, but 30 percent is enough; somewhere between 17 and 30 percent is a phase transition. Of course, we have no idea why there should be a transition at exactly this point, and one might turn for help to the literature on complex systems (e.g. Kauffman 1995).

Second, as those northern children came into contact with southerners, they would have heard verb-third forms like (14), because southerners treated pronouns as clitics, according to most current analyses (e.g. van Kemenade 1987). However, Kroch and Taylor argue that pronouns are not clitics in northern grammars. Therefore, forms with the structure of (14a), i.e. the sentences of (14b, c), would not have been consistent with the verb-second grammars of the earlier generation; they could not be analyzed with initial $_{SpecCP}$[XP]. Prima facie they might have triggered some special adjunction operation, but evidently that didn't happen, because they do not occur in northern grammars.

14 (a) XP – subject pronoun – Vf
(b) [Æfter his gebede] he $_{V}$ahof þæt cild up (*Homilies of the Anglo-Saxon Church* (Thorpe) 2.28)
"After his prayer he lifted the child up."
(c) [Þis] he $_{V}$dyde eal for þes biscopes luuen (*AS Chronicle* (Clarke (Peterborough) 1123.73))
"This he did all for this bishop's love."

Third, the V-to-I operation was being lost from English grammars, as we shall discuss in the next section. If verbs did not raise out of the VP to I, then finite verbs could not raise further to C (*pace* Vikner 1994 and others; see n. 8 below), and there could be no verb-second effect with finite verbs. In particular, if the finite verb could not get to C, nothing could occur in the specifier of CP, given the UG condition requiring anything in the specifier of CP to be licensed by a lexical C. Consequently, sentences corresponding to (9b–e) could not be generated, and it is such sentences which expressed the cue for a verb-second system. This would further reduce the degree to which the cue for a verb-second system was manifested according to the analysis we have offered.

If grammars emerge as children scan for cues, and if $_{SpecCP}$[nonsubject] is the cue for the grammar yielding a verb-second system, then we can understand why English lost verb-second at the time and in the way it did, and we can understand why the change seems to have taken place rapidly, indeed catastrophically (van Kemenade 1987). On the other hand, if learners simply seek grammars which match the input, there is no reason why verb-second sentences should have been lost, least of all why they were lost rapidly.

I have shown how it is plausible to assume that the trigger experience that northern children had came to differ in critical ways from the trigger experiences that their parents and ancestors had, as there was more contact with the south. This produced a distributional shift in utterances manifesting the cue for the verb-second system. The primary linguistic data (PLD) changed critically, and children converged on new grammars which generated a very different output. We know from acquisition studies that children are sensitive to statistical shifts in input data. For example, Newport, Gleitman and Gleitman (1977) showed that the ability of English-speaking children to use auxiliaries appropriately results from exposure to noncontracted, stressed forms in initial position in yes–no questions: the greater the exposure to these subject–auxiliary inversion forms, the earlier the use of auxiliaries in medial position. Also Richards (1990) demonstrated a good deal of individual variation in the acquisition of English auxiliaries as a result of exposure to slightly different trigger experiences. The question is: When do trigger experiences differ critically? When do they differ in such a way as to trigger a different grammar? That is where work on historical change is so illuminating. Sometimes we can identify points at which there have been clear grammatical shifts, and sometimes we can also identify prior changes in the PLD. What is relevant now is shifts affecting specific elements of I-language, the designated

cues. By hypothesis, this shows us changes in PLD having critical effects. In fact, for the immediate future it is work on language change that is likely to be the major source of insight into what triggers particular grammars.

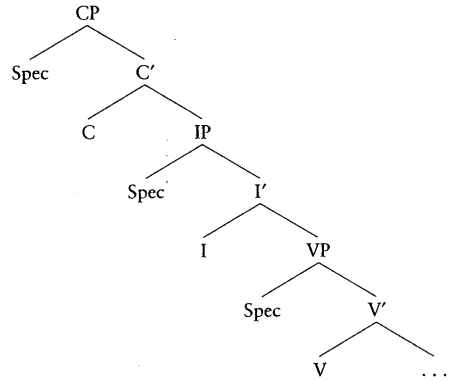
This story about the loss of verb-second systems is based on language contact. Northern children lost verb-second grammars when their PLD changed under the influence of contact with southerners. One may ask why northern children did not simply become bilingual, with two distinct systems internalized, but I have nothing enlightening to say about this beyond the discussion in chapter 4. There we discussed how the Blocking Effect militates against coexisting analyses (section 4.5). We know that children may attain distinct systems, having some kind of English grammar and a French grammar. However, we know little about why distinct systems should emerge under certain circumstances, as opposed to what happened in this case, one grammar being triggered by diverse data generated by more than one prior grammar. The account of the loss of verb-second systems happens to involve language contact, but the expression of the $_{\text{SpecCP}}$[XP] cue might drop for other reasons. If early Germanic was verb-second, then southern English must have lost the verb-second system earlier. We have no records of that change, but it may not have been the result of language contact. In the next section we shall discuss the loss of V-to-I movement, showing that the cue changes its distribution, but for reasons which have nothing to do with language contact.

6.3 V-to-I Raising and its Cue

Let us consider now a second case of grammatical change where a cue-based account is illuminating and where language contact plays no role. Operations which associate inflectional features with the appropriate verb appear to be parameterized, and this has been the subject of a vast amount of work covering many languages (see the collection of papers in Lightfoot and Hornstein 1994). We can learn about the shape of the cue(s) by considering how the relevant grammars could be attained, and that in turn is illuminated by how some grammars have changed.

Again, I adopt the familiar clause structure. Since we are now going to discuss English after the change to I-initial and V-initial order, the basic structure is (15), slightly different from the I-final and V-final Dutch system discussed in the last section.[7]

15



As we noted in the last section, subjects occur in the specifier of IP and wh-elements typically occur in the specifier of CP in the grammars of speakers of English and those of many other languages. Heads raise from one head position to another, so verbs may raise to I and then further to C. In fact, many grammars, like those of French speakers, raise their verbs to the position containing the inflectional elements (16c, d); but English grammars, unusually, do not; instead, they have an operation which lowers I onto an adjacent verb ((16a) but not (16b)). We know this because English finite verbs do not occur in some initial C-like position (17a) and cannot be separated from their complements by intervening material (17b), which strongly suggests that verbs do not move out of their VP in modern English.

16 (a) Jill $_{\text{VP}}$[leave + past]
(b) Jill $_{\text{I}}$[leave$_i$ + past] $_{\text{VP}}$[e$_i$]
(c) Jeanne $_{\text{I}}$[lit$_i$] $_{\text{VP}}$[toujours e$_i$ les journaux]
Jean reads always the newspapers
(d) Lit$_i$ $_{\text{IP}}$[elle e$_i$ $_{\text{VP}}$[toujours e$_i$ les journaux]?
Reads she always the newspapers?

17 (a) *Visited$_i$ $_{\text{IP}}$[you e$_i$ London last week]?
(b) *The women visited$_i$ not/all/frequently $_{\text{VP}}$[e$_i$ London last week]

The question I want to raise here is: What is it that forces French children to have the V-to-I operation, and what forces English children to lack the operation?

It is reasonable to construe the English lowering operation as a morphological phenomenon. It doesn't look like a syntactic operation: in general, lowering operations are unusual in the syntax, and a syntactic lowering operation here would leave behind an illicit trace, neither bound nor properly governed (Aoun et al. 1987; Rizzi 1990). Furthermore, one would expect a morphological operation, but not a syntactic operation, to be subject to a condition of adjacency. Therefore, the representation in (16a), reflecting a morphological operation, contains no trace of the lowered I. In any case, the English lowering needs to be taken as the default setting (as argued in Lightfoot 1993; Lasnik 1995; Roberts 1998); there is no nonnegative evidence available to the child which would force him or her to select an I-lowering analysis over a V-raising analysis (16b) for English, if both operations could be syntactic and subject to an adjacency requirement: children would need to know that (17a, b) do not occur (negative data, therefore unavailable as input to children). In that case, let us take the morphological I-lowering analysis as the default setting.

Now one can ask what triggers the availability of a syntactic V-to-I raising operation in grammars where it may apply. Some generalizations have emerged over the last few years. One is that languages with rich inflection may have V-to-I operations in their grammars, and rich inflection could be part of the trigger (Rohrbacher 1994). Indeed, it is striking that the verb *be* is richly inflected in standard English (*am/is/are*) and moves to I and then to C, yielding forms like *Is Bill president?*, while other forms of English (Black English Vernacular, or "Ebonics," and some forms of children's speech) do not inflect *be*, and the uninflected *be* does not move. So alongside (18a) in these types of English, one finds (18b, c), but not (18d).

18 (a) Bill be angry.
(b) Bill don't be angry.
Do Bill be angry?
What do Bill be?
(c) I don't be angry?
Do clowns be boys or girls?
Did it be funny?
(d) *Bill ben't angry.
*Be Bill angry?
*What be Bill?

However, the presence of V-to-I raising cannot be linked with rich inflection in a simple one-to-one fashion. It may be the case that if a language has rich inflection, then V-to-I raising is available (Lightfoot 1991; Roberts 1998). However, the opposite does not hold: if there is no rich inflection, a grammar may have the raising operation (Danish, Swedish,[8] and, if Otani and Whitman (1991) are correct, Chinese, Japanese, and Korean) or may lack it (English). Indeed, English verb morphology was simplified radically and that simplification was complete by 1400; however, overt V-to-I movement disappeared only in the eighteenth century, so there was a long period when English grammars had very little verbal inflection but did have V-to-I movement. In that case, there needs to be a *syntactic* trigger for V-to-I movement. So, for example, a finite verb occurring in C, i.e. to the left of the subject (as in a verb-second language or in interrogatives), could only get there by raising first to I, and therefore inversion forms like (16d) in French could be syntactic triggers for V to I.

On a cue-based acquisition approach, one would say that the cue for grammars raising V to I is a finite verb in I: i.e. $_{I}$[V], an element of I-language. Children seek this cue in the representations resulting from their (partial) parses. One unambiguous instance of $_{I}$[V] is an I containing the trace of a verb which has moved on to C, as in the structure of (16d). I would guess that (16d) would be a very important expression of the cue, and I doubt that structures like (16c) or (17b) would be robust enough to trigger V to I in isolation; this can be tested (see below). Again adopting Clark's terminology, one can ask how robustly the cue is "expressed." It is *expressed* robustly if there are many simple utterances which can be analyzed by the child only as $_{I}$[V]. So, for example, the sentences corresponding to (16c, d) can only be analyzed by the French child (given what the child has already established about the emerging grammar) if the V *lit* raises to I, and therefore they express the cue.[9] A simple sentence like *Jeanne lit les journaux*, "Jeanne reads the newspapers," on the other hand, could be analyzed with *lit* raised to I or with the I lowered into the VP in the English style; therefore it does not express the cue for V-to-I movement.

Early English grammars manifested the V-to-I operation, but later grammars do not; the operation was lost at some point. From the perspective adopted here, the operation ceased to be cued. The cue for V-to-I raising, $_{I}$[V], came to be expressed less in the PLD in the light of three developments in early Modern English.

First, the modal auxiliaries (*can*, *could*, *may*, *might*, *shall*, *should*, *will*, *would*, *must*), while once instances of verbs that could raise to I, were recategorized such that they came to be base-generated as instances of I; they were no longer verbs, so sentences with a modal auxiliary ceased to

include $_{I}$[V] and ceased to express the cue for V-to-I movement. Sentences with a modal auxiliary, *Kim must leave*, are very common in ordinary speech addressed to young children, and the recategorization meant that they no longer expressed the cue. Sentences of this form existed at all stages, of course, but they came to be analyzed very differently after the change in category membership. The evidence for the recategorization is the obsolescence of (19), which follows if the modal auxiliaries are no longer verbs but are generated in I and therefore can occur only one per clause (19a), without an aspectual affix (19b, c), and mutually exclusively with the infinitival marker *to*, which also occurs in I (19d).

19 (a) John shall can do it.
(b) John has could do it.
(c) Canning do it.
(d) I want to can do it.

This change has been discussed extensively in Lightfoot 1979, 1991; Kroch 1989; Roberts 1985, 1993b; and Warner 1983, 1993. Some of these studies have focused on the initiation of the change, others on its spread, and others on its completion. There is consensus that the change was complete by the early sixteenth century.

Second, as periphrastic *do* came to be used in negatives like *John did not leave* and interrogatives like *Did John leave?*, so there were still fewer instances of $_{I}$[V]. Before periphrastic *do* became available, sentences like *Visited you London last week?* (17a) and *The women visited all London last week* (17b) expressed the $_{I}$[V] cue. Periphrastic *do* began to occur with significant frequency at the beginning of the fifteenth century and steadily increased in frequency until it stabilized into its modern usage by the mid-seventeenth century. Ellegård (1953) shows that the sharpest increase came in the period 1475–1550. For every instance of *do*, there is no verb in I.

Third, in early grammars with the verb-second system all matrix clauses had a finite verb in C, as we just discussed. Therefore, many matrix clauses expressed the cue for V to I, $_{I}$[V] (on the assumption that V could move to C only by moving first to I). As these grammars were lost, and as finite verbs ceased to occur regularly in C, so the expression of the cue for overt V-to-I raising was reduced.

By quantifying the degree to which a cue is expressed, we can understand why English grammars lost the V-to-I operation, and why they lost it after the modal auxiliaries were reanalyzed as nonverbs, as the periphrastic *do* became increasingly common, and as the verb-second system

was lost. We can reconstruct a plausible history for the loss of V to I in English. What we are doing here is identifying when grammars changed and how the available triggering experiences, specifically those expressing the cue, seem to have shifted in critical ways prior to the grammatical change.[10]

Our conclusion in earlier work was that V-to-I movement was lost in the seventeenth century, much later than suggested by Kroch (1989), Roberts (1993b), and others. Warner (1997) now argues that the operation may have been lost as late as the eighteenth century. He offers some statistics from Ellegård (1953) and Tieken-Boon van Ostade (1987). Ellegård shows that interrogative inversion with nonauxiliary in positive clauses (i.e. *Came he to London?* as opposed to *Did he come to London?*) occurred 27 percent of the time for 1625–50; 26 percent for 1650–1700. Tieken-Boon van Ostade shows a drop to 13 percent in the eighteenth century. Forms like *Came he to London?* could be generated only by grammars with V to I, so those grammars are still well attested in the seventeenth century, persisting into the eighteenth. The figures show a gradual development at the level of the language, reflecting changes in the PLD and the decreasing use of grammars with V-to-I raising. Negative declaratives with a nonauxiliary (*He came not to London*, as opposed to *He did not come to London*), which are also generated only by V-to-I grammars, show a very different pattern of development. They occur 68 percent of the time in 1625–50, 54 percent in 1650–1700, dropping sharply to 20 percent in the eighteenth century.

These figures show emphatically that V-to-I grammars were still very much in use throughout the seventeenth century and into the eighteenth. The drop in the negative patterns is actually sharper than these figures suggest; Tieken-Boon van Ostade's figures for the later period include a high proportion of recurrent items (*know, doubt*, etc.) which tend to be distortingly conservative and which Ellegård omitted for that reason. A particularly interesting feature of these figures is the discrepancy between the interrogatives and the negatives, which lends some support to the hunch (above) that structures like those underlying (16d) and (17a) are a more effective expression of the cue $_{I}$[V] than structures like those of (16c) and (17b). We see that structures like (17b) were robust and widely attested in the texts of the late seventeenth century and then disappeared rapidly – the kind of bumpiness that the abstractness of the cues leads us to expect. In any case, we also see that V-to-I grammars were prevalent in the seventeenth century and survived into the eighteenth.[11]

The historical facts, then, suggest that lack of rich subject–verb agreement cannot be a sufficient condition for absence of V to I, but it may be a necessary condition. On this view the *possibility* of V to I not being triggered first arose in the history of English with the loss of rich verbal inflection; similarly in Danish and Swedish. That possibility never arose in Dutch, French, and German, where verbal inflections remained relatively rich. Despite this possibility, V to I continued to be triggered, and it occurred in grammars well after verbal inflection had been reduced to its present-day level. However, with the reanalysis of the modal auxiliaries, the increasing frequency of periphrastic *do*, and the loss of the verb-second system, the expression of $_{I}$[V] in English became less and less robust in the PLD. That is, there was no longer anything very robust in children's experience which had to be analyzed as $_{I}$[V], i.e. which *required* V to I, given that the morphological I-lowering operation was always available as the default. In particular, common, simple sentences like *Jill left* could be analyzed with I-lowering (16a); meanwhile, sentences like (17b) with post-verbal adverbs and quantifiers had to be analyzed with the V in I, but these instances of the cue were not robust enough to set the parameter, and they disappeared quickly, a by-product of the loss of V to I in another domino effect.[12]

This suggests that the expression of the cue dropped below some threshold, leading to the elimination of V-to-I movement. The next task is to quantify this generally, but we should recognize that the gradual reduction in the expression of $_{I}$[V] is not crucial. What is crucial is the point at which the phase transition took place, when the last straw was piled onto the camel's back. This can be demonstrated by building a population model, tracking the distribution of the $_{I}$[V] cues in the PLD, and identifying the point at which the parameter was reset and V to I ceased to be triggered (differing, of course, from one individual or one dialect area to another). This work remains to be done (see below), but one would expect to find correlations between the changing distribution of the cue and the change in grammars.

So children scan the environment for instances of $_{I}$[V]. This presupposes prior analysis, of course. Children may scan for this cue only after they have identified a class of verbs and when their grammars have a distinct inflection position, I.

Consequently, very young children may have difficulty in recognizing elements as such which are in fact instances of $_{I}$[V] in the adult language. This would explain why young children produce the optional infinitives noted by Ken Wexler (1994): children acquiring grammars with V-to-I

movement produce uninflected (infinitival) verbs in base-generated positions despite never hearing such forms from adults – another mystery for a pure input-matching learning model. Wexler cites some Dutch examples found by Mieke Weverink (20). At this "optional-infinitive" stage, children know that finite verbs may not occur (more technically, may not be "checked") in clause-final position in matrix clauses, because they do not produce clause-final inflected verbs: **Pappa schoenen wasst.* But they do not consistently know that finite verbs are instances of I.

20 (a) Pappa schoenen wassen.
Daddy shoes wash.
(b) Pappa nieuwe scooter kopen.
Daddy new scooter buy.
(c) Ik ook lezen.
I also read.
(d) Pappa kranten weg doen.
Daddy newspapers away do.

The difficulty in recognizing $_{I}$[V] also explains the tendency of young Dutch children (noted by Evers and van Kampen 1995) to produce *do* support forms (21).

21 (a) Ik doe ook praten.
"I do also talk."
(b) Dat doe ik spelen.
that do I play
"I am pretending that."

These forms also do not occur in adult speech and do not represent input matching; the adult language uses *do* support only in VP topicalization (22), and the adult equivalents of (21) are (23). That is, (23) is the kind of thing that Dutch children hear, not (21); but there is a stage at which many children are not able to generate the adult forms.

22 $_{VP}$[De roos treffen] doet hij zelden.
the mark hit does he seldom
"Hit the mark he seldom does."

23 (a) Ik $_{I}$[praat]$_{i}$ ook e_{i}
(b) Dat $_{I}$[speel]$_{i}$ ik e_{i}

This grammatical approach to diachrony explains changes at two levels. First, the cues postulated as part of UG which embody the points of parametric variation explain the unity of the changes, why superficially unrelated properties cluster in the way they do. Second, the cues permit an appropriately contingent account of why the change took place, why children at a certain point converged on a different grammar: the expression of the cues changed in such a way that a threshold was crossed and a new grammar was acquired. That is as far as this model goes, and it has nothing to say about why the distribution of cues should change. This may be explained by claims about language contact or socially defined speech fashions, but it is not a function of theories of grammar, acquisition, or change – except under one set of circumstances, where the new distribution of cues results from an earlier grammatical shift; in that circumstance, one has a "chain" of grammatical changes. One example would be the recategorization of the modal auxiliaries (above), which contributed to the loss of V to I.

Notice that this approach to change is independent of any particular grammatical model. Warner (1995) offers a persuasive analysis of parametric shift, using a lexicalist head-driven Phrase Structure Grammar model, quite different from the one assumed here. Interesting diachronic analyses have been offered for a wide range of phenomena, invoking very different grammatical claims (see Fontana 1993; van Kemenade 1987; Pearce 1990; Roberts 1993a, b, 1994; Sprouse and Vance 1998; Vance 1995; etc.).

The cue-based account we have offered here for the loss of verb-second phenomena and the loss of V to I works quite differently from the account offered by Clark and Roberts (1993) for the loss of verb-second and the account offered by Roberts (1998) for the loss of V-to-I movement. Using the learnability model of Clark (1992), Clark and Roberts account for the loss of verb-second in Middle French by having children track data which change somewhat. Given the introduction of XP subject V . . . forms, earlier grammars were no longer as successful in their fit with the input data and were therefore replaced by a new grammar which did not generate a verb-second language. The supposition here is that new forms were introduced for unspecified reasons, and that therefore the child tracking input data in accordance with Clark's genetic algorithm is forced to adopt a new grammar which fits the new data better. This amounts to saying that verb-second was lost because non-verb-second forms were introduced into the language. The nature of this explanation and its circularity are intimately linked to the input-matching nature of the learning model. The problem is that the model, at least in the way it is presently articulated, is crucially an input-matching device. Because input matching plays such an important role, it is difficult to see why there should be bumps and catastrophes, points

where children converge on grammars which generate sets of structures and sentences quite different from earlier grammars. One can imagine how the model might be revised, but that would involve reducing the role of input matching, which makes acquisition too heavily data-driven.

Roberts (1998) considers the loss of V to I in English and invokes an elegance condition from Clark and Roberts (1993), "which, *all other things being equal*, favors those parameter-settings which generate relatively simple representations over those which generate relatively more complex ones" (emphasis added). He takes simplicity to be in part a function of movement operations, and therefore movement operations are "marked." This element of UG, for Roberts, is an active, causal factor in the loss of V-to-I movement in English, biasing learners intrinsically against movement operations; to that extent, it is not merely a default option of the type that I invoked earlier (our idea that morphological lowering is a default option did not explain the loss of V-to-I movement). For Roberts, the UG bias explains why V to I is not acquired when *all other things are equal*. This raises two obvious questions: if UG has a bias against V-to-I operations, why and how did V-to-I operations ever develop in grammars? Second, when exactly are "all other things equal"? Presumably the answer to the latter question is that the UG bias is effective when the PLD do not demand a movement operation. But if the PLD do not demand a movement operation, then that is a sufficient explanation for the lack of movement, and the UG bias is unnecessary.[13] We shall return to the explanatory force of these UG biases in chapter 8.

Other historians have accounted for these changes in reverse, claiming that periphrastic *do* was introduced as a function of the loss of V-to-I raising. Since periphrastic *do* first occurred very early and spread most rapidly in the fifteenth–sixteenth century, while V-to-I grammars persisted into the eighteenth century, that story seems to be even more far-fetched than Odysseus's escape from the cave of Polyphemus.

6.4 Creolization and Signed Languages

Roberts (1998) goes on to illustrate the UG bias against V-to-I operations by turning to creoles. He argues that generally creoles have unmarked or default values of parameters, and that specifically they lack V-to-I movement because the movement represents a marked value (a strong feature), and this despite movement in the lexifier language.[14] It is to be expected, of course, that English-based creoles will lack V to I, since English lacks it. Also, it is not surprising that French-based creoles may lack it, because the

most robust evidence for V to I in French, namely V in C, is limited to contexts where the subject is a pronoun: *Lisez-vous ces journaux?*, but not **Lisent les ouvriers ces journaux*? If V to I is less common in creoles than one might expect, that might be a function of problems in recognizing the cue, $_{I}$[V], in the kind of restricted input that early speakers have. We noted at the end of the last section that children have some difficulty in recognizing these structures even in languages where they seem to be somewhat transparent.

On the matter of UG biases and creole evidence for them, consider Berbice Dutch, a better-worked-out example where the debate can be sharpened. A striking property of this Guyanese creole is that it has subject–verb–object order, while its lexifier languages, Dutch and Eastern Ijo, are both underlyingly object–verb and verb-second (as discussed in Lightfoot 1991: ch. 7). Roberts (1998) takes this to illustrate the fact that subject–verb–object order represents an unmarked option, and that creoles generally adopt unmarked parameter settings. A cue-based, degree-0 approach to learnability would tackle things differently.

Dutch and Ijo have underlying object–verb order, but verb–object order often occurs in matrix clauses because of a verb-movement operation that moves the verb to I and then to an initial C position, yielding verb-second order in the way we discussed earlier. Our theory of constituent structure demands that verbs be generated adjacent to their complements, either right-adjacent or left-adjacent. A degree-0 learner resolves the verb-order option on the basis of unembedded data which reveal the position of the verb (as argued in Lightfoot 1991). In Dutch these data are the position of separable particles (24a), negation elements (24b), certain adverbs (24c), and clause-union structures (24d), each of which mark the underlying position of the verb to the right of its complement.

24 (a) Jan belt de hoogleraar op.
"John calls the professor up."
(b) Jan bezoekt de hoogleraar niet.
John visits the professor not.
(c) Jan belt de hoogleraar soms/morgen op.
John calls the professor up sometimes/tomorrow.
(d) Jan moet de hoogleraar opbellen.
"John must call up the professor."

Furthermore, there are uninflected, infinitival constructions in colloquial Dutch, which manifest object–verb order directly in unembedded contexts (25).

25 (a) En ik maar fietsen repareren.
"I ended up repairing bicycles."
(b) Hand uitsteken
hand outstretch "signal"
(c) Jantje koekje hebben?
"Johnnie has a cookie?"
(d) Ik de vuilnisbak buiten zetten? Nooit.
I the garbage-can outside set? Never.
"Me put the garbage out? Never."

We can understand this in terms of cue-based acquisition: the cue for object–verb order is $_{VP}$[NP/DP V], where the V may be a trace. In each language, children set the verb-order parameter on the basis of evidence in unembedded domains; the evidence may be indirect and show that the *trace* of the moved verb is to the right of the direct object. The verb will necessarily be a trace for a degree-0 learner in a consistently verb-second language, because in simple, unembedded domains verbs are consistently moved to C and are pronounced in that position. The verb is often a trace in a language like Dutch, where verbs generally (but not always) move to C. So (24a) contains the structure *Jan belt$_i$ $_{VP}$[de hoogleraar op e$_i$]*, and the child knows that this is the structure by virtue of knowing that *opbellen* is a phrasal verb and that *belt* must therefore have originated in a position to the right of *op* and moved from there. Similarly, in (24b) the negative *niet* occurs to the right of the direct object and marks the position from which the verb has moved. In this way the child finds instances of the cue for object–verb order in unembedded domains in Dutch. If the evidence for the position of the verbal trace is obscured in some way, the PLD would fail to some extent to express the cue for object–verb order. Lightfoot 1991: 179 noted:

> In the case of Berbice Dutch, if the first speakers did not have robust evidence about the distribution of separable particles, or if negative elements were no longer retained in their D-structure position (marking the D-structure position of the verb), or if the verb-raising (clause union) operation was not triggered in the same way as in Dutch, then there would arise a situation comparable to that of late Old English: there would no longer be adequate data to trigger the object–verb setting. Negation, for example, works differently in Ijo and Dutch. In Dutch the negative element occurs to the right of an object NP, marking the position from which the verb moves, but in Ijo the negative particle "is adjoined directly to the verb in its proposition-negating role" (Smith et al. 1987) and moves with it, as in Old English:

26 Á nimi-ɣá.
I know not.

Ijo provided the negative for the creole, *kane*, which is a clitic attached to the verb, and because Ijo provided the basis for negation patterns, one of the Dutch indicators of the position of the verbal trace was obscured.

We lack good records for the early stages of Berbice Dutch; therefore it is hard to be more precise and to show exactly how the PLD failed to express sufficiently the cue for object–verb order. Lack of good records is a general problem for work on the early stages of creoles. However, the negation example is suggestive and shows that one indicator of underlying object–verb order may be nullified if the other language is dominant in the relevant aspect. Conversely, Dutch may have been dominant in an area of grammar that expressed the cue for the position of the verbal trace in Ijo. Of course, if children are not degree-0 learners, then the cue for object–verb order would be expressed robustly, because this is the standard order for embedded clauses in both Dutch and Ijo. In fact, early learners of Berbice Dutch were unaffected by this evidence, as expected if they were degree-0 learners, searching for the cue only in unembedded domains.[15]

So it is not difficult to see how a degree-0, cue-based learner might acquire a verb–object grammar when the lexifier grammars are object–verb and verb-second, and we do not need to say that there is a UG bias or that creoles always manifest unmarked parameter settings.

Creole children, just like all other children, scan their environment for cues. They interpret what they hear, impoverished though it may be, as expressing cues, and they converge on grammars accordingly. They are not subject to any bias built into UG of the type that Roberts has suggested. So new languages may emerge rapidly and fully formed, despite very impoverished input.

Work on creoles is limited by the sketchiness of the data available for the earliest stages, but the view that new languages emerge rapidly and fully formed despite impoverished input receives striking support from work on signed languages. The crucial fact here is that only about 10 percent of deaf children in the USA are born to deaf parents, who can provide early exposure to a conventional sign language like ASL (American Sign Language). This means that the vast majority of deaf children are exposed initially to fragmentary signed systems which have not been internalized well by their primary models. This is often some form of Manually Coded English (MCE), which maps English into a visual/gestural modality and is very different from ASL (which is not English-based). Goldin-Meadow and Mylander (1990) take these to be artificial systems and offer a useful review of work on how deaf children go beyond their models in such circumstances and "naturalize" the system, altering the code and inventing new forms which are more consistent with what one finds in natural languages. Goldin-Meadow and Mylander show that children exposed to people who use morphological markers irregularly and spasmodically nonetheless regularize the markers, using them consistently and "in a system of contrasts . . . akin to the system that characterizes the productive lexicon in ASL" (p. 341).

Newport (1998) extends these ideas by reporting work on a single child, Simon, showing how he comes to use morphology consistently and "deterministically," where his models used it inconsistently and "probabilistically." She notes that Simon does not create "an entirely new language from his own innate specifications," as the language bioprogram hypothesis of Bickerton (1984) would imply. "Rather, he appears to be following the predominant tendencies of his input, but sharpens them, extends them, and forces them to be internally consistent." Inconsistent input, then, presents no problem for young children, who simply generalize across the board. Newport reports that adult learners, on the other hand, are seriously impeded by inconsistent input and often perform even more inconsistently than their models.

Work by Supalla (1990) on MCE casts more light on this and on themes that I have been addressing in this chapter. MCE systems were invented by educators to teach English to deaf children. They rely on a lexicon borrowed heavily from ASL. However, while ASL morphology is generally "nonlinear," with simultaneous spatial devices serving as morphological markers, MCE morphology is generally "linear," using invented signs which reproduce the morphological structure of English; those signs precede or follow the root word. The English *take–took* alternation is an example of nonlinear morphology, and *walk–walked* is an instance of a linear alternation, where the verb stem precedes the tense marker. Supalla studied Signed Exact English (SEE2), the dominant version of MCE, where all bound morphemes are invented and based on English. For example, the SEE2 suffix representing English -ing involves the single hand-shape "I"; the suffix -S (for singular present tense or plural) is a static upright S hand-shape in the neutral signing space; the -ment, -tion, -ness, and -age suffixes are all syllabic, /M/, /S/, /N/, and /G/ respectively. Of the 49 English affixes that have an equivalent in SEE2, 44 consist of at least one syllable. They are strictly linear and, importantly, phonologically independent of the root.

Supalla cites several studies showing that SEE2 morphology is not attained well by children, who do not use many of the markers that they are exposed to and use other markers quite inconsistently and differently from their models. He focuses particularly on deaf children who are exposed only to SEE2 with no access to ASL, and he found that they restructure SEE2 morphology into a new system. The SEE2 "bound morphemes were rejected and replaced with devised forms. Moreover, in the devised forms,

the affixation type was predominantly non-linear in nature . . . not exactly like that of ASL, [but] formationally within the constraints of affixation in ASL" (Supalla 1990: 46). Unlike in Newport's study, children did not simply generalize markers which were used inconsistently in the input. Rather, there were particular problems with inflectional morphemes, and children invented a new system.

Supalla's approach to this was to postulate a Modality–Constraints Model, which limits signed languages to nonlinear morphology, whereas spoken languages tend to have linear morphology. However, this approach seems suspect. First, the correlation does not hold reliably: spoken languages often have nonlinear morphology (e.g. the *take–took* alternation of English above), and nonlinear morphology is comprehensive in Semitic and other languages; and Supalla (ibid. 20) points out that ASL has some linear morphology, e.g. agentive (analogous to the English *-er* suffix) and reduplicative markings. Second, the model fails to account for the fact that SEE2-type morphology does not exist even in spoken languages. What is striking about the inflectional morphemes of SEE2 is that they "are produced in terms of timing and formation as separate signs" (ibid. 52). Supalla shows that they are not subject to assimilation; they are phonologically independent, articulated distinctly, and even emphasized. In general, this kind of phonological independence is characteristic of free morphemes but not of inflectional, bound morphemes, and the system seems not to be learned by children.

Clearly this cannot be modeled by an input-matching learning device of the kind discussed earlier, because the input is not matched. It doesn't even come close. Furthermore, it is not enough simply to say that SEE2 morphology violates UG constraints, because that would not account for the way in which children devise new forms. Nor is it enough to appeal to some UG characterization of functional categories. More is needed from UG. The unlearnability of the SEE2 morphology suggests that children are cue-based learners, programmed to scan for clitic-like, unstressed, highly assimilable inflectional markers. That is what they find standardly in spoken languages and in natural signed languages like ASL. If the input fails to provide such markers, then appropriate markers are invented. Children seize appropriate kinds of elements which can be interpreted as inflectional markers. In signed languages there seems to be at least a strong statistical tendency to reinterpret linear elements in this fashion.[16] It would be interesting to see work that examines how this reinterpretation takes place and how new morphology is devised when children are exposed to unlearnable systems like SEE2. This would flesh out the general perspective of Goldin-Meadow and Mylander (1990) and of Newport (1998).

Deaf children are often exposed to artificial input, and we know a good deal about that input and about how it is reanalyzed by language learners. Therefore, the acquisition of signed languages under these circumstances offers a wonderful opportunity to understand more about abrupt language change, creolization, and cue-based acquisition. One particular case of great interest is the emergence of Nicaraguan Sign Language, as described by Kegl, Senghas, and Coppola (1998).

I submit that work on abrupt creolization, the acquisition of signed languages, and catastrophic historical change shows us that children do not necessarily converge on grammars which match input. This work invites us to think of children as cue-based learners, who scan the environment for certain elements of I-language in unembedded domains. These elements are not explicit in the input, but they are derived from the input, in the mental representations yielded as children understand and "parse" their input. So a cue-based learner acquires a verb-second grammar not by evaluating grammars against sets of sentences, but on exposure to structures commencing $_{\text{SpecCP}}$[XP]. This requires analyzing the XP as in the specifier of CP: that is, i.e. knowing that the XP needs to have no fixed grammatical or thematic function and knowing that it is followed by a finite verb. $_{\text{SpecCP}}$[XP] is the cue for a verb-second system, and the cue must be represented robustly in the mental representations resulting from parsing the PLD.

The name is new, but the cue-based approach to language acquisition is *implicitly* assumed in some earlier work, and it comports well with work on the visual system, which develops as organisms are exposed to very specific visual stimuli – horizontal lines, for example (Hubel 1978; Hubel and Wiesel 1962; Sperry 1968). Current theories of the immune system are similar: specific antigens amplify preexisting antibodies. In fact, this kind of thing is typical of selective learning quite generally (Piattelli-Palmarini 1986). The cue-based approach has been productive for phonologists concerned with the parameters for stress systems (Dresher 1998; Fikkert 1994, 1995), and a similar approach has been invoked independently for some syntactic problems by Fodor (1998).

Cue-based acquisition is a radical departure from much current work on learnability, which postulates various forms of input matching. It is striking that so much of this work has children dealing with elements of E-language (Chomsky 1986), often requiring that the system in effect perform elaborate calculations. For example, one of the best-known results of work on learnability, the Subset Principle of Berwick (1985), is usually construed as calculating subset relations among sets of E-language and choosing among grammars accordingly. Dresher and Kaye (1990) show

that the Subset Principle can be defined intensionally with respect to cues. The model advocated here plays down the centrality of E-language for a good account of acquisition, and postulates children seeking elements of I-language in the input and selecting grammars accordingly; the model makes no reference to elements of E-language or to the output of the grammar.

The cue-based approach assumes (with Lightfoot 1989) that there is a learning path, an order in which parameters are set. We have seen that a child cannot determine whether specifiers precede heads until some analytical vocabulary has been developed. Similarly, the child cannot determine whether the specifier of CP is filled (in a verb-second language) until he or she has identified phrasal categories, learned that initial categories do not have any fixed grammatical or thematic role and (therefore) are followed directly by a finite verb. All of this represents prior stages of acquisition. Representations are elaborated step by step in the course of acquisition, and the cues needed become increasingly abstract and grammar-internal. On this model the learning path is part of linguistic theory, a function of the way in which the parameters and their cues are stated.

Gibson and Wexler's TLA and Clark's genetic algorithms are learning algorithms quite distinct from the grammars assumed. However, the cue-based approach suggests that there is no relevant learning algorithm beyond the information provided specifically by UG.

Cue-based learning is "deterministic" in the sense of Berwick (1985): the learner may not backtrack or undo parameter settings that have already been set. This represents a strong and testable claim about how acquisition proceeds. Of course, there is no theory of cues yet: that is, no theory of what constitutes a possible cue. But then we also have no very substantive theory of parameters. These are topics for ongoing research.

On this view, one would expect there to be grammatical changes which are abrupt, and also that languages would differ from each other in bumpy ways. We may seek to quantify the degree to which cues are represented in the PLD, showing that abrupt, catastrophic change takes place when expression of those cues falls below some threshold of robustness.

If we devise productive models for historical change along these lines, relating changes in simple cues to large-scale grammatical shifts, our results will have consequences for the way in which we study language acquisition. In particular, we shall not be surprised that changes sometimes occur abruptly. With the development of computer corpora, Niyogi and Berwick's results, and an explicit cue-based theory of acquisition, we have all the ingredients for success in the historical domain, synthesizing work on language change, acquisition, and variation.

Notes

1 This chapter incorporates material from Lightfoot 1997a, which was a cue-based rendering of some similar material in 1997b (first published in 1994, in the University of Maryland Working Papers in Linguistics). I received helpful comments on that earlier paper from Stephen Crain, Michel DeGraff, Norbert Hornstein, Ceil Lucas, Jairo Nunes, and Juan Uriagereka.

2 Clark's Fitness Measure has a serious technical problem. There is no reason to suppose that a grammar with more parameters set correctly will be more successful in parsing/generating incoming data. Dresher (1998) illustrates this by considering the settings needed to generate the phonological stress system of Selkup, computing the relative score which the Fitness Measure would give them when applied to eight representative words. It isn't obvious what criterion the Fitness Measure should use, so he tried three different criteria: words correct, syllables correct, and main stress correct. Some results are shown in table 6.2.

Table 6.2 Fitness of grammars for Selkup stress

	Parameters correct		Words correct		Syllables correct		Main stress correct	
(a)	4/10	40%	2/8	25%	7/20	35%	3/8	37.5%
(b)	6/10	60%	1/8	12.5%	7/20	35%	5/8	62.5%
(c)	7/10	70%	4/8	50%	12/20	60%	4/8	50%
(d)	8/10	80%	5/8	62.5%	14/20	70%	5/8	62.5%
(e)	9/10	90%	5/8	62.5%	14/20	70%	5/8	62.5%
(f)	9/10	90%	3/8	37.5%	10/20	50%	3/8	37.5%

Candidates (e) and (f) are each correct in all but one (different) parameter, but they are very different in their apparent fitness. (e) scores high, but no higher than (d), which has fewer correct settings. Candidate (f), with only one parameter wrong, scores worse in every category than (c), which has three parameters wrong. And (a) does better than (b) in one category, despite having only four correct parameter settings. Dresher also points out that these results can be influenced in unpredictable ways by the chance occurrence of various types of words. As a result, there is no simple relationship between success and the number of parameters set correctly, which is a problem for Clark's Fitness Measure.

3 The UG condition that lexical material in the specifier of CP needs a lexically filled C entails that in constructions where a wh- phrase is moved to the front of a clause but the finite verb does not move to second position, the wh-phrase occurs not in the specifier of CP, but rather in some adjoined position or in the specifier of some other functional category: Brazilian Portuguese *O que a Maria comeu?*, "What did Maria eat?"; colloquial French *Où il est parti?*, "Where did he go?" Jairo Nunes (personal communication) points out

that Brazilian Portuguese also allows a fronted wh- element with an overt complementizer: *O que que a Maria comeu?* But the complementizer may not occur if there is no wh- fronting: **Que a Maria comeu?* This suggests that Brazilian Portuguese fronts wh- items by adjunction and by movement to the specifier of CP, the latter requiring the C position to be filled.

4 I do not adopt the details of Kroch and Taylor's analysis, for reasons discussed in Lightfoot 1997b, particularly in the Appendix.

5 (11d) (verb–object and I-final) probably does not exist as an actual option. Nothing in the texts seems to require it. This suggests that (10a, b) do not represent two independent parameters.

6 Van Kemenade reported 73 instances of subject-initial clauses, where the subject was followed by a finite verb, 32 instances of verb-third (i.e. XP-subject-Vf), 6 wh- questions, and 23 instances of Vf preceded only by *þa* or the negative particle *ne*. Of the 26 cases of XP-Vf-subject, 19 involved a full noun in subject position, and 7 a subject pronoun.

7 There is an extensive literature now on "exploded inflectional systems," involving agreement phrases, tense phrase, aspect phrase, etc.; but happily this is not relevant to our concerns here, and I ignore it.

8 Swedish is sometimes analyzed as lacking the V-to-I operation. So Vikner (1994) has verbs moving directly to C, because negatives precede finite verbs in embedded clauses: . . . *om Jan inte köpte boken*, "if John didn't buy the book." But this indicates that *inte* "not" and other such adverbs occur to the left of I and does not provide evidence against the application of V to I. Presence of verbs in C is strong evidence of movement through I, given almost any version of the proper government condition on traces. Also, Swedish allows VP fronting, and a dummy finite verb, analogous to English *do*, occupies the position of I (i).

(i) [läser boken] kanske Allan inte gör
reads the book maybe Allan not does

This provides direct evidence that the negative marker is left-adjacent to I even in matrix clauses.

In addition, Roberts (1998) invokes V-to-I movement for the Kronoby dialect of Swedish, citing work by Platzack and Holmberg (1989). Holmberg (1986) and Platzack (1986) argued that mainland Scandinavian has V-to-I raising, and that negatives and adverbs are adjoined to I′. Kyle Johnson (1997) argues something similar.

9 I ignore here the very plausible suggestion of Iatridou (1990) that infinitival counterparts to (16c) may not be direct evidence for movement of V across an intervening adverb, if French allows complex verbs of the form [V Adv].

10 Warner (1995) also adopts this methodology explicitly.

11 In fact, Kroch's own figures from Ellegård show several sentence types (positive intransitive questions, negative declaratives, and positive wh- object questions) with *do* less than 40 percent of the time at the very end of the sixteenth century, showing that V-to-I grammars were still very much in use.

12 Some readers balk at the notion that sentences like (17b) were too subtle and not robust enough to trigger a V-to-I operation. However, the fact of the matter is that these forms did not trigger V to I or anything else, because they dropped out of the language – compelling evidence, it seems to me. And they dropped out of the language at the same time as other putative reflexes of the V-to-I operation. This shows not only that they had no triggering effect, but also that they were incompatible with the grammatical operations that *were* being triggered – hence their disappearance. I know of no alternative account of this particular change.

Furthermore, it also seems reasonable to take the periphrastic *do* forms as robust enough to act as a trigger for grammatical development. They appear in interrogative and negative statements and imperatives. There are several statistical studies showing that *most* of the speech directed at young children consists of interrogatives and imperatives (see Newport et al. 1977).

13 Norbert Hornstein (personal communication) points out that an input-matching model of language acquisition might incorporate an analogue of our notion of a threshold for the expression of a cue by requiring the child to match the input to, say, 70 percent. In fact, Clark and Roberts's learner ignores triggers which fall below a frequency threshold. However, if the partial match is not keyed to specific parameters, this would predict random oscillation in grammars. By contrast, our thresholds are keyed to specific cues.

It is worth noting that our account would allow a single grammar to have both V-to-I movement and periphrastic *do*, while other accounts do not, notably those within Optimality Theory. Since V-to-I movement and periphrastic *do* coexisted for several hundred years in the history of English, it is implausible that these structures were incompatible and were manifested in separate, competing grammars; competition between grammars is generally more short-lived (see section 4.5).

14 Bickerton (1998: n. 13) notes that he "briefly adopted" the position that creoles have unmarked values for parameter settings, but "rapidly rejected" this viewpoint. Roberts claims that my objection to this position (Lightfoot 1991: 175–7) disappears, but he does not address the point made there: that Saramaccan allows wh- movement out of embedded clauses, and that possibility represents a marked parameter setting.

15 My account of Berbice Dutch is based largely on Smith et al. 1987, but Silvia Kouwenberg (1992) has followed this up with interesting results.

16 Supalla (1990: 50–1) hints at a possible explanation for this tendency to nonlinear morphology. He points to studies showing that individual ASL signs take about 50 percent longer to produce than English words, but comparable propositions take about the same amount of time. This is achieved by having signs with more morphemes per sign and nonlinear morphological structure. This could be explained if there is a particular rhythm to natural language, if language processing takes place naturally at a certain speed, and if a language with ASL-type signs and a linear, affixal morphology is just too slow.