The Variational Dynamics of Natural Language: Acquisition and Change

Charles D. Yang
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square, NE43-812
Cambridge, MA 02139
charles@ai.mit.edu

April 21, 1999

The Variational Dynamics of Natural Languages: Acquisition and Change

Abstract

This report is the *précis* of an ongoing project that develops a unified framework in which the dynamics of language acquisition and language change can be formalized and studied. The approach is inspired by Darwinian evolutionary thinking, and the formalism is modeled after the theory of population genetics. We view language acquisition as a population of human grammars, made available by UG, that compete in a selectionist process. Selection is made possible by the *variational* properties of individual grammars; specifically, their differential compatibility with the primary linguistic data in the environment.

We first present developmental facts that are problematic for both statistical learning approaches and traditional learnability models. We then propose a formal model of acquisition in the variational framework, and derive convergence results. The model receives empirical support from developmental evidence in child language, including the subject drop phenomenon and the late acquisition of V2, that a learner is best modeled as a population of grammars in competition and co-existence, rather than a single grammar supplemented with auxiliary assumptions. Finally, we extend the variational model of acquisition to language change, and formalize linguists' intuition of grammar competition in individual learners, who constitute an evolving linguistic population over time.

Contents

One hundred years without Darwin are enough.

H. J. Muller (1959) on the centennial
celebration of *Origin of Species*.

A central issue in linguistics and cognitive science is the problem of language acquisition: How does a human child come to acquire her language with such ease, yet without high computational power or favorable learning conditions? In terms of the familiar distinction between I-language and E-language (Chomsky 1986), language acquisition can be framed as the process in which the child forms an I-language knowledge on the basis of the primary linguistic data, the E-language expressions present in her linguistic environment.

Intimately related is the problem of language change. Language change can be observed in the distributional changes of linguistic expressions produced by generations of speakers over time. Such expressions are E-language tokens, but nevertheless are the reflection and product of the individual speakers' I-language knowledge. As Lightfoot (1979; 1991) pointed out, an adequate theory of language change must incorporate a theory of language acquisition. Thus the locus of language change is the individual speakers, whose I-language changes over time. The collective behavior of a population of learners constitute the primary linguistic data for the next generation, and therefore determines the dynamics of language change. This iterative process is put most succinctly by Battye & Roberts (1995):

(1)

$$\begin{tikzcd} \text{Parents' I-language} \arrow[r] & \text{Parents' E-language} \arrow[dl] \\ \text{Children's I-language} \arrow[r] & \text{Children's E-language} \end{tikzcd}$$

This paper attempts to develop a unified framework in which the dynamics of language acquisition and language change can be formulated and examined. We start with language acquisition.

1 Learnability and Development

As articulated by Noam Chomsky (1965, 1975), the language faculty of the child's mind/brain undergoes changes in the course of language acquisition, from an initial state S_0, the *Universal Grammar* (UG), to a relatively stable state S_L, that of a competent speaker of the language:

(2) $S_0 \rightarrow S_1 \rightarrow \ldots \rightarrow S_L$

A theory of language acquisition aims to identify and model the process in (2) with sufficient details and clarity. It is evident that any adequate theory must meet the following empirical conditions:

(3) a. **Learnability**:[1] such a theory must accurately characterize the asymptotic properties of the learner in a particular linguistic environment, under plausible assumptions about the learner's computational machinery, the nature of the input data, sample size, and so on.

b. **Developmental compatibility**: the learner modeled in such a theory must exhibit behaviors that are analogous to the actual course of language development (Pinker 1979).

The developmental compatibility condition has been ignored by virtually all formal models of language acquisition. One of the major themes of this paper is to bring developmental considerations to the foreground with the attention they deserve. I first present developmental facts that are problematic for both statistical learning (e.g. connectionism) and traditional learnability models, thus building the motivation for the variational framework developed in the rest of the paper.

1.1 Against statistical learning

An empiricist approach to language acquisition has (re)gained popularity in computational linguistics and cognitive science (e.g. Stolcke 1994; Charniak 1995; Chen 1995; de Marcken 1996; Bates & Elman 1996; Seidenberg 1997). The child is viewed as an inductive and "generalized" data processor such as a neural network, designed to derive structural regularities from the statistical distribution of patterns in the input data, without *prior* (innate) specific knowledge of natural language. This approach draws support from findings such as Saffran et al. (1996), where it is shown that human children are capable of associating word boundaries with conditional entropy computed from input data. Most concrete proposals of statistical learning employ expensive and specific computational procedures such as compression, Bayesian inferences, propagation of learning errors, and usually require a large corpus of (sometimes pre-processed) data. These properties immediately challenge the psychological plausibility of the statistical learning approach, which falls under the *learnability condition*. In the present discussion, however, we are not concerned with this but simply grant that someday, someone might devise a statistical learning scheme that is psychologically plausible and also succeeds in converging to the target language. We show that even if such a scheme were possible, it would still face serious challenges from the important but often ignored requirement of *developmental compatibility*.

[1]Before proceeding, a terminological note. As Noam Chomsky pointed out in many places, notions such as "learnability" or "convergence" carry some misleading connotations. For example, the learnability condition, in the standard sense of (Gold 1967), requires the learner to converge to the target language in the environment. This is not an empirically meaningful position. As noted at the beginning of this paper, language acquisition is simply the process in which the learner forms an internalized knowledge of the linguistic expressions present in her linguistic environment – there is no external target, hence no convergence. In fact, the final state attained may not be identified with any unique "grammar" such as "English" or "Chinese" in the informal, E-language sense. In section 6, we show that in non-uniform linguistic environments, learners crucially do *not* converge to a unique grammar but rather reach a stable equilibrium of multiple grammars in co-existence, and this is the underlying force in language change.

For historical reasons, I will continue to use terms such as convergence, when specifically referring to acquisition in a *uniform* linguistic environment. A *uniform* linguistic (e.g. monolingual) environment is one which supplies the child with sufficient evidence to unambiguously determine a unique I-language grammar that is thoroughly compatible with the primary linguistic data.

One of the most significant findings in child language research of the past decade is that different aspects of syntactic knowledge are learned at different rates. For example, consider the acquisition of finite verb placement in French, where inflected verbs precede negation (4a) and adverbs (4b):

(4) a. Jean (n')*aime* pas Marie
Jean (Prt) *loves* not Marie
"Jean does not love Marie."

b. Jean *voit* souvent Marie.
Jean *sees* often Marie.
"Jean often sees Marie."

This property of French is mastered as early as the 20th month, as evidenced by the extreme rarity of incorrect verb placement in child speech (Pierce 1992). Similar results have been found in many other verb-raising languages (see Phillips 1995 for a survey).

In contrast, some aspects of language are acquired relatively late. For example, the requirement of using a sentential subject is not mastered by English children until as late as the 36th month (Valian 1991), when English children stop producing a significant number of subjectless sentences.

When we examine the adult speech to children (transcribed in the CHILDES corpus; MacWhinney & Snow 1985), we find that an overwhelming majority of English input sentences contain an overt subject, whereas only 7-8% of all French input sentences contain an inflected verb followed by negation/adverb.[2] A statistical learning approach, one which builds knowledge purely on the basis of the distribution of the input data, predicts that English obligatory subject use should be learned (much) earlier than French verb placement[3] – exactly the opposite of the actual findings in child language.

Further evidence against statistical learning comes from the Root Infinitive (RI) stage (Weverink 1989, Wexler 1994, *inter alia*) in children acquiring certain languages. Children in the RI stage produce a large number of sentences where root verbs are not finite – ungrammatical in adult language and thus appear infrequently in the primary linguistic data. It is not clear how a statistical learner will "induce" such non-existent patterns from the training corpus. Section 4 presents further evidence from the acquisition of verb-second (V2) in Dutch grammar, where at an early stage, children use a large proportion (50%) of verb-initial (V1) sentences, a pattern that is restricted to Topic Drop and also quite sparse in adult speech. Again, an inductive learner purely driven by corpus data has no explanations for these disparities between child and adult languages.

Empirical evidence as such poses a serious problem for the statistical learning approach. Once the developmental compatibility condition is taken seriously, it seems a mistake to view language acquisition as an inductive procedure that constructs linguistic knowledge, directly and exclusively, from the distributions of input data.

[2] See footnote 19 for details of the counting procedure.

[3] It would be hard to justify a claim that one acquisition (English subject use) is "harder" than the other (French verb placement): in either case, the decision is binary. Nor is there an obvious reason why one acquisition but not the other should be affected by performance factors, as proposed by Pinker (1984), Bloom (1990, 1993), and Valian (1991) for subject drop in English children. In section 5, we will consider the subject drop phenomenon in detail.

1.2 The transformational approach

Another leading approach to language acquisition, largely in the tradition of generative linguistics, postulates that the human language faculty is constrained in some *a priori* space: there is a finite range of linguistic variations. The Principles and Parameters (P&P) framework (Chomsky 1981) expresses this idea, although different theories of modern generative grammar have different implementations (GB, HPSG, LFG, TAG, etc.). A constrained grammar space is on the one hand motivated by the fact that child language, although different from adult language, is still highly restrictive and does not vary along arbitary but inductively possible dimensions (Chomsky 1975; Crain & Nakayama 1987; *inter alia*). On the other hand, a constrained grammar space is motivated by the remarkable similarities found in comparative studies of the world's languages.

Virtually all previous formal models of language acquisition in the UG framework (Chomsky 1965; Wexler & Culicover 1980; Berwick 1985; Gibson & Wexler 1994; *inter alia*) are *transformational*,[4] borrowing a term from evolutionary biology (Lewontin 1983). In transformational models, the learner moves from one hypothesis/grammar to another as input sentences are processed, and at any time, the state of the learner is identified with a single hypothesis/grammar.

With the advent of the P&P framework, the *triggering* model, most clearly articulated in Gibson & Wexler (1994), has received considerable attention.[5] The learner in a triggering model moves from one parameter setting to another upon successive presentation of input sentences; hence a transformational model. In general, these models are concerned with the learnability condition, which requires the learner to converge to the target grammar under plausibly psychological assumptions. This is not the place to review the learnability properties of the triggering model; see, e.g. Dresher (1999) and the references in footnote 5. I will only highlight a conceptual problem with the triggering model pointed out by Valian (1990) and others. Valian (1990) questioned the robustness of the triggering mechanism, pointing out that the non-uniformity of linguistic evidence might mislead the learner. For example, English Wh-questions are formed by fronting the Wh word. In colloquial speech, adults occasionally do leave Wh words in-situ: *go where?*, *say what?* A triggering model of acquisition (Gibson & Wexler 1994, *inter alia*), or more generally, a transformational approach in which the learner switches from grammar to grammar on the basis of linguistic evidence, does not explain why children could ever unambiguously converge to the target parameter setting, rather than swinging back and forth between two settings like a pendulum (Valian 1990; Randall 1992).

Furthermore, the language development condition, which has largely been ignored in formal models of language acquisition, poses a more serious problem for the triggering model and the transformational approach in general. We now turn to some developmental facts.

1.3 An explanatory gap in transformational models

Since the state of the learner is identified with a particular hypothesis/grammar in transformational models, it is expected that (i) at any time, the child learner will produce expressions that are

[4] With the exception of Clark's (1992) model of acquisition using Genetic Algorithms (GA). It shares some features with the theory proposed here, but there are some important differences. See section 3.5 for details.

[5] See (Niyogi & Berwick 1996a, 1996b; Frank & Kapur 1996; Fodor 1998) for extensions and variations.

compatible with a *single* grammar, and (ii) the child will demonstrate sudden changes when the state of their language faculty switches from one grammar (or parameter setting) to another. Neither of these two predictions are empirically supported.

It is known that some aspects of grammar are acquired very early and are attested robustly: recall Roger Brown's well-known remark that in children learning English, word order errors are "triflingly few" (1973: 156). These aspects can indeed be described in terms of a single grammar (the target). However, many other aspects of child language are irregular, inconsistent, and can hardly be attributed to any single grammar. Consider the well-known dropped subject phenomenon in English children. English children start by dropping about 30% of all subjects, and this figure gradually declines over time to adult level at around age 3 (Valian 1991). It has been proposed that the pattern of subject drop in English children is due to a mis-set parameter – either the child's grammar is set to *pro*-drop, like Italian (Hyams 1986), or the child's grammar is set to topic-drop, like Chinese (Hyams 1991). However, comparative studies have demonstrated conclusively that the child's grammar cannot be identified either with Italian (Valian 1991) or with Chinese (Wang et al. 1992). This constitutes an example in which the child's grammar does not resemble any single adult grammar. Another such example will be discussed in section 4.2, where we shall see that in the acquisition of V2 grammars, Dutch children produce almost 50% of verb-initial sentences, in combination with V2 sentences – this peculiar pattern cannot be attributed to a single grammar either.

Moreover, a transformational theory entails that as the child shifts from grammar to grammar, child language undergoes discrete and abrupt changes. This prediction receives no empirical support. For example, there is no evidence for a sudden or sharp decline in the rate of dropped subjects in English children (Bloom 1990, 1993; Valian 1991; Wang et al. 1992). Longitudinal study of language development consistently shows patterns of smooth change in child language: the gradual increase in the use of V2 in German (Clahsen 1986) and Dutch (Wijnen 1999) children,[6] the gradual decline in the use of root non-finite verbs in a number of languages (Phillips 1995), to name a few examples. A transformational approach, however, leaves no possibility for the smooth patterns found in language development.

At least two theories have been put forth to complement the standard acquisition model, as to reconcile this tension between language learnability and language development. It is not possible here to give a comprehensive review;[7] I will only highlight some of their key features.

One influential line of thinking is the *Continuity Hypothesis* (Pinker 1984), according to which the child's acquisition device does not change over time and does not generate developmental patterns that are incompatible with in the adult language. The Continuity Hypothesis is attractive because it constitutes the null hypothesis for language acquisition research (Macnamara 1982; Pinker 1984): "it posits the least amount of change (i.e., none at all) in the mechanisms essential to language acquisition" (O'Grady 1997: 333). A common application of the Continuity Hypothesis is to assume that the child and adult languages differ not in linguistic competence but only in performance factors.[8] For example, Bloom (1990, 1993), Gerken (1991) and Valian (1991) have attributed

[6] See section 4.2 for a detailed discussion on the acquisition of V2 in Dutch children.

[7] See O'Grady (1997) for an excellent summary.

[8] That is, language acquisition is *literally instantaneous*, as linguists have idealized.

subject dropping to extragrammatical factors such as processing, pragmatics, prosody, and other performance constraints that are argued to affect children. Caution must be taken when explaining child language with performance factors; independent evidence must be provided so that the performance factors invoked are not an *ad hoc* patch for the facts it tries to explain. For empirical critique of the performance-based theories cited above, see Wexler & Hyams (1993); Roeper & Rohrbacher (1994)' Bromberg & Wexler (1995); Waller (1996); among many others. Another empirical problem with the Continuity Hypothesis, as Borer & Wexler (1987) point out, is that it does not explain why some aspects of grammar develop before others. (Recall the sharp time course contrast in the acquisition of French verb placement and that of English obligatory subject use.)

Another influential idea of language development is the *Maturational Theory* (Gleitman 1981; Borer & Wexler 1987; Felix 1987; Wexler 1994). It holds that various components of UG become operative in the child in specific, maturationally (and thus biologically) determined orders. A maturational analysis has been used to account for the late development of passives (Borer & Wexler 1987); see Pinker, Lebeaux, & Frost (1987), Demuth (1989), Fox & Grodzinksy (1997) for critical discussion. It has also been used (e.g. Wexler 1994) to explain the Root Infinitive (RI) stage, during which children acquiring a number of languages use both finite and non-finite root verbs. Wexler proposes that both finite and non-finite root verbs are grammatical to children, before the maturation of the UG knowledge that matrix sentences require finite verbs. At the same time, it is assumed that the major parameter values of the target grammar are correctly set very early on (Hyams 1996; Wexler 1996; Rizzi 1996; *inter alia*). A cluster of properties in child language, including the dropped subject phenomenon, has been attributed to the RI stage and the subsequent maturational escape out of it. Much like explanations that invoke performance factors, a maturational proposal must be accompanied by independent evidence so that it is a genuine explanation rather than a simple recapitulation of the empirical facts. This is not the place to discuss the maturational theory and the RI stage. Rather, in section 5, we will discuss new observations in the subject drop phenomenon and point out the idea of very early parameter setting cannot be correct.

Briefly summarizing, we have seen that traditional models of acquisition, the transformational approach, cannot readily accommodate the non-uniformity of child language and the gradualness of language acquisition. Additional assumptions about the child learner, such as the Continuity Hypothesis and the Maturational Theory, are required. Note that I am not suggesting that these assumptions are wrong; e.g. it might turn out that some aspects of grammatical knowledge *are* maturational.[9] I am only suggesting that they are powerful assumptions which we appeal to only when we cannot explain child language otherwise. This is obviously a case-by-case decision. As I will show later, at the evidence for a maturational (specifically, RI-based) explanation of subject drop is weak.

Having noted the problems with the standard models of acquisition, we turn to develop a theory that meets language learnability and language development conditions simultaneously, *without* auxiliary assumptions like those reviewed above. Our theory draws inspiration from Darwinian evolutionary biology.

[9]The framework developed here is fully compatible with such a discovery; see section 3.5.

2 The Variational Dynamics of Darwinian Evolution

Essential to Darwinian evolutionary theory is the concept of *variational* thinking. Although Darwin was not the first to propose a theory of evolution (transmutation), it was Darwin's genius to see the uniqueness of each individual biological organism, and thus the genuine variation in natural populations. This is what Ernst Mayr (1963, 1982, 1991) calls the *population thinking*: the differences among individuals in a population are "real", rather than deviations from some idealized archetype as in pre-Darwinian *essentialist* thinking. Such individual differences induce differences in the individuals' relative abilities to transmit their genes to the next generation, therefore allowing forces of evolution such as natural selection to operate.

The recognition of the genuine variation among individuals led Darwin to realize that the process of evolutionary change is fundamentally different from other known historical processes. His formulation of evolution was a complete epistemological break from the previous views. In an important paper, Richard Lewontin (1983) discusses this conceptual breakthrough:

> Before Darwin, theories of historical change were all *transformational*. That is, systems were seen as undergoing change in time because each element in the system underwent an individual transformation during its history. Lamarck's theory of evolution was transformational in regarding species as changing because each individual organism within the species underwent the same change. Through inner will and striving, an organism would change its nature, and that change in nature would be transmitted to its offspring. (1983: 65; italics original)

Darwin's conception of evolution is radically different and revolutionary, as Lewontin elaborates:

> In contrast, Darwin proposed a *variational* principle, that individual members of the ensemble differ from each other in some properties and that the system evolves by changes in the proportions of the different types. There is a sorting-out process in which some variant types persist while others disappear, so the nature of the ensemble as a whole changes without any successive changes in the individual members. (1983: 66; italics original)

It is difficult to overestimate the importance of the variational thinking in the study of evolution. Since Darwin, the study of evolution has been the study of a conversion process: the conversion in which individual variations, under evolutionary forces such as natural selection, genetic drift, developmental constraints, etc., are mapped into variations among populations in space and time. One of the greatest triumphs of modern biology lies in Fisher, Haldane, and Wright's population genetics theory, the unification of Mendelian genetics and Darwinian evolution. In population genetics theory, *discrete* Mendelian variation among individual organisms, which results in phenotypical properties, is translated into *continuous* distributions in a population of organisms, species, and the great diversity of life.[10]

[10] See Depew & Weber (1995) for a history of the development of population genetics in relation to the statistical revolution in mathematics and physics of the late 19th century. See Maynard Smith (1989), Hartl & Clark (1996) for good introductions to population genetics.

The variational thinking is therefore one of the most profound features of Darwinian evolution. But one might wonder why I have burdened the reader, most probably a linguist, with such a lengthy discussion. This is justified because the variational thinking teaches us an important lesson in making scientific observations and constructing explanatory theories. Lack of uniformity in data does not entail lack of regularity: variations in a collection of samples can be interpreted as a collection of individuals with distinct properties. In such cases, variations are real, expected, and should be taken seriously, rather than being swept under the rug as noise, performance, and other kinds of deviations from a single archetype, or, in the case of language acquisition, deviations from a single grammar.

3 A Variational Model of Language Acquisition

3.1 A population of grammars

To study a complex system under change, the fundamental problem is to identify an appropriate set of state variables that describes the system's properties. Principles that govern the dynamics of the system can then be formulated and examined. In the study of biological evolution, transformational and variational theories are sharply contrasted in the state variables they identify. A transformational theory views the state of evolution as properties of a collection of genotypes and evolutionary process is the direct change in the properties of the *genotypes*. A variational theory views the state of evolution as distribution of a collection of genotypes, each of which has different properties. Evolutionary process is thus the change of *distribution* over time.[11]

Let's extend Darwinian variational thinking to the study of language acquisition. As depicted in (2) and repeated in (5), language acquisition can most broadly be characterized as a sequence of states in which the learner moves from one state to another.

(5) $S_0 \rightarrow S_1 \rightarrow \ldots \rightarrow S_L$

Virtually all previous models of language acquisition are transformational: they identify the state of the learner as a *single* grammar/hypothesis, which switches from one to another over time.[12] As noted before, its problem lies in the explanatory gap between the learnability model and the actual patterns of language development.

I propose that the learner be modeled as a *population* of "grammars", the set of all principled language variations made available by the biological endowment of the human language faculty. (In section 3.5, I will review some related proposals.) Each grammar G_i is associated with a *weight* p_i,

[11]While a transformational theory assumes direct changes in genotypes, i.e. new genotypes are constantly generated, a variational theory assumes changes in genotypical distributions, i.e. no new genotypes are generated. The point of interest is obviously the evolutionary processes in a rather short time frame, where the effect of mutation (thus new genotypes) can in general be ignored.

[12]The statistical learning models are *transformational* as well: at any time, the learner is represented by a state of knowledge, be it a set of rules, or a collection of neural network weights. This state of knowledge directly undergo changes as more input data are processed.

$0 \leq p_i \leq 1$, and $\sum p_i = 1$. In a linguistic environment E, the weight $p_i(E, t)$ is a function of E and the time variable t, the time since the onset of language acquisition. We say that

Definition: Learning *converges* if

$$\forall \epsilon, 0 < \epsilon < 1, \forall G_i, \mid p_i(E, t+1) - p_i(E, t) \mid < \epsilon$$

.

That is, learning converges when the composition and distribution of the grammar population are stabilized. Particularly, in a monolingual environment E_T in which a target grammar T is used, we say that learning *converges to* T if $\lim_{t \to \infty} p_T(E_T, t) = 1$.

3.2 An acquisition algorithm

Write $E \to s$ to indicate that a sentence s is an utterance in the linguistic environment E. Write $s \in G$ if a grammar G can analyze s, which, in a narrow sense, is parsability (Wexler & Culicover 1980, Berwick 1985). Suppose that there are altogether N grammars in the population. For simplicity, write p_i for $p_i(E, t)$ at time t, and p'_i for $p_i(E, t+1)$ at time $t+1$. Learning takes place as follows:

The Algorithm L: Given an input sentence s, the child

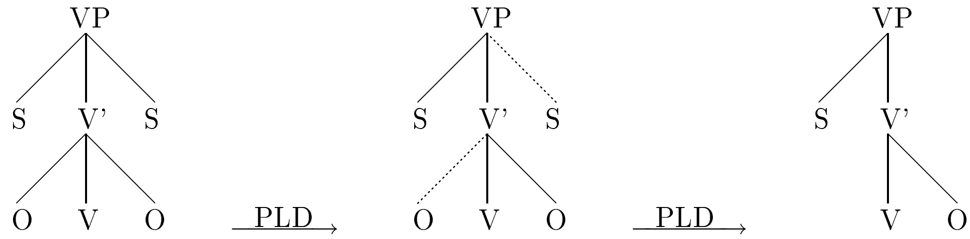
1. with the probability p_i, selects a grammar G_i
2. • if $s \in G_i$ then $\begin{cases} p'_i = p_i + \gamma(1 - p_i) \\ p'_j = (1 - \gamma)p_j & \text{if } j \neq i \end{cases}$

   • if $s \notin G_i$ then $\begin{cases} p'_i = (1 - \gamma)p_i \\ p'_j = \frac{\gamma}{N-1} + (1 - \gamma)p_j & \text{if } j \neq i \end{cases}$

Comment: The algorithm is the Linear reward-penalty (L_{R-P}) scheme (Bush & Mostellar 1958), one of the earliest and most extensively studied stochastic algorithms in the psychology of learning. It is real-time and on-line, and thus reflects the rather limited computational capacity of the child language learner, by avoiding sophisticated data processing and the need for a large memory to store previously seen examples. Many variants and generalizations of this scheme are studied in Atkinson et al (1965), and their thorough mathematical treatments can be found in Narendra & Thathachar (1989). See also Yang & Gutmann (1999) for a Hebbian learning algorithm and its analytic properties.

The algorithm operates in a selectionist manner: grammars that succeed in analyzing input sentences are rewarded, and those that fail are punished. As noted by Lightfoot (1991), there is neurological evidence (Hebb 1954; Hubel & Wiesel 1962; Changeux 1983; Edelman 1987; Deacon 1996; *inter alia*) that the development of neural substrate is guided by the exposure to specific stimulus in the environment in a Darwinian selectionist fashion. Note that the grammars in the population need not be represented explicitly. For example, UG might provide the child learner with a template of syntactic structures, for which all possible parametric variations are available. The learner uses this template to match input sentences, and gradually revising the structure as learning proceeds:

(6)



(6) illustrates how the word order in a language (SVO) is acquired. The learner initially has all the possible VP word order combinations available. Each combination, specified by two X-bar parameters, is probabilistically accessed by the learner to analyze incoming sentences in the primary linguistic data, and the weights of failed parameters are punished. Over time, incorrect parameter combinations (i.e. grammars) are weakened and eventually wiped out, with only the target grammar surviving. There is no need for all the four grammars (SVO, SOV, OVS, and VOS) to be explicitly represented; all the learner needs is a universal template and some parametric configurations. For convenience in deriving formal results, we have chosen to treat grammars as explicitly represented entities.

3.3 The asymptotic behavior of the learner

For simplicity but without loss of generality, assume that there are two grammars ($N = 2$), the target grammar T_1 and a *pretender* T_2. The results presented here generalize to the N-grammar case; see Narendra & Thathachar (1989).

For the two grammar case, the algorithm **L** in section 3.2 becomes **L'**:

The Algorithm L': Given an input sentence s, the child

1. with the probability p_i, selects a grammar T_i
2. - if $s \in T_i$ then $\begin{cases} p_i' = p_i + \gamma(1 - P_i) \\ p_j' = (1-\gamma)p_j & \text{if } j \neq i \end{cases}$
   - if $s \notin T_i$ then $\begin{cases} p_i' = (1-\gamma)p_i \\ p_j' = \gamma + (1-\gamma)p_j & \text{if } j \neq i \end{cases}$

Definition: The *penalty probability* of grammar T_i in a linguistic environment E is

$$c_i = \Pr(s \notin T_i \mid E \to s)$$

In other words, c_i represents the probability that the grammar T_i fails to analyze an incoming sentence s and gets punished as a result. Making an analogy to biological evolution, c_i denotes the *un*-fitness of grammar T_i. Notice that the penalty probability is an *intrinsic* property of a grammar, in a particular linguistic environment E, determined by the distributional patterns of linguistic expressions in E. It is not explicitly computed, as in Clark (1992); see section 3.5.

It is clear that in a monolingual environment for which T_1 is the target grammar, $C_{T1} = 0$. In a multi-lingual environment where the child is consistently exposed to expressions from more than one grammar, every grammar has a positive penalty probability. In section 6, we show that this is exactly the case for learners, who receive non-uniform linguistic input during the period of language change.

Our main result is as follows:

Theorem:

$$\lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1 + c_2} \quad \text{if} \quad \mid 1 - \gamma(c_1 + c_2) \mid < 1 \tag{1}$$

Proof: See Appendix A.

Comment 1: The convergence condition $\mid 1 - \gamma(c_1 + c_2) \mid < 1$ is met if the learning rate γ is fairly small. That is, the learner does not radically alter her knowledge of language, in contrast to triggering models of (Gibson & Wexler 1994) where the learner adopts or abandons an entire grammar on the basis of one input sentence. It is easy to see that $p_1 \to 1$ (and $p_2 \to 0$) when $c_1 = 0$ and $c_2 = u > 0$; that is, the learner converges to the target grammar T_1 in a monolingual environment.

A special case of the algorithm L (L') is the *error-driven* learner (Wexler & Culicover 1980; Berwick 1985; Gibson & Wexler 1994): if a selected grammar fails to analyze a sentence, it is punished; if it succeeds, nothing happens. It is easy to see that the target grammar, which is never punished (or rarely, see comments below on the robustness of our learner), will eventually eliminate all other competing grammars.

Learning is *robust*. Suppose that there is a small amount of noise in the input, i.e. sentences such as speaker errors which are not compatible with the target grammar. Then $c_1 > 0$. If $c_1 \ll c_2$, convergence to T_1 is still ensured by the Theorem [1]. Recall that in section 1.3, we discussed the pendulum effect raised by Valian (1990) and Randall (1992), a problem in the triggering framework. If there is a certain level of "noise", sentences that are not compatible with the target grammar (e.g. questions with in-situ Wh words, for languages like English that in general do front Wh words). This is not a problem in the variational model: as long as the "noise" level is quite low compared with the "signal", i.e. $c_1 \ll c_2$, convergence to the target grammar T_1 is still ensured by the theorem.

Comment 2: In the present model, one can directly measure the rate of *change* in the weight of the target grammar, and compare with developmental findings. Suppose T_1 is the target grammar, hence $c_1 = 0$. The expected increase of p_1, Δp_1 is computed as follows:

(7) A sentence s is presented to the learner, whose internalized weights for T_1 and T_2 are p_1 and p_2 respectively.

a. If, with probability p_1, T_1 is chosen,
$\Delta p_1 = \gamma(1 - p_1)$

b. If, with probability p_2, T_2 is chosen,

- with probability c_2, T_2 fails to analyze s,
  $\Delta p_1 = (1-\gamma)p_1$
- with probability $(1-c_2)$, T_2 analyses s,
  $\Delta p_1 = -\gamma p_1$

Thus,

$$E[\Delta p_1] \quad = \quad c_2 p_1 p_2 \tag{2}$$

Since $p_2 = 1 - p_1$, Δp_1 [2] is obviously a quadratic function of $p_1(t)$. Therefore, the growth of p_1 will produce the familiar S-shape curve. The target grammar will first go through a stage of slow growth, followed by a stage of rapid development, before gradually reaching the stable dominance (i.e. having wiped out all competing grammars). This is schematically shown below: There is

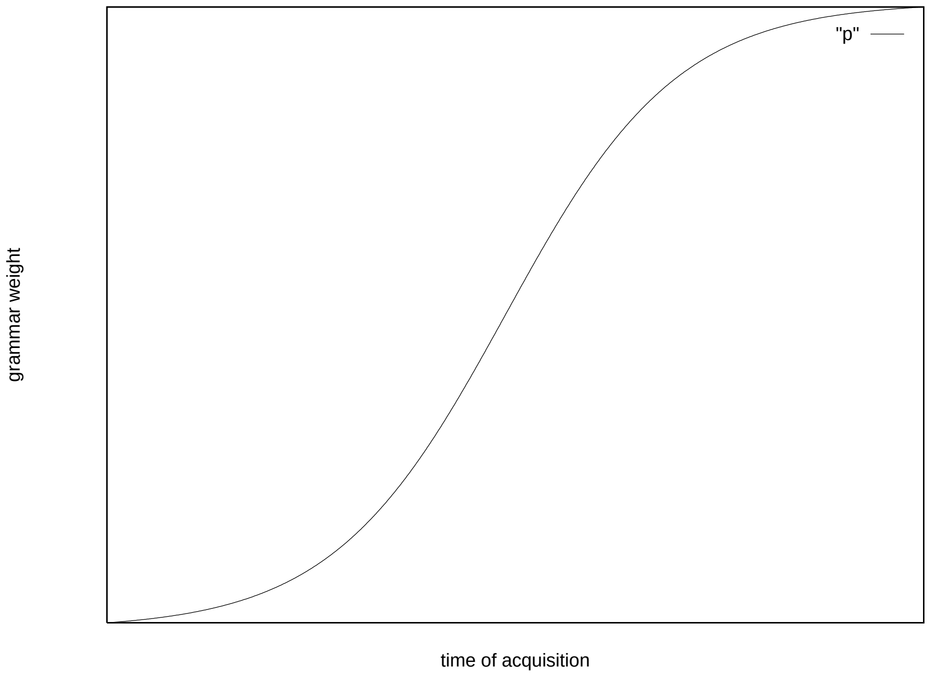


Figure 1: The development of the target grammar in a variational learner.

evidence for an S-shape pattern in child language development (Clahsen & Smolka 1986; Lightfoot 1991; Wijnen 1999; *inter alia*), which, if true, suggests that a variational learning algorithm might indeed be what the child learner employs. [13]

[13] It is imperative that the learner punish grammars that fail to analyze incoming sentences. A special case of the algorithm L (L') can be called a *tolerant* learner, in which if a grammar succeeds in analyzing a sentence, it is rewarded; otherwise nothing happens. A tolerant learner corresponds to the *Linear-Inaction* scheme (Atkinson et al. 1965), which has quite different stochastic properties, e.g. convergence is not always guaranteed.

3.4 (Un)fitness of grammar and the nature of input data

In the present model, the course and outcome of language development are determined by the relative (un)fitness of the grammars in the population, expressed in terms of their penalty probabilities. The theorem in section 3.3 states that as long as the penalty probability for the target grammar is 0 or close to 0, convergence is guaranteed. There are a number of ways to achieve this. We will discuss them in turn.

3.4.1 Unambiguous evidence

One way to ensure convergence is to assume the existence of *unambiguous evidence*, which has been formulated as *cues* in Dresher & Kaye (1990), and as *unambiguous triggers* by Fodor (1998) in the triggering framework of Gibson & Wexler (1994). It expresses the idea that input sentences can be grouped into two classes with respect to the target grammar T:

(8) a. $\nexists G, G \neq T,$ such that $s \in G.$

b. $\exists G, G \neq T,$ such that $s \in G.$

Sentences belonging to type (8a) are therefore unambiguous evidence for T, i.e. sentences that can only be analyzed by the target grammar. For example, suppose that English is the target grammar, which requires a sentential subject. Languages like Italian, Spanish, and Chinese, on the other hand, have the option of dropping the subject:

(9) a. (lei) ha parlato [*Italian*]
(he) has(3sgm) spoken

b. (ta) hui lai [*Chinese*]
(he) will come

c. he has spoken [*English*]

Therefore, sentences with an overt subject are not necessarily useful in distinguishing English from optional subject languages. However, there exists a certain type of English sentence that *is* indicative (Hyams 1986; Jaeggli & Safir 1989):

(10) a. There is a man in the room.

b. Are there toys on the floor?

The subject of these sentences is "there", a non-referential lexical item that is present for purely structural reasons – to satisfy the requirement in English that the pre-verbal subject position must be filled. Optional subject languages (Italian, Spanish, Chinese, etc.) do not have this requirement, and do not have expletive-subject sentences.[14] Thus, "there" sentences constitute the true unambiguous evidence for obligatory subject use.

[14] See, e.g. Jaeggli & Safir (1989) for extensive discussion of optional subjects in human language. Note also that by expletives I mean *overt* (phonologically realized) expletives, to be distinguished from "null" expletives that are sometimes posited in the literature.

3.4.2 Acquisition in a parametric space

Suppose that natural language grammars vary in a parametric space, as modern linguistic theories suggest. We can then study the dynamical behaviors of grammar *classes* that are defined in these parametric dimensions. Following Clark (1992) and Clark & Roberts (1993: p317), we say that a sentence s *expresses* a parameter α if a grammar must have set α to some definite value in order to assign a well-formed representation to s. Convergence to the target value of α can be ensured by the existence of evidence (s) defined in the sense of parameter expression. The convergence to a single grammar can then be viewed as the intersection of parametric grammar classes, converging in parallel to the target values of their respective parameters.

3.4.3 Unambiguous evidence is not necessary

It is an empirical problem whether unambiguous evidence exists for any given grammar. Unambiguous evidence is, however, not necessary for the proposed model to converge. It follows from the theorem [1] that even if no evidence can unambiguously identify the target grammar from its competitors, it is still possible to ensure convergence as long as all competing grammars fail on *some* proportion of input sentences; i.e. they all have positive penalty probabilities. To give a concrete example, consider the following four grammars: a SVO grammar (English), a VSO grammar (Irish), an OVS grammar (Hixkaryana: Derbyshire 1977, Derbyshire & Pullum 1986), and the target, a V2 grammar (German) in which the verb obligatorily appears in the second position of the matrix clause and the first position can be occupied by virtually any category. The four grammars admit the following word order variations for matrix clauses, where only patterns relevant to the present discussion are listed:

(11) a. German: SVO, OVS, XVSO
b. English: SVO, XSVO
c. Irish: VSO, XVSO
d. Hixkaryana: OVS, XOVS

We have used X to denote non-argument categories such as adverbs, adjuncts, etc., which can quite freely appear in sentence-initial positions. Note that none of the patterns in (11a) could conclusively distinguish German from the other three grammars. Thus, no unambiguous evidence appears to exist. However, if SVO, OVS, and XVSO patterns appear in the input data at positive frequencies, the other three grammars will have positive penalty probabilities. In other words, the German grammar has a higher overall "fitness value" than other grammars by the virtue of being compatible with *all* input sentences. As time goes by, German will eventually emerge to dominate the population. Note that crucially, the target grammar's rise to dominance is strictly realized in an on-line Markovian manner (algorithms L and L'), with *no* memory to keep track of, say, the frequency of particular types of input data, as in cue-based models.

The "unlearnable" language is of course the subset grammar (Berwick 1985). If a grammar δ has the penalty probability of 0, then its superset grammar Δ also has the penalty probability of

0 in the monolingual environment of E_δ, and they are therefore indistinguishable for the proposed algorithm (or, *any* algorithm, for that matter). It is however an empirical (and as far as I know, open) question whether true case of superset-subset relation exists in human language. In section 5, I show that the standard assumption that Italian/Chinese is a superset of English with respect to subject use is incorrect: there is abundance of evidence available to the child learner to distinguish English, Italian, and Chinese from each other, and in fact, they do.

3.5 Some related proposals

The idea that the child learner be described by multiple grammars has been independently suggested by a number of researchers. For example, in a critique of the parameter mis-setting theory of language acquisition, Valian (1990), citing Berwick & Weinberg (1984), suggested that both values of a parameter might be available to the learner at the same time. Weinberg (1990) suggested that parameter values might be associated with probabilities, which change over time (cf. Bloom 1993). These researchers did not develop this idea.

Roeper & Rohrbacher (1994) suggest that child learners form grammar hypotheses by projecting lexical heads or specifiers that have independent semantic or phonetic content, following a theory of the economy of projection (Speas 1994). They go on to propose that children acquiring a non-null-subject language (such as English) initially have access to a Japanese-type grammar, which allows subject omission, as well as an English-type grammar. This idea is similar to and can be formulated in the variational framework developed here. I have been assuming that the initial hypothesis space of the learner is the set of all possible (parametric) variations of human grammars. If the theory of economy of projection is correct and indeed applies to child language acquisition,[15] then the space of competing grammars can then be defined as such and all of our theoretical results carry over. Similarly, the variational model can also incorporate a maturational view of language development (if correct): the maturation of some aspect of grammatical knowledge will simply inject new grammars into the existing population, and the variational dynamics operates as before.

Kroch and his colleagues (Kroch 1989; Pintzuk 1991; Santorini 1992; Kroch & Taylor 1997) have shown that the linguistic knowledge of speakers during the course of language change should be explained as co-existing grammars, rather than a single grammar with peripheral assumptions. It will be shown in section 6 that the proposed model of acquisition straightforwardly extends to their view of language change.

The cue-based model (Dresher & Kaye 1990; Dresher 1999; Lightfoot 1997; also Fodor 1998, which can be viewed as a cue-based system, cf. Bertolo et al. 1996) consists of an ordered set of parameters and corresponding cues. Based on the cues, the learner keeps track of patterns in the input and set the parameter when a sufficient amount of evidence has been accumulated. The cue-based model shares an important feature with the theory proposed here, in that the learner does not radically alter her hypothesis space based on a single piece of data (in the triggering models). However, as pointed out earlier (section 3.2, comment 2), the cue-based learner essentially keeps a memory storage of previously seen input, an assumption that is psychologically dubious.

[15] Or, if there is some general principle that governs child's hypothesis formation in language acquisition, e.g. Roeper's (1996) merger theory; see also Wijnen (1999).

Furthermore, the cue-based model requires that cues (unambiguous evidence, in our definition) exist. As noted in section 3.4.3 for the acquisition of V2 grammar, unambiguous evidence may not in general be available; the present model does not require ambiguous evidence for convergence. It is further unclear how a cue-based model addresses the developmental compatibility condition.[16]

Clark's (1992) model using Genetic Algorithm (GA) is perhaps closest to the current model. Specifically, Clark assumes that grammars are represented by binary parametric vectors that undergo sexual reproduction via crossover, i.e. parts of a grammar vector are swapped/combined with others. A mutation process is also assumed, which, with some associated probability, flips bits in the grammar vector. Candidate grammars are evaluated against input data, hence some measure of "fitness" is defined, which then translates into differential reproduction. The two models share an important feature: the emergence of the target grammar is is driven by some notion of variation in competing hypotheses, i.e. their relative fitness as determined by the linguistic data in the environment.

However, some assumptions of the GA model are questionable. Firstly, it is not clear how the processes of crossover and mutation are empirically motivated. Secondly, as is common in GA applications, evaluation of individual fitness in Clark's model is computationally expensive (specifically, counting and comparing the number of UG violations of each grammar in the population) – what Dresher (1999) calls global evaluation – a psychologically questionable assumption. Thirdly, it is well-known in the optimization literature (e.g. Baluja & Caruana 1995) that GA does not guarantee convergence to the optimal solution, i.e. the target grammar may not be acquired.[17] Lastly and perhaps most importantly, in Clark's model, learning stops when the target grammar arises as a result of differential reproduction and mutation. That is, even just prior to the algorithm converging on the target grammar, the composition of the grammar population may not be close to the target grammar. Therefore, the problems with the transformational models remain in Clark's theory; for instance, the gradual (rather than abrupt) emergence of the target grammar is not accounted for. In the proposed theory, as it will become more clear in sections 4 and 5, the target grammar is in co-existence with other grammars from the very beginning. It gradually rises to dominance, hence capturing the smoothness of language development.

This said, Clark's model has an important insight of relating the fitness of individual grammars to the linguistic expressions present in the environment. The choice of Genetic Algorithm, on my view, is unfortunate.

[16]Other properties of the cue-based system are also suspect: for example, it is assumed that once a parameter value is set, all other parameter are reset to their default values (Dresher & Kaye 1990; Dresher 1999). Conceptually, the assumption of default parameter values and the idea of parameter value resetting must be justify, particularly in comparison with the variational model proposed here, for which neither of these assumptions is not needed. Empirically, it is not clear how to evaluate the mechanism of all-or-none parameter setting (and resetting): recall that child language shows no patterns of sudden restructuring of grammatical knowledge.

[17]Neither does natural selection (Lewontin & White 1960; Cohan 1984), from which GA draws inspiration, contrary to some vulgar claims in the popular press.

4 Some Longitudinal Predictions

We now start testing some empirical predictions made by the variational model proposed here. In this section, we compare the longitudinal development of verb raising in French, English obligatory subject use, and the acquisition of matrix V2 in Dutch. In section 5, we will consider the case of subject drop in detail.

The present model makes two predictions that cannot be made in the standard transformational theories of acquisition:

(12) a. As the target gradually rises to dominance, the child entertains a number of co-existing grammars. This will be reflected in distributional patterns of child language, under the null hypothesis that the grammatical knowledge (in our model, the population of grammars and their respective weights) used in production is that used in analyzing linguistic evidence. For grammatical phenomena that are acquired relatively late, child language consists of the output of more than one grammars.

b. Other things being equal, the rate of development is determined by the penalty probabilities of competing grammars relative to the input data in the linguistic environment (see section 3.3, comment 2).

To evaluate longitudinal predictions, we must estimate the the penalty probabilities of the competing grammars in a particular linguistic environment. We will measure the frequency of unambiguous evidence[18] from corpora of adult-to-child speech (CHILDES, MacWhinney & Snow 1985).

4.1 The early acquisition of the verb raising parameter

As shown in section 1.1, unambiguous evidence for French verb placement are sentences in (4) such as "Jean *voit* souvent Marie", where the inflected verb precedes negation or a VP-adjoined adverb. Unambiguous evidence for obligatory subject use in English is the *there*-type sentences (Hyams 1986; Jaeggli & Safir 1989; *inter alia*). Using naturalistic caretaker speech transcribed in CHILDES, we estimate that the frequencies of these triggers are 8% and 1%, respectively.[19]

As reviewed in section 1.1, we know that French verb raising is a very early acquisition, as evidenced by the extreme rarity of errors in tense/position correlation (Pierce 1992). Similar finding have been reported in child German (Poeppel & Wexler 1993), and Dutch (Krámer 1993) (among

[18]When it exists (but see section 3.4), because they are easy to measure and estimate.

[19] The corpus of a French child Gregoire from CHILDES is used. Unambiguous triggers are sentences where a finite verb proceeds the negation marker (*pas*) or an VP-adjoined adverb, as shown in (4). (Finite) auxiliaries, which also proceed negation and adverbs, are excluded. This is because in non V-raising languages such as English, auxiliaries nevertheless proceed negation, showing that auxiliary to T raising is an independent phenomenon. We counted the number of unambiguously V-raising sentences (193), and then divided it by the total number of sentences (2443) in the speech of the caretaker. I would like to thank Julie Legate for her assistance in this experiment.

The English *there* expletive sentences were retrieved with the UNIX command `grep` based on the mother speech of all three Harvard children. We found 562 *there* sentences like those in (10), of a total of 55670 sentences.

others languages; see Phillips 1995 for a review). If the current model is valid, we conclude that unambiguous evidence as 8% of all input data is sufficient for a very early acquisition: in this case, the target value of the verb-raising parameter is correctly set. The magic number of 8% will be used as a benchmark to compare developmental rates of other aspects of grammar. If this is true, we have an immediate explanation for Brown's (1973: 156) observation that word order errors in English children are "triflingly few" – virtually all English sentences display rigid word order, e.g. verb always (immediately) precedes object, which give a very high (perhaps close to 100%) rate of unambiguous evidence, sufficient to drive out other word orders very early on.

It is important to note that the early acquisition of French verb-raising certainly does not imply other aspects of French grammar are acquired early. All we have shown is that the property of verb raising is an early acquisition. Or, in sense of acquisition in a parameter space, we have shown that French children have identified their target language is a member of the V-raising parameter grammar *class*; other parameters might be set quite late.[20] As noted in section 3.4.2, a parametric grammar is successfully acquired only when all the relevant parameters are set to the correct values: grammars are learned in a parameter space as intersection of grammar classes.

English obligatory subject use, in comparison, is a late phenomenon: Valian (1991) showed that at around age 3, children start using subjects at adult-level frequency.[21] If the developmental rate is determined by the frequency of unambiguous evidence (more precisely, penalty probabilities of competing grammars), then we conclude that a 1% level of unambiguous evidence, like the expletive sentences in English, results in a relatively late acquisition. This will also be used as a benchmark when we consider the developmental rates of other grammars (like that of Dutch V2, in the section below).

4.2 The late acquisition of the V2 parameter

Dutch, like other West Germanic languages, has the finite verb in second position of the matrix clause. The first position can be occupied by virtually any phrase. Therefore, the following patterns are possible in Dutch: SVO, OVS, and XVSO, where X stands for adverbs, prepositional adjuncts, etc. As discussed in section 3.4.3, the SVO pattern is compatible with a SVO grammar,(e.g. English) the OVS pattern is compatible with a OVS grammar (e.g. Hixkaryana), and the XVSO pattern is compatible with a VSO grammar (e.g. Irish). It appears that no unambiguous evidence exists that *alone* will distinguish a true V2 grammar from other competing grammars. However, also discussed in section 3.4.3 is that only a true V2 grammar is compatible with *all* the patterns described above. In other words, only a true V2 grammar has the penalty probability of 0. The theorem [1] ensures that the V2 grammar will eventually win the grammar competition.

Observe that despite of its compatibility with *some* input patterns, an OVS grammar can not survive long in the population of competing grammars. This is because an OVS grammar has an extremely high penalty probability. Examination of CHILDES shows that OVS patterns consist of

[20] And indeed, obligatory use of subject is a late acquisition (Pierce 1992), much like in English children.

[21] In section 5, we will dispute the recent claim (Hyams 1996; Rizzi 1996; Wexler 1996) that, contra earlier proposals of parameter mis-setting (Hyams 1986, 1991), English children know their language requires subjects at least as early as the point they begin speaking.

only 1.3% of all input sentences to children.[22] Corpus analysis of adult Dutch conversations and written documents (Nieuwborg 1966; Jansen 1981) shows similar patterns: SVO sentences constitute more than 60% of all sentences (cf. Lightfoot 1997), and OVS sentences are consistently rare (less than 2%). Therefore, only SVO and VSO grammar, members of the [-V2] class, are "contenders" alongside the (target) V2 grammar, by the virtue of being compatible with significant portions of input data. But notice that OVS patterns penalize *both* SVO and VSO grammars, and are only compatible with the V2 grammar. Therefore, OVS patterns are *effectively* unambiguous evidence (among the contenders) for the V2 parameter, which eventually drive SVO and VSO grammars out of the population.

In the variational model, the rarity of OVS sentences predicts that the acquisition of the V2 parameter in Dutch is a relatively late phenomenon. Furthermore, because the frequency (1.3%) of Dutch OVS sentences is comparable to the frequency (1%) of English expletive sentences, we expect that Dutch V2 grammar is successfully acquired roughly at the same stage when English children have adult-level subject use (around age 3). Although I am not aware of any report on the timing of the correct setting of the Dutch V2 parameter, there is evidence in the acquisition of German, a similar language, that children are considered to have successfully acquired V2 by the 36-39th month (Clahsen & Smolka 1986).

The variational model also predicts that in the course of development, the target grammar is in co-existence with competing grammars for quite some time before gradually and eventually winning out. This is contrary to the claim that children set V2 parameter early on: for instance, Poeppel & Wexler (1993) noted that overwhelmingly, finite verbs appear in first or second position of the matrix sentence, and took this as evidence for the correct setting of V2. However, mere presence of finite verb in the second position does not establish the child has acquired V2 grammar – English sentences with SVO patterns also have finite verb in the "V2" position.[23]

Co-existence of competing grammars implies that that Dutch children will produce superficially "V2" patterns, which are actually a mixture of sentences that are generated with a true V2 grammar, a SVO grammar, or a VSO grammar with a modifier preceding the verb. Furthermore, a VSO grammar will also produce a (non-trivial) number of verb-initial (V1) sentences. This prediction is born out in a number of independent studies of Dutch acquisition. In a longitudinal studies of 5 Dutch children from age 2 to 3, Wijnen (1999) reports that 50% of sentences with finite verbs do not have a pre-verbal topic, and are therefore V1 sentences. In Table 1, we show the the longitudinal development of verbal syntax in a Dutch child, Hein, reported in Haegeman (1996). Table I is computed from her findings (1996: table 8 & 9) for finite matrix sentences that contain an overt subject:

At a younger age, Hein produced a significant amount (almost 50%) of V1 patterns, close to

[22] We examined a random sample of 5,000 sentences in a Dutch child Hein's mother's speech. Patterns that have an object in the initial position followed by a finite verb (auxiliary or lexical) are counted as OVS sentences. I would like to thank Edith Kaan for her help with this experiment.

[23] Poeppel & Wexler's data, along with Pierce (1992), Deprez & Pierce (1993), Krämer (1993), and see Phillips 1995 for a review, do demonstrate, convincingly on my view, the existence of functional categories in very early stages of language acquisition, contra Guilfoyle (1984), Lebeaux (1988), Radford (1990), among others, *and* that children have very good good knowledge of the tense/position relation in verbal syntax: almost all finite verbs are in first or second positions of the sentence, and almost all nonfinite verbs are in the final position.

age	V_{FIN} initial sentences	all V_{FIN} sentences	%
2;4	76	170	45%
2;5	66	132	50%
2;6	147	411	36%
2;7	93	201	46%
2;8	94	292	32%
2;9	98	269	36%
2;10	93	321	28%
2;11	36	259	14%
3;0	56	246	22%
3;1	101	268	35%

Table 1: Longitudinal V1 patterns in Hein.

the figure reported in Wijnen (1999). Haegeman (1996: p288) notes that "(w)hen the subject does not occupy the sentence-initial position, the position often remains empty", which she assumes is filled by a zero operator or a zero topic. Notice also that all the sentences in table I contain an overt subject (V1 sentences have post-verbal subjects), so the "zero topic" cannot be due to subject drop, which V2 language learners are known to do. While it is true that in adult Dutch, German, and other V2 languages, topic drop is occasionally attested in colloquial speech, there is however no reason to believe that as many as 50% of all topics are dropped by children. In a theory which assumes the V2 parameter is set correctly very early on (Poeppel & Wexler 1993; Wexler 1998), V1 patterns receive no explanation.[24] When V1 sentences constitute 50% of child language, appealing to "topic drop" while maintaining that the child has correctly set the V2 parameter is as good as saying the child uses V2 grammar *randomly*.

Also note that the very existence of V1 patterns poses another serious problem for statistical learning approach to language acquisition: since V1 sentences are in general ungrammatical in adult language and thus appear only sporadically in (true) topic drop sentences, how does statistical learning "induce" a statistically significant amount of patterns from a statistically insignificant amount of evidence?

Under the variational model, V1 patterns are expected, as the child learner has access to a V-initial grammar at the same time she entertains SVO and V2 grammars. Furthermore, as shown in table I, the proportion of V1 utterances gradually declines[25] as the child grows older, due to the accumulative effect of OVS patterns. Under the variational model, no additional assumption about the learner is need to explain the statistically significant V1 sentences in child language and its decline over time.

[24] See also Clahsen et al. (1994), where they argued that the evidence for early V2 setting is statistically weak.

[25] The relatively high proportion of V1 utterances at 3;1 is probably due to a sampling effect: the recording at all the other stages were carried out in 5 to 10 sessions, and that at 3;1 had only one session.

5 Explaining Subject Drop

We now turn our attention to the phenomenon of subject drop in English children. Before sketching out an analysis in the variational framework, let's briefly review the facts of child subject drop and some previous proposals.

5.1 Subject drop in children

Subject drop is one of best-known and most extensively studied cases in language acquisition. Children acquiring obligatory subject grammars (e.g. English) are known to produce a significant number of sentences without overt subjects (Bloom 1970; Braine 1973; Brown 1973; Hyams 1986; among numerous others), and their subject usage reaches adult level at around age 3 (Valian 1991). These properties of child language must be explained by any adequate theory of language acquisition.

In a classic study, Hyams (1986) proposed that English children initially mis-set the *pro*-drop parameter and therefore use an optional-subject grammar like Italian. The correct setting of the parameter is triggered, in the sense of Gibson & Wexler (1994), by unambiguous evidence: "there"-type expletive sentences that are not compatible with *pro*-drop grammars. As briefly mentioned in section 1.3, there are both empirical and conceptual problems with this position. On the empirical side, Valian (1991) reports that American children with the lowest MLU (hence highest rate of null subjects) drop subjects in 30% of all utterances whereas Italian children at the same MLU drop subjects in 70% of all utterances. Furthermore, over 70% of overt subjects in American children are pronouns, whereas only 20-35% of overt subjects in Italian children are pronouns. These findings suggest that dropped subject phenomena in children acquiring a *pro*-drop grammar and a non-*pro*-drop grammar are quite different matters.

Hyams (1991) also suggested that English children in subject drop stage use a Chinese-like grammar, a discourse-oriented topic-drop grammar in which null subjects are also allowed (Huang 1984, 1989). Since Chinese also allows null objects, which are licensed and identified by discourse (we will discuss the grammar of Chinese null pronominals in section 5.2), this proposal predicts that Chinese and English children should produce comparable rates of dropped object utterances. Wang et al. (1992) carried out an distributional analysis of Chinese and American child speech. Examining 2-3 year old children, they found that overall, Chinese children drop objects in 22.5% of all sentences, and American children only drop objects in 3.6% of all sentences. They also reported that Chinese children also drop more subjects (46.5%) than American children (25.9%).

Note that the subject drop figures in Valian (1991; 31%) and Wang et al. (1992; 25.9%) are considerably lower than others found in the literature. For example, Bloom (1990) and Hyams & Wexler (1993) both reported that the rate of null subject of Brown's Harvard children (Adam, Eve, Sarah) is around 50%. There are a number of reasons why the figures Valian and Wang et al. reported are more accurate. First, one must exclude imperatives and repetitions in child language. Second, as will be discussed in section 5.2, adult English allows subject omission in the sentence-initial position in certain contexts, and this null subject is licensed by a discourse topic (Haegeman 1997, forthcoming). Wang et al (1992: p233) took this fact into consideration: a sentence is counted as subject drop only if it cannot be used in adult speech in the same discourse context. Phillips

(1995: Appendix 2) used a similar criterion (slightly more restricted) and produced similar figures for the Harvard children (e.g. 25% of Adam's sentences are "true" subject drop). The statistical results that survive such stringent counting criteria are genuine and more conclusively show that English children cannot be using an Italian or a Chinese type grammar with respect to sentential subjects.

So what kind of grammars *do* English children use? As reviewed in section 1.3, two leading approaches have been taken to address children's apparent deviance from adult language. One approach (Bloom 1990, 1993; Valian 1991; Gerken 1993; *inter alia*), developing along the lines of the Continuity Hypothesis (Pinker 1984), assumes that the cognitive ability and grammatical knowledge are identical in children and adults, and the apparent difference (subject drop) between child and adult language is due to performance factors: pragmatics, prosody, processing (production) load, etc. This position has been shown to be empirically problematic; see Wexler & Hyams (1993), Roeper & Rohrbacher (1994), Bromberg & Wexler (1995), Waller (1996), among many others.

The other, competence-based approach assumes that there is *discontinuity* between child and adult grammars, which is usually supposed to be maturational in nature: absence of functional categories (Guilfoyle 1984; Lebeaux 1988; Radford 1990), truncated clause structures (Rizzi 1994), and maturation of INFL/Tense node (Wexler 1994) are some of the well-known proposals. Contrary to earlier ideas of parameter mis-setting (Hyams 1986, 1991), it is now standardly assumed that parameters are set correctly very early on (Hyams 1996; Rizzi 1996; Wexler 1998), thus essentially in agreement with the Continuity Hypothesis.

The claim of early parameter setting forces these proposals to attribute the null subject phenomenon to other aspects of the child's grammar. As noted above, even using very stringent counting criterion which specifically *excludes* topic drop sentences that would be acceptable in adult language, English children still drop a significant amount (about 30%) of subjects. Therefore, subject drop cannot simply be attributed to topic or diary drop in adult English (Hyams & Wexler 1993; Haegeman 1996), without offering an independent theory why children drop topics more frequently than adults and how this tendency disappears. Furthermore, assuming children drop subject like adults but simply more frequently does not explain the infrequent but not statistically insignificant amount of object drop (Bloom 1990; Valian 1991; Wang et al. 1992), which is impossible in adult English.

Another proposal of recent years links dropped subject to the Root Infinite (RI) stage. At the same time, it is assumed that the subject parameter is correctly set from early on (Hyams 1996; Rizzi 1996; Wexler 1996, 1998). According to Wexler (1994), non-finite root verbs are, for a variety of reasons, grammatical in children. Null subject has been linked to non-finite tense (Sano & Hyams 1994; Hyams 1996; Wexler 1998).

It is not my main concern here whether the RI-based theories are in general correct or not. Rather, I argue that it cannot be the whole story. I will show the idea of early parameter setting is untenable. In section 4, we have already seen that in face of statistically significant (50%) of V1 sentences, it is difficult to maintain that the V2 parameter is set correctly. In section 5.4, I will present two observations in child English subject drop that cannot be be captured by an RI-based approach.

5.2 A typology of null subjects

Before we attempt to give an explanation for child subject drop, we first review the nature of adult subject drop, including null subject (NS) languages such as Chinese and Italian. From now on, I will simply say Italian (Chinese) grammar when referring to an Italian (Chinese) *type* grammar with respect to subject use.

Perhaps all languages allow some degree of subject omission in informal contexts, even in English which is generally considered an obligatory subject language:

(13) a. Got milk?

b. Seems okay to me.

c. Have been working on this paper for months.

Haegeman (1990; 1997; forthcoming) argues extensively that English subject omission such as above is restricted to the initial NP position of the root clause. Following Rizzi (1994, 1996), Haegeman argues that the empty subject does not have a clause-internal antecedent, and therefore is identified in the discourse. English is standardly considered as an obligatory subject grammar as seen in its use of expletives in the grammatical subject position:

(14) a. There is a book on the shelf.

b. There seems to be a war in Europe.

c. It is raining outside.

True NS languages like Chinese and Italian do not have expletives (Jaeggli & Safir 1989). They also have far more extensive patterns of pronominal drop than the sentence-initial subject omission in English described above. We will examine them in turn.

Following Huang (1984, 1989), we distinguish two kinds of NS languages: the Italian type, which identifies/recovers the empty subject through morphological agreement (including number and gender) in Agr, and the Chinese type, which identifies the empty subject through a discourse topic operator. As shown in (15), both Italian and Chinese allow null subjects.

(15) a.

e	Ho	telefonato.
(I)	have(1sg)	telephoned.

b.

TOPIC$_i$	e_i	Kanjian-le	Lisi.
TOPIC$_i$	(he$_i$)	saw-ASP	Lisi.

Both Chinese and Italian allow subject drop in embedded clauses:

(16) a.

Gianni	dice	che	*e*	ho	telefonato.
Gianni	says	that		have(1sg)	telephoned.

"Gianni says that I have telephoned."

b.	TOPIC$_i$	Zhangsan	shuo	e_i	hui	la.
	TOPIC$_i$	Zhangsan	said	(he/she)	will	come.
	"Zhangsan said that he/she will come."					

Because of the difference in the licensing and identification of null arguments, Chinese and Italian have quite different distributional properties. In Italian, null objects are in general not allowed.[26] Chinese, on the other hand, does allow null objects freely:

(17)	TOPIC$_i$	Zhangsan	kanjian-le	e_i.
	TOPIC$_i$	Zhangsan	saw-ASP	him$_i$.
	Zhangsan saw him.			

The empty pronominal e in (17) is identified with the discourse topic.

However, if the topic position is occupied by a nominal phrase, Chinese no longer allows an empty subject. This can be seen in (18), part of a conversation about John:

(18)	Speaker A:	... John xihuan Mary.
		... John likes Mary.
	Speaker B:	* **Sue**, *e* xihuan!
		Sue, John likes!

Speaker B topicalizes Sue to emphasize that it is Sue that John likes, not Mary. Since Sue has become the prominent topic, the subject, which is intended to refer to the discourse topic (John), cannot be dropped.

When the overt topic phrase is *not* nominal, nothing prevents subject from being dropped. Again, consider the following example embedded in a conversation about John:

(19)	Speaker A:	John zai xuexiao da-le ren.
		John at school beat-ASP someone.
		"John beat people up at school".
	Speaker B:	**Zai gongyuan**, *e* da-le ren.
		In the park beat-ASP someone.
		"In the park, he beat people up."

When the prepositional (thus non-nominal) phrase *zai gongyuan*, "in the park", is topicalized, nothing prevents the subject from being dropped, which can be identified with the discourse topic, John.

Italian, on the other hand, identifies null subjects through Agr morphology, and does not have the restrictions on subject drop seen above in Chinese. Subjects can be dropped freely in nominal and non-nominal Wh questions,[27] as shown in (20):[28]

[26] At least not with non-arbitary reference (Rizzi 1986; Campos 1986; Raposo 1987; Huang 1989).

[27] Following Chomsky (1977) and many others, I assume that topicalization is Wh-movement, and has essentially identical properties.

[28] I thank Michela Ippolito for these examples.

(20) a.	Chi	*e*	ha	baciato?		
	Who		has(3sgm)	kissed?		
	"Who has he kissed?"					
b.	Chi	*e*	credi	che	*e*	ami?
	Who		think(2sg)	that		loves(3sgf)?
	"Who do you think she loves?"					
c.	Dove	hai	*e*	visto	Maria?	
	Where	have(2sg)		seen	Maria?	
	"Where have you seen Maria?"					

We summarize the key differences in distributional properties of Chinese, English, and Italian with respect to subject use:

English	expletives sentence-initial subject drop
Chinese	object drop no subject drop with nominal topic phrase
Italian	rich Agr morphology unrestricted subject drop

Table 2: Key differences in pronominal drop among Chinese, English, and Italian.

In sections 5.3 and 5.4, I will present evidence that Chinese, English, and Italian children are sensitive to such properties of their respective languages, which lead to quite different outcomes in the courses of acquisition.

5.3 The early acquisition of Chinese and Italian

Suppose that, as discussed in section 5.2, UG makes available three kinds of grammars with respect to subject use: the obligatory type of English, the discourse-based optional type of Chinese, and the Agr-based optional type of Italian. We now consider how Chinese, Italian, and English children come to know the properties of their respective languages. The first claim (this section) is that Chinese and Italian children acquire the knowledge of subject use very early on. The second claim (section 5.4) is that English children use a mixture of a Chinese (but not Italian) type, discourse-based optional subject grammar *and* an English type, obligatory subject grammar.

Table 2 gives a summary of the differences among Chinese, English, and Italian. It is easy to see that a child acquiring these languages have unambiguous evidence to distinguish their respective grammars from competing ones. Table 3 illustrates: C_X stands for children acquiring language X, G_Y stands for the grammar of Y, and the entry at (C_X, G_Y) stands for the unambiguous evidence (there might be more than those listed) that can eliminate G_Y from the population of grammars that C_X is entertaining.

	G_{Chinese}	G_{Italian}	G_{English}
C_{Chinese}	—	null objects	null objects
C_{Italian}	NS in questions	—	NS in questions
C_{English}	expletives	Agr morphology	—

Table 3: Unambiguous evidence for Chinese, English, and Italian

Consider first a Chinese child, and how she rules out English and Italian on the basis of unambiguous evidence. Recall that object drop (17) is not possible in English, which only allows sentence-initial subject drop, nor is it possible in Italian. Hence, whenever an object drop sentence is presented to the child, an English or Italian grammar fails to analyze it and will be punished as a result. A study by Wang et al. (1992) shows that Chinese adults use a fair amount of object drop sentences in speech to children (11.6%; computed from their Appendix B) as well as among themselves (18%; computed from their Appendix D). Recall that in section 4, we found that 8% of French adult utterances are of the pattern V_{FIN} Adv/Neg, and that is the unambiguous evidence for finite verb raising. If an 8% level of unambiguous evidence suffices for an very early acquisition of the target parameter value (Pierce 1992), we can conclude that from very early on, Chinese children know their language is not English or Italian with respect to subject use, based on input sentences that contain dropped objects.[29]

This therefore predicts that Chinese children, from very early on, know that their language is a discourse-based optional subject language. This prediction is borne out in Wang et al. (1992), where they found that children's use of subject and object drops is not only fairly stable for all age groups, but also comparable to that of adults, as summarized in Table IV (based on their Appendix C, D):[30]

Turning now to Italian children, let's see how they eliminate English and Chinese grammars. Note that in *declarative* sentences without topicalization, Chinese is essentially a superset of Italian in the range of pronominal drop: both allow productive subject drop, in matrix (15) and embedded clauses (16), and Chinese furthermore allows object drop (17). However, as shown in section 5.2, Chinese does not allow subject drop when an overt nominal phrase assumes the sentence-initial topic position (18). In an overt Wh-movement language such as Italian, this means that whenever an nominal Wh-word is fronted, a Chinese grammar (and of course an English grammar too) does not allow subject drop; an Italian grammar, on the other hand, *does* (20). Every subjectless question with a nominal Wh-phrase punishes the Chinese and English grammars.

[29] As shown in section 5.4, young children have very good morphological knowledge of their knowledge. If that is the case, Chinese children could rule out Italian grammar almost immediately on the basis of morphology; see section 5.4 for details. In any case, Italian cannot survive beyond the earliest stage of child's multiple word utterances, either on the basis of (the lack of) Agr morphology or on the basis of object drop discussed in the text.

[30] The ratio of object drop is computed as:

$$\frac{\#\text{ of object drop sentences with transitive verbs}}{\text{total }\#\text{ of sentences}}$$

The adult subject/object drop ratios are computed from adult-to-adult conversations, which is more accurate reflection of their grammar than when talking to children. See Wang et al. (1992: section 3.1.3) for details.

Age	Subject Drop (%)	Object Drop (%)
2	55.728	20.192
3	45.650	21.376
4	38.252	26.031
overall	46.543	22.533
adults	45.852	18

Table 4: Subject/object drop in Chinese children and adults.

It is known that in adult speech to children, 70% of all utterances have dropped subjects (Bates 1976; Caselli et al. 1995). We also know that Wh questions are one of most frequent constructions children are exposed to. A cursory look at the CHILDES corpus shows that there is an abundance of Wh questions with dropped subjects, far more than the magic number of 8% that is required for a very early setting of a parameter. We thus conclude that Italian children know from very early on that their language is Italian, not English or Chinese. This prediction is confirmed in Valian's findings (1991): at both of the two developmental stages (1;6 to 1;10, and 2;0 to 2;5) Italian children drop subjects in about 70% of sentences, identical to the figures in adult speech reported in Bates (1976).

5.4 Two grammars are better than one

Finally, we consider the acquisition of English. We employ the same methodology as in earlier sections, by considering how unambiguous evidence eliminates competing grammars – in this case, how it drives Chinese and Italian grammars out of an English child.

As alluded to in footnote 29, there is strong evidence that very young children have virtually perfect knowledge of the agreement morphology of their languages. For example, Phillips (1995), reviewing a number of cross-linguistic studies, observes that "in languages with overt agreement morphology, children almost always use the agreement morphemes appropriate to the argument being agreed with" (p3). For example, Guasti (1994) found that three young Italian children (of the age 1;8 - 2;7, 1;10 - 2;6, 2;2 - 2;7, respectively) correctly use agreement morphology in more than 95% of all contexts. Furthermore, almost all errors in morphology use are errors of *omission*, i.e. failing to use a correct agreement affix and using a null form instead. Errors of *substitution*, i.e. using incorrect agreement morphology, are exceedingly rare ($<$ 2%). Clahsen & Penke (1992) had similar findings in a young German child during the period of 1;7 to 2;8.

Children's near flawless knowledge of morphology can be used in the acquisition of syntax. Recall that Italian requires the verb to agree with subject (null or overt), and that English and Chinese do not. If an Italian grammar is chosen by an English child for grammatical analysis, it fails to check the Agr feature due to English's lack of Agr morphology, and is punished as a result. Given children's acute sensitivity to agreement morphology in their languages, almost every English sentence counts as unambiguous evidence against an Italian grammar. We can be sure that English children must have ruled out Italian as a possible grammar of subject use since very early on. Valian's comparative study (1991) showed just this: starting from a very early stage, English children show very different properties with respect to subject use compared to Italian children, who we know from section 5.3 are using adult-like Italian grammar with respect to subject use.

Chinese, on the other hand, cannot be ruled out immediately. Chinese, like English, is morphologically impoverished, so that morphology cannot tell them apart. All English sentences that contain overt *referential* subjects are compatible with Chinese, and the relatively small amount of English subject-initial drop sentences is compatible with Chinese as well. However, non-referential expletives such as *there* are unique to obligatory subject grammars (Hyams 1986; Jaeggli & Safir 1989; *inter alia*), and they indeed constitute unambiguous evidence against a Chinese grammar. As noted in section 4.1, such sentences are quite rare in adult speech to children (about 1%). Therefore, much like the late acquisition of V2 in Dutch children, we predict that a Chinese (but not Italian) grammar and an English grammar will co-exist for an extended period of time before English eventually takes over on the accumulative effect of expletive sentences.

This conclusion makes a number of predictions that cannot be made by previous analyses of subject drop. First, since a Chinese grammar is available to the child and is in co-existence with English grammar, we predict that English child will produce a statistically significant number of subject drop sentences, but at a rate *lower* than that of Chinese children, who come to know the nature of subject use in their language very early on (see section 5.3). Wang et al.'s (1992) result immediately confirms this: after ruling out subject-drop sentences that would be possible in adult language, English children at age 2-3 drop subjects in 25.885% of all sentences, whereas Chinese children drop subjects in 46.543% of all sentences, comparable to adult level.

A second and stronger prediction concerns the use of object drop. Since a Chinese grammar is available to an English child, we *expect* a non-trivial level of object drop when the Chinese grammar is probabilistically used by the children, without having to appeal to performance factors (Bloom 1993). Furthermore, if both genuine subject drop (i.e. after ruling out subject omission sentences that are possible in adult language) and object drop are attributed to the existence of Chinese grammar in the child, we expect that at the same age group, the *relative frequency* of subject and object drop should be more or less identical between English children and Chinese children (the latter, as shown earlier, use a Chinese-type grammar exclusively). This prediction is borne out in the comparative study of subject/object drop of Wang et al. (1992). Table V shows the subject and object drop frequencies in Chinese and American children of the same age group (2-3):

Language	Null Subject (NS) %	Null Object (NO) %	NO/NS %
Chinese	55.728	20.192	36.233
English	25.885	8.308	32.096

Table 5: Comparison of Chinese and American subject/object drop.

Phillips (1995), using slightly more stringent criteria than Wang et al. (1992), found that Adam, an English child (Brown 1973), drops 25% of subjects in all sentences in the subject drop stage. Taking this with the 8% of object drop rate reported in (Bloom 1990), we see that Adam's NS/OS ratio also patterns very closely with the children summarized in Table V.

A final prediction has to do with the phenomenon of subject drop in Wh-questions. Recall that as discussed in section 5.2, Chinese does not allow subject drop when a nominal phrase undergoes topicalization (21a), but does when the topic phrase is non-nominal (21b):

(21) a.
Speaker A: ... John xihuan Mary ...
... John likes Mary ...
Speaker B: * **Sue**, *e* xihuan!
Sue, John likes!

b.
Speaker A: John zai xuexiao da-le ren.
John at school beat-ASP someone.
"John beat people up at school".
Speaker B: **Zai gongyuan** *e* da-le ren.
In the park beat-ASP someone.
"In the park, he beat people up.".

English children produce a large number of Wh-questions with subject drop (Roeper & Rohrbacher 1994; Bromberg & Wexler 1995). However, they are predominately *adjunct* Wh-questions such as "where going", or "why laughing at me". Null subject nominal Wh questions is virtually non-existent. In fact, examining the Adam's recordings from 2;3 to 3;0, the period of subject drop, I found only 6 null subject sentences in 506 *what* questions (cf. Roeper & Rohrbacher 1994), and none in 106 *who* questions.[31] The complete list is given in (22):

(22)

What looking for?	Adam06
What doing?	Adam07
What getting?	Adam08
What looking for?	Adam09
What call it?	Adam15
What think?	Adam18

Also note that in Chinese, it is possible to drop subject when an *inanimate* nominal phrase is topicalized, so that the null subject could still be linked to the discourse topic:

(23)

Zheben shu,	*e*	zuotian	jiu	kanguo le.
this book,	(I)	yesterday	already	read ASP.

"This book, I already read yesterday."

Hence the utterances in (22) are actually compatible with a Chinese grammar, which makes the subject drop asymmetry in argument/adjunct Wh-questions even stronger.

Therefore, virtually all of Adam's subjectless Wh-questions reported in (104 in all, Roeper & Rohrbacher 1994; Bromberg & Wexler 1995 counted 120) are adjunct *where*, *why*, and *how* questions. This very strong asymmetry in argument/adjunct subject-less questions receives a straightforward explanation in the variational framework, which predicts that young English children use English and

[31] I have excluded the formulaic *what (is) dat* and *who (is) dat* questions, which have an overt subject in any case.

Chinese grammars at the same time. Since the sentence-initial position is occupied by a Wh phrase, null subject in questions cannot be generated by an English grammar, which only allows sentence-initial subject drop. It therefore must be generated with a Chinese grammar. Since Chinese does not allow null subject with a nominal (argument) topic phrase, but does allow null subject with non-nominal (adjunct) phrases, the subject drop asymmetry in argument/adjunct Wh-questions falls out.

This asymmetry is not predicted, as far as I know, by any other account of subject drop, e.g. the tense-based theory (Sano & Hyams 1994; Roeper & Rohrbacher 1994; Bromberg & Wexler 1995; Hyams 1996; *inter alia*) which does not make a distinction between argument and adjunct questions.

In sum, the relative frequency of null subject and null object, and the asymmetry in subject-less Wh argument/adjunct questions receive no explanation in a theory that assumes English children know their language is an obligatory subject grammar early on. They are readily accounted for in the variational model which suggests that English children are using an English-type and a Chinese-type grammar at the same time.

5.5 The dynamics of acquisition and evolution

Before moving on to the study of language change, we summarize some key features of the variational framework for language acquisition:

(24) a. The irregularity in child language and the gradualness of language development can not and should not be attributed to a single grammar supplemented with auxiliary assumptions.

b. Language acquisition is a selectionist process in which grammars form a population and compete to match linguistic expressions present in the environment.

c. The course and outcome of acquisition are determined by the relative compatibilities of the grammars with input data; such compatibilities, expressed in *penalty probabilities*, are quantifiable and empirically testable.

d. Certain patterns in child language, such as the acquisition of V2 and subject use discussed here, can be explained as co-existing grammars in competition.

The biologically endowed linguistic knowledge enables the learner to go beyond unanalyzed distributional properties of the input data. We argued in section 1.1 that it is a mistake to model language acquisition as directly learning the probabilistic distribution of the linguistic data. Rather, language acquisition is guided by particular input evidence that serves to disambiguate the target grammar from the competing grammars. The ability to use such triggers is based on the learner's linguistic knowledge. Future research will focus on more cross-linguistic examination of language development and the consequences of the variational model in the acquisition of phonology and morphology.

Lewontin (1983) points out that in a transformational theory, individual elements are the *subjects* of the evolutionary process: influenced by their *internal* forces, changes take place in individuals themselves, which result in evolution. In a variational theory, on the other hand, individual organisms are the *objects* of evolutionary process. Differences among individuals are manifested, often fatefully, as affected by *external* forces such as the environment. Similarly, in transformational models of language acquisition, it is the child that makes changes to her language faculty by switching from one grammar to another. In the variational theory, the learner is the object of the forces of acquisition: linguistic experience triggers, constrains, and shapes the process of language acquisition. In a specific linguistic environment, the target grammar unfolds, from one of the many undifferentiated blueprints, to the stable state of a mature speaker.

6 Toward a Variational Theory of Language Change

We now turn to the study of language change. Generative linguists, since Halle (1962), have long realized that the locus of language change is the individual learner. Language changes when a population of individual learners form an I-language that is different from that of the previous generation. The E-language expressions produced by the present generation constitute the linguistic environment for the generation following it. This iterative process is depicted in (1) (from Battye & Roberts 1995), and repeated below:

(25)
$$\begin{tikzcd} \text{Parents' I-language} \arrow[r] & \text{Parents' E-language} \arrow[dl] \\ \text{Children's I-language} \arrow[r] & \text{Children's E-language} \end{tikzcd}$$

As Lightfoot (1979, 1991) pointed out, genuine explanations of language change must incorporate a model of language acquisition. One must be explicit about how child learners, on the basis of linguistic experience, attain the knowledge of language that is different from their parent generations. A description of language change, even very good one (which, for example, unifies a number of grammatical changes to the change of a single parameter value), is not a complete theory of language change – one must explain *how* the parameter value changed.

What makes the I-language of the generation $n + 1$ different from that of the generation n? Consider two *prima facie* possibilities:

(26) a. generations n and $n + 1$ are exposed to primary linguistic data that are sufficiently different, and as a result, form different I-language knowledge.

b. generations n and $n + 1$ are exposed to identical primary linguistic data, and the generation $n + 1$ mis-learns.

(26b) amounts to saying that two learners with more or less identical linguistic experience attain different languages. This is untenable position, in my opinion. Language acquisition research shows that children are highly competent and robust learners; it seems improbable that given sufficiently similar experience, children will attain languages that are substantially different (e.g. a major parameter is set to different value). Even if some individual learners, small in numbers, do mis-learn, it is unlikely that they would have much impact on the composition and dynamics of the linguistic population.[32] Furthermore, while the idea of mis-learning aims to explain historical change in some languages, it has difficulty dealing with the relative stability in some other languages, say, the word order in German.

We therefore concentrate on (26a): because of the difference in linguistic experience, different generations of learners acquire different languages. The question then becomes, What makes the primary linguistic data of generation $n+1$ different from that of previous generation? There are many possibilities here. For example, migration of foreign speakers might introduce novel expressions that were previously unseen; social and cultural factors might also eschew the distributional patterns of linguistic expressions used in a population. The *causes* for such changes are of course interesting topics of research, but are not relevant for a theoretical study of language change. We are chiefly concerned with the predictive *consequences* of such changes: what happens to language learners when the primary linguistic data is altered, and how does it affect the composition of the linguistic population?

In this section, we first develop a mathematical model of language acquisition, based on our variational model of language acquisition. We then use the model to explain the loss of V2 in the history of French and English, and discuss their similarities and differences.

6.1 Uniformity, punctuation, and language change

Consider a *uniform* linguistic environment E, in which speakers in the generation $n+1$ have the same language as speakers in the generation n. A special case is a monolingual linguistic environment, in which the linguistic evidence available to the learner unambiguously identifies a single language, L_1, a particular instantiation of human grammar in the parametric space.[33] If learners do not mis-learn, as suggested above, then the linguistic population is stable and language does not change.

Now suppose that some novel linguistic expressions, by the means of migration or genuine innovation as suggested above, are introduced into E_{L1}, and become part of the primary linguistic data of the generation $n+1$. Such expressions, by definition, are not completely compatible with L_1. For example, if a German population migrates to England, the V2 grammar they use introduces expressions such as XVSO and OVS that are not compatible with the English grammar, although the latter can indeed analyze a certain portion of German sentences, namely, SVO patterns that are nevertheless generated by a V2 grammar. Or, if the speakers in a monolingual environment E_{L1}, for whatever reason, start making substantial use of a marked or stylistic construction. Note that the case of genuine innovation is *formally equivalent* to the case of language contact: from the learner's perspective, an increase in the use of a (marked) construction can be viewed as introducing a new grammar, L_2, for which the marked construction is grammatical. In both cases, primary linguistic data is no longer exclusively L_1, but a mixture of L_1 and L_2; the uniformity of primary linguistic

[32] We will formalize this later on.

[33] As will become clear later on, it is *possible*, at least theoretically, for two grammars to co-exist stably over an extended period of time.

data is punctured.

Now consider a non-uniform linguistic environment, which consists of expressions from two languages L_1 and L_2. Suppose a proportion α of L_1 expressions are incompatible with L_2, and a proportion β of L_2 expressions are incompatible with L_1. Call α (β) the *advantage* of L_1 (L_2). Figure 2 schematically illustrates.

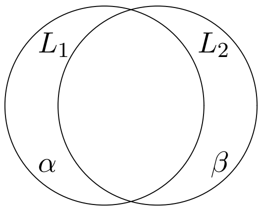


Figure 2: Two mutually incompatible grammars.

The variational model views language acquisition as competition and selection among I-language grammars, internally in the learner's mind. Recall that the fitness of individual grammars is defined in terms of their *penalty probabilities*:

Definition: The *penalty probability* of grammar T_i in a linguistic environment E is

$$c_i = \Pr(s \notin T_i \mid E \rightarrow s)$$

The penalty probabilities ultimately determine the outcome of language acquisition:

Theorem:

$$\lim_{t\rightarrow\infty} p_1(t) = \frac{c_2}{c_1 + c_2} \quad \text{if} \quad \mid 1 - \gamma(c_1 + c_2) \mid < 1$$

Suppose that at generation n, the linguistic environment $E_{L1,L2}$ consists of $pL_1 + qL_2$, where $p + q = 1$. That is, in $E_{L1,L2}$, proportion p of expressions are generated by L_1, and a proportion q of expressions are generated by L_2, and they collectively constitute the primary linguistic data to the learners in generation $n+1$. The penalty probabilities of L_1 and L_2, c_1 and c_2, are obviously βq and αp. The theorem above allows us to compute p' and q', the weights of L_1 and L_2 respectively, that are internalized in the learners of the generation $n+1$:

(27) The dynamics of a two grammar system:

$$\begin{aligned} p' &= \frac{\alpha p}{\alpha p + \beta q} \\ q' &= \frac{\beta q}{\alpha p + \beta q} \end{aligned}$$

(27) shows that an individual learner in generation $n+1$ might internalize a combination of two grammars L_1 and L_2 at a different set of weights than the parental generation n.[34] p'/q':

(28)

$$\begin{aligned}\frac{p'}{q'} &= \frac{\alpha p/(\alpha p+\beta q)}{\beta q/(\alpha p+\beta q)}\\ &= \frac{\alpha p}{\beta q}\end{aligned}$$

If L_2 is to overtake L_1, then $q' > q$, which implies that $p'/q' < p/q$. From (28), it is clear that

(29) L_2 takes over L_1 if $\beta > \alpha$: the *advantage* of L_2 is greater than that of L_1.

Recall that α and β are presumably constants, which characterize the distributional patterns in the use of the respective languages. For example, when one mixes a German grammar (L_2) and an English grammar (L_1), β, the advantage of L_2, is simply the proportion of German sentences, namely, VS patterns, that are not compatible with L_1, English. Conversely, the advantage of English is from the V>2 sentences, e.g. "he really likes to see a movie", that are incompatible with the German V2 grammars. Such constants are be estimated from corpus studies. (We will do so when we study the loss of V2 in French and English).

Plotting the $q(t)$, the weight of L_2, as a function of time t (see Figure 3), we obtain the familiar S-shape curve that has been commonly observed in language acquisition (Osgood & Sebeok 1954; Weinreich, Labov, & Herzog 1968; Bailey 1973; Kroch 1989): This patterns has been noted in historical texts, which showed a mixed patterns of the "old" language and the emerging "new" language.[35]

The present model bears some similarities with two formal models of language change, put forth by Niyogi & Berwick (1995) and Clark & Roberts (1993) respectively. The methodology employed here is exactly that of Niyogi & Berwick's: a formal model of individual learners is extended to a population of learners, and the dynamics of the population over time can then be studied analytically. The difference lies in the conception of language acquisition. Niyogi & Berwick assume a variant of Gibson & Wexler's (1994) triggering model, for which the learner, at any time, is identified with a single grammar. As a result, N&B assume that learners mis-learn, by converging to grammars other than the target, and that the patterns of grammar mixture in language change are literally interpreted as some speakers using, monolingually, L_1, and others, L_2. Aside from

[34] Suppose, in a uniform linguistic environment E_{L1} a small number, say, n, of learners *do* mis-converge to a non-target grammar L_2. Suppose the total number of speakers in the population is N. The effect of the mis-learners on the next generation can be quantified:

$$E_{L1,L2} = L_1\frac{N-n}{N} + L_2\frac{n}{N}$$

If $n << N$, then the linguistic environment for the next generation is virtually identical to a uniform environment without mis-learners. That is, making an analogy to evolution, the mis-learners suffer from the "genetic drift" effect.

[35] This gradual change is observable, of course, only if β does not thoroughly overwhelm α, in which case L_2 takes over the population rather dramatically.

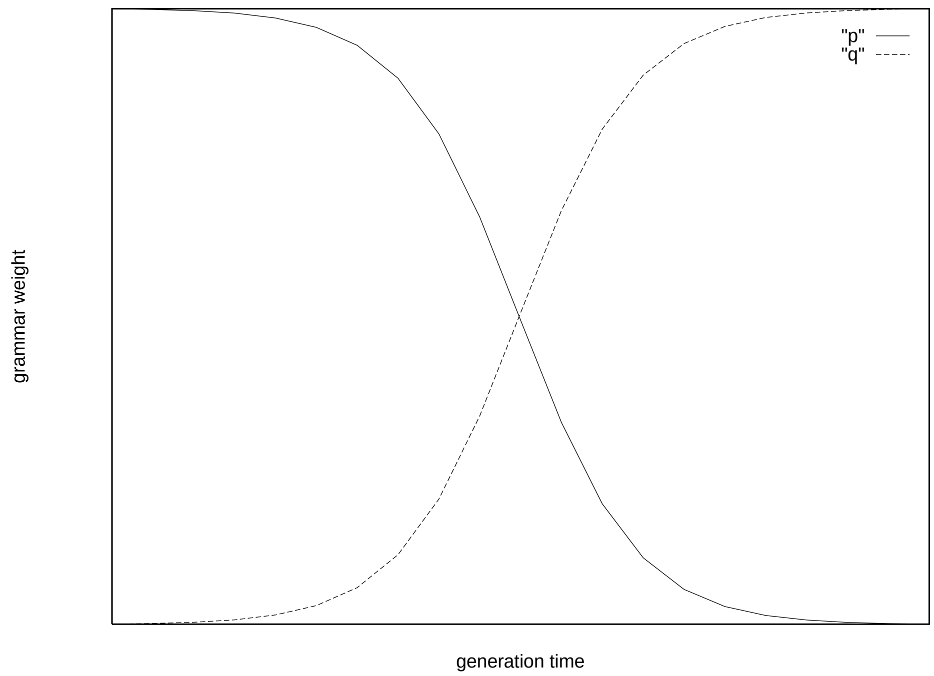


Figure 3: One language (q) replacing another (p) over time.

the conceptual and empirical problems of the triggering model pointed out in the acquisition part of this paper, there is strong evidence from Kroch and his colleagues' work (Kroch 1989; Pintzuk 1991; Santorini 1992; Kroch & Taylor 1997) that individual speakers in the period of language change should be viewed as using multiple grammars at the same time – this is a situation that the triggering model cannot explain.[36]

Clark & Roberts (1993) extend the Genetic Algorithm model of acquisition (Clark 1992) to the study of language change.[37] Their model and the present one share an important feature: the outcome of language acquisition is determined by the compatibilities of grammars with linguistic evidence, in a selectionist manner. However, they, like Niyogi & Berwick, also identify the final state of acquisition with a single grammar. Therefore, when the linguistic evidence does not unambiguously identify a single grammar, they posit some general constraints on the learner, e.g. the *elegance* condition, which requires the learner to select the simplest hypothesis of grammar. Aside from such auxiliary assumptions,[38] the position of learner converging to a single grammar is difficult to maintain in face of the empirical evidence of grammar co-existence in speakers during the time of language change.

Briefly summarizing the theoretical considerations in the section, we have extended the variational model of language acquisition to a population of learners and presented some analytical results concerning the dynamical system thus construed. We reject the notion of mis-learning as a

[36]In fact, historical linguists commonly use terms such as "erosion" to indicate the gradual appearance of a grammatical construction, or "optional applications" to indicate the apparent co-existence of multiple grammars used by speakers.

[37]See section 3.5 for a discussion of the GA acquisition model.

[38]Which may be correct but remain to be shown, and are in any case not needed in the present model.

viable driving force for language change. Rather, we suggest that non-uniformity in the primary linguistic data, however introduced, is the prerequisite for language change. One the uniformity is punctured, individual learners form internal representations of co-existing I-language grammars. The propagation of such grammars in successive generations of individual learners defines the dynamics of language change. We now apply the variational model of language change to the loss of V2 in the history of French and English.

6.2 Loss of V2 in French

Old French (OF) had a cluster of properties, including V2 and *pro*-drop, that are lost in Modern French (ModF). The following examples are taken from Clark & Roberts (1993):

(30) Loss of null subjects:

a.

* Ainsi	s'amusaient bien	cette	nuit.	(ModF)	
thus	(they) had fun	that	night		

b.

Si	firent pro	grant	joie	la	nuit.	(OF)
thus	(they) made	great	joy	the	night.	

(31) Loss of V2:

a.

*Puis	entendirent-ils	un	coup	de	tonnerre.	(ModF)
then	they heard	a	clap	of	thunder.	

b.

Lors	oïrent ils	venir	un	escoiz	de	tonoire.	(OF)
then	they heard	come	a	clap	of	thunder	

In this section, we will explain the loss of V2 in the variational model.

Recall that in order for a ModF SVO grammar to take over a V2 grammar, it is required that SVO has a greater *advantage*. That is, there are more sentences in the primary linguistic data that are incompatible with the V2 grammar than with the SVO grammar. (32) shows the advantage patterns of V2 over SVO, and vice versa:

(32) a. $s \in$ V2 but not SVO (i.e. VS):
XVSO, OVS

b. $s \in$ SVO but not V2 (i.e. V>2):
SXVO, XSVO

If the distribution patterns in modern V2 languages are indicative of those of ancient times, we can see that the V2 constraint is in general very resilient to erosion. In languages like German, the V2 constraint is very strongly manifested. V > 2 patterns, such as (33) (from Roberts 1992: 145) are restricted to a small number of adverbs, and are quite rare in distribution:

(33)

....	denn	Johann	hat	gestern	das	Buch	gelesen.
....	so	Johann	had	yesterday	the	book	read.

Statistical analysis of Dutch, German, Norwegian, and Swedish (cited in Lightfoot 1997) shows that about 70% of all sentences in V2 languages are SVO, and about 30% are VS patterns, which include XVSO and OVS. Therefore, even if an English-speaking population is merged with a German-speaking population, and thus a SVO grammar is explicitly introduced into the primary linguistic data, it is still very likely that the German V2 grammar will win over the population quite rapidly. In a rough analysis of the Penn Treebank, a corpus of modern English, I found less than 10% of all sentences having V>2 patterns such as (34):

(34) a. He always reads newspapers in the morning.
b. After dinner they played backgammon.

Therefore, the 10% advantage of SVO grammar, expressed in V>2 patterns, cannot throw off a V2 grammar, which has 30% of VS patterns to counter.

If the V2 constraint is so resilient, how on earth did Old French lose it? The reason, on my view, is that OF is also a null subject language.

Recall that the advantage of V2 grammar over SVO grammar is expressed in VS patterns. However, this advantage would be considerably diminished if the subject is dropped to yield [X V *pro*] patterns: an null subject SVO grammar can analyze such patterns as [X (*pro*) V]. (35) is taken from Roberts (1992: 155), which shows the prevalence of subject drop in early Middle French (MidFr):

(35)

	SV	VS	NS
Froissart, *Chroniques* (c. 1390)	40%	18%	42%
15 Joyes (14esme Joye) (c. 1400)	52.5%	5%	42.5%
Chartier *Quadrilogue* (1422)	51%	7%	42%

The 30% advantage in non-*pro*-drop V2 languages has been reduced to 5-18% in the *pro*-drop MidFr. As the same time, V>2 patterns have gone from fairly sparse (about <5%) in OF (Roberts 1992: 95) to 11-15% in early MidFr, as the class of sentence-initial XPs that do not trigger SV inversion was expanded (Vance 1989; Roberts 1992). (36), taken from Roberts (1992: 146, citing Vance (1989)), shows some representative examples:

(36) a. Lors la royne *fist* Santré appeller.
then the queen made Santré to-call.
"Then the queen had Saintré called."

b. Et a ce parolles le roy *demanda* quelz prieres ilz faisonient.
And at these words the king asked what requests they made.

c. Apres disner le chevalier me *dist* ...
after dinner the knight to-me said ...
"After dinner the knight said to me ...

(37) (Roberts 1992: 148), based the examination of the three texts in (35), shows the frequency of V>2 patterns in MidFr:

	Froissart, *Chroniques* (c. 1390)	12% (of 73)
(37)	*15 Joyes (14esme Joye)* (c. 1400)	15% (of 40)
	Chartier *Quadrilogue* (1422)	11% (of 45)

Comparing (37) with (35), we see that at early MidFr stage, there were more V>2 sentences than VS sentences, due to the effect of subject drop. Thus, an SVO grammar (plus *pro*-drop) has an advantage over an OF V2 grammar (plus *pro*-drop) – V2 in French was then destined to extinction.

Our analysis of the loss of V2 in French crucially relies on the fact that null subject was lost *after* V2 was lost. Roberts (1992) shows that this is the case. In late 15th century and early 16th century, when SVO orders had already become "favored", there was still significant use of null subjects. (38), taken from Roberts (1992: 155 and 199) shows this:

		SV	VS	NS
(38)	Anon., *Cent Nouvelles Nouvelles* (1466)	60.2%	10%	12%
	Anon., *Le Roman de Jehan de Paris* (1495)	60%	10%	30%
	Vigneulles, CNN (1505-15)	60%	11%	29%

Overall, the mean figures for the relevant patterns are shown below (Roberts 1992: 199):

		SV	VS	NS
(39)	15th century	48%	10%	42%
	16th century	77%	3%	15%

The decline, and eventually, disappearance, of VS patterns are the result of the SVO grammar winning over the V2 grammar. We see that in the 16th century, when V2 almost entirely evaporated, there was still considerable amount of subject drop. In fact, Roberts (1992) goes on to argue that the loss of null subject was partially due to the loss of V2. This pattern is consistent with our interpretation of V2 loss.

If French lost V2 because of its prevalent use of subject drop, which reduced the advantage of the V2 grammar, what happened to Old English, particular the northern dialect (Kroch & Taylor 1997), which was very much like a (stable) modern Germanic language but had no subject drop? We now turn to a tentative analysis of the loss of V2 in the history of English.

6.3 The erosion of V2 in Middle English

Old English (OE) was, generally speaking, a West Germanic language. Its peculiarities lie in the distribution of its V2 patterns, which are different from modern West Germanic languages such as Dutch and German (van Kemenade 1987; Pintzuk 1991; Kroch & Taylor 1997; Kroch et al. 1997; *inter alia*). In OE, as shown below,[39] when the subject is an NP, sentence-initial topicalized phrase triggers subject-verb inversion, hence the finite verb is in the second position:

(40)	a.	þæt	hus	hæfdon	Romane	to	ðæm	anum	tacne	geworht
		that	building	had	Romans	with	the	one	feature	constructed.

[39] All examples, unless specified otherwise, are taken from (Kroch, Taylor, & Ringe 1997).

b. þær wearþ se cyning Bagsecg ofslægen
there was the king Bagsecq slain

When the subject is pronoun, no inversion takes place, creating superficially V3 patterns:

(41) a. Ælc yfel he mæg don.
each evil he can do.

b. scortlice ic hæbbe nu gesæd ymb þa þrie dælas ...
briefly I have now spoken about the three parts.

c. ðfter his gebede he ahof þæt cild up ...
after his prayer he lifted the child up

The subject pronoun is typically analyzed as a clitic (van Kemenade 1987; Pintzuk 1991; *inter alia*).

There are also genuine V3 patterns, as a certain class of temporal adverbs and adjuncts do not trigger subject-verb inversion:[40]

(42) a. Her Oswald se eadiga arcebisceop forlet þis lif.
in-this-year Oswald the blessed archbishop forsook this life

b. On þisum geare Willelm cyng geaf Raulfe eorle
In this year William king gave (to) Ralph earl
Willelmes dohtor Osbearnes sunu
William's daughter Osborn's son.

In constructions (e.g. questions) where the verb raises to C,[41] the subject, be it pronoun or NP, is in the post-verbal positions:

(43) a. hwi sceole we oþres mannes niman?
why should we another man's take

b. þa ge-mette he sceaðan.
then met he robbers

c. ne mihton hi nænigne fultum æt him begitan.
not could they not-any help from him get

d. hæfdon hi hiora onfangen ær Hæsten to Beamfleote come.
had they them received before Hæsten to Benfleet came

[40]Although genuine V3 patterns are also possible in modern West Germanics, they are restricted to "it-then" sentences and left-dislocation, and their distributions are not as wide as in OE. See Kroch & Taylor (1997) and Kroch et al. (1997) for details

[41]Hence, Kroch & Taylor (1997) and Kroch et al. (1997) argue that OE should be analyzed as an IP-type V2 grammar with the verb in T or some position lower than C, like modern Yiddish (Santorini 1992). It suffices for our purpose to distinguish OE from CP-type grammars like modern German (and of course, from a SVO grammar which OE eventually became).

A recent study (Kroch & Taylor 1997; Kroch et al. 1997) shows that in the period of early Middle English (ME), there was considerable dialectical variation with respect to the V2 constraint. Specifically, the southern dialect essentially preserved the V2 of Old English: proposed XPs, with exception of a a certain class of adverbs and adjuncts noted earlier, generally trigger subject-verb inversion with full NP subjects but rarely with pronoun subjects. Table 6, taken from Kroch et al. (1997: table 1), illustrates:

	NP subjects			Pronoun subjects		
Preposed XP	# inverted	# uninverted	% inverted	# inverted	# uninverted	% inverted
NP complements	50	4	93	4	84	5
PP complements	12	4	75	0	11	0
Adj. complements	20	1	95	7	14	33
þa/then	37	2	95	26	10	72
now	12	1	92	8	22	27
PP adjuncts	56	19	75	2	99	2
adverbs	79	59	57	1	181	1

Table 6: V2 in southern early Middle English.

Given the distributional patterns in table 6, we have a plausible explanation for the loss of V2 in the southern dialect of early Middle English (essentially, Old English). Following van Kemenade (1987: Chapter 6) among others, we relate the eventual loss of V2 in English to loss of its morphological case system: clitics are case affixes and are therefore dependent on inflectional morphology. Both noun and verb morphologies became impoverished during the Middle English period (cf. Roberts 1997), which is strongly correlated to the loss of V2. Recall that in the southern dialect of early ME, a non-topicalized pronoun subject and verb are not inverted, resulting in superficially V3 patterns, as well as genuine V3 patterns with topicalized adverbs and adjuncts:

(44) TOPIC subject-pronoun V_{FIN} ...

With the impoverishment and eventual loss of the morphological case system, clitics are no longer possible. Therefore, patterns such as (44) can no longer be analyzed by an Old English type V2 grammar anymore, but *can* be analyzed by SVO grammar with the subject-pronoun treated a DP, as in modern English. Examining table 6, which was based on texts in the Penn-Helsinki Parsed Corpus of Middle English, we can see that in a total of 3064 matrix sentences, 511 (16.7%) are of the V>2 pattern `TOPIC subject-pronoun V`, and 314 (10.2%) are of the VS pattern `TOPIC V subject`. When subject pronoun could not be analyzed as clitics any more but only as NP, the SVO grammar would have had a greater advantage (16.7%) than the V2 grammar (10.2%), and eventually rose to dominance. The loss of morphological case system makes the loss of V2 possible, and this process is related to language acquisition in the variational model.

Notice that we also have an immediate account for the so-called "residual V2" in modern English, manifested in questions. Recall that in (43), we have seen that when V raises to C, both pronoun and NP subjects are in post-verbal position. In other words, the primary linguistic data *for questions* has been uniform with respective to a V2 grammar throughout the history of English. Therefore, the V2 character in question formation is preserved.

In contrast to the southern dialect, Kroch & Taylor (1997) showed that the northern dialect, under heavy Scandinavian influence, was very much like modern Germanic languages. The V2 constraint was uniformly and rigidly enforced, and there is no asymmetry between pronoun and NP subjects as seen in Old English and southern early Middle English. The following table, also from Kroch et al. (1997), illustrates:

	NP subjects			Pronoun subjects		
Preposed XP	# inverted	# uninverted	% inverted	# inverted	# uninverted	% inverted
NP complements	7	0	100	58	3	95
PP complements	18	0	100	10	0	100
Adj. complements	1	0	100	4	2	67
then	15	0	100	28	1	97
PP adjuncts	42	5	89	73	7	91
adverbs	25	1	96	51	5	91

Table 7: V2 (before language contact) in the Northern Prose Rule of Saint Benet.

As noted in section 6.2, the V2 constraint as exhibited in West Germanic languages is difficult to overthrow. This is due to the advantage a V2 grammar has over competing grammar, e.g. an SVO grammar: V2 grammar generates VS sentences which punish SVO grammar, SVO grammar generates V>2 sentences which punish V2 grammar, and VS sentences usually outnumber V>2 sentences. In discussing the loss of V2 in Old French, I argued that subject drop in Old French considerably diminished V2's advantage, to a point where an SVO grammar, aided by an increase in V>2 patterns, eventually won out. How did the northern early Middle English, a rigid V2 language *without* subject drop, evolve into an SVO language?

Kroch & Taylor (1997) and Kroch et al. (1997) show that the extensive contact between the northern and southern populations in the period of Middle English was essential to the eventual loss of V2 in English. They insightfully attributed the erosion of V2 to the competition of grammars in learners during language contact. This position and the variational framework developed here are highly compatible and mutually supportive. The northern V2 dialect, when mixed with the southern (essentially OE) language, constituted a non-uniform linguistic environment for later generations of learners, who, instead of converging to a single grammar, attained a mixture of co-existing grammars. Table 7, taken from Kroch et al. (1997) show the consequences of language contact in the northern dialect.

The effect of language contact is clear. In table 7, pronoun subject inverted with finite verb in 62% of all occurrences, compared to the prior-to-contact, much stronger V2 constraint which had pronoun subject verb inversion in 92% of all occurrences in table 6. This indicates that as a result of language contact and mixing, the V2 constraint in the Northern dialect was considerably weakened. When the V2 constraint is sufficiently weakened, and if the morphological case system of the mixed language got lost, then an SVO grammar would have gradually taken over, in the manner described earlier for the loss of V2 in OE.

	NP subjects			Pronoun subjects		
Preposed XP	# inverted	# uninverted	% inverted	# inverted	# uninverted	% inverted
NP complements	8	0	100	16	9	64
PP complements	21	3	88	48	21	70
Adj. complements	10	0	100	2	6	25
then	6	1	86	24	23	51
now	4	0	100	14	3	82
adverbs	20	5	80	35	26	57

Table 8: V2 (after language contact) in the Northern ms. (Thornton) of the Mirror of St. Edmund.

Notice that for the northern dialect, the initial contact with the southern dialect was crucial in the loss of V2, for otherwise V>2 patterns (44) would have never been introduced. That is, a West Germanic V2 language similar to the northern dialect, would not lose V2 even if its morphological case is lost; this might be the case of mainland Scandinavian languages such as Swedish and Danish. Once language contact was made, the uniformity of primary linguistic data was punctured, which resulted in co-existence of two distinct grammars internalized in the learners. The loss of morphological case system further altered the dynamics of the two grammars; the rest is history.

7 Conclusion

This preliminary paper has sketched out a model of language acquisition and discussed its extensions and consequences. We have been carrying the banner of Darwinian variational thinking throughout. Once the variation in the non-uniformity in child language is recognized as genuine, one is led to a different perspective on some traditional problems in language acquisition and language change. In the variational framework, we have attempted to construct a model of language acquisition that meets the conditions of language learnability and language development simultaneously. There seems to be strong empirical evidence in child language that the child learner during language development must be modeled as co-existing grammars instead of a single grammar supplemented with auxiliary assumptions.

Another aim of this paper is to bring some formal rigor to the empirical studies of natural language. Of course, we are not formalizing for the sake of being "formalized". However, I think that 40 years of generative research has taught us a sufficient amount about the basic structures of human language such that a formalization of certain aspects of linguistics study is indeed possible and perhaps necessary. Again, an analogy to the study of evolution is appropriate. Both biological organisms and human languages are constrained and indeed defined in a finite space of possibilities: for the former, the specification of genes, and for the latter, the Universal Grammar. Furthermore, I think that we understand UG far deeper than we understand genetics: for one thing, the structure of UG, expressed in Principles and Parameters, is perhaps far simpler than the structure of genetics; for another, the mapping from I-language knowledge to E-language expressions is far more direct than that from genotype to phenotype and linguistic data is far more amenable to direct examination. The dynamical properties of biological systems and linguistic systems are necessarily an interplay between the underlying composition and the environment (ecological or linguistic). In evolution, some of such dynamical properties are formulated in the theories of population genetics; it is my hope that the present paper contributes to the understanding the dynamics of natural language, and to situating the study of natural language in a broad biological framework.

Acknowledgments

I am most indebted to Julie Legate, who made important suggestions and comments in numerous aspects of this paper, and to Sam Gutmann, who insisted on getting the math right and who collaborated on an earlier paper. I am also grateful to the following people for helpful comments and discussion: Stefano Bertolo, Bob Berwick, Noam Chomsky, John Frampton, Ted Gibson, John Goldsmith, Dick Lewontin, Enrst Mayr, Partha Niyogi, Fritz Newmeyer, David Pesetsky, Bill Philip, Carson Schütze, Bill Turkel, Ken Wexler, and the audiences at MIT and Microsoft Research Institute. An NSF graduate fellowship provided the financial support. Special thanks go to Stephen Jay Gould for arranging a trip to the American Museum of Natural History, where this project was conceived.

Appendix A: The Convergence Proof

Theorem:

$$\lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1+c_2} \quad \text{if} \quad \mid 1-\gamma(c_1+c_2) \mid < 1 \tag{3}$$

Proof: From the algorithm in 2, it follows that $\{p(t)\}$ is a discrete-time Markov process on the state space which is the unit simplex $S_2 : \{p : p_1 + p_2 = 1, p_1, p_2 > 0\}$ with the stationary state transition function. Given the learner L' in an environment with penalty probabilities of $\{c_1, c_2\}$, let's consider the $p_1(t+1)$ given $p_1(t)$:

$$\begin{aligned} E[p_1(t+1) \mid p_1(t)] &= [p_1(t) + \gamma(1-p_1(t))][p_1(t)(1-c_1) + (1-p_1(t))c_2] \\ &\quad +(1-\gamma)p_1(t)[p_1(t)c_1 + (1-p_1(t))(1-c_2)] \\ &= [1-\gamma(c_1+c_2)]p_1(n) + \gamma c_2 \end{aligned}$$

Taking expectations on both sides, we have

$$E[p_1(t+1) = [1-\gamma(c_1+c_2)]E[p_1(t)] + \gamma c_2$$

which is a linear difference equation in terms of $E[p_1(t)]$. Solving it gives

$$E[p_1(t)] = [1-\gamma(c_1+c_2)]^n p_1(0) + \frac{1-[1-\gamma(c_1+c_2)]^n}{\gamma(c_1+c_2)}\gamma c_2$$

or

$$\lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1+c_2} \quad \text{if} \quad \mid 1-\gamma(c_1+c_2) \mid < 1$$

References

Atkinson, R., G. Bower, and E. Crothers. (1965). *An Introduction to Mathematical Learning Theory*. New York: Wiley.

Baluja, S. and R. Caruana (1995). Removing the genetics from the Standard Genetic Algorithm. Technical Report, Department of Computer Science, Carnegie Mellon University.

Bailey, C.-J. (1973). *Variation and linguistic theory*. Washington, DC: Center for Applied Linguistics.

Bates, E. (1976). *Language and context : the acquisition of pragmatics*. New York: Academic Press.

Bates, E. and J. Elman. (1996). Learning rediscovered: A perspective on Saffran, Aslin, and Newport. *Science* 274: 5294.

Battye, A. and I. Roberts. (1995). Introduction. In Battye, A. and I. Roberts (eds.). *Clause structure and language change*. New York: Oxford University Press.

Berwick, R. (1985). *The acquisition of syntactic knowledge.* Cambridge, MA: MIT Press.

Bloom, L. (1970). *Language development: Form and function in emerging grammar.* Cambridge, MA: MIT Press.

Bloom, P. (1990). Subjectless sentences in child languages. *Linguistic Inquiry* 21: 491-504.

Bloom, P. (1993). Grammatical continuity in language development: the case of subjectless sentences. *Linguistic Inquiry* 24: 721-34.

Borer, H. and K. Wexler (1987). The maturation of syntax. In Roeper, T. and E. Williams (eds.). *Parameter setting*. Dordrecht: Reidel.

Braine, M. (1973). Three suggestions regarding grammatical analyses of children's language. In Ferguson, C. and D. Slobin (eds.) *Studies of child language development*. New York: Holt, Reinhart, and Winston.

Bromberg, H. and K. Wexler (1995). Null subjects in child wh-questions. MIT Working Papers in Linguistics 26.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Bush, R. and F. Mostellar. *Stochastic models for learning*. New York: Wiley.

Campos, H. (1986). Indefinite object drop. *Linguistic Inquiry* 17: 354-359.

Caselli, M.C., Bates E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., and Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development* 10: 159-199.

Charniak, E. (1995). *Statistical language learning.* Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon.

Chomsky, N. (1977). On Wh-movement. In Culicover, P., T. Wasow, and A. Akmajian (eds.) *Formal syntax*. New York: Academic Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

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

Changeux, J.-P. (1983). *L'Homme Neuronal*. Paris: Fayard.

Clahsen, H. (1986). Verbal inflections in German child language: Acquisition of agreement markings and the functions they encode. *Linguistics* 24: 79-121.

Clahsen, H. (ed.) (1996). *Generative perspectives in language acquisition*. Amsterdam: John Benjamins.

Clahsen, H. and M. Penke (1992). The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone-corpus. In Meisel, J. (ed.) *The acquisition of verb movement: Functional categories and V2 phenomena in language acquisition*. Dordrecht: Kluwer.

Clahsen, H., M. Penke, and T. Parodi. (1994). Functional categories in early child German. *Language Acquisition* 3: 395-429.

Clahsen, H. and K. Smolka (1986). Psycholinguistic evidence and the description of V2 in German. In Haider, H. and M. Prinzhorn (eds.) *Verb second phenomena in Germanic languages*. Dordrecht: Foris.

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

Clark, R. and I. Roberts (1993). A computational model of language learnability and language change. *Linguistic Inquiry* 24: 299-345.

Cohan, F. (1984). Can uniform selection retard random genetic divergence between isolated conspecific populations? *Evolution* 38: 495-504.

Crain, S. and M. Nakayama (1987). Structure dependency in grammar formation. *Language* 63: 522-543.

Deacon, T. (1996). *The symbolic species: the co-evolution of brain and language.* New York: Norton.

Depew, D. and B. Weber (1995). *Darwinism evolving*. Cambridge, MA: MIT Press.

Dëprez, V. and A. Pierce (1993). Negation and functional projections in early grammar. *Linguistic Inquiry* 24: 25-68.

Derbyshire, D. (1977). Word order universals and the existence of OVS languages. *Linguistic Inquiry* 8: 590-599.

Derbyshire, D. and G. Pullum (eds.) (1986). *Handbook of Amazonian languages*, Vol. 1. Berlin: Mouton de Gruyter.

Dresher, E. and J. Kaye (1990). A computational learning model for metrical phonology. *Cognition*, 34: 137-195.

Dresher, E. (1999). Charting the learning path: cues to parameter setting. *Linguistic Inquiry.*

Edelman, G. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Elman, Jeffrey, Elizabeth Bates, Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett (1996). *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Felix, S. (1987). *Cognition and language growth.* Dordrecht: Foris.

Frank, R. and S. Kapur (1996). On the use of triggers in parameter setting. *Linguistic Inquiry* 27: 623-660.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry* 29: 1-36.

Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language* 30: 431-51.

Gibson, E. and K. Wexler (1994). Triggers. *Linguistic Inquiry* 25: 355-407.

Gleitman, L. (1981). Maturational determinants of language growth. *Cognition* 10: 115-26.

Guasti, M. (1994). Verb syntax in Italian child grammar: finite and nonfinite verbs. *Language Acquisition* 3: 1-40.

Guilfoyle, E. (1984). The acquisition of tense and the emergence of lexical subjects in child grammar. McGill Working Papers in Linguistics 2.

Gold, M. (1967). Language identification in the limit. *Information and Control* 10: 447-474.

Haegeman, L. (1996). Root infinitives, clitics, and truncated structures. In Clahsen (1996).

Haegeman, L. (1997). Register variation, truncation and subject omission in English and French. *English Language and Linguistics* 1: 233-270.

Haegeman, L. (forthcoming). Adult null subjects in non *pro*-drop languages. In Friedemann, M. & L. Rizzi (eds.) *The acquisition of syntax.* London: Longman.

Halle, M. (1962). Phonology in generative grammar. *Word* 18: 54-72.

Harris, A. and K. Wexler (1996). The optional-infinitive stage in child English: Evidence from negation. In Clahsen, H. (ed.) (1996).

Hartl, D. and A. Clark (1996). *Principles of population genetics.* Boston, MA: Sinauer.

Hebb, D. (1949). *The organization of behavior.* New York: Wiley.

Huang, J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15: 531-574.

Huang, J. (1989). Pro-drop in Chinese: A generalized control theory. In Jaeggli, O. and K. Safir (eds.) *The null subject parameter*. Dordrecht: Kluwer.

Hubel, D. and T. Wiesel (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160: 106-54.

Hyams, N. (1986) *Language acquisition and the theory of parameters*. Reidel: Dordrecht.

Hyams, N. (1991). A reanalysis of null subjects in child language. In Weissenborn, J., H. Goodluck, and T. Roeper (eds.) *Theoretical issues in language acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum.

Hyams, N. (1996). The underspecification of functional categories in early grammar. In Clahsen (1996).

Jaeggli, O. and K. Safir (1989). *The null subject parameter*. Boston: Kluwer.

Jansen, F. (1981). *Syntaktische konstrukties in gesproken taal*. Amsterdam: Huis aan de drie grachten.

van Kemenade, A. (1987). *Syntactic case and morphological case in the history of English*. Dordrecht: Foris.

van Kemenade, A. and N. Vincent (eds.) (1997). *Parameters and morphosyntactic change*. New York: Cambridge University Press.

Krämer, I. (1993). The licensing of subjects in early child language. MIT Working Papers in Linguistics 19.

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change* 1: 199-244.

Kroch, A. and A. Taylor (1997). Verb movement in Old and Middle English: dialect variation and language contact. In van Kemenade, A. and N. Vincent (eds.) *Parameters of morphosyntactic change*. Cambridge: Cambridge University Press.

Lebeaux, D. (1988). Language acquisition and the form of the grammar. Ph.D. dissertation. UMass, Amherst.

Lewontin, R. (1983). The organism as the subject and object of evolution. *Scientia* 118: 65-82.

Lewontin, R. and M. White. (1960). Interaction between inversion polymorphisms of the two chromosome pairs in the grasshopper, *Moraba scurra*. *Evolution* 14: 116-129.

Lightfoot, D. (1979). *Principles of diachronic syntax*. Cambridge, UK: Cambridge University Press.

Lightfoot, D. (1991). *How to set parameters*. Cambridge, MA: MIT Press.

de Marcken, C. (1996). Unsupervised language acquisition. Ph.D. dissertation, MIT.

Macnamara, J. (1982). *Names for things: A study of human learning*. Cambridge, MA: MIT Press.

MacWhinney, B. and C. Snow (1985). The Child Language Date Exchange System. *Journal of Child Language* 12, 271-296.

Maynard Smith, J. (1989). *Evolutionary genetics*. New York: Oxford University Press.

Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.

Mayr, E. (1991). *One long argument: Charles Darwin and the genesis of modern evolutionary thought.* Cambridge, MA: Harvard University Press.

McCloskey, J. (1996). Subjects and subject positions in Irish. In Borsley, R. and I. Roberts (eds.) *The syntax of the Celtic languages: A comparative perspective*. Cambridge University Press.

Narendra, K. and M. Thathachar (1989). *Learning automata*. Englewood Cliffs, NJ: Prentice Hall.

Niyogi, P. and R. Berwick (1996a). A language learning model for finite parameter space. *Cognition* 61: 162-193.

Niyogi, P. and R. Berwick (1996b). Learning from triggers. *Linguistic Inquiry* 27: 605-622.

O'Grady, W. (1997). *Syntactic development.* Chicago: University of Chicago Press.

Osgood, C. and T. Sebeok. (1954). Psycholinguistics: A survey of theory and research problems. *Journal of Abnormal and Social Psychology* 49: 1-203.

Phillips, C. (1995). Syntax at age 2: Cross-linguistic differences. MIT Working Papers in Linguistics 26.

Pierce, A. (1992). *Language acquisition and syntactic theory: a comparative analysis of French and English child grammar.* Boston: Kluwer.

Pinker, S. (1979). Formal models of language learning. *Cognition* 7: 217-283.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pinker, S., D. Lebeaux, and L. Frost. (1987). Productivity and constraints in the acquisition of the passive. *Cognition* 26: 195-267.

Pintzuk, S. (1991). Phrase structure in competition: variation and change in Old English word order. PhD thesis. University of Pennsylvania.

Poeppel, D. and K. Wexler (1993). The full competence hypothesis. *Language* 69: 1-33.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax.* Oxford: Basil Blackwell.

Radford, A. (1996). Towards a structure-building model of acquisition. In Clahsen (1996).

Randell, J. (1991). The catapult hypothesis: an approach to unlearning. In Weissenborn, J., H. Goodluck, and T. Roeper (eds.) *Theoretical issues in language acquisition.* Hillsdale, New Jersey:

Lawrence Erlbaum.

Rizzi, L. (1986). Null object in Italian and the theory of pro. *Linguistic Inquiry* 17: 501-557.

Rizzi, L. (1994). Early null subjects and root null subjects. In Hoekstra, T. and B. Schwartz (eds.) *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins.

Rizzi, L. (1996). A parametric approach to comparative syntax: properties of the pronominal system. In Haegeman, L. (ed.) *The new comparative syntax*. London: Longman.

Roberts, I. (1993). *Verbs and diachronic syntax: A comparative history of English and French.* Dordrecht: Kluwer.

Roberts, I. (1997). Directionality and word order change in the history of English. In van Kemenade and Vincent (eds.).

Roeper, T. and B. Rohrbacher (1994). Null subject in early child English and the theory of economy of projection. Ms. UMass Amherst.

Raposo, E. (1987). Case theory and Infl-to-Comp: the inflected infinitives in European Portuguese. *Linguistic Inquiry* 18: 85-109.

Saffran, J.R., R. N. Aslin, and E. L. Newport. (1996) Statistical learning by 8-month old infants. *Science* 274, 1926-1928.

Sano, T. and N. Hyams (1994). Agreement, finiteness, and the development of null arguments. NELS 24.

Santorini, B. (1992). Variation and change in Yiddish subordinate clause word order. *Natural language and linguistic theory* 10: 595-640.

Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275: 1599-1604.

Speas, M. (1994). Null arguments in a theory of economy of projection. Ms. UMass, Amherst.

Stolcke, A. (1994) Bayesian Learning of Probabilistic Language Models. Ph.D. thesis, University of California at Berkeley, Berkeley, CA.

Stromswold, K. (1996). Paper presented at Boston University conference on language development.

Valian, V. (1990). Null subjects: A problem for parameter-setting models of language acquisition. *Cognition* 35: 105-122.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* 40: 21-82.

Waller, B. (1996). Against a metrical basis for subject drop in child language. Ms. MIT.

Wang, Q., D. Lillo-Martin, C. Best, and A. Levitt. (1992). Null subject vs. null object: Some evidence from the acquisition of Chinese and English. *Language Acquisition* 2: 221-54.

Weinberg, A. (1990). Markedness versus maturation: the case of subject-auxiliary inversion. *Language Acquisition* 2: 165-194.

Weinreich, U., W. Labov, and M. Herzog. (1968). Empirical foundations for a theory of language change. In W. Lehmann (ed.) *Directions for historical linguistics: A symposium.* Austin: University of Texas Press.

Weverink, M. (1989). The subject in relation to inflection in child language. M.A. thesis. University of Utrecht.

Wexler, K. (1994). Optional infinitives, head movement, and the economy of derivation in child language. In Lightfoot, D. and N. Hornstein (eds.) *Verb movement.* Cambridge: Cambridge University Press.

Wexler, K. (1996). The development of inflection in a biologically based theory of language acquisition. In Rice, M. (ed.) *Toward a genetics of language.* Mahwah, NJ: Erlbaum.

Wexler, K. and P. Culicover (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Wijnen, F. (1999). Verb placement in Dutch child language: A longitudinal analysis. Ms. University of Utrecht.

Yang, C. and S. Gutmann (1999). Learning by martingales. Ms. MIT and Northeastern University.