Language Variation and Change, **12** (2000), 231–250. Printed in the U.S.A.

Internal and external forces in language change

CHARLES D. YANG
Yale University

ABSTRACT

If every productive form of linguistic expression can be described by some idealized human grammar, an individuals's variable linguistic behavior (Weinreich, Labov, & Herzog, 1968) can be modeled as a statistical distribution of multiple idealized grammars. The distribution of grammars is determined by the interaction between the biological constraints on human grammar and the properties of linguistic data in the environment during the course of language acquisition. Such interaction can be formalized precisely and quantitatively in a mathematical model of language learning. Consequently, we model language change as the change in grammar distribution over time, which can be related to the statistical properties of historical linguistic data. As an empirical test, we apply the proposed model to explain the loss of the verb-second phenomenon in Old French and Old English based on corpus studies of historical texts.

Language change is observed when a generation of speakers produces linguistic expressions that differ from those of previous generations, either in form or in distribution. Language change is explained when its causal forces are identified and their interactions are made clear.

At least two components are essential for any causal theory of language change. One component, long recognized by historical linguists, is a theory of language acquisition by child learners: ultimately, language changes because learners acquire different grammars from their parents. In addition, as children become parents, their linguistic expressions constitute the acquisition evidence for the next generation. Following Battye and Roberts (1995), this iterative process can be stated in terms of the familiar distinction between E-language and I-language (Chomsky, 1986), as shown in Figure 1.

The other crucial component in language change has become clear through the generative linguistics research of the past half century. Modern linguists and psychologists have drawn attention to an important fact of child language: namely, although child language differs from adult language, it differs in highly restrictive ways. Given the input to children, there are logically possible and computationally simple inductive rules to describe the data that are, however, never attested

I would like to thank Morris Halle, Julie Legate, and an anonymous LVC reviewer for discussion and comments. I am also indebted to Ian Roberts, Tony Kroch, and Ann Taylor for their scholarship on the history of French and English, which provides the empirical foundation for the formal model presented here. This work was partially supported by a National Science Foundation graduate fellowship.

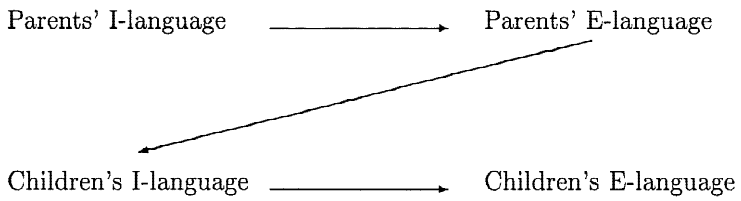

FIGURE 1. The dynamics of language acquisition and language change.

in child language (Chomsky, 1975; Crain & Nakayama, 1987). The restrictiveness of the human language space, coupled with the similarities revealed in comparative studies of the world's languages, have led linguists to conclude that human languages are delimited in a finite space of possibilities. A Universal Grammar (UG) is proposed as part of our biological endowment, which consists of discrete rules and constraints that interact in infinite, yet nonarbitrary, ways. Therefore, language acquisition and hence language change are determined by both internal and external factors. The internal knowledge of UG determines the space of languages that learners can attain, and the external linguistic experience in the environment determines what language children do attain. Their interactions over time, as depicted in Figure 1, in turn determine the space of language change.

In this article, we develop a model of language change that characterizes the dynamic interaction between the internal UG and the external linguistic evidence, as mediated by language acquisition. We borrow insights from the study of biological evolution, where internal and external forces (i.e., genetic endowment and environmental conditions) interact in a similar fashion. Our starting point is a model of language acquisition (Yang, 1999a, 1999b, 2000) which is able to relate directly and quantitatively the statistical properties of linguistic evidence to the grammatical knowledge that a learner attains. Next, we give the motivation for and a brief summary of the acquisition model. After spelling out our model of language change in detail and deriving a number of formal results, we use the proposed model to explain the loss of verb-second (V2) in Old French and the erosion of V2 in Old English.

UNIVERSAL GRAMMAR AND LEARNING

One concrete instantiation of UG is the Principles and Parameters (P&P) framework (Chomsky, 1981), which broadly encompasses a number of grammatical formalisms. In the P&P framework, the space of linguistic variability is defined by a set of parametric choices interacting with universal principles. Hence, acquiring a language has been formulated as setting the parameters to the appropriate values of the target grammar.[1]

An influential approach to parameter setting is the triggering model (Gibson & Wexler, 1994). In this model, the learner changes the value of a parameter if the

present grammar cannot analyze an incoming sentence and the grammar with the changed parameter value can. Hence, the learner changes hypotheses in an all-or-none manner: parameters are "triggered" (switched on and off) on the basis of relevant input sentences. However, this discrete view of language learning is at odds with the facts of language development.[2]

In general, child language development is gradual, thus providing no support for the on/off view of parameter setting. Consider, for example, the extensively studied null subject stage in child language, during which children often drop subjects in sentences that require subjects. It has been suggested (Hyams 1986; Hyams & Wexler, 1993) that the null subject stage is due to missetting the subject parameter to the value of an optional subject grammar. The null subject stage ends when children later reset the parameter to the target value. However, Bloom (1993) found no sharp change in the frequency of subject use throughout the null subject stage of Adam and Eve, two American children studied by Brown (1973). Behrens (1993) reported similar findings in a longitudinal study of German children's null subject stage. Furthermore, English children's subject drop is distributionally quite different from that of children acquiring true optional subject grammars such as Italian (Valian, 1991) or Chinese (Wang, Lillo-Martin, Best, & Levitt, 1992), contrary to the proposal of parameter missetting and triggering.

At a more fundamental level, there is a tension between the discrete and symbolic basis of grammatical competence and the continuous and variable patterns in linguistic performance; the gradualness of language acquisition is just a case in point. Furthermore, Weinreich, Labov, and Herzog, in a classic paper, argued that it is unrealistic to study language as a "homogeneous object," and that the "nativelike command of heterogeneous structures is not a matter of multidialectalism or 'mere' performance, but is part of unilingual linguistic competence" (1968:101). Subsequent work in this research tradition has amply demonstrated the inherent variability of linguistic expressions, challenging the idealized conception of knowledge of language as a unique and abstract "grammar."

We believe that this tension can be resolved once the study of language is situated in a broad biological framework. The tension between UG and language performance (acquisition and use) is reminiscent of the tension between the discrete basis of Mendelian genetics and the continuous distribution of genotypes in populations, which was resolved in the modern synthetic theory of evolution.

The variational approach to biology and language

As Ernst Mayr (1963) remarked, a premise of understanding biological evolution is the recognition of the uniqueness of individuals: variation among individuals is real; it is not due to individuals' imperfect realizations of some idealized archetype. Furthermore, evolution is a variational process, in which the distribution of individual genotypes changes over time rather than the individuals themselves changing directly (Lewontin, 1983). By associating probabilities with the distri-

bution of variant individuals, modern population genetics is able formally to characterize the effect of evolutionary forces, thus bridging the gap between the discrete variation at the individual level and the continuous variation at the population level.

We may take the variational approach to the study of language acquisition and use. The central observation is again variation: the variation across languages and the variation in child language en route to adult language. In light of the variational thinking in biology, the linguistic difference between children and adults may not be children's imperfect grasp of adult language, as traditional approaches have suggested. Rather, it may reflect principled hypotheses of language that children entertain before conclusively settling on the target language. Hence, language acquisition can be viewed as a variational process in which the distribution of grammars changes as an adaptive response to the linguistic evidence in the environment.

In what follows, we summarize a formal model that embodies the variational approach to language acquisition. Our model of language change, which is presented in the next section, is a fairly straightforward derivative of the acquisition model.

Language acquisition as grammar competition

The variational approach to language acquisition can be realized as a competition process among a population of grammars. It is significant that competition-based learning has been extensively studied as a general model of learning in mathematical psychology, and that it enjoys considerable support in human and animal learning and decision making (Atkinson, Bower, & Crothers, 1965; Bush & Mosteller, 1951, 1958). To be sure, such models were intended to describe simple associative conditioning between stimulus and response, and the learning hypothesis space for learning is usually limited to a number of simple choices. Nevertheless, its key component of hypothesis competition seems quite general across behavioral domains. For our purpose, we supply the general competition learning model with a domain-specific hypothesis space: UG, a finite collection of possible human grammars.

To account for the gradualness of language acquisition and, ultimately, of variability in language use, we explicitly introduce statistical notions to our competition-based learning model. Each grammar G_i is associated with a weight p_i, which denotes the probability with which the learner accesses that grammar. Write $s \in E$ to indicate that a sentence s is a possible utterance in the linguistic environment E. We assume E is a fixed environment, from which s is drawn independently and randomly. Write $G \rightarrow s$ if a grammar G can analyze s, which, in a narrow sense, can be interpreted as parsability (Berwick, 1985; Wexler & Culicover, 1980). Suppose that there are altogether N grammars in the population. For simplicity, write p_i for the weight of G_i at time t and p_i' for that at time $t + 1$. Each time instance denotes the presentation of an input sentence. Learning is modeled as in (1).

(1) Given an input sentence s, the child with the probability p_i selects a grammar G_i

$$\text{if } G_i \rightarrow s \text{ then } \begin{cases} p_i' = p_i + \gamma(1 - p_i) \\ p_j' = (1 - \gamma)p_j & \text{if } j \neq i \end{cases}$$

$$\textit{if } G_i \not\rightarrow s \text{ then } \begin{cases} p_i' = (1 - \gamma)p_i \\ p_j' = \dfrac{\gamma}{N - 1} + (1 - \gamma)p_j & \text{if } j \neq i \end{cases}$$

The model given in (1) is the linear reward–penalty (L_{R-P}) scheme (Bush & Mosteller, 1951, 1958), one of the earliest and most extensively studied learning algorithms in mathematical psychology.[3] Informally, it says that, when an input sentence is presented, a grammar is selected; the probability with which it is selected is determined by its weight. The grammar is then used by the learner to analyze the sentence. If the analysis is successful (i.e., the sentence is successfully parsed), the selected grammar is rewarded, and all the other grammars are indirectly punished; otherwise, the selected grammar is punished, and all the other grammars are indirectly rewarded. Reward and punishment are linear functions of a learning parameter γ.

In any selectionist process, some measure of fitness of individuals must be defined. Following Bush and Mosteller (1958), we have (2).

(2) The penalty probability of grammar G_i in a linguistic environment E is

$$c_i = \Pr(G_i \not\rightarrow s | s \in E)$$

Penalty probability is an intrinsic property of a grammar relative to a fixed linguistic environment E, determined by the distributional patterns of expressions used in E. For example, consider a Germanic V2 environment. A V2 (target) grammar, of course, has the penalty probability of 0. An English SVO grammar, although not compatible with all V2 sentences, is nevertheless compatible with a certain proportion of them. For example, according to a corpus analysis cited in Lightfoot (1997:265), about 70% of Dutch, German, Norwegian, and Swedish sentences have the surface order SVO in matrix clauses.[4] Since the grammars in the delimited UG space are fixed (i.e., it is only their weights that change), their fitness values, defined as penalty probabilities, are also fixed in a particular linguistic environment. Note that the learner does not access or compute penalty probabilities, which are merely formal devices used to study the dynamics of learning. However, these fitness measures can be estimated from corpora of (historical) texts, making quantitative predictions on the rate and direction of language acquisition and language change. This is what we do later in the article.

The asymptotic properties of competition learning models have been extensively studied in both mathematical psychology (Norman, 1972) and machine learning (Narendra & Thathachar, 1989). For simplicity but without loss of generality, suppose that there are two grammars in the population, a target grammar G_1 and a "pretender" G_2, which are associated with penalty probabilities c_1 and c_2, respectively. If the learning rate γ is sufficiently small (i.e., if learners do not alter their "confidence" in grammars too radically), one can show (see Narendra

& Thathachar, 1989:162–165) that the asymptotic distributions of $p_1(t)$ and $p_2(t)$ will be essentially normal and can be approximated as in (3).

$$(3)\quad \lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1 + c_2}$$

$$\lim_{t\to\infty} p_2(t) = \frac{c_1}{c_1 + c_2}$$

(3) shows that grammars more compatible with the input data (i.e., with lower penalty probabilities) are better represented in the population than those less compatible with the input data (i.e., with higher penalty probabilities). It follows that a unique target in an idealized homogeneous learning environment will eliminate all competing grammars by virtue of having the penalty probability of 0. Thus, the acquisition model meets the standard learnability condition (Gold, 1967). It is also clear that the success of learning depends not on any theory-internal assumptions about UG, but on the assumption of a finite number of possible grammars.

Because penalty probabilities of competing grammars can be directly estimated from text corpora such as the CHILDES database (MacWhinney & Snow, 1985), the present model allows quantitative predictions about the rate of language development. The gradualness of language acquisition is interpreted as the target grammar's gradual rise to dominance. In the meantime, competing nontarget grammars are probabilistically accessed by the child learner, resulting in variation (nonuniformity) in child language. For instance, the null subject phenomenon is attributed to an obligatory subject grammar in coexistence with an optional subject grammar. The reader is referred to Yang (1999a, 1999b, 2000) for developmental evidence of this sort in support of the variational model of acquisition.

More pertinent to language change, we conclude this section by considering an immediate consequence of an acquisition model situated in a realistic learning environment, with the sort of inherent linguistic variability identified by Weinreich et al. (1968). If linguistic expressions, however variable, can be attributed to some possible human grammars, then such a learning environment can be formally viewed as consisting of expressions generated by a combination of multiple sources, each of which is an idealized grammar under our assumption of UG. It is therefore clear from (3) that, in such environments, the learner converges to a stable equilibrium of grammar distributions. This result is crucial for our model of language change, which we lay out next.

GRAMMAR COMPETITION AND LANGUAGE CHANGE

The role of linguistic evidence

With the competition-based acquisition model in hand, we return to the problem of language change. The fundamental question is, what makes generation $n + 1$ attain knowledge of language that differs from that of generation n?

Recall that the language attained by a learner is the product of internal knowledge of UG and external linguistic evidence present in the environment, which are mediated by the algorithm of language acquisition. If we assume, as there is no reason not to, that the biological endowment of UG is held constant from generation to generation, we may conclude that the source for the discrepancy between two generations of speakers lies in the linguistic evidence: generation n and $n + 1$ are exposed to sufficiently different linguistic evidence and thus form different knowledge of language as a result.

This conclusion is only warranted under some further justifications. We argue that language change cannot take place without sufficiently different linguistic evidence across generations. With a generation of speakers viewed as a population of individuals, it remains a theoretical possibility that, in spite of comparable linguistic evidence, some members of generation $n + 1$ attain a different grammar from generation n, as a result of mislearning. However, this position is untenable for three empirical reasons.

First, language acquisition research has shown that children are highly competent and robust learners. It seems improbable that, given sufficiently similar experience, children would attain languages that differ substantially (e.g., a major syntactic parameter is misset to a wrong value). Second, historical linguistics has shown that language change occurs on the scale of the entire population, not scattered individual members. As Bloomfield (1927, cited in Hockett, 1968:25) remarked,

> It may be argued that change in language is due ultimately to the deviations of individuals from the rigid system. But it appears that even here individual variations are ineffective; whole groups of speakers must, for some reason unknown to us, coincide in a deviation, if it is to result in a linguistic change. Change in language does not reflect individual variability, but seems to be a massive, uniform, and gradual alteration, at every moment of which the system is just as rigid as at every other moment.

Third, while one might attempt to invoke the idea of individual mislearning to explain historical change in some languages, it leaves mysterious the relative stability in other languages (e.g., the rigidity of word order in West Germanic languages).

We therefore reject individual mislearning under sufficiently similar linguistic evidence as a possible mechanism of language change. A question immediately arises. What makes the linguistic evidence for generation $n + 1$ different from that of the previous generation? There are many possibilities. Migration of foreign speakers may introduce novel expressions that were previously unseen; linguistic innovation may modify the linguistic evidence for the next generation of learners; or social and cultural factors may influence the distributional patterns of linguistic expressions used in a population. These are interesting topics of research and are an integral part of a complete explanation of language change. However, they are not directly relevant to a formal model of language change. We are chiefly concerned with the predictive consequences of such changes: what

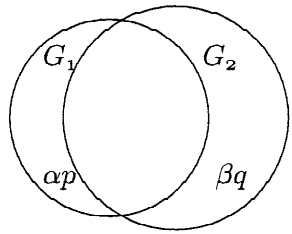

FIGURE 2. Two mutually incompatible grammars constitute a heterogeneous linguistic environment.

happens to language learners when the linguistic evidence is altered, and how does it affect the composition of the linguistic population as a result? This is much like the population genetic theory of evolution, which concerns the predictable changes in the population once some new genotypes are introduced. The precise manner in which new genes arise, which could be mutation, migration, and so on, is a separate question that often contains too much contingency to give a firm answer.

A variational model of language change

Suppose that, as a result of migration, genuine innovation, or other sociological and historical factors, a linguistic environment is established for a generation of language learners that differs substantially from the one for the previous generation. The expressions used in such an environment, call it $E_{G1,G2}$, can formally be viewed as a mixture of expressions generated by two independent sources: the two grammars G_1 and G_2. Suppose a proportion α of G_1 expressions are incompatible with G_2 and a proportion β of G_2 expressions are incompatible with G_1. Call $\alpha(\beta)$ the advantage of $G_1(G_2)$. Figure 2 illustrate this.

The variational approach views language acquisition as competition and selection among grammars. Recall that the fitness of individual grammars is defined in terms of their penalty probabilities (2), repeated in (4).

(4) The penalty probability of a grammar G_i in a linguistic environment E is

$$c_i = \Pr(G_i \not\rightarrow s | s \in E)$$

The penalty probabilities ultimately determine the outcome of language acquisition.

(5) $$\lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1 + c_2}$$

$$\lim_{t\to\infty} p_2(t) = \frac{c_1}{c_1 + c_2}$$

Suppose that at generation n the linguistic environment $E_{G1,G2} = pG_1 + qG_2$, where $p + q = 1$. That is, in $E_{G1,G2}$, a proportion p of expressions can be viewed

as generated by G_1 and a proportion q of expressions can be viewed as generated by G_2, and they collectively constitute the linguistic evidence to the learners in generation $n + 1$. The penalty probabilities of G_1 and G_2, c_1, and c_2, are thus βq and αp. The results in (5) allow us to compute p' and q', the weights of G_1 and G_2 respectively, that are internalized in the learners of generation $n + 1$.

(6) The dynamics of a two grammar system

$$p' = \frac{\alpha p}{\alpha p + \beta q}$$

$$q' = \frac{\beta q}{\alpha p + \beta q}$$

(6) shows that an individual learner in generation $n + 1$ may form a combination of two grammars, G_1 and G_2, at a different set of weights than do those in the parental generation n.[5] From (6), we have (7).

(7) $$\frac{p'}{q'} = \frac{\alpha p/(\alpha p + \beta q)}{\beta q/(\alpha p + \beta q)} = \frac{\alpha p}{\beta q}$$

In order for G_2 to overtake G_1, the weight of $G_2(q)$ internalized in speakers must increase in successive generations and eventually drive the weight of $G_1(p)$ to 0. That is, for each generation, it must be the case that $q' > q$, which is equivalent to $p'/q' < p/q$. Thus, we obtain a sufficient and necessary condition for grammar competition in a linguistic population.

(8) The fundamental theorem of language change
G_2 overtakes G_1 if $\beta > \alpha$: the advantage of G_2 is greater than that of G_1

Recall that α and β are presumably constants, which characterize the distributional patterns in the use of the respective languages. Note that we may not be able to estimate α and β directly from historical context, which only reflects the penalty probabilities of the competing grammars (i.e., αp and βq). However, (8) says that, if $q' > q$ (G_2 is on the rise), it must be the case that $\beta > \alpha$, and that, if $\beta > \alpha$, G_2 will necessarily replace G_1. Hence, we have the corollary in (9).

(9) Corollary: Once a grammar is on the rise, it is unstoppable.

Plotting $q(t)$, the weight of G_2, as a function of time t, as in Figure 3, we obtain the familiar S-shaped curve that is often observed in language change (Bailey, 1973; Kroch, 1989; Weinreich et al., 1968), as the new linguistic form gradually replaces the old form.

The present model shares an important feature with Clark and Roberts' (1993) work, which extended the use of Genetic Algorithms in acquisition (Clark, 1992). In both models, the outcome of language acquisition is determined by the com-

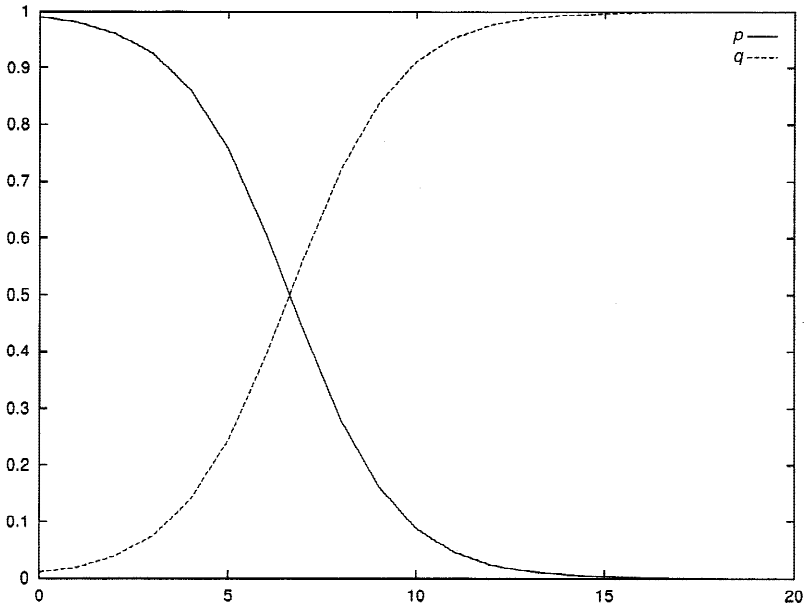

FIGURE 3. One grammar (q) replacing another (p) over time. The X-axis denotes the number of generations, and the Y-axis denotes the weights of the two competing grammars.

patibilities of grammars with linguistic evidence, in a Darwinian selectionist manner. However, Clark and Roberts identified the final state of acquisition with a single grammar.[6] Therefore, when the linguistic evidence does not unambiguously identify a single grammar as a realistic, inherently variable environment, they posited some general constraints on the learner (e.g., the elegance condition, which requires the learner to select the simplest among conflicting grammars). Aside from such auxiliary assumptions that require independent justification from child language research, the position of learners converging on a single grammar cannot be defended in face of the evidence found by Kroch and his colleagues (Kroch, 1989; Kroch & Taylor, 1997; Kroch, Taylor, & Ringe, 2000; Pintzuk, 1991; Santorini, 1992). They showed that it is possible for historical texts during a period of language change to reflect a combination of multiple grammars. In fact, historical linguists commonly use terms such as "erosion" or "optional application" to indicate the gradual disappearance of a grammatical construction. These facts and, more generally, linguistic variability of the sort noted by Weinreich et al. (1968) can straightforwardly be modeled as the coexistence of multiple UG grammars, as in the approach taken here.

For the purpose of this article, we assume that all speakers in a linguistic community are exposed to identical linguistic experience, and that a speaker's linguistic knowledge is stable after the period of language acquisition (i.e., there is no generational overlap). It is possible to incorporate such spatially and tem-

porally varying factors into the dynamics of language change, which may be aided by the well-established models of population genetics and evolutionary ecology. We leave these options for future research.

To summarize, we extended the variational model of language acquisition to a population of learners and presented some analytical results concerning the dynamical system thus construed. We concluded that heterogeneity in the linguistic evidence, however introduced, is a prerequisite for language change. Once the homogeneity is punctured, language learners form internal representations of coexisting 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, drawing comparisons and connections to previous analyses.

THE LOSS OF V2 IN FRENCH

Old French (OF) has a cluster of properties, including V2 and *pro*-drop, that are lost in Modern French (ModF). The examples in (10) and (11) are taken from Clark and Roberts (1993).

(10) Loss of null subjects

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

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

(11) Loss of V2

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

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

In this section, we provide an analysis for the loss of V2 under the variational model. All examples and statistics cited in this section are taken from Roberts (1993).

Recall that, in order for the ModF SVO grammar to overtake the OF V2 grammar, the SVO grammar must have a greater advantage. That is, there must be more sentences in the linguistic evidence that are incompatible with the V2 grammar than with the SVO grammar. (12) shows the advantage patterns of V2 over SVO, and vice versa.

(12) a. Advantage of V2 grammar over SVO grammar

$$V2 \rightarrow s \text{ but } SVO \not\rightarrow s\text{: VS (XVSO, OVS)}$$

b. Advantage of SVO grammar over V2 grammar

$$SVO \rightarrow s \text{ but } V2 \not\rightarrow s\text{: V}>2 \text{ (SXVO, XSVO)}$$

If the distribution patterns in modern V2 languages are indicative of those of older times, we can see that the V2 constraint is in general very resilient to ero-

sion. In languages such as German, the V2 constraint is very strongly manifested. V>2 patterns are restricted to *if*... *then* clauses, left dislocation, a small number of adverbs, and so on and are quite rare in frequency (cf. Hawkins, 1985).

(13) Rare V>2 patterns in modern German

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

Statistical analysis of Dutch, German, Norwegian, and Swedish corpora (cited in Lightfoot, 1997:265) shows that about 70% of all sentences in V2 languages are SVO and about 30% are VS patterns, which include XVSO and OVS. Our own counts, based on a Dutch sample of adult-to-child speech (MacWhinney & Snow, 1985), are similar: 66.8% SVO, 23% XVSO, and 1.2% OVS. In contrast, based on the Penn Treebank, a corpus of modern English, we found that less than 10% of all sentences have V>2 word order.

(14) V>2 patterns in modern English

a. He always reads newspapers in the morning.
b. Every night after dinner Charles and Emma Darwin 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 did OF lose it? The reason, in our view, is that OF is a null subject language. Recall that the advantage of the V2 grammar over the SVO grammar is expressed in VS patterns. However, this advantage would be considerably diminished if the subject were dropped to yield [X V *pro*] patterns: a null subject SVO grammar (e.g., modern Italian) can analyze such patterns as [X (*pro*) V]. (15) shows the prevalence of subject drop in early Middle French (MidFr) (Roberts, 1993:155).

(15)

Text	SV	VS	*pro*
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 is reduced to 5%–18% in *pro*-drop MidFr. During the same period of time, V>2 patterns go from less than 5% in OF (Roberts, 1993:95) to 11%–15% in early MidFr, as the class of sentence-initial XPs that do not trigger SV inversion is expanded (Vance, 1989). (16) shows some representative examples.

(16) V>2 patterns in early MidFr

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 . . .'

(17), which is based on the examination of the three texts in (15), shows the frequency of V>2 patterns in MidFr (Roberts, 1993:148).

(17)

Text	V>2
Froissart, Chroniques (c. 1390)	12%
15 Joyes (14esme Joye) (c. 1400)	15%
Chartier, Quadrilogue (1422)	11%

Comparing (17) with (15), we see that at the early MidFr stage there are more V>2 sentences than VS sentences, due to the effect of subject drop. Thus, an SVO grammar (with *pro*-drop) has an advantage over a V2 grammar (with *pro*-drop). V2 in French was destined to extinction, as predicted by the theorem in (8) and its corollary in (9).

Our analysis of the loss of V2 in French crucially relies on the fact that null subjects were lost after V2 was lost. Statistics in Roberts (1993) show that this was indeed the case. In the late 15th century and early 16th century, when SVO orders were already favored, there was still significant use of null subjects, as the statistics in (18) demonstrate (Roberts, 1993:155, 199).

(18) The lasting effect of *pro*-drop in MidFr

	SV	VS	*pro*
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 in (19) (Roberts, 1993:199).

(19)

	SV	VS	*pro*
15th century	48%	10%	42%
16th century	77%	3%	15%

The decline and eventual 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 had almost completely vanished, there was still a considerable amount of subject drop. This diachronic pattern is consistent with our explanation for the loss of V2 in OF.

We believe that the present analysis may extend to other Western Romance languages, which, as is well known, all had V2 in medieval times. Under the present model of grammar competition, it is no accident that such languages at one time had *pro*-drop, as in OF, and that many still do, as in Italian and Spanish. It appears that the combination of *pro*-drop and V2 is intrinsically unstable and

necessarily gives away to an SVO (plus *pro*-drop) grammar. Without concrete statistics from the history of these languages, we can only extrapolate from their modern forms. It is reported (Bates, 1976, cited in Caselli, Casadio, & Bates, 1999:73) that modern Italian employs *pro*-drop in 70% of all sentences; as a result, the 30% advantage of a V2 grammar over an SVO grammar (in VS sentences) would be reduced to 30% × 30% = 9%. This is a figure already lower than 10%, our estimate of the advantage that an SVO grammar has over a V2 grammar, which would lead to the demise of V2.

THE EROSION OF V2 IN MIDDLE ENGLISH

We now turn to the erosion of V2 in Middle English (ME). Unless otherwise specified, all our examples and statistics are taken from Kroch and Taylor (1997). Our interpretation of the historical facts they reported supports and formalizes their analysis.

Word order in Old English

Kroch and Taylor (1997) showed that Old English (OE) is, generally speaking, a Germanic language similar to Yiddish and Icelandic. Its peculiarities lie in the distribution of its V2 patterns, which differs from that of modern West Germanic languages such as Dutch and German (Kroch & Taylor, 1997; Pintzuk, 1991; van Kemenade, 1987).

In OE, when the subject is an NP, the finite verb occupies the second position in matrix clauses.

(20) V2 with NP subjects in OE

a. þæt hus hæfdon Romane to ðæm anum tacne geworht
that building had Romans with the one feature constructed
b. þær werþ se cyning Bagsecg ofslægen
there was the king slain

In contrast, a pronominal subject precedes the verb, creating superficially V3 patterns with a nonsubject topic phrase.

(21) V3 with pronoun subjects in OE

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 often analyzed as a clitic (Pintzuk, 1991; van Kemenade, 1987).

Furthermore, there are genuine V3 patterns when the topic position is occupied by a certain class of temporal adverbs and adjuncts. In these constructions, the subject, whether pronominal or phrasal, precedes the verb.[7]

TABLE 1. *V2 in southern early ME*

	NP Subjects	Pronoun Subjects
Preposed XP	% Inverted	% Inverted
NP complements	93 (50/54)	5 (4/88)
PP complements	75 (12/16)	0 (0/11)
Adjective complements	95 (20/21)	33 (7/21)
þa/then	95 (37/39)	72 (26/36)
now	92 (12/13)	27 (8/30)
PP adjuncts	75 (56/75)	2 (2/101)
Adverbs	57 (79/138)	1 (1/182)

(22) V3 with XP topics in OE

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 Willelmes dohtor
In this year William king gave (to) Ralph earl William's daughter
Osbearnes sunu
Osborn's son

In certain constructions, illustrated in (23), the verb invariably raises to C, and the subject, be it pronoun or phrasal, is postverbal.

(23) Verb raising to C in OE

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

The southern dialect

Kroch and Taylor (1997) showed that there was considerable dialectical variation with respect to the V2 constraint during early ME. Specifically, the southern dialect essentially preserved the V2 of OE. Preposed XPs, with exception of a certain class of adverbs and adjuncts noted earlier, generally trigger subject–verb inversion with full noun phrase subjects, but rarely with pronoun subjects. Table 1, taken from Kroch et al. (2000), illustrates.

Following van Kemenade (1987), we relate the eventual loss of V2 in OE to the loss of subject cliticization. The loss of subject cliticization (and the loss of word order freedom in general) can be linked to impoverishment of the morphological case system of pronouns (for a possible theoretical formulation of this

traditional idea, see Kiparsky, 1997). Recall the V3 patterns in the southern dialect of early ME, which are manifested in sentences with pronominal subjects (21) and with certain adverb and adjunct topics (22), schematically shown in (24).

(24) XP subject pronoun V_{fin} . . .

With the impoverishment and eventual loss of the morphological case system, clitics were no longer possible. Therefore, patterns such as (24) were no longer compatible with an OE-type V2 grammar. However, they were compatible with an SVO grammar, with the subject pronoun treated as a noun phrase, as in modern English. Examining Table 1, we can see that 62% (511/825) of all matrix sentences are of the V>2 pattern and 38% (314/825) are of the VS pattern. When subject pronouns could no longer be analyzed as clitics but only as noun phrases, the SVO grammar gained a greater advantage than the V2 grammar and eventually rose to dominance. The loss of the morphological case system makes the loss of V2 possible, and the competition between the SVO grammar and the V2 grammar is straightforwardly captured in the present model of language change.

Notice that we also have an immediate account for so-called residual V2 in modern English questions. In (23), we saw that, when V raises to C, both pronoun and phrasal subjects occupy postverbal position. In other words, the linguistic evidence for these constructions has been homogeneous with respect to a V2 grammar throughout the history of English. Therefore, their V2 character has been preserved.[8]

The northern dialect

In contrast to the southern dialect, Kroch and Taylor (1997) showed that the northern dialect, under heavy Scandinavian influence, very much resembled other modern Germanic languages. The V2 constraint was uniformly and rigidly enforced, and one does not find the almost categorical asymmetry between pronoun and phrasal subjects in OE and southern early ME.

As noted earlier, the V2 constraint exhibited in West Germanic languages is difficult to overthrow. This is due to the advantage a V2 grammar has over competing grammars such as an SVO grammar. The V2 grammar generates VS sentences that punish the SVO grammar, whereas the SVO grammar generates V>2 sentences that punish the V2 grammar, but VS sentences usually outnumber V>2 sentences. In discussing the loss of V2 in OF, we argued that subject drop in OF considerably diminished the advantage of the V2 grammar to the point where an SVO grammar, aided by an increase in V>2 patterns, eventually won out. How did northern early ME, a rigid V2 language without subject drop, evolve into an SVO language?

Kroch and Taylor demonstrated that the extensive contact between the northern and southern populations in the period of ME was essential to the eventual loss of V2 in English, insightfully attributing the erosion of V2 to the competition

TABLE 2. *V2 (after language contact) in the northern manuscript (Thornton) of the Mirror of St. Edmund*

	NP Subjects	Pronoun Subjects
Preposed XP	% Inverted	% Inverted
NP complements	100 (8/8)	64 (16/25)
PP complements	88 (21/24)	70 (48/69)
Adjective complements	100 (10/10)	25 (2/8)
then	86 (6/7)	51 (24/47)
now	100 (4/4)	82 (14/17)
Adverbs	80 (20/25)	57 (35/61)

of grammars in learners during language contact. Their linguistic analysis is identical to that required by the present model of language change. The northern V2 dialect, when mixed with the southern (essentially OE) language, constituted a heterogeneous linguistic environment for later generations of learners, who, instead of converging on a single grammar, acquired a mixture of coexisting grammars. Table 2, taken from Kroch et al. (2000), shows the consequences of language contact in the northern dialect.

The effect of language contact is clear. Recall that, prior to contact, the northern dialect was much like Germanic languages, in which V2 is strongly enforced. Kroch et al. (2000) found subject–verb inversion in 93% of all sentences containing subjects. After contact, as shown in Table 2, while phrasal subjects still follow subjects in general, the overall subject–verb inversion rate drops to 68% (208/305). This indicates that, as a result of language contact and mixing, the V2 constraint in the northern dialect was considerably weakened. Once the V2 constraint was sufficiently weakened, and if the morphological case system of the mixed language was lost, then an SVO grammar would have gradually taken over, in the manner described earlier for the loss of V2 in OE.

For the northern dialect, the initial contact with the southern dialect was crucial in the loss of V2. That is, a V2 language similar to the northern dialect would not lose V2 without language contact, which would introduce a substantial amount of V>2 patterns for the learner, even if its morphological case were lost. Northern Germanic languages such as Swedish, Danish, and Norwegian, with impoverished morphological case systems but nevertheless strongly V2, fall into this category. Once language contact was made, the homogeneity of linguistic evidence was punctured, which resulted in the coexistence of two distinct grammars in the learners. The loss of morphological case resulted in the loss of clitics, which further favored the SVO grammar and eventually culminated in its dominance. Kroch and Taylor's idea that language contact is the prerequisite for the loss of V2 in the northern dialect dovetails with our theoretical model rather nicely.

CONCLUSION

We now summarize this preliminary investigation of an acquisition-based model of language change. Our approach is motivated by Darwinian variational thinking and is founded on two observations. First, the deviation of child language from adult language is not simply noise or imperfection; it is the reflection of actual grammatical hypotheses. Second, language use is inherently variable, and there is evidence of multiple grammars in mature speakers during the course of language change. The model formalizes historical linguists' intuition of grammar competition and directly relates the statistical properties of historical texts (hence, acquisition evidence) to the direction of language change. It is important to recognize that, while sociological and other external forces clearly affect the composition of linguistic evidence, grammar competition as language acquisition (the locus of language change) is internal to the individual learner's mind/brain. We hope that the present model, by directly linking the statistical properties of historical texts and the predictable outcome of language acquisition, will contribute to a framework in which the problems in language change can be studied formally and quantitatively.

NOTES

1. For present purposes, we assume that the principles and parameters are innate. There is now a large body of literature that documents children's very early knowledge of language, which supports the assumption of innateness. For example, Pierce (1989) showed that, as early as the eighteenth month (i.e., around the beginning of the two-word stage, at which such tests are possible), children have virtually mastered the placement of finite verbs in so-called verb raising languages such as French, in which finite verbs precede negation and adverbs. This shows that their knowledge of tense and basic clausal structures must be available from very early on. This observation has been duplicated in many other languages and in many other constructions (for a summary, see O'Grady, 1997). In addition, for aspects of grammar that are acquired rather late, such as the null subject phenomenon, children in the course of learning show deviations from the target form in specific ways, and these deviations can be explained by appealing to innate principles and parameters (see Yang, 2000). It is possible that these developmental facts could be explained by extralinguistic means, say, with reference to learning strategies, language processing constraints, and so on. However, those cases would have to be argued for (and would certainly be interesting, if true) and would have to be compatible with the empirical studies of child language (in particular, very early knowledge of grammar).

2. For now, we put aside the learnability problems with the triggering model. Both theoretical analysis (e.g., Berwick & Niyogi, 1996; Dresher, 1999) and computer simulations (Kohl, 1999) have demonstrated that convergence to a target grammar cannot be guaranteed in the triggering model.

3. We chose the $L_{R\text{–}P}$ model primarily because it allows the learner to obtain a weighted combination of multiple grammars when learning stops, as in (3). This appears to be precisely the case that Kroch and his colleagues (Kroch, 1989; Kroch & Taylor, 1997; Pintzuk, 1991; Santorini, 1992) identified for speakers during language change and hence must be incorporated in any correct learning model. Alternative models may also satisfy this requirement, although at the present level of understanding we cannot be sure exactly how human children carry out grammar competition via updating grammar weights. Hence, what remains is our commitment to a competition approach to language learning, supported by the evidence reported in Yang (2000), and a concrete realization of this approach, as embodied in (1).

4. For simplicity, we follow Lightfoot (1991) in considering only degree-0 sentences as linguistic input, mainly because of the availability of the relevant corpus statistics. In addition, matrix V2 is presumably independent of the underlying word order in the language, which usually shows up in the embedded clause. In any event, it is clear that the utility of the model does not hinge on this assumption.

5. Suppose that, in a uniform linguistic environment E_{L1}, a small number (n) out of a total of N learners do misconverge on a nontarget grammar L_2. The effect of the mislearners on the next generation can be quantified as follows:

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

If $n \ll N$, then the linguistic environment for the next generation is virtually identical to a uniform environment without mislearners. Thus, the impact of the mislearners on the next generation is negligible.

6. The same is true for Niyogi and Berwick (1995), another formal and acquisition-based model of language change. In their model, some speakers converge on one grammar and others converge on another. That is, language change is viewed as the change in the proportion of speakers with one grammar versus the other.

7. Although genuine V3 patterns are also possible in modern West Germanic languages, they are restricted as discussed earlier in connection with (13). Their distribution is not as wide as in OE (see Kroch & Taylor, 1997, for details).

8. More precisely, what has been preserved are the parametric choices that OE made in dimensions such as question formation, which the residual V2 is attributed to.

REFERENCES

Atkinson, Richard, Bower, Gordon, & Crothers, Edward. (1965). *An introduction to mathematical learning theory.* New York: Wiley.

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

Bates, Elizabeth. (1976). *Language and context: The acquisition of pragmatics.* New York: Academic.

Battye, Adrian, & Roberts, Ian. (1995). Introduction. In A. Battye & I. Roberts (eds.), *Clause structure and language change.* New York: Oxford University Press. 1–30.

Behrens, Heike. (1993). *Temporal reference in German child language.* Doctoral dissertation, University of Amsterdam.

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

Berwick, Robert, & Niyogi, Partha. (1996). Learning from triggers. *Linguistic Inquiry* 27:605–622.

Bloom, Paul. (1993). Grammatical continuity in language development: The case of subjectless sentences. *Linguistic Inquiry* 24:721–734.

Bloomfield, Leonard. (1927). Review of J. O. H. Jespersen, *The philosophy of grammar. Journal of English and Germanic Philology* 26:244–246.

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

Bush, Robert, & Mosteller, Frederic. (1951). A mathematical model for simple learning. *Psychological Review* 68:313–323.

_____ (1958). *Stochastic models for learning.* New York: Wiley.

Caselli, Cristina, Casadio, Paola, & Bates, Elizabeth. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* 26:69–111.

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

_____ (1981). *Lectures on government and binding.* Dordrecht: Foris.

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

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

Clark, Robin, & Roberts, Ian. (1993). A computational model of language learnability and language change. *Linguistic Inquiry* 24:299–345.

Crain, Stephen, & Nakayama, Mineharu. (1987). Structure dependency in grammar formation. *Language* 63:522–543.

Dresher, Bezalel Elan. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30:27–67.

Gibson, Edward, & Wexler, Kenneth. (1994). Triggers. *Linguistic Inquiry* 25:355–407.

Gold, E. Mark. (1967). Language identification in the limit. *Information and Control* 10:447–474.

Hawkins, John. (1985). *A comparative typology of English and German.* Austin: University of Texas Press.

Hockett, Charles. (1968). *The state of the art.* The Hague: Mouton.

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

Hyams, Nina, & Wexler, Kenneth. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry* 24:421–459.

Kiparsky, Paul. (1997). The rise of positional licensing. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters and morphosyntactic change.* New York: Cambridge University Press. 460–494.

Kohl, Karen. (1999). *An analysis of finite parameter learning in linguistic spaces*. Master's thesis, MIT.
Kroch, Anthony. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
Kroch, Anthony, & Taylor, Ann. (1997). Verb movement in Old and Middle English: Dialect variation and language contact. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters and morphosyntactic change.* New York: Cambridge University Press. 297–325.
Kroch, Anthony, Taylor, Ann, & Ringe, Donald. (2000). The Middle English verb-second constraint: A case study in language contact and language change. In Susan Herring et al. (eds.), *Textual parameters in older language.* Philadelphia: Benjamins. 353–391.
Lewontin, Richard. (1983). The organism as the subject and object of evolution. *Scientia* 118: 65–82.
Lightfoot, David. (1991). *How to set parameters.* Cambridge, MA: MIT Press.
_____ (1997). Shifting triggers and diachronic reanalyses. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters and morphosyntactic change.* New York: Cambridge University Press. 253–272.
MacWhinney, Brian, & Snow, Catherine. (1985). The Child Language Date Exchange System. *Journal of Child Language* 12:271–296.
Mayr, Ernst. (1963). *Animal species and evolution.* Cambridge, MA: Harvard University Press.
Narendra, Kumpati, & Thathachar, Mandayam. (1989). *Learning automata.* Englewood Cliifs, NJ: Prentice Hall.
Niyogi, Partha, & Berwick, Robert. (1995). The logical problem of language change. Manuscript, MIT Artificial Intelligence Laboratory.
Norman, Frank. (1972). *Markov processes and learning models.* New York: Academic.
O'Grady, William. (1997). *Syntactic development.* Chicago: University of Chicago Press.
Pierce, Amy. (1989). *On the emergence of syntax: A crosslinguistic study.* Doctoral dissertation, MIT.
Pintzuk, Susan. (1991). *Phrase structure in competition: Variation and change in Old English word order.* Doctoral dissertation, University of Pennsylvania.
Roberts, Ian. (1993). *Verbs and diachronic syntax: A comparative history of English and French.* Dordrecht: Kluwer.
Santorini, Beatrice. (1992). Variation and change in Yiddish subordinate clause word order. *Natural Language and Linguistic Theory* 10:595–640.
Valian, Virginia. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* 40:21–82.
Vance, Barbara. (1989). *Null subjects and syntactic change in Medieval French.* Doctoral dissertation, Cornell University.
van Kemenade, Ans. (1987). *Syntactic case and morphological case in the history of English.* Dordrecht: Foris.
Wang, Qi, Lillo-Martin, Diane, Best, Catherine, & Levitt, Andrea. (1992). Null subject vs. null object: Some evidence from the acquisition of Chinese and English. *Language Acquisition* 2:221–254.
Weinreich, Uriel, Labov, William, & Herzog, Marvin. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkeil (eds.), *Directions for historical linguistics: A symposium.* Austin: University of Texas Press. 95–188.
Wexler, Kenneth, & Culicover, Peter. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.
Yang, Charles D. (1999a). Formalizing language development. In *Proceedings of the 35th meeting of the Chicago Linguistics Society.* Chicago: Chicago Linguistic Society.
_____ (1999b). A selectionist theory of language acquisition. In *Proceedings of the 37th meeting of the Association for Computational Linguistics*. Washington, DC: The Association for Computational Linguistics. 272–279.
_____ (2000). *Knowledge and learning in natural language.* Doctoral dissertation, MIT.