21

Grammar Competition and Language Change

CHARLES D. YANG

21.1. A FORMAL APPROACH TO LANGUAGE CHANGE

The purpose of this chapter is to develop a general mathematical model that formalizes linguists' insight of grammar competition driving language change. As an empirical test, we apply the model to account for the loss of verb second (V2) in Old French.

A formal model of language change is to be understood in the following two senses. First, it is *causal*. As has long been recognized by linguists (Halle 1962, Chomsky and Halle 1968, Lightfoot 1979*a*, among others), the causal force of language change is language acquisition: language changes as a group phenomenon because individual learners acquire different grammars from their parents (Lightfoot, this volume). In addition, as children become parents, their linguistic production constitutes the primary linguistic evidence for the next generation of language learners. Following the traditional ideas of Klima (1965) and Anderson (1973), among others, this process can be viewed as an interactive exchange of I- and E-languages, mediated by language acquisition. Figure 21.1, adapted from Battye and Roberts (1995: introduction), illustrates this.

Hence, a causal model of language change must incorporate a theory of language acquisition, which in turn demands that diachronic explanations be compatible with what has been discovered about child language. When a linguistic change is observed, for example, the shift from one word order to another, the model must give an explicit account of how such a change took place, from independent perspectives of language learnability and language development.

Second, a formal model of language change must be *quantitative*. Here we may benefit from the (more mature) study of another sort of historical change, that of

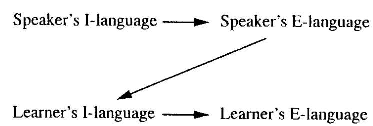

FIGURE 21.1. The dynamics of language acquisition and language change

biological evolution. The modern theory of evolution was not complete until the establishment of population genetic theories, which unified the Mendelian theory of genetics and the Darwinian theory of evolution. In population genetics, the fitness of Mendelian genes is explicitly measured and then directly fed into a quantitative model so that evolutionary changes (i.e. the changes in the distribution of genes in a population) can be predicted. It seems that the time is ripe for a similar synthesis in the study of language change. On the one hand, with the development of the Principles and Parameters framework and its refinement (Chomsky 1981, 1995), a 'genetic landscape' of language change—the Universal Grammar, the space in which language *could* vary and change—is being mapped out with increasing clarity and detail. On the other hand, recent diachronic studies in generative grammar have taken a quantitative spin: the rate and magnitude of language change can now be rather precisely measured from careful analysis of historical corpora. The next step, naturally, is to situate statistical measures of language change into a quantitative model that allows definite predictions. For example, much like predicting the direction of evolutionary change, one would like to say, when such and such statistical patterns are observed in historical text, such and such changes are bound to occur.

With these two goals in mind, we proceed to develop a formal model of language change. Our foundation is a model of language acquisition developed in a series of papers by the present author (see Yang 2000 for a summary). In particular, we propose that the idea of grammar competition, which has been a prominent topic in language-change research, appears to be the correct description of child language acquisition. Furthermore, the idea of grammar competition can be made precise in a formal model, which allows us to quantitatively relate the statistical properties of linguistic evidence to the grammatical knowledge that a learner attains. Section 21.2 gives a brief summary of the acquisition model. Section 21.3 lays out our model of language change in detail and derives a number of formal results. Section 21.4 puts the model into use, as we attempt to explain the loss of V2 in Old French (OF) based on the statistics reported in Roberts (1993*a*).

21.2. LANGUAGE ACQUISITION AS GRAMMAR COMPETITION

21.2.1. *Against triggering*

Our starting point is a particular view of Universal Grammar, the Principles and Parameters framework (Chomsky 1981), which broadly encompasses a number of grammatical formalisms. In the Principles and Parameters framework, the space of linguistic variations is defined by a set of parametric choices, interacting with universal principles. Consequently, acquiring a grammar is viewed as setting the parameters to the appropriate values.

An influential approach to parameter setting is the 'triggering model' (Gibson and Wexler 1994). In the triggering 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 literally 'triggered' (switched on and off) on the basis of relevant input sentences. However, this view of language-learning is empirically inadequate in two major ways.

First, it is by now well known that the triggering model has serious learnability problems. Both theoretical analysis (Berwick and Niyogi 1996, Frank and Kapur 1996, and Dresher 1999) and computer simulation in a linguistically realistic space (Kohl 1999) have demonstrated that the convergence to a target grammar cannot be guaranteed in the triggering model.

Second, and far more seriously, the developmental patterns of child language provide no support for the on-and-off triggering view of parameter-setting.[1] 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 and Wexler 1993) that the null-subject stage is due to missetting the subject parameter to the value of an optional subject grammar such as Italian or Chinese, and that 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). Furthermore, English children's subject drop is distributionally quite different from children acquiring true optional subject grammars such as Italian (Valian 1991) and Chinese (Wang *et al.* 1992), contrary to the proposal of parameter missetting and triggering. In fact, the gradualness of language development has been a major objection to the parameter-setting view of language acquisition (Valian 1991).

However, there needs not to be a conflict between the conception of grammar as discrete parameters and and the gradual development of children's language, once the study of language is situated in a broad biological framework. In what follows, we present a summary of such an approach developed in Yang (2000), which draws inspiration from the theories of population genetics.

21.2.2. *A variational approach to language*

As Ernst Mayr remarked in *Animal Species and Evolution* (1963) and many other places, a premise to understand biological evolution is the recognition of the uniqueness of individuals: variation among individuals is real, 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 distribution of variant individuals, modern population genetics can formally characterize the effect of evolutionary

[1] More generally, *any* theory of acquisition that posits all-or-none changes in the learner's grammar is faced with the same problem.

forces, thus bridging the gap between discrete variations at the individual genetic level and continuous evolutionary changes at the population level.

We may take the variational approach to the study of language acquisition and change. 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 a *single* grammar, as the triggering approach suggests. Rather, they may be reflections of principled hypotheses of language that children entertain before conclusively settling on the target language, an idea dating back to Jakobson (1941 [1968]). Language acquisition can then be viewed as a variational process in which the *distribution* of grammars changes as an adaptive response to the linguistic evidence in the environment: a model that implements this idea is presented in Section 21.2.3. Similarly, linguistic variations observed during historical change, which lead to the idea of grammar competition, can be viewed as distributional changes of grammars over successive generations of speakers: this will be elaborated in our model presented in Section 21.3.

21.2.3. *The acquisition model*

To account for the gradualness of language acquisition, we embrace the idea of statistical learning. Unlike models, say, connectionist networks, which assume statistical learning is all there is, we follow the Principles and Parameters framework and assume that there is a finite number of non-arbitrary grammars allowed by UG. These grammars form the learning space. Each grammar G_i is associated with a weight p_i, which denotes the probability with which the learner can access that grammar. Language-learning can be modeled schematically as follows:

(1) Upon the presentation of an input sentence *s*, the child
  (*a*) selects a grammar G_i with the probability p_i
  (*b*) analyzes *s* with G_i
  (*c*) • if successful, rewards G_i by increasing p_i
  • otherwise, punishes G_i by decreasing p_i

Metaphorically speaking, the learning hypotheses, the grammars given by UG, *compete*: grammars that succeed in analyzing a sentence are rewarded and those that fail are punished. As learning proceeds, grammars that over all have more success with the data will be more prominently represented in the learner's hypothesis space. At the same time, non-target grammars, despite their decreasing weights, still have non-zero probabilities of being accessed: this then explains the deviation in children's language from adults'. The gradualness of language development is also explained, as the rise of the target grammar is modeled by a gradually increasing variable, its weight/probability.

To quantify a competition-based learning model, we need to introduce some formal devices. In any competition process, some measure of fitness of individuals must be defined. A most straightforward definition is given in (2).

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

$$c_i = \Pr(G_i \nrightarrow s \mid s \in E)$$

where $s \in E$ means s is a possible sentence in E and $G_i \rightarrow s$ means s is analyzable (e.g. parsable) with G_i.

Note that penalty probability is an *intrinsic* property of a grammar relative to a linguistic environment, determined by the distribution of linguistic expressions used in that environment. 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.[2] Note also that the learner does not compute penalty probabilities, which are merely formal devices used to study the dynamics of language-learning (and later, language change). However, these fitness measures can be empirically estimated from text corpora, making quantitative predictions on the rate of language acquisition and the direction of language change.

It is not important here to give the precise realization of (1). While we are committed to the view that models language acquisition as grammar competition, we know virtually nothing about the exact manner in which children update grammar weights in (1).[3] For concreteness, we adopt a classic learning model from mathematical psychology, the Linear Reward–Punishment (L_{R-P}) scheme (Bush and Mosteller 1958), which is supported by substantial evidence from the psychology of learning (see Atkinson *et al.* 1965 for a summary).

For simplicity but without loss of generality, suppose that there are two grammars in the population, G_1 and G_2, which are associated with penalty probabilities c_1 and c_2 respectively. It is possible to show (see Narendra and Thathachar 1989: 162–5) that the weights of the two grammars converge to the following values:

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

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

(3) says that when learning stops, 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, in some linear combination of their respective penalty probabilities. It is worth pointing out that no matter what initial values p's may assume, the L_{R-P} model guarantees that, when learning stops, a multiple-grammar combination will be formed.

[2] About 70% of Dutch, German, Norwegian, and Swedish sentences have the surface order of SVO, according to corpus analysis cited in Lightfoot (1997).

[3] As a result, much of the acquisition evidence assembled in Yang (2000) attempts to demonstrate the reality of grammar competition, by showing the presence of coexisting grammars before the learner converges on the target grammar, and to make *relative* comparisons of the rate of cross-linguistic development.

The result in (3) has two desirable properties, which justify the choice of the L_{R-P} as our acquisition model. First, it is obvious that a unique target, in an idealized homogeneous learning environment, will eliminate all competing grammars, by the virtue of having the penalty probability of 0. Thus, the acquisition model meets the standard learnability condition, an improvement over the standard triggering model. Second, and more relevant to language change, it follows from (3) that in a non-homogeneous linguistic environment where no single grammar is compatible with every input sentence, the learner converges to a stable combination of two grammars. This is a welcome result. Statistical studies of historical texts, particular those by Kroch and his colleagues (Kroch 1989, Pintzuk 1999, Santorini 1992, Kroch and Taylor 1997), reveal that speakers during the transitional stages of language change are best modeled as a combination of more than one grammars. The L_{R-P} model gives a straightforward accommodation for these findings.

The reader may consult Yang (2000) for a formal analysis of the L_{R-P} model as applied to the Principles and Parameters framework, along with evidence from child language (including syntax and morphophonology) to demonstrate its utility as a general model of language acquisition. In what follows, we will extend the acquisition model to the study of language change.

21.3. GRAMMAR COMPETITION AND LANGUAGE CHANGE

21.3.1. *The role of linguistic evidence*

With the acquisition model in hand, we return to the problem of language change. The fundamental question is, again, What makes a generation attain a language different from the previous generations?

We will assume that language change cannot take place without sufficiently different linguistic evidence across generations. This assumption requires some justification. 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*. Indeed, this is the approach taken by Niyogi and Berwick (1995) in their model of language change based on Gibson and Wexler's (1994) triggering model. However, this position is empirically untenable in three ways. First, 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 syntactic parameter is misset to a wrong value). Second, historical linguistics shows that language change occurs on the scale of the entire population, not on that of scattered individual members, as Bloomfield (1927, cited in Hockett 1968) remarks:

> 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, say, the rigidity of word order in Western 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 for the previous generation? There are many possibilities. For example, migration of foreign speakers might introduce novel grammatical constructions that were previously unseen; social and cultural factors might also obscure the distributional patterns of linguistic expressions used in a population. These are interesting and important topics of research, but are not relevant for a formal model of language change.[4] We are chiefly concerned with the *predictable consequences* of the change in linguistic evidence: what happens to language learners when their linguistic evidence is altered, and how does it affect the composition of the linguistic population over time?[5]

21.3.2. *A formal model of language change*

Suppose that, resulting from migration, genuine innovation, and other sociological and historical factors, a linguistic environment is established for a generation of language learners that is substantially different 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: two idealized grammars G_1 and G_2. Further, 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 α (β) the *advantage* of G_1 (G_2). Figure 21.2 illustrates this.

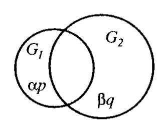

FIGURE 21.2. Two mutually incompatible grammars constitute a heterogeneous linguistic environment

[4] In other words, the approach taken here is based on the principle of parsimony: a model that explains a problem of language change without appealing to social and cultural factors is superior to one that does appeal to such factors. We will show that the model proposed here suffices for the case of the loss of V2 in French, invoking only an independently motivated model of UG and language acquisition.

[5] A similar scenario arises in the population-genetic theory of evolution, which concerns the predictable outcome in a population once some new genes are introduced. The precise manner in which new genes arise, which could be mutation, migration, etc., is a separate question, which usually involves a good deal of (unpredictable) historical contingencies.

Recall that when neither of the two grammars is wholly compatible with the linguistic evidence, the $L_{R\text{-}P}$ model converges to a combination of them, with the weights determined by their penalty probabilities. We repeat the theorem in (4) below:

(4)
$$\lim_{t\to\infty} p_1(t) = \frac{c_2}{c_1 + c_2}$$
$$\lim_{t\to\infty} p_2(t) = \frac{c_1}{c_1 + c_2}$$

Suppose that at generation n, the linguistic environment $E_{G_1,G_2}=pG_1 + qG_2$, where $p + q=1$. That is, in E_{G_1,G_2}, a proportion p of expressions are generated by G_1, and a proportion q of expressions are generated by G_2, and they collectively make up the linguistic evidence for the learners in generation $n+1$. For 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 (4) allow us to compute p' and q', the weights of G_1 and G_2 respectively, that are internalized in the learners of generation $n+1$:

(5) The dynamics of a two-grammar system:

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

(5) 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 from the parental generation n. (5) further leads to:

(6)
$$\frac{p'}{q'} = \frac{\alpha p/(\alpha p + \beta q)}{\beta q/(\alpha p + \beta q)} = \frac{\alpha p}{\beta q}$$

In order for G_2 to overtake G_1, the weight of G_2 (q) 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 $q' > q$, i.e. $p'/q' < p/q$. From (6), some simple algebra yields a sufficient and necessary condition for language change in a linguistic population:

(7) *The fundamental theorem of language change*
G_2 replaces G_1 if $\beta>\alpha$: the *advantage* of G_2 is greater than that of G_1.

Note that under this model, language change requires the input to be critically altered just once, e.g. speakers of a new language move into a local population, influence a generation of learners, and then move out. If the condition in (3) is met,

the new grammar retained by the younger generation will result in a self-sustaining cycle of production and learning (cf. Figure 21.1), and eventually drive the old grammar to extinction.

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 directly α and β from historical texts, which only reflect the penalty probabilities of competing grammars, i.e. at generation n, p, and q, and at generation $n+1$, $p'=\alpha p$ and $q'=\beta q$. However, (7) says that *if* $q'>q$ (G_2 is on the rise), it must be the case that $\beta>\alpha$. And, if $\beta>\alpha$, G_2 will necessarily replace G_1. Thus, we have the following corollary:

(8) **Corollary**. Once a grammar is on the rise, as seen in the increasing number of sentences incompatible with its competitor, it is unstoppable.

Plotting the $q(t)$, the weight of G_2, as a function of time t, we derive the familiar S-shape curve (see Figure 21.3) that has often been observed in language change (Weinreich *et al.* 1968, Bailey 1973, Kroch 1989, Keenan 1998, among others), as the 'new' linguistic form gradually replacing the 'old' form (Figure 21.3).[6]

The present model shares an important feature with Clark and Roberts's (1993) work, which extends the use of Genetic Algorithms in acquisition (Clark 1992). In both models, the outcome of language acquisition is determined by the compatibilities of grammars with linguistic evidence, in a Darwinian selectionist manner. However, they identify the final state of acquisition with a single grammar.[7] Therefore, when the linguistic evidence does not unambiguously identify a single

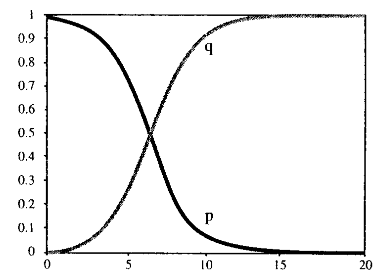

FIGURE 21.3. One language (q) replacing another (p) over time

[6] As pointed out by an anonymous reviewer, while the S-shape curve seems to be the most common pattern in language change, there are exceptions. For example, Kroch (1989: 229ff.) shows that between 1500 and 1575, the frequency of unemphatic *do* in English rose from 1.37% to 8% but dropped back to a negligible level in another seventy-five years. The reader is referred to Kroch's paper for details, as well as his analysis that tries to reconcile this exceptional pattern with the general S-shape curve in language change.

grammar, they posit some general constraints on the learner, e.g. the 'elegance' condition, which requires the learner to select the simplest among conflicting grammars. Aside from the fact that these auxiliary assumptions would require independent justification from child-language research, the position of a learner converging to a single grammar cannot be defended in face of the empirical evidence found by Kroch and his colleagues (Kroch 1989, among others). They have shown that historical texts during the period of language change must be analyzed as a combination of multiple grammars. In fact, historical linguists commonly use terms such as 'erosion' or 'optionality' to indicate the gradual appearance of a grammatical construction. These facts, and more generally, linguistic variability of the sort noted by Weinreich *et al.* (1968), Labov (1969), and numerous others in sociolinguistic research, can straightforwardly be modeled as coexistence of multiple grammars in the approach taken here.

For the purpose of this chapter, we have assumed 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 spatial and temporal factors into the dynamics of language change, which can be aided by the well-established models of population genetics and evolutionary ecology. We leave these options for further research.

To summarize the theoretical considerations in this section, we have extended the variational model of language acquisition to a population of learners and derived some analytical results concerning the dynamical system thus construed. We conclude that conflicting 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. To demonstrate its utility, we apply the model to the loss of V2 in the history of French.

21.4. THE LOSS OF V2 IN FRENCH

Old French had a cluster of properties, including V2 and *pro*-drop, which are lost in Modern French (ModF).[8] The following examples are taken from Clark and Roberts (1993):

[7] The same is true for Niyogi and Berwick (1995), another formal and acquisition-based model of language change. In their model, some speakers converge to one grammar and others converge to another; language change is viewed as the change in the *proportion* of speakers with one grammar vs. the other. This does not appear to be an empirically tenable position, as discussed in the text.

[8] As documented by Roberts (1993*a*) and others, null subjects in OF were more restricted than in Spanish and Italian, *pro*-drop par excellence. In particular, while null subjects were quite free in matrix clauses, they were not possible in embedded clauses. This peculiarity need not concern us. As will become clear in a moment, we are interested in the impact of null subjects on the loss of V2, a matrix phenomenon.

(9) *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.

(10) *Loss of V2*
(*a*) *Puis entendirent-ils un coup de tonnerre. (*ModF*)
then heard-they a clap of thunder.
(*b*) Lors oïrent ils venir un escroiz de tonoire. (*OF*)
then heard they come a clap of thunder.

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

Recall that in order for a ModF SVO grammar to overtake a V2 grammar, it is required that the SVO grammar has 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. (11) shows the advantage patterns of V2 over SVO, and vice versa:[9]

(11) (*a*) *Advantage of V2 grammar over SVO grammar*
V2 $\rightarrow$ s but SVO $\not\rightarrow$ s: VS (XVSO, OVS).
(*b*) *Advantage of SVO grammar over V2 grammar*
SVO $\rightarrow$ s but V2 $\not\rightarrow$ s: V>2 (SXVO, XSVO).

If the distribution patterns in modern V2 languages are indicative of those in the past, 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 are restricted to a small number of constructions, and are in general quite rare:

(12) *Rare V>2 patterns in modern German*
denn Johann hat gestern das Buch gelesen.
so Johann has 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. Our own counts based on a Dutch sample of adult-to-child speech are similar: 66.8% SVO, 23% XVSO, and 1.2% OVS (see Yang 2000 for detail). In contrast, based on the Penn Treebank, an annotated corpus of modern English, we found that only about 10% of all declarative sentences have V>2 word order:

[9] Because of the availability of the relevant corpus statistics, we follow Lightfoot (1991) to consider only degree-0 sentences as linguistic input, although nothing hinges on this assumption.

(13) *V>2 patterns in modern English*
  (*a*) He always reads newpapers 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 on earth did Old French lose it? The reason, on our view, is that OF was 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: a null-subject SVO grammar (like modern Italian) can analyze such patterns as [X–(*pro*)–V]. (14) shows the prevalence of subject drop in early Middle French (MFr):

(14)

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

(Roberts 1993*a*: 155)

The 30% advantage in non-*pro*-drop V2 languages has been reduced to 5–18% in the *pro*-drop MFr. At the same time, V>2 patterns have gone from fairly sparse (about <5%) in OF (Roberts 1993*a*: 95) to 11–15% in early MFr, as the class of sentence-initial XPs that do not trigger SV inversion was expanded (Vance 1989). (15) shows some representative examples:

(15) *V>2 patterns in early MFr*
  (*a*) Lors la royne *fist* Santré appeller.
  then the queen made Santré to call
  'Then the queen had Saintré called.'
  (*b*) Et a ces parolles le roy *demanda* quelz prieres ilz faisaient.
  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 . . .'

(16), which is based on the examination of the three texts in (14), shows the frequency of V>2 patterns in MFr:

(16)

Text	V>2
Froissart, *Chroniques* (*c.*1390)	12% (of 73)
15 Joyes (*14esme Joye*) (*c.*1400)	15% (of 40)
Chartier *Quadrilogue* (1422)	11% (of 45)

(Roberts 1993*a*: 148)

Comparing (16) with (14), we see that at the early MFr stage, there were more V>2 sentences than VS sentences, due to the effect of subject drop. Thus, following the corollary in (8), it must be the case that an SVO grammar (+*pro*-drop) has an advantage over an OF V2 grammar (+*pro*-drop). V2 in French was then destined to extinction, as predicted.

Our analysis of the loss of V2 in French crucially relies on the fact that null subjects were lost *after* V2 was lost. Roberts (1993*a*) shows that this was indeed the case. In the late fifteenth century and early sixteenth century, when SVO orders had already become 'favored', there was still significant use of null subjects, as the statistics in (17) demonstrate:

(17)

	SV	VS	NullS
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%

(Roberts 1993*a*: 155, 199)

Overall, the mean figures for the relevant patterns are shown below (Roberts 1993*a*: 199):

(18)

	SV	VS	NullS
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 sixteenth century, when V2 almost completely evaporated, there was still a considerable amount of subject drop.[10] This diachronic pattern is consistent with our explanation for the loss of V2 in Old French.

We believe that the present analysis may be extended to other Western European Romance languages, which, as is well known, had V2 in the medieval times. Under the present model, it is no accident that all those languages at one time had *pro*-drop, as in Old French, and many still do, as in Italian, Spanish, etc. It appears that the combination of *pro*-drop and V2 are *intrinsically unstable*, and will necessarily give way to a SVO (+*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) 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%. Now this is a figure already

[10] See Vance (1989, 1997) for a discussion of how the diminishing V2 constraint resulted in a new system of *pro*-drop.

lower than the approximately 10% of $V>2$ sentences by which an SVO grammar has an advantage over a V2 grammar, which would lead to the demise of V2.[11]

21.5. 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 independent observations: (i) the deviation of child language from adult language is not simply noise or imperfection—it is the reflection of competing grammars in changing distributions, and (ii) multiple grammars are evidenced in mature speakers during the course of language change. It is important to recognize that while sociological and other external forces clearly affect the composition of linguistic evidence, language acquisition via grammar competition is an I-language process, which occurs internally to the individual learner's mind/brain.

The model not only formalizes historical linguists' insight of grammar competition, but also quantitatively relates the statistical properties of historical texts (hence acquisition evidence) to the direction of language change. We hope that the present model will contribute to a formal framework in which problems of language change can be stated precisely and studied quantitatively.

[11] *Pro*-drop can only help knock out V2 if it is *abundant*. Plenty of VS sentences can then be reduced to [X–V–(*pro*)], which can be analyzed as [X–(*pro*)–V], leading to the rise of the SVO grammar. This requirement was met in medieval Western European Romance languages, which had free *pro*-drop in the matrix clauses. In contrast, languages may retain V2 if they had more restrictive and less frequent subject drop such that the critical point of V > 2 patterns outnumbering XV was never reached. Old Icelandic (Sigurðsson 1993) and Surselvan (Sprouse and Vance 1999), both of which had V2 and limited subject drop, and still do, seem to confirm the expectation. I am indebted to an anonymous reviewer for pointing out the relevance of these languages to the present model.