D.E.L.T.A., vo. 8, n.º Especial, 1992 (53-103)

A COMPUTATIONAL MODEL OF LANGUAGE LEARNABILITY AND LANGUAGE CHANGE

Robin Clark (Université de Genève)*
Ian Roberts (University of Bangor)

ABSTRACT: O presente artigo desenvolve a idéia de que a mudança lingüística é guiada pela reestruturação da gramática-alvo, durante a aquisição da linguagem pela criança. A idéia é formalizada nos termos do modelo gerativista de Princípios e Parâmetros. Nesse quadro teórico, a aquisição é vista como um processo de fixação de valores paramétricos e a mudança lingüística ocorre quando a população converge de um valor v para ao menos um parâmetro P enquanto a gramática adulta apresenta P e v. Ainda que a idéia básica seja muito atraente, há uma questão importante sem resposta: como e por que o valor "adulto" de P pode ser desconsiderado pela população que adquire a língua? Este artigo tenta responder a essa questão, a partir dos trabalhos de Clark (1990a,b,c) sobre a aplicação de algorítmos genéticos na fixação paramétrica e a análise das mudanças paramétricas sofridas pelo francês no séc. XVI, proposta por Roberts (a sair).

1. Introduction

Darwin's (1859) theory of natural selection had an important influence on the Neogrammarians. As was the case with Darwin, they believed that diachronic change was the result of selective pressures on organisms from the environment operating on random variation within a population (see Haldane, 1990 for a classic exposition of natural selection as the motive force underlying evolution). While Darwin proposed that natural selection was accounted for by the greater reproduction rates of fitter organisms, Paul (1920) proposed that language change is driven by restructuring of the target grammar that may take place during language acquisition. If the input to language acquisition is taken to be the environment and language acquisition is taken to be the linguistic correlate of biological reproduction, a clear parallelism between Darwin's view of natural selection and Paul's view of the selection of grammars emerges. Despite the appeal of this notion, no successful evolutionary theory of the relationship between language acquisition and language change has been developed in the 130 years since Darwin's <u>On the Origin of Species</u>. The purpose of this paper is to relate natural selection, language acquisition and language change in light of current computational models of learning.

The basic problem for the hypothesis that language change is driven by acquisition concerns the relationship between the adult input, which is generated by one grammar, and the learner's hypotheses, which differ at certain points from the adult grammar. Under the appropriate idealizations, acquisition is a process of accurately fixing parametric values; in other words, the learner sets parameter p^n to the value v_i where the target grammar has p_n set to v_i. Language change, on the other hand, presupposes that a population must converge on a value v_i for at least one parameter, p, where the adult grammar has $p(v_2)$ and $v_1 \neq v_2$. We will show that this idea can be formalized in light of current thinking on language learnability and that doing so sheds light both on the processes that underlie diachronic change and on those that drive learning. The problem here is as much one for the theory of language acquisition as it is one for the theory of diachronic change[1]. A central problem for acquisition theory is that of characterizing how the learner formulates and retracts hypotheses in light of its linguistic environment. Equally, one of the central problems for language change concerns how a population of learners can converge on a grammar that is systematically different from the adult grammar in the sense defined above. In both cases, hypothesis formation and retraction by learners appears to be a crucial mechanism.

We will adopt the <u>genetic algorithm</u> (GA) approach to learnability developed in Clark (1990)[2]. This approach treats the learning problem as a problem in population genetics. Roughly, parameter settings are taken as phenotypes which may be expressed by parsing devices. The parsers are run against an input text and their relative fitness is measured by a simple metric. Those hypotheses which are judged most fit are then combined via a special mating operation. This technique allows the learner to search the hypothesis space efficiently while optimizing the learner's computational resources.

Crucially, the GA technique presupposes that the input text expresses each parameter with sufficient frequency that the learner's hypotheses are placed under pressure to bear that parameter setting. Hypotheses that carry a parameter value corresponding to a parameter setting frequently expressed in the input text will be strongly selected for by the fitness metric. As a result, hypotheses containing ''favorable'' parameter settings will tend to reproduce more frequently, while the ''unfavorable'' setting will disappear from the population. If, on the other hand, a parameter is not expressed frequently in the input text, the learner will be under less pressure to set that parameter in accordance with the target setting. In this case, the fitness metric environment. The fitness metric plays a crucial role in mediating between the learner and the input text.

We will propose that parametric change occurs when the target of acquisition contains parameter values which cannot be uniquely determined on the basis of the linguistic environment. This can occur when the evidence presented to the learner is formally compatible with a number of different, and conflicting, parameter settings. In these cases, the learner must evaluate its hypotheses using criteria that are not purely a response to the external environment; in particular, the learner must consider factors like the Subset Condition (Berwick, 1985) and elegance of derivations (the Least Effort Principle; Chomsky, 1989). Thus, the consideration of language change from a learnability perspective gives us access to how learners evaluate the relative merit of their hypotheses. Our goal here will be to characterize, in a precise manner, the conditions under which a learner arrives at a grammar distinct from the target, thus fueling diachronic change.

Intuitively, our argument will be that, due to various factors, the input data does not put pressure on the learner to set certain parameters to a definite value; several alternative grammars can adequately account for the input stream; the appropriate choice of grammar is underdetermined by the linguistic environment, even given the learner's rich internal structure. Since external pressures do not force the learner to select a particular grammar, it will turn in on itself, abandoning external pressure, and rely on grammar internal properties, to select from the alternatives at hand. If this is correct, then diachronic change can provide crucial information on those factors that learners rely on to select hypotheses. Since the external environment is not decisive in these cases, diachronic change reflects pure learnability considerations. Since diachronic change reflects what is, in some sense, ''pathological'' learning, a careful study of its properties can reveal a great deal about how learning transpires in non-pathological cases (a similar idea is developed for phonological change by Kiparsky, 1982).

We will argue that parametric change can involve a variety of factors. Change in one component, for example the phonology, can obscure syntactic parameter expression. The resulting text will not uniquely drive the learner toward the target. At this point, the learner appeals to the fitness metric to select an appropriate parameter setting and factors such as the Subset Condition or general economy of representations come into play rather than pure selective pressure from the input text. This type of change is exemplified by the introduction of subject clitics in 15th-century French. A second important factor is instability due to independent parametric changes within a component; change in one parameter setting can trigger a number of changes to other parameter settings. As we shall show, parametric change in 16th-century French provides a case study on how parametric change can cascade through a system (see Roberts, forthcoming). During this period, French ceased to be both a null subject language and a verb second language. We will show that, due to innovations in the 15th-century, the system became unstable and deep parametric change was forced on the learner via the fitness metric.

Fundamental to this analysis is the formalization of the notion of stability relative to a particular parameter-setting: A parameter-setting is stable to the degree that its expression in the input data is unambiguous. Following Clark (1990), we will say that a parameter value, $p(v_j)$ is expressed by an input sentence, s_i, just in case a grammar must have p set to value v_j in order to assigned a well-formed representation to s_i. An unstable parameter-setting, then, is one whose expression is ambiguous. We will show that, through a variety of independent changes, 16th-century French became highly unstable, resulting in the loss of null subjects and verb second phenomena.

2. Genetic Algorithms and Language Learnability

The basic problem faced by a language learner is to discover a target grammar based on an input text consisting of short, simple, grammatical sentences[3]. A principles and parameters (P&P, see Chomsky, 1981) approach to grammar provides a powerful way of limiting the problem of discovering the appropriate target grammar given the impoverished nature of the input data. Parameters can be viewed as finite vectors along which natural languages may vary; the learner is faced with the problem of searching a finite space of possible grammars rather than the more difficult problem of inducing a set of rules which lies at an undetermined point in an infinite hypothesis space.

Formally, the learning problem can be characterized by the following (see Clark, 1990 for a more detailed exposition):

$$(1)\gamma[\emptyset_n(\varphi(\sigma_i))] = P_m$$

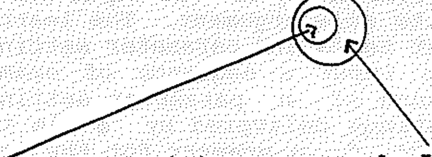
In (1), the input text is represented by σ. The input text is taken by the learner, φ , as its sole argument. The learner, φ , is a relation between input texts and parameter settings; in other words, φ takes as input a sequence of example sentences and returns a sequence of n parameter settings. The sequence of parameter settings delivered by φ is then mapped by the function $\emptyset_n$ onto G_i, a grammar for the input text, σ_i . Once a grammar is delivered by $\emptyset_n$, it can be used to determine a parser, P_m, for the grammar Gi. Roughly, γ can be thought of as a compiler which takes high-level grammars and implements them as parsing machines. The learning problem for natural languages can then be thought of as a relation between input texts and parsing devices.

In considering the learning problem, it is important to recall that the learner is computationally bounded. In other words, the learner has finite resources in terms of time and memory. It cannot take indefinite periods of time before converging to the target grammar nor does it have a perfect memory for past sequences in the input text or past (unsuccessful) hypotheses. Furthermore, the learner is given little information about the proper analysis to be accorded to the input data. It has only limited information about the proper structural analysis for any given datum[4] , and little to no access to input which is ill-formed with respect to the target.

The claim that the hypothesis space, under a P&P approach, is finite is not, in itself, sufficient to guarantee that the learner can converge in a reasonable amount of time. Finite problems can be sufficiently large that their solution might take an impractical amount of time to compute. Suppose, for example, that the hypothesis space is determined by 30 binary parameters. In this case, there are 2^{30}, or 1,073,741,824, possible grammars. If the learner could test each of these grammars at the rate of one per second, it might in the worst case take the learner over thirty-four years to converge on the target. Clearly, the learner must be capable of searching the hypothesis space in a more efficient manner.

Beyond efficiency considerations, it is clear that the learner cannot use a brute-force search technique to converge on the target since certain parameters may fall into subset relations; that is, the language that results when a certain parameter, p_x, is set to 0 is a proper subset of the language that results when p_x is set to 1:

(2) $L\ [p_1, ..., p_{x-1}, p_x(0), p_{x+1}, ..., p_z] \subset$
$L\ [p_1, ..., p_{x-1}, p_x(1), p_{x+1}, ..., p_z]$



$L[p_1, ..., p_{x-1}, p_x(0), p_{x+1}, ..., p_z] \subset L[p_1, ..., p_{x-1}, p_x(1), p_{x+1}, ..., p_z]$

All the sentences that are grammatical in the subset will also be grammatical in the superset language. If the learner guesses the superset language, then no further evidence will contradict its hypothesis. Thus, the learner will never have grounds to retract this (incorrect) hypothesis. Thus, the learner must guess the minimal language compatible with the input sequence σ_i . Given that the learner has no reliable access to negative evidence, it appears that the learner must guess the smallest possible language compatible with the input at each step of the learning procedure. This is, in essence, the Subset Condition of Berwick (1985) which is intended to circumvent the sort of trap posed by subset parameters.

A further possibility arises if we consider that sets of parameters might interact in such a way as to generate superset languages. That is, when considered individually the parameters in question may not necessarily generate superset languages, but when they act in a group they do generate a superset language. This is the <u>shifting</u> relation of Clark (1990)[5] :

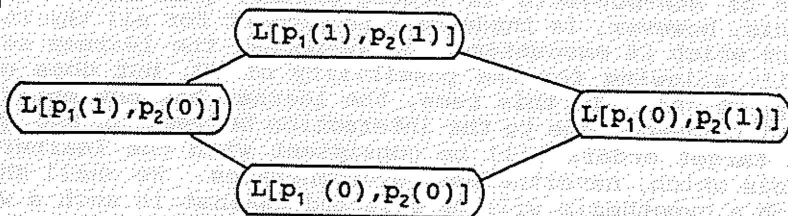
(3) Shifting
Two parameters, x_i and x_j, cause a <u>shift</u> at values $x_i(1)$ and $x_j(1)$ just in case:

(a) $L[\ \emptyset_n\ (x_1, ..., x_i(1), ..., x_j(0), ..., x_n)] \not\subset$
$L[\ \emptyset n\ (x_1, ..., x_i(0), ..., x_j(1), ..., x_n)]$
(b) $L[\ \emptyset n\ (x_1, ..., x_i(0), ..., x_j(1), ..., x_n)] \not\subset$
$L[\ \emptyset n\ (x_1, ..., x_i(1), ..., x_j(0), ..., x_n)]$
(c) $L[\ \emptyset n\ (x_1, ..., x_i(1), ..., x_j(0), ..., x_n)] \subset$
$L[\ \emptyset n\ (x_1, ..., x_i(1), ..., x_j(1), ..., x_n)]$
(d) $L[\ \emptyset n\ (x_1, ..., x_i(0), ..., x_j(1), ..., x_n)] \subset$
$L[\ \emptyset n\ (x_1, ..., x_i(1), ..., x_j(1), ..., x_n)]$

In other words, a shift occurs given two parameters which generate superset languages when they are both set to some particular value. Notice, crucially, that if the language generated by setting x_i to 0 is a subset of the language generated by setting x_i to 1, this relationship is preserved in the shifted language. In brief, a learner could obey the Subset Condition on the microscopic level (with respect to a single parameter) while violating it on the macroscopic level (due to shifting interactions between parameters). In order for the learner to avoid these higher-level violations of the Subset Condition, it would have to calculate interactions between parameter settings. But this would become increasingly difficult as the number of parameters that could ''conspire'' to generate a shifted language increased; given n parameters, the learner may have to consider n! possible interactions.

The graph in (4) illustrates a case of shifting involving superset parameters. In this example, we have two parameters p_1 and p_2 which interact to generate a shifted language, $L[p_1(1),p_2(1)]$ In the following graph, dominance indicates the subset/superset relation:

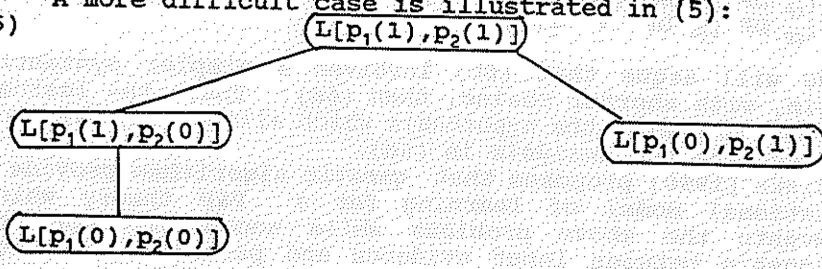
(4)



In this case, both p_1 and p_2 are superset parameters; any language with p_1 set to '0' is a subset of a language with p_1 set to '1' and any language with p_2 set to '0' is a subset of a language with p_2 set to '1'. Note that $L[p_1(1),p_2(0)]$ and $L[p_1(0),p_2(1)]$ are not in the superset relation with each other. The language, $L[p_1(1),p_2(1)]$, however, properly contains the other three possible options. As we shall see, below, the learner will be reluctant to posit the language $L[p_1(1),p_2(1)]$ and will only do so if faced with a significant amount of empirical prodding in the form of failed parses.

A more difficult case is illustrated in (5):

(5)



In this case, only one of the parameters, p_1, is a superset parameter. One might imagine that p_1 regulates the option of having left-dislocation of a constituent. The parameter, p_2, does not generate languages in the superset relation. For example, one might take p_2 to be a parameter which regulates verb-second phenomena in matrix clauses. Suppose that p_1 and p_2 interact in such a way that, when both are set to '1', the language allows left-dislocation of a constituent over the V2 structure of the root clause; the resulting language has all of the normal V2 orders plus clauses with an additional constituent left-dislocated before the normal V2 order. Such a language would be a shifted language.

Take the case where the target language is V2 without left-dislocation. Suppose that the learner, during an early phase of the learning cycle, erroneously set p_1 to '1', allowing left-dislocation of a constituent in response to the presence of non-subjects in clause-initial position. This hypothesis, however, is inadequate to account for all the root V2 orders which it encounters. In response, the learner sets p_2 to '1', allowing for the possibility of V2, but does not reset p_1 to '0'. In this case, the learner has now entered a shifted language; due to the interaction between p_1 and p_2, all the target orders will be consistent with the learner's hypothesis which, nevertheless, overgenerates. We shall show that such a hypothesis will be selected against in such a way that the learner can retract its overgeneral hypothesis without access to direct negative evidence. Such a shifted language, although a possibility empirically, will tend to be unstable diachronically with loss either of V2 or of left-dislocation. Notice that a learner will have two analyses available for ''V3'' structures (structures with two constituents before the tensed verb); either such a structure involves left-dislocation with a standard V2 structure, as in (6a) or the structure involves simple left-dislocation, as in (6b):

(6)a. $[_{CP}$ DP $[_{CP}$ DP $[_{C'}$ V $[_{IP}$...]]]]
b. $[_{CP}$ DP $[_{IP}$ DP V ...]]

We will argue that (6b) involves a simpler syntactic analysis, with shorter chains, than (6a). Thus, the learner will tend to prefer the hypothesis which allows (6b) over one which requires the analysis in (6a).

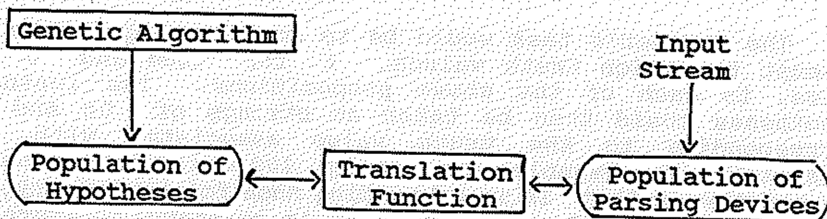
Clark (1990) proposes that genetic algorithms provide a computational model of learning for a P&P theory which circumvents the above problems while accounting for the relationship between input evidence and parameter setting.

Genetic algorithms mimic natural selection by representing hypotheses about a problem in a way that is similar to the way in which genetic material is represented. Hypotheses are then tested against the problem space with the most fit hypotheses contributing to the formation of new hypotheses via reproduction (the combination of pre-existing hypotheses to form new hypotheses in a way that is similar to the biological recombination of DNA present in mating). By ''breeding'' the most fit hypotheses, testing them against the problem space and pruning the least fit, a genetic algorithm can efficiently search large spaces and find optimal solutions[6].

A genetic algorithm consists of the following components:

- A representation of hypotheses in terms of strings, similar in structure to genetic material.
- A crossover mechanism. This mechanism combines two hypotheses and produces a new hypothesis by combining parts of each to the parent's genetic material.
- A mutation operator. This mechanism randomly alters an offspring's genotype to produce a new hypothesis close to, but not identical with, the parent's genetic endowment.
- A measure of fitness of hypotheses in terms of their performance in an environment.
- A reproductive mechanism which allows a hypothesis to produce offspring; in general, the more fit a hypothesis is, the greater the likelihood that it will reproduce.

Most crucial for our purposes are the representation of hypotheses in terms of strings and the notion of a fitness metric. Crucially, we must be capable of mapping between our problem space and a representation of that hypothesis space in terms of strings. Fitness will be measured in terms of the performance of parsers relative to a stream of input data; the actual algorithm will operate on the string representation of the hypotheses. We must, then, have a translation function which relates our hypotheses (strings) to the parsers that they represent:

(7)


We will conceive of the learner, ϕ , as operating on strings of parameter settings; the translation function in (7) then maps the learner's hypothesis strings onto parsing devices; thus, the translation function is comparable to the functions $\emptyset_n$, which maps sequences of parameter settings onto grammars, and γ , which maps grammars onto parsers. In a sense, the hypothesis strings represent genotypes for parsing devices while the translation function ($\emptyset_n$ and γ) maps genotypes onto phenotypes. These are then tested against the linguistic environment for relative fitness and the results are used as a basis for forming new hypotheses.

Let us turn to a more careful consideration of the representation of hypotheses. It is common to think of parameters as variables in Universal Grammar which range over a limited set of values. The bounding nodes for classical subjacency (Chomsky, 1977) provide a good example of such a parameter. Here subjacency is taken as an invariant property of natural languages while the bounding nodes may be contingently selected from a restricted set:

(8) Subjacency
No rule may involve X and Y in the configuration:
$...X...[_{\alpha}...[_{\beta}...Y..._{\beta}]..._{\alpha}]...$
(order irrelevant)
where α and β are bounding nodes;
$\alpha, \beta \in \{NP, IP, CP\}$.

Parameters can equally be viewed as variant properties of natural languages; in other words, we can think of a parameter as a descriptive statement which may be either true or false of a given grammatical system. From this perspective, we could rewrite the parameter for the bounding nodes as a series of three statements:

(9) a. IP is a bounding node for subjacency.
b. CP is a bounding node for subjacency.
c. NP is a bounding node for subjacency.

The learner's task would be to scan the input data and attempt to assign truth values, 1 for true and 0 for false, to each of the above propositions. The learner's hypotheses could then be taken as strings of 0s and 1s corresponding to the truth value associated with each parameter. For example, the string 100 could correspond to the hypothesis that IP is a bounding node for subjacency, but neither CP nor NP are. Thus, it is relatively natural to represent parameter settings in terms of strings.

The crossover operator combines two hypothesis strings to create new hypotheses. For example, suppose that the two hypotheses below have been selected for reproduction:

(10)a. 000111
b. 101000

Suppose we ``cut'' both strings after the third position in the bit string:

(11)a. 000 - 111
b. 101 - 000

The first part of string (a) is then recombined with the second part of string (b) and the first part of string (b) is recombined with the second part of string (a):

(12)a. 000 - 000
b. 101 - 111

And, thus two new ``offspring'' hypotheses which have inherited genetic material (hypotheses about settings of particular parameters) from each parent are created. It should be noted that fitness interacts in a crucial way with the crossover operation. Highly fit hypotheses are more likely to be selected to take part in crossover and, therefore, are more likely to pass the parameter settings which made them fit on to new generations of hypotheses.

The mutation operator similarly creates new hypotheses on the basis of existing ones. In essence, it must slightly alter a hypothesis string in order to create a new, but ``nearby'', hypothesis. We can do this by flipping a randomly selected bit position in a hypothesis string by the following rules:

(13)a. 0 -> 1
b. 1 -> 0

Thus, selecting the second position of the following hypothesis for mutation would yield a ``mutant'' which is nearly identical to its parent structure:

(14) 000111 -> 010111

The mutation operation can be viewed as a means of searching the immediate hypothesis space surrounding a parameter string. Thus, the learner can, in a sense, experiment with near-optimal hypotheses which approximate, but do not correspond to, the target.

In terms of an actual parsing framework, there would be a fixed central algorithm, corresponding to UG. Within this algorithm would be various flags, indicating points where code must be inserted for the parser to function. The `0's and `1's in the hypothesis strings could be interpreted as pointers to the parameterized code. Upon receiving an hypothesis string, the machine would look up the various pieces of code indicated by the '0's and '1's and systematically substitute the code it finds for the flags in the main algorithm. The result would be a special parsing device designed to analyze the language enumerated by the hypothesis string. Thus, a ''self-constructing'' parser would be the ensemble of the core algorithm, the parameterized code and a learning device which would select the appropriate hypothesis string in response to the input text. We then have a straightforward model of the translation function required by the genetic algorithm to relate hypothesis strings to parsing devices. Recall that this translation function, itself, corresponded to the functions γ and $\emptyset_n$ in the formalization of the learning problem, above.

Having seen how hypotheses can be represented in terms of strings and how these can be combined systematically to form new hypotheses, there remains the problem of defining the relative fitness of a hypothesis with respect to a linguistic environment. Ultimately, we want the learner to become better able to represent the input data. In other words, the learner should change its hypothesis on the basis of evidence from the external environment and its new hypothesis must be better able to account for this evidence. In some sense, new hypotheses must be an improvement over the old hypotheses.

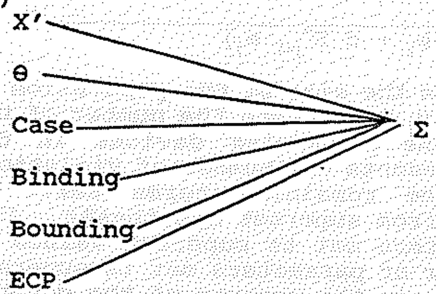
Clark (1990) provides a crude definition of improvement based on the ability to parse input sentences in terms of failed parses. We will modify his treatment by supposing that the crucial property of a failed parse is that it violates at least one principle of core grammar. In particular, we will suppose that a parser consists of a number of modules (Case, binding, X'-theory, and so on) which operate in tandem to produce a full syntactic representation. When a principle in one of these modules is violated, when the current grammar cannot assign a well-formed representation to some input, the offended component will signal a violation. With this in mind, we adopt the following notion of improvement of one hypothesis with respect to another:

(15) A hypothesis A is an improvement over a hypothesis B if, given an input datum, s_i, A signals m violations of core grammatical principles while B signals n violations and $m < n$.

Intuitively, a parser which signals 3 violations on a parse is rather better than one which signals 4 violations and a parser which signals 2 violations is superior to one which signals 3. Crucially, parsers need not perform perfectly in order for the performance to be distinguished.

We will suppose, then, that the various modules of the parser are connected to a summation function, Σ , in the following manner:

(16)



Each module can signal a violation to the function Σ, which then sums up the number of violations and passes the number on to the learner. Notice that the learner has no access to which grammatical principles have been violated; it only receives a number representing the sum of the violations for each parse.

As noted above, the learner must be able to distinguish between hypotheses which generate a superset language and those which do not. If a superset hypothesis and a subset hypothesis can both account for an input datum, then, all things being equal, the learner should prefer the latter to the former. Thus, any fitness metric should be such that it generally rates a subset hypothesis more highly than superset hypothesis just so long as the subset hypothesis is empirically adequate (does not fail to parse the input data).

Finally, we will assume that the learner can take into account the overall ``elegance'' of its hypotheses. That is, the learner will, all else being equal, prefer hypotheses which lead to more compact representations. Compactness, here, can be defined in terms of such factors as the number of nodes required to cover the input string, the length of the chains associated with arguments and operators, or both. For the moment, we will assume that the measure of elegance is a raw node count from each parse.

With these factors in mind, we suggest the following as a fitness metric, defined over a population of parsing devices relative to an input sentence (see Clark, 1990 for an earlier version of this metric). It should be noted that hypotheses are judged indirectly by means of the parsing devices which they determine, just as a genotype is judged through its expression as a phenotype:

(17) The Fitness Metric

$$\frac{(\Sigma^{n}_{j=1} v_j + b\, \Sigma^{n}_{j=1} s_j + c\, \Sigma^{n}_{j=1} e_j) - (v_i+bs_i+ce_i)}{(n-1)(\Sigma^{n}_{j=1} v_j + b\, \Sigma^{n}_{j=1} s_j + c\, \Sigma^{n}_{j=1} e_j)}$$

In the above equation, n is the size of the population of parsing devices. The term $\Sigma^{n}_{j=1} v_j$ represents the number of the violations signaled by all the n parsing devices in the population and the term vi represents the number of violations signaled by parser P_i. The remaining terms serve to distinguish subset hypotheses from superset hypotheses. $\Sigma^{n}_{j=1} s_j$ is the number of the parameters set to superset values in the population and s_i the number of parameters set to superset values in hypothesis h_i. $\Sigma^{n}_{j=1} e_j$ is the measure of the general elegance of the analyses returned by the entire population, in terms of a measure of the size of each parse tree, a raw node count, while ei is the measure for the parser P_i. The constant, b, a ''superset penalty'' for guessing too large a language. This term serves to scale Subset Condition violations so that they are less severe than raw failure to parse. The learner is, then, still able to hypothesize superset languages if necessary. Similarly, the constant c is a scaling factor for the elegance of the representation, again serving to make elegance a less important factor than grammatical violations. We will leave the question of the exact values of these constants open, assuming only that 1>b, c>0.

It is, perhaps, useful to consider the contribution of each of the above factors, using some hypothetical examples. Let us turn first to the way in which the fitness metric treats grammatical violations. For the population, this is the term $\Sigma^{n}_{j=1} v_j$ in the fitness metric; for the individual parsing device, it is the term v_i. Suppose we have the three parsing devices, p_1, p_2 and p_3. Running these on an input sentence yields:

(18)a. p_1 returns 1 violation, covering the input with 15 nodes.

b. p_2 returns 2 violations, covering the input with 15 nodes.

c. p_3 returns 3 violations, covering the input with 15 nodes.

Running the above results through the fitness metric gives the following results, with $b = 0.02$ and $c = 0.05$ (we ignore, here,the contribution of the subset factor by assuming that none of the hypotheses underlying the parser contain superset settings):

(19)a. p_1 receives a fitness rating of 0.393939.
b. p_2 receives a fitness rating of 0.333333.
c. p_3 receives a fitness rating of 0.272727.

Thus, parser p_1 is judged the most fit, p_2 the next most fit and p_3 the least fit. Notice that the learner does not receive information about which grammatical principles are violated. It has no need of such information in order to distinguish between the hypotheses at hand. Instead, it need only observe the performance of its hypotheses in an external manner, without information as to their inner-workings. The learner will base its new hypotheses on those old ones that are relatively more fit, thus passing on the parameter settings which made those hypotheses fit to future generations. Those parameter settings which avoid grammatical violations relative to the input text will be preserved, and those which tend to generate violations will gradually disappear.

Let us turn, now, to the contribution of the superset penalty, the term $\Sigma^n_{j=1}\ s_j$ for the entire population and the term s_i for a single parsing device. Suppose that p_1 and p_2 both signal no violations of any grammatical principles and both cover the input in 20 nodes. Suppose further that p_2 contains a superset setting for one parameter and p_1 contains no superset settings. The fitness metric will then return:

(20)a. p_1 receives a fitness rating of 0.50495.
b. p_2 receives a fitness rating of 0.49505.

Notice that the ``smallest hypothesis'', in this case the one underlying p_1, is judged more fit than the one which violates the Superset Condtion. Thus, the fitness metric can distinguish both between hypotheses that are unequal in their parsing powers and between hypotheses that are equal in parsing power but differ with respect to the Subset Condition.

We turn, finally, to the contribution of the ``elegance'' factor; this is the term $\Sigma^n_{j=1}\ e_j$ for the entire population and e_i for individual parsing devices. Consider two hypotheses, p_1 and p_2, which both return no violations, contain no superset settings but cover the input with trees of different elegance. Suppose that p_1 is able to cover the input with 17 nodes while

p_2 covers the input with 18 nodes. The results of the fitness metric are then:

(21)a. p_1 receives a fitness rating of 0.514286.
b. p_2 receives a fitness rating of 0.485714.

The first hypothesis is preferred by the fitness metric since it is able to span the input in a more elegant way than the second hypothesis.

In order to see the importance of this factor, consider the case where the target is SVO. Suppose that hypothesis h1 treats the subject as being in the Spec of IP at S-Structure while hypothesis h_1 treats the subject as having moved to the Spec of CP, attracting the main verb with it. For a simple clause, h_1 and h_2 will return the following structures:

(22)a. h_1: $[_{IP}$ DP $[_{I'}$ I VP]]
b. h_2: $[_{CP}$ DP_i $[_{C'}$ V_j $[_{IP}$ t_i $[_{I'}$ t_j VP]]]]

By assumption, both h_1 and h_2 can account for the input stream. Notice, however, that h_2 involves systematically longer chains than h_1 since the former always involves movement of the subject to the Spec of CP, with subsequent attraction of the verb to C^o, while the latter does not. The representations returned by h_1 are simpler than those returned by h_2. Since the learner, via the fitness metric, can take into account the general elegance of representations, it can successfully distinguish between h_1 and h_2. Notice, however, that elegance is defined quite simply as a count of the nodes in the tree covering an input item plus the lengths of the chains in the representation.

The fitness metric can be seen as working as follows. The population of parsing devices specified by the learner's hypothesis strings is run against each input item. The term

$$\Sigma^n_{nj=1}\ v_j + b\ \Sigma^n_{nj=1}\ s_j + c\ \Sigma^n_{nj=1}\ e_j$$

yields the total number of violations, superset settings and the total elegance of representations of the entire population, with the various factors weighted appropriately by the constants b and c. Dividing this term by n, the size of the population, would give the average number of undesirable properties for the entire population. Consider, next, the term $v_i + bs_i + ce_i$. This yields the number of unhealthy properties each individual parsing device carries. As this term grows in relation to the population average, the relative fitness of the parsing device decreases. If this term decreases with respect to the population average, then the parsing device is judged relatively more fit[7] .

The opportunity to reproduce (that is, be selected for the crossover operation and mutation) is a direct function of relative fitness. The simulation developed in Clark (1990) assumes that the fitness associated with a hypothesis corresponds transparently to its proportion of the general population. In an environment with random mating, then, those hypotheses with a high proportion in the population are more likely to meet and reproduce. The fitness ratings are used to simulate a weighted roulette wheel, the results of which undergo the crossover and mutation operations. In other words, successful hypotheses will receive a high fitness rating. The fitness rating corresponds to the probability that the hypothesis will get to reproduce. Thus, the fittest hypotheses will reproduce more frequently and pass on their parameter settings to new hypotheses. Cumulatively, then, the population will tend toward the optimal set of parameter settings for the target.

Crucially, the most fit hypotheses are the most likely to contribute to the formation of new hypotheses. These hypotheses have the greatest opportunity to pass on the parameter settings which made them fit to new hypotheses. By pruning weak hypotheses at random intervals, these are ultimately prevented from contributing their inferior parameter settings to the general pool. Thus, fit parameter settings tend to take over while unfit parameter settings are purged. By iterating the process of parsing, judging fitness, reproduction and ''death'', the learner is able to incrementally approach the target grammar.

Before turning to the diachronic data, two other definitions are required. Consider a simple example like:

(23) John loves Mary.

Notice that certain parameters must be set in a particular way if the sentence is to be parsed. Both John and Mary must receive θ-roles and Case, the verb love must be capable of picking up its inflectional affix, and so on. Any parsing device which can successfully account for these features of the sentence in (23) will return a well-formed representation. Other parameters (e.g., bounding nodes and those that regulate conditions on anaphora) are irrelevant to the representation of the above sentence. It won't matter what values for these parameters the parsing device presupposes. This suggests that any given input sentence expresses certain parameters and that a set of distinct parsing devices can account for the above sentence:

(24) Parameter Expression
A sentence σ expresses a parameter p_i just in case a grammar must have p_i set to some definite value in order to assign a well-formed representation to σ.

When a given datum expresses some parameter value, the learner will be under pressure to set that parameter to the value expressed by the datum. This is because the fitness metric will prefer hypotheses with the correct setting to those which lack that setting. This provides a simple definition of the intuitive notion of 'triggering datum':

(25) Trigger
A sentence σ is a trigger for a parameter P_j if σ expresses P_j.

Given the above interpretation of the input data, we can imagine a method of encoding the data in string form. Suppose we have a function Ψ which maps a sentence onto the set of sequences of parameter settings which are compatible with that sentence. For example, a given input sentence, s_m can be accounted for by grammars with the second and third parameters set to 0 and the fifth parameter set to 1. Applying Ψ to s_m would give the following set of parameter strings:

(26) $\Psi(s_m) = \{00001, 10001, 00011, 10011\}$

Using ``*'' as a variable to range over 0 and 1, we could replace the above set of strings with a cover term:

(27) {00001, 10001, 00011, 10011} = [*00*1]

We will refer to the sequence [*00*1] as the p-encoding for s_m; the p-encoding of a sentence may be thought of as a ``pure'' representation of the parameters expressed by the sentence. Notice that, in principle, one could replace the sentences in an input text with their p-encodings and, so, study the frequency of expression for various parameters and the overall structure of the text relative to parameter expression.

There is an important relationship between parameter expression and the fitness metric. Ultimately, the fitness associated with a hypothesis governs its probability of being selected for reproduction. The more fit a hypothesis is, the more likely it is to pass on those parameter settings which made it fit. Consider, now, parameter expression. When a parameter is expressed, those hypotheses which have the correct value for that parameter will be judged more fit than those which lack the proper value. If a parameter is expressed frequently, then those hypotheses bearing the correct value will have more opportunity to be selected for reproduction and the appropriate parameter setting will tend to dominate in the population. Furthermore, those hypotheses bearing the incorrect will have a lower fitness rating and will tend to reproduce less frequently to the point where the parameter values which made them unfit are washed from the population. Thus, parameter settings which are expressed frequently will tend to be set quickly and efficiently by the learner. Parameters which are expressed infrequently, however, will tend not to affect a hypotheses fitness in the same way. The learner will have correspondingly less stake in setting the parameter correctly and it will take the learner longer to set the parameter.

Consider, now, the case where parameters are ambiguously expressed. In our terms, there might be several contradictory p-encodings associated with a class of data, for example. Here the learner has several possible solutions available which can account for the input without generating grammatical violations. In this case, frequency of parameter expression will not aid the learner in distinguishing between its hypotheses. Instead, the learner will have to rely on the structure of the hypotheses themselves, and not their empirical coverage, in order to select a winning hypothesis. These internal factors are the overall elegance of representations and the number of superset settings in each hypothesis, both of which are factors in the fitness metric. We argue, here, that it is this sort of case which provides the fuel for core diachronic change in a parameter setting. We will turn, in the next section, to a case where learners were faced with just such an ambiguity.

3. A Case Study in Diachronic Change

3.1 The History of French

Roberts (forthcoming) analyses three major syntactic changes in the history of French as reflexes of a single underlying parametric change. The three changes are as follows:

(28) Loss of ``simple inversion'' in interrogatives:
    (a) *A Jean pris le livre? ModF
        " Has Jean taken the book.''

(b) Comment fu ceste lettre faitte? OF
``How was this letter made?''

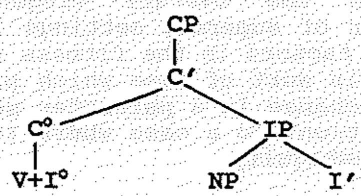
(29) Loss of null subjects:
(a) *Ainsi s'amusaient bien cette nuit. ModF
``Thus, they had fun that night.''
(b) Si firent pro grant joie la nuit. OF
``Thus, they made great joy the night.''

(30) Loss of V2:
(a) *Puis entendirent-ils un coup de tonnerre. ModF
(b) Lors oïrent ils venir un escoiz de tonoire. OF
``Then heard they a clap of thunder.''

As Roberts shows in some detail, each of these constructions was lost in the early 16th century. Roberts argues that these changes reflect an underlying change in the value of the parameter determining Nominative-assignment proposed by Koopmann & Sportiche (1990): Nominative may be assigned (by I) either under government, or under agreement or both. The central idea of this account of the history of French is OF allowed Nominative-assignment under government, while ModFr does not.

More precisely, all of the OF constructions depend on the possibility of the inflected verb, V+I, assigning Nominative Case to the subject in SpecI' from C:

(31)



This situation was allowed the grammar of OF (and is still allowed in, for example, the contemporary Germanic languages). In a grammar where this configuration of Case-assignment is not allowed, no lexical NP can survive in subject position in inversion contexts; this is the situation in ModF where (28) is thus a violation of the Case Filter. Following Kayne (1983), Rizzi & Roberts (1989), we assume that clitics can survive in subject position in this context since they are able to pass the Case Filter in other ways (cf, also, Baker, 1988; Everett, 1986).

Adopting Rizzi's (1986a) proposal that the necessary condition on formal licensing of pro is that it occupy a Case-marked position, Roberts accounts for the change illustrated in (31b) by extending the Nominative parameter to the pro module; it is well-known that OF null subjects were licensed only in contexts of inversion (cf. Thurneysen, 1892; Price, 1971; Einhorn, 1974; Foulet, 1982; Vanelli, Renzi & Benincà, 1983; Adams, 1987a,b), and so a natural interpretation of this is that null subjects could only be licensed where Nominative was assigned under government, i.e., in the configuration (30). In this way, we see why null subjects were lost when Nominative-assignment under government was lost.[9] Regarding (31c), V2 also depends on the capacity of I to assign Nominative Case to the subject under government after being raised to C with the verb. Note that Nominative-under-government is a necessary, not a sufficient, condition for V2. So a system without this possibility cannot have V2. However, a system with this possibility does not have to have V2 (Modern English is probably such a system). In fact, as we shall see, obligatory V2 was already eroding in MidF---this was a crucial factor in the instability which led to the change in the Nominative-assignment parameter.

The principal trigger for the change in the possibilities of Nominative assignment was the introduction of new word-orders that were not strictly in conformity with V2, notably XSVO (where 'X' could be a topic or an adverb). This innovation was probably caused by the development of a series of subject clitics in MidF (see below). The cumulative effect of the new word-orders was to destabilize the system in such a way that setting the parameter for Nominative-assignment under government positively became impossible by about 1500, and learners converged on a grammar lacking this property. The result was the elimination of the structures in (28-30) in 16th century texts: a major change in the grammar of French. Note that we do not consider the null-subject parameter or the V2 parameter as in any sense subsumed by the Nominative parameter; however, the particular circumstances of French at the time the change took place were such that the loss of Nominative-assignment under government entailed the loss of null subjects and the elimination of V2. Our proposal is that the initial weakening of V2 combined with the development of a series of subject clitics created a system which ultimately eliminated V2, and in doing so eliminated null subjects and simple inversion. In particular, the weakening of V2 had the effect that hypotheses which allowed an input datum to be analyzed as a V2 structure become more costly relative to the fitness metric; thus, the learner was under pressure from fitness to eliminate the V2 fitness to eliminate the V2 hypothesis.

Although we concentrate exclusively on French here, there is also evidence (cf. in particular Vanelli, Renzi & Benincà (1983)) that many of the Gallo-Italian dialects of Northern Italy have undergone the same parametric change, since in their recorded history, simply inversion, V2 and, arguably, null subjects have been lost (although the contemporary dialects in fact have a kind of ''disguised'' null-subject system which probably represents an independent diachronic innovation; cf. Poletto (1990), Renzi & Vanelli (1983), Rizzi (1986b)). Moreover, Renzi (1983) argues that Modern Standard Italian has undergone the same changes as French regarding inversion while retaining null subjects.

In all, five parameters are relevant to our account of the historical development of French. These are given in

(32)a. NOM is assigned (by I) under agreement.
{1,0}
b. NOM is assigned (by I) under government.
{1,0}
c. Clitic nominative pronouns.
{1,0}
d. Null subjects licensed canonically (Case-dependently).
{1,0}
e. Obligatory V-movement to C in matrix declaratives (V2).
{1,0}

Note that we split Koopman & Sportiche's parameter for nominative Case assignment into two separate parameters in order to preserve a basically binary vocabulary for parameters (see the discussion of subjacency and bounding nodes in Section 2). We take it that (32a) has been constant at 1 throughout the entire period (but see Section 3.5). As just mentioned, (32 b,d, e) changed together in the 16th-century. The shift in (32d) and (32e) was forced by the change in the value of (32b). This is presumably quite a standard situation with parametric change: changes in parametric values interact. Moreover, parameter values can be affected by non-syntactic factors, notably phonological changes. This is the case with (32c); properties connected to the stress system may cause a class of pronouns to cliticize and thereby trigger a shift in the value of this parameter.

We now review the relevant data from the different periods of French. We show how the data trigger parameter settings. To illustrate the general technique, we will first consider modern French. Then we will consider Old French and finally the period of greatest ''structural instability'' (and, hence, of greatest interest), Middle French.

3.2. Learning Modern French

Before considering the earlier periods of French, let us first consider the situation in the contemporary language. What are the parameter values for ModF? It is clear that Nominative Case is assigned by I to its specifier position, hence the first position in the string must be set to 1. On the other hand, Rizzi & Roberts (1989) argue that ModF does not allow Nominative-assignment under government; this is what leads to the restriction to clitics in contexts where the inflected verb, a complex head which contains I, moves to C (e.g. in interrogatives or conditionals; cf. also (31a)):

(33)a. Ont-ils/*les enfants vu ce film?
''Have they/the children seen this film?''
b. Aurait-elle/*Marie fait cela...
"Had she/Marie done this...''

Once moved to C, I must Case-mark the subject position under government; the ungrammaticality of these examples with a non-clitic subject shows this is impossible. In terms of this analysis, I does not assign Nominative under government in ModF, and so we set the second position in the string to 0.[9] It is well-known that ModF has a class of clitic nominative pronouns (cf. Kayne (1975), Rizzi (1986b)); the contrasts in (33) in fact illustrate that these elements interact with Case theory in a manner distinct from non-clitics. In Rizzi & Roberts (1989) it is proposed that clitics can satisfy Case theory by incorporating with the verb in C (cf. Baker (1988), Everett (1986, 1989)). So we take it that in ModF parameter (32c) is set to 1. Both parameters (32d) and (32e) are set to 0: ModF is not a null-subject language, as a comparison with contemporary Italian shows, and neither is it a V2 language, as a comparison with German shows.[11]

These remarks on the grammar of ModF (which we of course cannot fully substantiate here; cf. the references cited for further arguments) lead to the following conclusion regarding the representation of the parameters in (32) as a string of binary units:

(34) The ''target string'' for ModF is 10100.
Nominative Case is assigned under agreement and subject clitics are allowed.

Let us now see how the parameter values for ModF are expressed in the input text. Recall that a sentence S expresses a parameter P_i iff a grammar must have P_i set to a particular value in order to assign a well-formed representation to S. In such a situation, S is a trigger for P_i (cf. (28). The following examples illustrate a significant part of the trigger for the parameter values of ModF:

(35)a. Jean aime Marie.
''Jean loves Marie.''
b. Hier Jean est parti.
''Yesterday Jean left.''
c. Où est-il allé?
''Where did he go?''

Recall that the conditions of acquisition are such that starred examples like (28a-30a), which can be used by the linguist to justify a particular analysis, are not available. Moreover, many sentences are amenable to differing structural analyses which can affect their status as triggers. This last point is crucial to understanding how change takes place, as we shall see.

Consider first (35a), a simple declarative sentence with canonical SVO order. In terms of the usual analysis of ModF, the relevant parts of this sentence are as follows:

(36) [$_{IP}$ Jean [$_{I'}$ aime ...]]

Parsed in this way, (35a) triggers Nominative-assignment under agreement and indicates that V-movement to C is not required in matrix declaratives, i.e. that ModF is not V2. So (36) is associated with the following P-encoding (cf. (29) for a general definition of P-encoding):

(37) [1 * * * 0]

Nominative assigned under agreement and V movement to C is not allowed in matrix declaratives.

(37) indicates that (35a) tells the learner that Nominative is assigned under agreement, and that French is not V2, but does not say anything about whether Nominative is assigned under government, whether subject clitics are allowed, or whether null subjects are allowed.

However, strings exactly equivalent to (35a) are grammatical in the Germanic V2 languages. In these languages, the relevant parts of the structure are as follows:

(38) [$_{CP}$ Jean [$_{C'}$ aime [$_{IP}$ t [$_{I'}$ t..]

Call this the ''V2 parse'' of an SVO sentence. Here I assigns nominative Case to the SpecI' (i.e., the position occupied by the trace of the subject) under government; we will refine this analysis in Section 3.5. So the P-encoding for this parse is:

(39) [* 1 * * 1]

The parser must have nominative Case assignment under government and V-movement to C is obligatory in matrix declaratives.

As (39) shows, (35a) remains silent regarding subject clitics and null subjects.

To sum up, SVO declaratives in ModF have the following P-encodings:

(40) SVO declaratives P-encode:

a. [1 * * * 0]
V remains in I in matrix declaratives and nominative Case is assigned under agreement

b. [* 1 * * 1]
V moves to C in matrix declaratives and nominative Case is assigned under government.

SVO sentences are thus associated with the different P-encodings depending on the parse they are given. We can characterize this situation in terms of the following notion of P-ambiguity:

(41) A sentence S is P-<u>ambiguous</u> wrt some parameter P_i just in case S has the set of well-formed representations $(R_1...R_n)$ and P_i must be set to some definite value v_1 in order to assign R_i to S (i.e. R_i triggers a P_i (v_1)) while P_i does not need to be set to v_1 in order to assign $R_j \neq R_i$ to S.

ModF SVO sentences are P-ambiguous, as (40) shows. As we will see, the representation where V is in C is disfavoured since it involves a more complex structure than that where V is in I.

Consider now (35b). In V2 languages generally, orders of this type are impossible (cf. Schwartz & Vikner (1989)). This can be interpreted in terms of a ban on adjunction to CP. Supposing that this is so, this example must be parsed with the adverb attached to IP, V in I and the subject in SpecI'. In other words, the relevant parts of the structure are like the parse of (35a) given in (36), and the triggering properties of the sentence are the same. More generally, we can conclude the following:

(42) XSV P-encodes [1 * * * 0]
Nominative Case is assigned under agreement and movement of V to C is not allowed in matrix declaratives.

Now consider the interrogative in (35c). (35c) provides evidence for the subject clitic (this evidence is probably morphological, given the existence of a separate paradigm of clitic pronouns), and therefore, given the fact that clitic pronouns do not obey the Case Filter in the same way as non-clitic NPs (see above), provides no evidence for either Case-assignment parameter. We take it that interrogative sentences by their nature provide no evidence regarding V2 in declaratives, and the null-subject parameter is not determined either. So we arrive at the following (where ''s'' indicates a subject clitic in the schematic word-order):

(43) Wh VsO P-encodes [* * 1 * *]
Subject clitics are possible.

If the subject clitic is not recognized as such, but treated as a full NP, this sentence would P-encode (44):

(44) [* 1 * * *]
Nominative is assigned under government.

We assume, however, that phonological and morphological evidence disfavours this possibility.

Putting these P-encodings together (and disregarding that in (44)), the following picture emerges:

(45)a. [1 * * * 0] SVO, XSV
Nominative under agreement; no V2 in declaratives.
b. [* 1 * * 1] SVO
Nominative under government; V2 in declaratives.
c. [* * 1 * *] Wh VsO
Subject clitics are possible.

The two parameters that are not positively set are nominative under government and null subjects. These are both set to 0 in the optimal case. Let us see why.

The two parameters determining Nominative-assignment by

I, (31a,b), are in a shifting relation. Although neither parameter directly determines a superset relation (a grammar which allows Nominative-assignment under agreement generates a language which intersects with one which does not; similarly for Nominative-assignment under government), if both parameters are set to 1 they together generate a language which is the superset of that which results from setting either parameter to 0. This is a classic case of shifting (of the type seen in Section 2). Now, as we have seen, (31a) is unambiguously expressed in the input for ModF and, so, is set to 1. In order to avoid shifting, a positive value for (31b) is strongly disfavored. Since there is no unambiguous evidence for nominative Case assignment under government, the pressure against shifting is decisive and the parameter is set to 0 in the optimal grammar.

It should be noted that the only evidence for nominative Case assignment under government are sentences with the order ''SVO'', with a V2 parse, which can also be analyzed more compactly under the assumption that nominative Case is assigned under agreement. In particular, the V2 parse for the ''SVO'' order must involve movement of the subject to the SpecCP, and thus entails a longer chain than would occur under the competing analysis. Thus, the non-V2 parse is again favored and the V2 parse is disfavored by the fitness metric. This provides further evidence to the learner in favor of setting the V2 parameter to 0, as well as disfavoring nominative Case assignment under government.

For the null-subject parameter, we could follow the reasoning in Berwick (1985) and invoke the Subset Condition (cf. Section 2). If null-subject languages are superset of non-null-subject languages, the lack of a trigger for a positive value of the null-subject parameter will guarantee that (31d) is set to 0. Alternatively, we could appeal to morphological conditions, and say that, although the syntactic evidence does not determine a value for (31d), the ''poverty'' of French verbal inflection determines a negative value. We will leave this question open here.

The above paragraphs demonstrate how the various factors we are concerned with work. On the basis of simple, plausible, positive evidence, the learner can converge on the correct parameter settings for Modern French. What we intend to show in what follows is how these same factors led to a major parametric change in French c1500.

3.3 Old French

We mentioned earlier that OF allowed Nominative-assignment under government (cf. (30a,c)). We assume that nominative could also be assigned under agreement, although we will return to this point in Section 3.5. (30b) shows that OF allowed null subjects, although it is well-known that these were possible only in contexts of inversion. Another well-known and much-discussed difference between OF and ModF is that the OF nominative pronouns je, tu, il, etc., were potentially tonic elements, unlike their ModF counterparts (cf. Kayne (1975) on ModF; Adams (1987a,b), Roberts (forthcoming, 2.2) and below on OF). These facts about OF syntax (which are discussed at length in Roberts (forthcoming)) lead to the following parametric settings, in terms of (31):

(46) The target string for OF is 11011.

Nominative Case assignment was possible both under agreement and under government; Null subjects were possible; V2 was obligatory in matrix declaratives.

As in the previous section, we now show how this string could be determined on the basis of simple, positive evidence[11].

The following kinds of sentence were available as evidence, where ''(S)'' indicates a null subject:

(47)a. XVS: (Et) lors demande Galaad ses armes.
''(And) then asks Galahad (for) his arms.'')
b. SVO: Aucassins ala par le forest.
''Aucassin went through the forest.'')
c. XV(S)O: Si firent grant joie la nuit.
''So they made great joy the night.'')

(48)a. WhVSO: (Mais) ou fu cele espee prise ... ?
''(But) where was that sword taken...?'')
b. WhVSO: Ne nos connoissez vos mie?
"Neg us know you not?''

(47a) is a V2 declarative (as in Modern Germanic languages, conjunctions like ''and'' and ''but'' do not count in the computation of V2; these elements can be external to CP when they conjoin CPs). The relevant parts of the structure of this sentence are:

(49) $[_{CP}$ lors $[_{C'}$ demande $[_{IP}$ Galaad ... $]]]$

Here the inflected verb in C assigns Nominative to the subject NP, Galaad, under government. Of the five parameters in (31), this example then positively triggers Nominative-assignment under government and V2. More generally, we have (50):

(50) XVS P-encodes [* 1 * * 1]; that is, nominative is assigned under government and V2 is obligatory in matrix declaratives.

OF also allowed SVO sentences like (47b). As in the case of the ModF SVO order, this kind of sentence is P-ambiguous in the following way:

(51) SVO P-encodes either [* 1 * * 1] (Nominative under government and V2) or [1 * * * 0] (Nominative under agreement and no V2)

We will return to this point below.

As we have already mentioned, OF allowed null subjects in V2 contexts. (47c) illustrates this. Such examples are also P-ambiguous from the point of view of the learner: if V is in C, then the null subject is licensed under government in SpecI; if V is in I, the null subject is licensed under agreement in SpecI. In the former case, Nominative under government and V2 are triggered; in the latter case, Nominative under agreement is triggered along with a negative value for V2. In both cases, the null-subject parameter is positively triggered. So we have the following P-ambiguity:

(52) XV(s)O P-encodes either [* 1 * 1 1] (as above, with null subject) or [1 * * 1 0] (as above, with null subject)

Consider now the interrogatives in (48). (48a) has the same trigger properties as a V2 declarative, except that by assumption interrogatives cannot trigger the V2 parameter. On the assumption that the Nominative pronouns were tonic,[12] (48b) involves Nominative-assignment under government to the clitic, just as with any other NP subject. These examples, then, have the following P-encoding:

(53) WH VSO P-encodes: [* 1 * * *]
Nominative under government.

Putting the above P-encodings together, we arrive at (54):

(54)a. [* 1 * * 1]
Nominative under government and V2
b. [1 * * * 0]
Nominative under agreement, V2 disallowed
c. [* 1 * 1 1]
Nominative under government, null subjects and V2
d. [1 * * 1 0]
Nominative under agreement, null subject and no V2
e. [* 1 * * *]
Nominative under government

We clearly see that both Nominative parameters are triggered positively. Notice that the positive evidence overrides the fact that these two parameters are in a shifting relationship (we return to this in Section 3.5). The null-subject parameter is also positively triggered. V2 is also triggered if we take it that the positive evidence for the more complex trigger weighs heavier than the pressure in favour of the simpler structure in the P-ambiguous cases; this is a matter which can be captured by the fitness metric. Finally, as mentioned in Note 12, there is no morphological evidence in favour of subject clitics, in that there was only one series of subject pronouns at this time. Phonological evidence presumably militates against treating the nominative pronouns as obligatory clitics; for example, these pronouns could stressed in OF, as their occurrence in topicalized position as in (55) indicates:

(55) Je, que sai?

Moreover, subject pronouns, unlike object pronouns, could appear first in V2 declaratives. This indicates that they ``count'' just like other XPs for the determination of V2; object pronouns did not ``count'', however:

(56)a. Tu es or riche et ge sui po proisié.
``You are now rich and I am little valued.''
b. Toutes ces choses te presta Nostre Sires.
``All these things to you lent our Lord.''

Moreover, there was only one series of subject pronouns in OF. On the basis of evidence of this kind, the subject-clitic parameter was set to 0.

So we see how simple, positive data could trigger the parameter-settings for OF. This discussion of the OF data brings out one important point: clear positive evidence overrides all other considerations. We saw this in two cases. First, OF had a shifted system with respect to the Nominative parameter, but learners nevertheless converged on this system since there was clear, positive evidence for it. Second, the P-ambiguities of SVO and V2/null-subject examples are resolved by the unambiguous V2 cases, and moreover this resolution is in the direction of the more complex structure. In other words, clear positive evidence can override both subset/shifting considerations and the pressure towards the simplest possible structure. In terms of our assumptions and definitions, ``clear, positive evidence'' means non-P-ambiguous evidence. The only non-P-ambiguous evidence for V2 is the XVS order, so we see that this type of sentence played a crucial role. This order was very frequent in OF matrix declaratives; Roberts (forthcoming; 2.3.1) gives the following percentages for (X)VS and SV(X) order (based on the first 100 matrix declaratives with overt subjects in six representative texts):

(57) (X)VS = 58%   SV(X) = 34%

Although a more sophisticated and exhaustive quantitative analysis is needed in order to fully demonstrate the point, we can conclude that (X)VS orders were sufficiently frequent to trigger a positive setting of the V2 parameter. This in turn means that SVO sentences could be analysed as V2, unlike in ModF. Thus, a shifted system is allowed because there is clear evidence for it; this situation is quite different for the case of modern French where the only evidence there is for the shifted system is P-ambiguous and is thus disregarded.

In Section 1, we introduced the notion of stability of parameter setting, proposing that a parameter -setting is stable to the degree that its expression in the input data is unambiguous. Was the V2 parameter stable in OF? The only non- P-ambiguous trigger for V2 is provided by XVS orders. The frequency of these orders positively sets the Nominative-under- government parameter, and so makes the V2 parse available for the P-ambiguous SVO and null-subject structures. The potential instability created by the ``non-V2 parses'' of these examples is eliminated in the optimal grammar of OF. Nevertheless, it is likely that the non-V2 parse for SVO and null-subject sentences was a close rival for the V2 parse, even in (later) OF, especially since elegance considerations always favour a non-V2 parse over a V2 parse where there is a choice. More explicitly, in terms of the fitness metric, the existence and frequency of an unambiguous trigger for V2 was sufficient to establish a positive setting for the V2 parameter. Recall that the relative elegance of a parse plays a less crucial role in judging fitness than real grammatical violations. This is because the elegance factor is scaled down by the constant c of the fitness metric while violations are not scaled down. Thus, a hypothesis which leads to slightly more inelegant representations without generating grammatical violations will ultimately drive a hypothesis which generates elegant violations out of the population.

In the next section, we will see how the MidF situation contrasts with what we have just described for OF. In particular, we will see that, in part because of the introduction of new word orders and in part because of the diminishing frequency of XVS, XVS orders were no longer able to trigger a positive value for the V2 parameter. As a result of this the V2 parameter became maximally unstable. The instability was resolved by a parametric change which led to the loss of the constructions in (28) - (30).

3.4. Middle French

In MidF, XSV is introduced, and SVO and V1 become more frequent. These factors together meant that the V2 constraint was less rigourously respected than it had been in OF (although V2 orders were still possible throughout this period, unlike ModF). Also, a separate series of nominative clitics emerged. For now, we will take the introduction of the new word orders as given, although we discuss possible causes for this change in the conclusion (cf. also Adams (1987a,b), Roberts (forthcoming)). We treat the cliticization of Nominative pronouns as a phonologically-driven change. Otherwise, MidF was like OF and different to ModF, in particular with respect to Nominative- assignment under government and null subjects. We do not present a target string for MidF, however, since we precisely wish to show how indeterminacy in one parameter (V2) created indeterminacy elsewhere (nominative-assignment under government and the possibility of null subjects).

Let us consider the types of evidence available in MidF. As with OF, we have the following kinds of declaratives:

(58)a. XVS: Or avoit nostre curé priez des aultres prebtres.
''Now had our priest asked the other priests.''
b. SVO: Les Anglais veulent un roi guerrier.
''The English want a warrior king.''
c. XV(S)O: Or ai eu plusseurs fois grant imagination.
''Now have (I) had several times great imagination.''

Also as in OF, these constructions have the following P-encodings, corresponding to (58a), (58b) and (58c), respectively:

(59)a. XVS: [* 1 * * 1]
Nominative under government; obligatory V2 in matrix declaratives.
b. SVO: [* 1 * * 1] (as above) or [1 * * * 0] (nominative under agreement; no V2)
c. XV(S)O: [* 1 * 1 1] (as in (a) with null subjects) or [1 * * 1 0] (as in (b) with null subjects)

The changes that took place in MidF created further possibilities, however. Consider the following (where ''s'' indicates a subject clitic):

(60)a. XVs: Or ai je proposé ensi que ...
''Now have I proposed thus that...''
b. XsV: Et ce conseil nous vous donnons.
''And this advice we to you give.

Taking these examples to positively trigger the subject-clitic parameter, these examples have the following P-encodings, corresponding to (60a) and (60b):

(61)a. XVs: [* * 1 * 1]
Subject clitics and V2
b. XsV: [* * 1 * 0]
Subject clitics and no V2

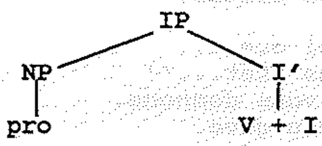
Since clitics can receive Case in ways unavailable to other nominal elements, sentences containing subject clitics provide no information about either Nominative-assignment parameter. The order verb-clitic in (60a) triggers a positive setting for the V2 parameter. On the other hand, since French subject clitics (then as now) do not attach to a verb and move with it (unlike object clitics), the order clitic-verb in (60b) triggers a negative value for the same parameter (but see below for further discussion of this kind of case).

As we mentioned above, MidF allows, with growing frequency, other word orders which are not found in OF. These are illustrated below:

(62)a. XSV: Lors la royne fist Saintré appeller.
``Then the queen had Saintré called.''
b. (S)VY: Se appensa de faire ung amy.
``(He) to himself thought to make a friend.''

(62a), combined with the greater frequency of SVO orders in MidF as compared to OF, shows that V2 began to ``erode'' at this period. Sentences of the kind in (62b) illustrate another phenomenon, noticed and analysed by Vance (1989): the fact that null subjects increase their distribution in this period. As Vance shows, null subjects are no longer licensed only in inversion contexts. In Roberts (forthcoming:2.3.5), this situation is analysed in terms of the idea that null subjects could be licensed under agreement as well as under government in MidF, while in OF they are licensed only under government. So MidF allowed a null subject in the following configuration:

(63)



The P-encodings for these orders are as follows:

(64)a. XSV: [1 * * * 0]
Nominative under agreement, no V2
b. (S)VY: [1 * * 1 0] (as in (a) with null subject) or [* 1 * 1 1] (Nominative under government, null subject and V2)

In interrogatives, we find the same general situation as in declaratives. On the one hand, the same kinds of examples are found as in OF:

(65)a. Wh VSO: Que voelt ceste parolle dire?
``What wants this word to say?''
b. Wh VsO: a qui estes vous?
``Whose are you?''

(65a) has the same P-encodings as its OF counterpart, viz.:

(66) WH VSO: [* 1 * * *]
Nominative under government

(^65b), on the other hand, no longer encodes Nominative under government, since the subject has cliticized:

(67) Wh VsO: [* * 1 * *]
Subject clitics

Let us now put together the MidF P-encodings:

(68)a. [* 1 * * 1]
Nominative under government and V2
b. [1 * * * 0]
Nominative under agreement and no V2
c. [* 1 * 1 1]
Nominative under government, null subjects and V2
d. [1 * * 1 0]
Nominative under agreement, null subjects, no V2
e. [* * 1 * 1]
Subject clitics and V2
f. [* * 1 * 0]
Subject clitics and no V2
g. [* 1 * * *]
Nominative under government
h. [* * 1 * *]
Subject clitics

In terms of P-encodings alone, the V2 parameter appears to be no more or less unstable than it was in OF. However, two factors distinguish the MidF situation from the OF one. First, the unambiguous trigger for V2 -- XVS order -- is much less frequent in MidF than in OF. According to Marchello-Nizia (1979), the mean orders for three texts from the late 15th century are as follows:

(69) (X)VS = 10% SV(X) = 60%

This is a significant difference in frequency as compared to OF (cf. the figures given in the previous section). The second factor concerns the status of SVO clauses. We saw in 3.2 how the ``V2 parse" for these clauses is disregarded in ModF, and in 3.3 we saw how the V2 parse was favoured in OF. In MidF, there is total indeterminacy on this point: there is (infrequent) evidence for V2 in the form of XVS order, and there is evidence against V2 in the form of XSV; any parsing device with a positive setting for V2 would engender a violation on this word order and would be disfavored by the fitness metric. A further factor which adds to the instability of V2 at this point is the development of left-dislocation with a resumptive pronoun (Priestley, 1955; Kroch, 1989). This is illustrated by the following example from Priestley:

(70) Les autres arts et sciences, Alexandre les honoroit bien.
".The other arts and sciences, Alexandre them-honored well.''

The development of this type of construction led to a situation of shifting of the type described above in (5) and (6). That is, the interaction between left-dislocation and V2 further obscured the latter due to surface ``V3'' orders. Kroch (1989; p. 215) shows that there is a real correlation between the rise of the construction in (70) and the loss of V2. The correlation results from the action of the fitness metric which will systematically judge a system of this type as relatively unfit.

Late MidF V2 provides an instance of the type of situation described in the Introduction: learners are unable to converge on a single value for a parameter. In other words, the V2 parameter is maximally unstable. Here we are in the ``pathological" situation for acquisition, then. Since the available data cannot decide between two parametric values, other aspects of the fitness metric come into play: the subset criterion and the elegance criterion.

We have seen above, that a language with both V2 and left-dislocation will be disfavored by the Subset Criterion, since it is a case of shifting. A further factor which can decide between competing parses, and therefore competing P-encodings and triggers, is the criterion of ``elegance.''

It is reasonable to suppose that learners follow a Least Effort Strategy in that they try to assign the simplest possible parse to the input string (this idea is discussed at length in the context of syntactic change in Roberts (forthcoming); cf. also de Vincenzi (1989), who proposes that something of this kind is a general parsing strategy, not limited to language learners. Note that the Least Effort Strategy as conceived here is not a principle of grammar; in this we differ from Chomsky (1989)). This idea can be instantiated in terms of counting nodes, traces or chain-positions; we will not attempt to choose between those possibilities here (Roberts (forthcoming) opts for chain-positions; for a formal statement of this, cf. his

Note 26, Chapter 2). What is important here is that any parse which represents the inflected verb as being moved to C is more costly in terms of Least Effort than one which represents the verb as moved only to I (by any of the above criteria).

Suppose, then, that the Least Effort Strategy plays a crucial role in resolving the instability in the data, by penalizing all P-encodings which depend on V-movement to C where there is a choice between this and V-movement to I. More technically, suppose that hypothesis h_1 is identical to hypothesis h_2 except that h_2 allows for V2 in matrix declaratives while h_1 does not. That is, h_1 and h_2 admit the same sentences, and contain the same number of superset settings to parameters, differing only as to the value for the V2 parameter. Hypothesis h_2, then, systematically includes more structure in its representation than h_1 since h_2 will represent the verb as having moved to C (as well as movement of the subject in SVO). In other words, if h_1 returns k nodes on a structure, h_2 will return <u>k+n</u> nodes. Letting m represent the number of superset settings in each hypothesis, running each of the above through the fitness metric will yield the following ratings:

(71)a. $h_1 : 1 - \frac{m+ck}{2m+c(2k+n)}$

b. $h_2 : 1 - \frac{m+c(k+n)}{2m+c(2k+n)}$

Since $1 - \frac{m+ck}{2m+c(2k+n)}$ is greater than $1 - \frac{m+c(k+n)}{2m+c(2k+n)}$, the fitness metric prefers h_1 over h_2 and the learner is under pressure to select h_1. This, then, effectively sets the V2 parameter to 0.

As in OF, MidF has one order where the V2 parameter was unambiguously P-encoded as 1: that of XVS orders, which unambiguously P-encode [* 1 * * 1]. In the situation of instability that reigned in Middle French, the fitness metric, formulated so as to take account of the way in which the Least Effort criterion resolves P-ambiguities, will lead to convergence on a grammar where such experience is simply disregarded (i.e. not parsed where no alternative analysis can be found; an alternative analysis is often available, Roberts (forthcoming:2.4.1) shows that many cases of V2 could be treated as ``free inversion"). Thus we see how an unambiguous trigger for a given property can be disregarded when the system is maximally unstable, even if the instability is located in another area of the grammar.

If the hypotheses where the V2 parameter has a positive value are penalized, the only remaining triggers for nominative Case assignment under government are Wh VSO orders. This order, too, is only weakly triggered in 15th-century French. The difference with OF is that several new constructions were available, notably complex inversion and (qu')est-ce que (cf. Foulet (1921) on the development of the latter as a non-emphatic interrogative). Nominative-assignment under government has very little triggering data, while Nominative-assignment under agreement receives strong support from the input data. Since the two parameters are in a shifting relationship, there is some pressure (built into the fitness metric, as we saw in Section 2) not to set them both to 1. In this situation, the fact that Nominative- assignment under government was only weakly triggered led to a change in the value of this parameter.

The change to a system with Nominative Case assignment under agreement entailed a change in the null-subject parameter (already only weakly triggered, as (70) show) for theory-internal reasons. As we said earlier, we follow Rizzi (1986a) in assuming that null subjects can only be licensed in positions where Case is assigned. Hence, once Nominative Case could no longer be assigned under government, null subjects could no longer be licensed under government. In this way, we see how French lost null subjects with no significant change in the verbal inflectional morphology. There is a complication here, however; we mentioned above that MidF, unlike OF, also allowed null subjects to be licensed in configurations of agreement. Why were these null subjects lost with those licensed in government configurations? Roberts (forthcoming) answers this question in terms of a postulate concerning the identification of null subjects which we can phrase as follows:

(72) Where null subjects are licensed only in configurations of agreement, they require a ``pronominal" Agr for identification.

A ``pronominal" Agr is an Agr which morphologically distinguishes at least five persons, i.e. an Agr of the kind found in languages such as Spanish and Italian. French Agr is not pronominal in this sense, and indeed has not been since early in the OF period. The intuition behind (72) is that a system where null subjects are licensed under government requires less inflectional morphology for the recovery of the content of those null subjects than one in which the only licensing configuration is agreement, since government is a closer syntactic relation than agreement. A system that licenses null subjects both under government and agreement, like MidF, tolerates a relatively poorer agreement morphology. So, once null subjects could no longer be licensed under government in French, the relative ``poverty" of the verbal morphology became crucial, and null subjects were lost also in contexts where they had been licensed under agreement. As Roberts shows, the parallel development of Gallo-Italian dialects, in particular Veneto, supports the postulation of (72).

Thus, at the beginning of the MidF period (c1300) the relevant parameter-settings were those in (73a) while, by the end of this period (c1500) they were as in (73b):

(73)a. 1 1 0 1 1 (= OF)
b. 1 0 1 0 0 (= ModF)

It is clear that the crucial element of instability was created by the gradual erosion of V2 as a rigid constraint on word-order in matrix declaratives. In particular, the introduction and spread of XSV orders brought about a situation which eliminated a crucial trigger for Nominative-assignment under government---XVS order. The previous discussion shows how the GA approach to learnability, and in particular the fitness metric, can shed some light on this. What seems to have happened is that V2 was mildly unstable in, say, 1300 (cf. the discussion at the end of the previous section) in the sense that non-V2 parses for certain types of sentence, e.g. SVO, were close competitors for V2. These competitors generated ``mutant" word orders, notably XSV, which were highly successful. The critical point is reached in the late 15th century, when V2 is eliminated. For completely contingent reasons (which concern the overall organization of the MidF grammatical system), the loss of V2 led to the loss of Nominative-assignment under government. And for reasons to do with the organization of UG, this entailed the loss of null subjects. Moreover, Roberts (in progress) argues that this in turn led to the loss of clitic-climbing (cf. also Kayne (1989)). This account of syntactic changes in the history of French illustrates how syntactic change can be internally driven: change in one parameter can destabilize another. We will see a further example of this in the next section.

However, we now find ourselves in a classic ''chicken-and-egg'' situation: how were XSV orders introduced into a V2 system? Recall that such orders are ungrammatical in Modern V2 Germanic languages. If we say that the weakening of V2 was a condition for this development, we run the risk of falling into an unproductive regress. It was in part for this reason that we avoided this issue above and simply
took this innovation as given. However, there are good reasons to think that the introduction of XSV order is related to the cliticization of subject pronouns. Adams (1987b) points out that the overwhelming majority of early cases of XSV involved a pronominal subject. As Adams suggests, it is possible that XSV originates from cases of V2 where the clitic subject pronoun is not counted in determining V2[13] If Adams' idea is correct then the initial stimulus for the erosion of V2 comes from a morpho-phonological change in the subject pronouns. As is frequently the case, syntactic change can be traced back to extra-syntactic factors, although the relationship between the extra-syntactic factors and the syntactic changes caused can be extremely indirect. This is because instability, once introduced, can propagate through a grammatical system.

3.5. Some Concluding Remarks

Here we wish to address some of the wider issues that are raised by our case study of language change. These concern the status of the shifting relationship between the Nominative parameters in Section 3.1 with respect to the OF data and what our approach has to say about the classic questions for diachronic linguistics concerning the nature of innovation and loss.

How is it the case that a massively unstable system of parameter settings, as we saw in the case of Middle French, can come into being in the first place? Of course, factors that are external to the syntax, like invasions or phonology, can destabilize a syntactic system, as the history of English amply illustrates. However, as we mentioned at the end of the previous section, we believe that instability can propagate internally to a syntactic system and that exactly this has happened in the history of French. Consider again the P-encodings we arrived at for the OF data:

(74)a. [* 1 * * 1]
b. [1 * * * 0]
c. [* 1 * 1 1]
d. [1 * * 1 0]
e. [* 1 * * *]

Bearing in mind that the correct grammar for OF does not contain non-V2 parses, i.e. that the P-encodings in (74b) and (74d) are discarded in the correct grammar, it seems that Nominative assignment under agreement has a quite precarious status in OF. There is another trigger for Nominative-assignment under agrement, however: the fact that subordinate clauses regularly had SVO order (assuming, contra Lightfoot (1989), that subordinate word-order can trigger parameter settings). Thus, it is the fact that OF had a root/embedded asymmetry with respect to V2 order that is crucial for triggering Nominative-assignment under agreement. Now, there is evidence that early OF (prior to c1200) allowed embedded V2 (cf. Cardinaletti & Roberts (1991), Dupuis (1989), Hirschbuhler (1990)). This means that Nominative-assignment under agreement was an OF innovation, emerging in subordinate clauses as V2 became a uniquely root phenomenon. This innovation started the chain of changes leading to the the MidF innovations that were crucial to our account in the previous section (and hence to the later changes discussed there).

Assume that an archaic stage of OF did not allow nominative Case assignment under agreement. How can Case-assignment under agreement arise? Notice that by innovating Nominative-assignment under agreement, a shifted system is introduced on the basis of a non-shifted one. Following an idea originally due to Cardinaletti (1990), suppose that expletive elements can never topicalize. In a V2 system, however, SpecC′ is a topic position: it is an A′ position and a position which does not receive Case. Cardinaletti proposes that where an expletive occupies this position, as frequently happens in the V2 Germanic languages, the position is able to count as an A-position in that (Nominative) Case can be assigned there. Thus, we can attribute the introduction of Nominative-assignment under agreement to t he introduction of a lexical expletive capable of occupying SpecCP in matrix declaratives. OF had a lexical expletive il which appeared in SpecC′ in exampleses like:

(75) Il ne me chaut.(Einhorn (1974:123))

Supposing that this construction emerged in archaic OF, we can then say that Nominative-assignment under agreement was triggered by this kind of example.

Finally, let us briefly consider what implications our proposals may have for traditional preoccupations of diachronic syntax: the nature of innovation and the nature of loss. Of course, it should be immediately clear that the conception of how grammatical systems differ from one another that lies at the heart of the P&P approach means that parameters themselves never change.[15] What changes over time are parametric values, as we said at the outset.

Nevertheless, at the level of constructions, e.g. available word-order types, it is clear that possibilities are both innovated and lost. In our terms, innovations may arise from one of two sources: either internally, when a parametric change makes new constructions available, or externally, when phonological or morphological change weakens evidence for certain hypotheses. The second type of innovation is likely to lead to instability at the level of the parameter settings; this is what we saw in the case of the introduction of XSV orders triggered by the cliticization of subject pronoun in MidF.[16]

Concerning loss it seems that only parametric change can truly eliminate a construction in the sense that construction C is accepted by native speakers of L at time T and rejected at T′ (T > T). This has been the fate of simple inversion, V2 and null subjects in French. In terms of the standard view of language acquisition, this situation seems problematic. Put very simplistically: why is one generation's trigger experience the next generation's fossil? Various answers to this question have been proposed, but we believe we have a new and interesting one.

An approach to learnability based on a GA comprising a version of the fitness metric makes it possible to see how a data point can be disregarded in a situation of instability (where instability can be formalized); this was what we saw in the case of XVS orders in 15th-century French. Although relatively infrequent and often parsable as some other construction, XVS is certainly found in 1500, and so, given the standard assumption that parameters can be set on the basis of quite impoverished experience, an account of loss based on frequency considerations alone will not answer the fundamental question. The fitness metric, properly formulated so that frequency and other considerations are taken into account, seems able to resolve this tension between the fact that structures are lost in the course of language change and standard views of acquisition since we are able to see why one class of input strings may be rendered unparsable. This can happen even where, as in the case XVS orders, the input in question is intrinsically simple and structurally ''transparent''; this is a major difference between our account and the approach to language change in Lightfoot (1979).

A further important consideration which emerges from our discussion is that exactly the same string S_i can be a successful trigger for a parameter setting $p(v_j)$ in one grammatical system G_i, but fail to succesfully trigger $P(v_j)$ in system $G_j \neq G_i$. French XVS order is a case of exactly this sort, where G_i is the grammar of OF and G_j that of late MidF. In terms of the GA, S_i can trigger a successful hypothesis or an unsuccessful one. As in the biological world, sucessful propagation depends as much on the external environment as on internal properties, and so little can be predicted purely on the basis of internal structural criteria. It is this aspect of the GA which makes possible a deeper understanding of language change. Note also that language change, in these terms, does not refer only to the ''limit cases" of innovation and loss, but also to the varying success of strings in encoding viable parameter settings.

Our approach also has implications for the theory of markedness. It is part of the classical concept of markedness that marked properties are both diachronically unstable and ''difficult'' in terms of acquisition. A shifted system of parameter settings can be thought of as a marked system. It is clear from our discussion that a shifted system is diachronically unstable. Consider again the shifted system that we discussed in Section 2, one featuring both V2 and left-dislocation. Neither V2 nor left-dislocation is marked on its own (cf. the stability of Germanic V2 and the fact that all periods of French feature left-
dislocation of one kind or another); however, their combined presence in a system leads to markedness -- witness the instability of MidF[16]. So we suggest that in general markedness, rather than being an inherent property of certain parameter values, is a property that derives from the interaction of parameters in a given grammatical system, relative to the fitness metric. This in turn implies that a given parameter-value can be marked in one grammatical system (or at one period), and unmarked in another system (e.g. at another period).

Diachronic studies of the type discussed here also have important implications for the study of language learnability and language acquisition. As discussed briefly above, diachronic change represents a type of ''pathological'' learning, where learners systematically arrive at the wrong grammar for the target language. Strictly speaking, these are cases where learners fail. We would argue that learners fail for reasons which reveal something important about their internal structure. As we have seen above, parametric change is the result of an input text which places indifferent pressure on the learner's hypotheses; several different grammars can provide an acceptable account for the input text. At this point, we have seen that other factors, always related to the learner's internal fitness metric, come into play to distinguish between the competing hypotheses. These factors involve the Subsect Condition and a measure of elegance. Let us return to the fitness metric, repeated here as (76):

(76) The Fitness Metric

$$\frac{(\Sigma^{n}_{j=1} v_j + b\, \Sigma^{n}_{j=1} s_j + c\, \Sigma^{n}_{j=1} e_j) - (v_i + bs_i + ce_i)}{(n-1)(\Sigma^{n}_{j=1} v_j + b\, \Sigma^{n}_{j=1} s_j + c\, \Sigma^{n}_{j=1} e_j)}$$

Our study of diachronic change reveals certain facts about the scaling constants b and c . We assume that empirical coverage of the input text is the learner's central interest; thus, violations (calculated by $\Sigma^{n}_{j=1} v_j$ for the population and by v_i for the individual) is the single most important factor in the equation. Both superset settings and elegance are scaled down by the constants <u>b</u> and <u>c</u> , respectively.

Let us now consider what the relative magnitudes of <u>b</u> and <u>c</u> are. We saw in section 3.4 that, at a certain point, French was a V2 language which allowed for left-dislocation (the latter associated with atonic pronouns) and that this was a shifted language which would be selected against by the fitness metric. Furthermore, the relative frequency of structures which would have required both V2 and left-dislocation were relatively low, placing little pressure on the learner in terms of violations. All else being equal, learners could have preferred either a language with matrix V2 and no left-dislocation or a language with left-dislocation and no matrix V2. Notice that left-dislocation is a superset parameter; a language which allows left-dislocation in addition to its basic word order is a superset of a language which allows only the basic word order. We argued, on the other hand, that matrix V2 led to more complex representations, relative to the input text, than a grammar without matrix V2.

Now, the changes we have seen in French involve the abandonment of matrix V2, a non-superset parameter, and the persistance of left-dislocation, a superset parameter. Given our premises, then, the fitness metric must have preferred a grammar which generated an elegant set of representations and a superset language over a grammar which generated inelegant representation and a subset language. Thus, learners appear to consider elegance a more important factor than superset settings when evaluating hypotheses:

(77) c > b

Thus, our study of diachronic change has enabled us to make a concrete hypothesis about how learners evaluate parameter settings. We can now test this hypothesis against actual child grammars, perhaps by attempting to characterize successive developmental stages in child language. In general, we should see children avoiding grammars which create inelegant representations. More to the point, children will resist grammars which force longer chains to the point of, temporarily at least, preferring grammars with superset settings if these latter grammars can approximate the target.

We have seen how a theory of language learning based on a genetic algorithm affords a novel and insightful account of language change, taking as our case study of language change the development of word order and null subjects in French. We believe that our account sheds light on the both the mechanisms of language change and those of language acquisition, and goes some way towards building a bridge between these two domains. Moreover, we have seen that it is possible to characterize the markedness of systems, and to clearly see the role played by such factors as elegance and frequency of input, and the interactions between these factors. We know of no other approach to language learnability and language change which achieves these results.

NOTES

* The first author received support from grant 11-25362.88 from the Fonds national suisse pour la recherche scientifique.
1. The first person to formulate this problem in terms of generative syntax was Lightfoot (1979).
2. Genetic algorithms were developed by Holland; see particularly Holland (1975). Goldberg (1989) provides a

comprehensive overview of the technique; see also Booker, Goldberg & Holland (1990). Clark (1990) develops a model of parameter setting in terms of GAs as an approach to demonstrating the learnability property.

3. We will not enter into an extensive discussion of this characterization of the input text; interested readers may consult Wexler & Culicover (1980) for an extensive discussion and defense of this position.

4. We will grant, following Morgan (1986) that intonation, for example, can provide the learner with some information about structural analyses. We would maintain, however, that the proper set of parameter settings are still massively underdetermined even if the learner has access to phonological cues.

5. As we shall see, below, shifting is more than a logical possibility and serves to force parametric change over time.

6. Space prevents a comprehensive discussion of this class of algorithms; see Goldberg (1989) for a general introduction and Clark (1990) for an application to the learnability problem for natural languages.

7. The results discussed here receive a more formal discussion in Clark (1991) where proofs of certain theorems entailed by the fitness metric are given. For present purposes, the important point is that, relative to an input text, the fitness metric drives the learner toward a hypothesis which minimizes the number of violations and the number of superset settings and which generates the most elegant syntactic representations possible, given that grammatical violations are avoided.

8. Vance (1989) in fact shows that 15th century MidF null subjects could also be licensed under agreement as well as government. Nevertheless, both null subjects licensed under government and null subjects licensed under agreement are lost with simple inversion in the 16th century. Roberts (forthcoming), section 2.4.3, proposes that the loss of null subjects where they were licensed under government also entailed their loss throughout the system on the basis of the idea that, for null subjects to be licensed only under agreement, a very rich ``pronominal'' morphology is required. This type of morphology is found in Italian or Spanish but not in MidF or ModF. Hence, the ``poverty'' of French agreement, combined with the change in the Nominative parameter, led to the loss of null subjects everywhere. We will discuss Vance's data further below.

9. In our presentation, we abstract away from the ``split Infl'' hypothesis of Pollock (1989), restricting ourselves to projections of I. To fully account for the facts of Modern

French inversion, however, it is necessary to split I into at least Agr and T (and their projections). In terms of the ''Agr over T'' system of Belletti (1990), our Nominative parameter refers to Agr. To account for stylistic inversion, we probably need to say that T can assign Nominative to a post-verbal subject under government (cf, Rizzi, 1990). See next note.

10. Literary Modern French allows strings which appear to be V2, eg, Dans cette maison vécut Racine. However, such examples should be treated as stylistic inversion. Stylistic inversion differs from V2 and subject-clitic inversion in that the subject appears in a position following the entire verbal complex in a compound tense and it is not sensitive to the root/embedded distinction, unlike true V2. See Kayne & Pollock (1978) and Pollock (1986). Pollock (1986), in fact, suggests that stylistic inversion may involve a non-referential null subject in SpecI'. If so, Modern French allows at least some highly restricted occurences of null subjects and, so, (32d) should be reformulated to refer to referential null subjects.

11. In the case of OF, as in all languages now lacking in native speakers, negative evidence in the form of grammaticality judgements is unavailable. The linguist working on such languages is in a situation which is almost analogous to that of the child acquiring his native language, although in fact the linguist's situation is worse since he has no access to UG and his data is seriously degenerate owing to dialect mixture, scribal error, etc. Unlike the child, however, the linguist has no access to a regular input text. The child is surrounded by native speakers producing grammatical utterances. The linguist is not, since all the native speakers are dead.

12. In fact, there are reasons to think that in the position immediately following the inflected verb, as in (48b), these pronouns did cliticize in OF (cf. Dupuis (1989:119f.), Roberts (forthcoming; 2.2.2) and Vance (1989:70ff.)). However, Roberts argues that the crucial step in the development of the system of subject pronouns in French was the emergence of complementary distribution between the <u>je</u>-series and the <u>moi</u>-series. This happened because the cliticization of the je-pronouns became obligatory in MidF. What the OF evidence shows is that these pronouns were optionally clitics in that they cliticized only in certain contexts. In other contexts, e.g. those in (55) and (56) below, these pronouns were clearly tonic. It may be, then, that the correct formulation of the parameter in (32c) should refer to obligatory cliticization of nominative

pronouns, or, more likely, to the existence of a special series of clitic pronouns. Note that in the latter case the trigger for the parameter is morphological: the learner must recognize two paradigms of subject pronouns.

13. We do not want to propose that preverbal subject pronouns in MidF or ModF are syntactic clitics; rather, following Kayne (1983), we believe that these pronouns cliticize only in PF. However, the ultimately unsuccessful hypothesis that these pronouns were indeed syntactic clitics could nevertheless have given rise to XSV orders at the time when the subject-pronoun system was undergoing change.

14. Except perhaps at the higher diachronic level of phylogenetic change; it is a reasonable assumption that the set of parameters available to modern homo sapiens is not the same as the set that was available to the first hominids with a language faculty. Of course, we are concerned in the text with changes in the recorded history of languages which by assumption fall within the set of human languages, so this question does not arise.

15. There is at least a metaphorical sense in which cases like XSV are successful rogue hypotheses, where success is determined by the least-effort criterion. This is mutation at the level of constructions, not at the level of parameters, so the mutation operator of Section 2 is presumably not relevant.

16. Modern German also has left- dislocation, but with a tonic resumptive pronoun. On the other hand, MidF left-dislocation featured atonic resumptive pronouns. So we see that a further way in which the clitic nature of pronouns in MidF created instability.

REFERENCES

Adams, M.(1987a) ''From Old French to the Theory of Pro-Drop,'' Natural Language and Linguistic Theory 5, 1-32.

Adams, M.(1987b) Old French, Null Subjects and Verb Second Phenomena, PhD Dissertation, UCLA

Baker, M. (1988) Incorporation: A Theory of Grammatical Function Changing, University of Chicago Press.

Belletti, A. (1990) Generalized Verb Movement: Aspects of Verb Syntax, Rosenberg & Sellier, Turin.

Berwick, R. (1985). The Acquisition of Syntactic Knowledge. The MIT Press, Cambridge, MA.

Booker, L. B., D. E. Goldberg & J. H. Holland (1990). ''Classifier Systems and Genetic Algorithms'' in J. Carbonell (ed) Machine Learning: Paradigms and Methods. The MIT Press,

Cambridge, MA.
Cardinaletti, A. (1990) Pronomi nulli e pleonastici nelle lingue germaniche e romanze: Saggio di sintassi comparata, Dottorato di Ricerca in Linguistica, Università di Padova.
Cardinaletti, A. & I. Roberts (1991) ``Clause Structure and X-second,'' to appear in W. Chao & G. Horrocks (eds) Proceedings of the Thirteenth GLOW Colloquium, Foris, Dordrecht.
Chomsky, N. (1977). ``On Wh-Movement'' in P. Culicover, T. Wasow & A.Akmajian (eds) Formal Syntax. Academic Press Inc., New York.
Chomsky, N. (1981). Lectures on Government and Binding. Foris Publications, Dordrecht, the Netherlands.
Chomsky, N. (1989) ``Some Notes on Economy of Derivations and Representations,'' in I. Laka and A. Mahajan (eds) MIT Working Papers in Linguistics, Volume 10, MIT, Cambridge, Mass.
Clark, R. (1990). Papers on Learnability and Natural Selection. Technical Reports in Formal and Computational Linguistics, No. 1. Université de Genève.
Clark, R. (1991). ``The Selection of Syntactic Knowledge''. manuscript, Université de Genève. in preparation.
Dupuis, F. (1989) L'expression du sujet dans les subordonnées en ancien français, thése de PhD, Université de Montréal.
Einhorn, E. (1974) Old French: A Concise Handbook, Cambridge University Press.
Everett, D.(1986) ``Piraha Clitic Doubling and the Parametrization of Nominal Clitics,'' in T. Rapoport, N. Fukui and B. Sagey (eds) MIT Working Papers in Linguistics Volume 8, MIT, Cambridge, Mass., 85-127.
Everett, D.(1989) ``Clitic Doubling, Reflexives and Word Order Alternations in Yagua,'' Language 65, 339-372.
Foulet, L.(1921) ``Comment ont évolué les formes de l'interrogation?'' Romania 47, 243-348.
Foulet, L. (1982) Petite syntaxe de l'ancien français, Editions Champion, Paris (3eme édition).
Goldberg, D. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Company, Reading, MA.
Haldane, J. B. S. (1990). The Causes of Evolution. Princeton Science Library, Princeton University Press, Princeton, NJ.
Hirschbuhler, P. (1990). ``La légitimation de la construction V1 à sujet nul dans la prose et le vers en ancien français,'' Revue québécoise de linguistique, 19, 32-55.
Holland, J. (1975). Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI.

Kayne, R. (1975) French Syntax, MIT Press, Cambridge, Mass.
Kayne, R. (1983) ``Chains, Categories External to S, and French Complex Inversion,'' Natural Language and Linguistic Theory 1, 109-37.
Kayne, R. (1989) ``Null Subjects and Clitic Climbing,'' in Jaeggli & Safir (1989), 239-261.
Kayne, R. & J.-Y. Pollock (1978) ``Stylistic Inversion, Successive Cyclicity and Move-NP in French,'' Linguistic Inquiry 9, 595-621.
Kiparsky, P. (1982). Explanation in Phonology. Foris Publications, Dordrecht, the Netherlands.
Koopmann, H. & D. Sportiche (1990) ``The Position of Subjects,'' to appear in Lingua.
Kroch, A. (1989). ``Reflexes of Grammar in Patterns of Language Change''. Language Variation and Change, 1, pp. 199-244.
Lightfoot, D. (1979) Principles of Diachronic Syntax, Cambridge University Press.
Lightfoot, D. (1989) ``The Child's Trigger Experience: Degree-0 Learnability,'' Behavioral and Brain Sciences 12, 321-334; commentary 334-375.
Marchello-Nizia, C. (1979) Histoire de la langue française aux XIVe et XVe siècles, Bordas, Paris.
Morgan, J. (1986). From Simple Input to Complex Grammar. The MIT Press, Cambridge, MA.
Paul, H. (1920) Prinzipien der Sprachgeschichte, Niemeyer, Halle (5th Edition).
Poletto, C. (1990) ``Diachronic Development of Subject Clitics,'' talk given at the Crucial Languages Workshop, Université de Genève.
Pollock, J.-Y. (1986) ``Sur la syntaxe de en et le paramétre du sujet nul,'' in M. Ronat and D. Couquaux (eds) La Grammaire Modulaire, Editions de Minuit, Paris, 211-246.
Pollock, J.-Y. (1989) ``Verb Movement, UG and the Structure of IP,'' Linguistic Inquiry 20, 365-424.
Price, G. (1971) The French Language: Present and Past. Edward Arnold, London.
Priestley, L. (1955). ``Reprise Constructions in French''. Archivum Linguisticum, 7, pp. 1-28.
Renzi, L. (1983) ``Fiorentino e italiano: storia dei pronomi personali soggetto,'' in F. Albano Leoni et al (eds) Italia linguistica: idee, storia, struttura, Bologna, 223-239.
Renzi, L. & L. Vanelli (1983) ``I pronomi soggetto in alcune varietà romanze,'' in Scritti Linguistici in onore di Giovan Battista Pellegrini, Pisa, 121-145.

Rizzi, L. (1986a) ``Null Objects in Italian and the Theory of pro," Linguistic Inquiry 17, 501-557.
Rizzi, L. (1986b) ``On the Status of Subject Clitics in Romance,'' in O. Jaeggli and C. Silva-Corvalàn (eds) Studies in Romance Syntax, Foris, Dordrecht, 391-419.
Rizzi, L. (1990) Relativized Minimality, MIT Press, Cambridge,Mass.
Rizzi, L. & I. Roberts (1989) ``Complex Inversion in French,'' Probus 1, 1-30.
Roberts, I. (forthcoming) Verbs and Diachronic Syntax, Kluwer, Dordrecht.
Roberts, I. (in progress) ``Head Government and the Local Nature of Head Movement,'' ms. Université de Genève/University of Wales, Bangor.
Schwartz, B. & S. Vikner (1989) ``All Verb Second Clauses are CPs,'' Working Papers in Scandinavian Syntax 43, 27-49.
Thurneysen, R.(1892) ``Die Stellung des Verbums im Altfranzosischen," Zeitschrift für Romanische Philologie 16, 289-371.
Vance, B. (1989) Null Subjects and Syntactic Change in Medieval French, PhD Dissertation, Cornell University.
Vanelli, L., L. Renzi & P. Benincà (1986) ``Typologie des pronoms sujets dans les langues romanes," Actes du XIe Congrés de Linguistique et Philologie Romanes, Aix-en-Provence.
Vincenzi, M. (1989) Syntactic Parsing Strategies in a Null Subject Language, PhD Dissertation, University of Massachusetts, Amherst.
Wexler, K. & P. Culicover (1980). Formal Principles of Language Acquisition. The MIT Press, Cambridge, MA.