Combinatorial Variability*

David Adger

1 Introduction

The purpose of this paper is to provide a plausibility argument for a new way of thinking about intra-personal morphosyntactic variation. The idea is embedded within the framework of the Minimalist Programme, and makes use of notions of feature interpretability and feature checking. Unlike many current approaches to this kind of variation (involving Multiple Grammars (Kroch 1994) or building stochastic weighting into the grammar; see Manning 2003 for an overview), the system attempts to predict (rather than capture) frequencies of variants. It does this by combining an evaluation metric for the acquisition of uninterpretable features with the standard

*This work is a development of ideas from an earlier paper by Jennifer Smith and myself, so I hereby include the acknowledgments from that paper (Adger and Smith 2005). In addition, I'd like to thank participants in my classes at Tromsø(October 2004) and at the LOT Winterschool in Groningen (January 2005), and members of the Linguistic Research Group at Queen Mary for comments which have really helped my thinking about these issues. More specific thanks are due to an anonymous JoL referee, Jenny Cheshire, Daniel Harbour, Janne Johannesen, Fritz Newmeyer, Jeff Parrott, Andrew Radford, Gillian Ramchand, Michal Starke and Rob Truswell, and of course to Jennifer Smith, without whose fantastic empirical work and socio-linguistically informed comments none of this would have been possible.

properties of features and syntactic operations in the Minimalist framework.

To show that all intra-personal morphosyntactic variation works in this way is too large a task for a single journal article. What I want to argue here is that the approach is at least plausible; I show this via an account of an interesting pattern of *was/were* variability in a Scottish Dialect (Smith 2000).

2 Variability

What does it mean for something to be variable? The usual notion is that a single unit (at some level of abstraction) can come in a variety of forms; so, for example, we might think of pea-plant seeds showing variation in whether they are smooth or wrinkled, or clover varying in whether it has three leaves or four. The variation in form can be thought of as involving categories which are either discrete (how many leaves) or continuous (perhaps level of wrinkledness).

The notion of variation in form in linguistics is similar: we have a single unit (say a phoneme) which has a range of forms (allophones). It is standardly assumed that variants arise as a function of their syntagmatic context: unvoiced stops in English are realised in two ways (aspirated or not) depending on whether they are preceded by a sibilant, and whether they are in a stressed syllable. The particular variant is the deterministic result of the featural specification of the phoneme and its context. Moreover, the context is a linguistic representation. Schematically, we can represent variation in this sense as follows:

(1)

$$\alpha \rightarrow \begin{cases} a\_\_C_1 \\ A\_\_C_2 \\ \aleph\_\_C_3 \end{cases}$$

If this were all one needed to say, then one could claim that there is no non-deterministic variation in the phonological component of the grammar. However, even restricting ourselves to the variants of unvoiced stops, we need to allow some optionality of aspiration, since in unstressed syllables which are not preceded by sibilants, aspiration is possible but not required, giving rise to what used to be termed “free” variation. I will call such non-deterministic variation **variability**, and it is what this paper is mainly concerned with.

The kind of rule based approach to variation (and variability) just exemplified has also been used to deal with variants in morphology: particular allomorphs are chosen depending on their morpho-syntactic (or morpho-phonological) contexts. I'll call this kind of variation in linguistic form, where a single category has a range of variants determined by rule, Variation in Exponence (VE). Most theories of variation in linguistics treat variability as VE.

The kind of viewpoint outlined above makes the assumption that either variants are *determined* by their context or they are in free variation. If we abandon this assumption, and assume instead that variants may merely be *influenced* by context, then we apparently require some notion of probability: what is the probability that we will find variant x in context C, and how does this interact with the influence of other contexts? At this point we have moved some way towards the kind of approach defended by Labov in his early work (under the rubric of variable rules—

Labov 1969, Labov 1972), which still, at an appropriate level of abstraction, takes a Variation in Exponence perspective. However, within that approach, empirical findings showed that it was not only properties of the linguistic representation that were relevant as influencing factors for the variants; social factors and processing factors also turn out to be crucial—that is factors pertaining to *use* enter into the functioning of linguistic rules (if we maintain this conception of what a linguistic rule is). The variable rule approach can be thought of as allowing an extension of the notion of context in a Variation in Exponence approach beyond purely linguistic representations. It quickly follows, on such assumptions, that the competence performance distinction is difficult to maintain. For this reason (and for other reasons), generativists, convinced by the explanatory and methodological efficacy of the competence performance distinction, have tended to maintain the idea that the grammar determines structure rather strictly, and that probabalistic factors are to be excluded from the grammar per se (see, for example, Newmeyer 2003).

The argument for including usage based factors in the grammar is based on the assumption that the real source of variation is the interaction of form and context, with the impact of context cast in probabalistic terms. However, there is another potential source of variation. Turning back to our discussion of variation in biology, it is certainly true that aspects of form are determined by the organism's interaction with its context (environment). However, as Mendel showed a few centuries back, variation in form also arises because of the combinatorial mechanics of discrete elements—genes. What Mendel further showed was, not only that variation was (at least partially) the result of genetic (re)combination, but also that the *frequencies* of the variants found in nature could be predicted by the ways that genes combined:

genetic heredity produces variants in a way that depends on the notion of recessive and non-recessive genes and on the mechanism of gene combination.

What I want to show in this paper is that the combinatorial mechanisms that create structure in language from discrete elements also give rise to variability in the sense we are interested in here: the non-deterministic choice of form. Moreover the analogies with Mendelian genetics are rather sharp: I will argue that the grammar produces variants in a way that depends on the notion of interpretable and uninterpretable features and on a major mechanism of syntactic combination (Agreement).

With this in place, I will show how, once one controls for the effect of context, not only do we predict the possibility of stable variability within the grammar, but, on the minimal assumption of random choice of equivalent lexical items, we predict the correct frequencies of variants found in corpora, even though we will not incorporate probabilities into the grammar itself. I will term this kind of variation Combinatorial Variability (CV). I will not make an argument here that CV is the only source of variation, but rather one for the weaker claim that we at least need something like CV. I make this argument by showing that, at least for the case study presented here, the VE model doesn't supply us with as satisfying an explanation as the CV approach. The question then is whether we need VE in addition to CV in the more general case. The genetic analogy suggests that we do, since phenotypic variation is dependent on both genetic and environmental factors—however an analogy is just an analogy, and I will not pursue this question further here.

If the approach I defend here is tenable, then we have a clear rapprochement between transformational generative grammar and variationist sociolinguistics: the grammar produces variants in a way that predicts particular probability distributions, however, those probabilities can be perturbed at the point of use by factors such as ease of lexical access, recency effects, meta-linguistic or social judgments on the form etc (see Kroch 1994, Henry 1995, Wilson and Henry 1998, Bender 2001, Bresnan, Dingare, and Manning 2001, Manning 2003, Anttila 1997, Anttila and Cho 1998, for previous approaches in Government and Binding Theory, Minimalism, HPSG, Stochastic OT and Stochastic LFG). Of course, the idea that the output of the grammar interacts with the performance mechanisms is what has always been assumed by generativists from the earliest work, but what I hope to show here is that we can embed variability into the grammar itself, making predictions about frequency of occurrence of particular forms purely as a function of the architecture of the grammatical theory postulated.

3 Some background assumptions

3.1 Lexical Items and Features

Current thinking within the Minimalist Program (Chomsky 1995, Chomsky 2000a, Chomsky 2001) assumes that complex syntactic objects are built up from the combination of atomic objects using the syntactic operation Merge (with Movement considered as a version of Merge which targets material already constructed in the derivation). These atomic objects are drawn from what is essentially a memorized set (the lexicon), and each element of the set (that is, each lexical item, LI) is itself built up out of a combination of atomic objects (features). The combination of features to make lexical items is assumed to be the construction of a simple set, rather than anything more complex. There is a debate about how to best think of features. I will adopt the idea here that they are always bivalent in nature, although I don't think anything turns on this (that is, I think that it is possible to do what I suggest here with a bigger set of privative features, perhaps structured into a geometry, or with multivalent features, as in Adger 2003).

The assumption that lexical items are unstructured sets of features, which are just memorized by the language learner raises a further question: how does the learner decide on the features? Assume the learner has a conceptual space that she needs to ascertain the grammar of (say the space of pronominals). The conceptual structure of human thought provides her with a range of possible analyses, in terms of semantically motivated notions such as number, participant of the speech act etc. Some subset of these will be available to reify grammatically as a set of features.

A bivalent feature is just a feature that captures contrasts: the child hears particular forms used for particular purposes and makes a decision about the relationship between conceptual structures and syntactic features. For example, the feature [singular:±] can be used to classify pronouns into singulars and plurals. The particular forms used in the language surrounding the acquirer will alert her to the contrasts that are relevant for the language (so if a language has a dual and paucal number, the feature [singular:±] will not be sufficient to capture the relevant number contrasts). It may be that the available features are completely determined by UG, or it may be that the space of possibilities is circumscribed by UG, and the actual features are determined on the basis of the evidence plus some general categorization algorithm. Either of these views is compatible with what I argue here.

3.2 Pronouns

I will assume the following three features for the analysis of pronouns in English (see, e.g. Harley and Ritter 2002):[1]

(2) [singular:±]; [participant:±]; [author:±]

The feature [singular:±] marks the number of the pronoun. Since there is no dual number syntactically or morphologically marked in English, I will assume that this feature is sufficient. The feature [participant:±] marks whether the pronoun refers to a participant of the speech act (the speaker or author) or not. The final feature [author:±] allows us to distinguish between speaker and author. Having a specification for [author:±] entails having a positive specification for [participant] (that is, it is not possible to be classified for whether one is an author or not if one is not a participant of the speech act). I will also assume that if a pronoun is specified as [participant:+], it bears a specification for [author], at least in English, since there are no pronominal forms in English which do not distinguish between speaker and author. So we have the following general restriction on ϕ-feature on pronouns:[2]

(3) Feature Co-occurrence Restriction: a lexical item is specified for [participant:+] iff it has a specification for [author].

[1] The set of features required in other languages may be larger, to handle cases of dual, inclusive/exclusive distinctions, etc. I ignore gender throughout, although interesting questions arise about the semantic interpretation of this feature, and how it fits into the general system outlined here.

[2] This FCR is intended to derive from the semantics of the features, and is hence irrelevant for the uninterpretable features discussed immediately below.

This means that third person pronouns in English are never specified for the feature [author]. It is as irrelevant to their syntax as, say, [past] is.

Given this, we have the following set of pronouns in English:

(4)

$\begin{bmatrix} \text{singular:}+ \\ \text{participant:}+ \\ \text{author:}+ \end{bmatrix}$ I	$\begin{bmatrix} \text{singular:}- \\ \text{participant:}+ \\ \text{author:}+ \end{bmatrix}$ we
$\begin{bmatrix} \text{singular:}+ \\ \text{participant:}+ \\ \text{author:}- \end{bmatrix}$ you	$\begin{bmatrix} \text{singular:}- \\ \text{participant:}+ \\ \text{author:}- \end{bmatrix}$ you
$\begin{bmatrix} \text{singular:}+ \\ \text{participant:}- \end{bmatrix}$ he/she/it	$\begin{bmatrix} \text{singular:}- \\ \text{participant:}- \end{bmatrix}$ they

3.3 Agreement

Certain lexical items carry features which are purely formal in nature: their 'job' is to establish syntactic dependencies. We will call such features **uninterpretable**, following Chomsky (1995) and notate them with a prefixed *u*, following Pesetsky and Torrego (2001). These purely formal features are not associated with a semantic interpretation directly, but they have to be in an agreement relation with semantically interpreted features, or else the structure is ill-formed. This idea is implemented in various ways in current syntactic theory. I will choose a rather neutral implementation of the idea which will be sufficient for our purposes here:

Let us define an object, an agreement-chain:[3]

(5) An agreement-chain is a pair of lexical items, where the uninterpretable features of one LI are a subset of the interpretable features of the other.

For example, the following pair of lexical items will be an agreement chain:

(6)

$$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{author:+} \end{bmatrix} \cdots \begin{bmatrix} u\text{singular:+} \\ u\text{participant:+} \\ u\text{author:+} \end{bmatrix}$$

We then state a filter over the syntactic representations that interface with the semantic systems:

(7) Full Interpretation: every uninterpretable feature must be in (a lexical item in) an agreement chain.

[3] I abstract away here from questions of representation and derivation (see, for example, Brody 1997). I also abstract away from the question of whether uninterpretability is a property that can be derived from whether a feature is lexically valued or not (that is, whether it bears a specification for plus or minus before it enters the syntactic systems). In current minimalist theory, the assumption is that features do not enter the syntax with a specification of their value, but rather that they receive this during the syntactic computation (Chomsky 2001). What I say here is compatible with this approach, but I have chosen to implement the ideas here using fully valued features for simplicity of presentation.

We also need to work out how two lexical items come to be in an agreement-chain: they must at least be in a c-command relation, and moreover they must be appropriately local to each other. The details of these restrictions are not relevant here and I will ignore them in what follows. These ideas will now allow us to rule out examples like the following, for particular individual grammars:

(8) *He were there.

This sentence is ungrammatical in my own idiolect (although not in many dialects of English). Its ill-formedness follows from what we have said so far together with the specification that *were* (in my grammar) has the specification [*u*singular:–]:

(9) He[singular:+, participant: -] were[*u*singular:–, ...] ...

The [*u*singular:–] specification on *were* is not in an Agreement chain, and hence violates Full Interpretation.

4 A schematic overview

Let's see how this approach to grammar allows us to capture variability in a different way from simply listing variant realisations of an underlying form (VE). Take a structure where we have a lexical item LI_1 bearing interpretable features F_1,F_2 and F_3. It's immediately clear that such a lexical item will be able to combine with a range of other items bearing different subsets of uninterpretable versions of these features:

(10)

$$
\begin{aligned}
& & LI_2\{uF_1\} &\rightarrow PF(LI_2) = x \\
LI_1\{F_1, F_2, F_3\} \ldots & & LI_3\{uF_2\} &\rightarrow PF(LI_3) = y \\
& & LI_4\{uF_3\} &\rightarrow PF(LI_4) = z
\end{aligned}
$$

LI_1 can combine with LI_2, with the result that the uninterpretable feature $u\mathrm{F}_1$ on LI_2 will be in an agreement chain. This is all that is required for the wellformedness of this structure, since all of the features of LI_1 are interpretable. If the final phonological form of LI_2 is x (symbolized above by $\mathrm{PF}(\mathrm{LI}_2) = \mathrm{x}$), then we have a final representation with x in it. However, exactly the same derivation holds for LI_3, except that in this case the relevant uninterpretable feature is F_2, and the phonological form associated with LI_3 is y. The final representation will contain a y, rather than an x. The same thing holds, mutatis mutandis, for LI_4 and z. It is important to see that the array of *interpretable* features in all three representations is exactly the same even though their phonological forms are different, so the meaning associated with both representations, which is determined by the semantically interpretable features, is exactly the same.[4] This system, then, allows variability to arise in the combinatorial system. What allows the variability is the possibility that particular lexical items may be underspecified for the uninterpretable agreement features that they contain. This underspecification is irrelevant to the semantic systems, since these features are not interpreted.

This system also predicts something about frequencies: if there is a random choice of which LI is entered into the system, that we should find x, y and z in equal

[4]In this sense we provide a definition of the Labovian notion of (morpho-syntactic) linguistic variable in terms of interpretable and uninterpretable features. See further Adger and Smith 2005.

proportions. However, if some of the PF outputs of the lexical items are the same, we predict a disproportionality in the final output:

(11)

$$LI_1\{F_1, F_2, F_3\} \ldots \begin{array}{l} LI_2\{uF_1\} \rightarrow PF(LI_2) = x \\ LI_3\{uF_2\} \rightarrow PF(LI_3) = x \\ LI_4\{uF_3\} \rightarrow PF(LI_4) = z \end{array}$$

Here we have two ways that the grammar can output an x, but only one way to make a z. We therefore predict a statistical variance in the output, such that we will find x more often than z.

In this kind of system, there is also another way to the result that x is more common than z. In (10) and (11), I have assumed that there is a random choice of lexical items (that is, that there is an equal probability that any of the three lexical items is chosen); however, this is almost certainly not true in all cases. Choice of a lexical item by a speaker in any particular utterance is potentially influenced by social and/or psychological factors, so tat a particular lexical item may have a higher probability of being chosen in a particular utterance (for example, if that lexical item has been recently accessed, it may be easier to access again; or if a lexical item is just more frequent, it may be easier to access). So we could have a situation just like that in (10), but with x being chosen more commonly than z for extra-grammatical reasons.

Under the model developed here, any measure of the frequency of a variant in a corpus will be a result of two different factors: the input probability of that lexical item, and its combinatorial effect. To determine an empirical measure of the latter, we need to control for the former. See the discussion in section 6.4.

Assuming we can control for input probabilities, what we have seen here is that the combinatorics of the syntactic system itself, working on the featural specifications of lexical items, predicts not only variability, but also particular frequencies of surface variants.

The system also allows us to maintain a fairly strict competence performance distinction while still allowing variability, and indeed while predicting the frequencies of variants. The argument that the existence of robust variability impacts negatively on the reasonableness of the competence performance distinction clearly does not hold on this view of the etiology of variability.

The underlying idea, then, is that the grammatical system produces variants as a result of the featural specifications of lexical items and their mode of combination in the syntax. In what follows, I show this idea at work in the analysis of *was/were* variability in a particular dialect, drawing on the important work of Smith 2000 (see also Adger and Smith 2005 for an earlier analysis of this data).

5 Combinatorial Variability in Action

5.1 Basic Data from Buckie

Buckie is a small fishing town situated on the coast 60 miles north of Aberdeen in Scotland. It is quite isolated in both geographic and economic terms and therefore remains relatively immune to more mainstream developments in English. As with similarly isolated communities, this is reflected in the linguistic behaviour of the community (Smith 2000). The data was collected using a standard sociolinguistic methodology (Labovian sociolinguistic interviews—again see Smith 2000 for this and other methodological details), is highly vernacular in nature and amounts to approximately forty hours of tape-recorded casual conversations; these were fully transcribed and the corpus consists of over 300,000 words. The speakers in the sample were born and raised in the community, and indeed the majority come from families who have been in the town for generations. They are working class and exhibit networks that were generally confined to the community in question. The speaker sample is shown in Table 1 (see further Smith, 2000).

Table 1: Speaker sample

age range	male	female
22-31	8	8
50-60	7	7
80+	4	5

The local vernacular displays a huge range of non-standard phonological, morphological and syntactic phenomena. I focus here on the analysis of variability in the form of the past tense copula/auxiliary *be* with pronominal subjects, reported by Smith. Examples are given below, which show categorial versus variable use of the form of past tense *be*:

(12) a. So when **I was** cleaning at Christmas, ... (g:165,10)

b. **I was** aie running about and dancing. (g:331.11)

I was always running about and dancing.

(13) a. He says 'I thocht **you were** a diver or somethin.' (7:262.41)

He said I thought you were a diver or something.

b. 'Aye, I thocht **you was** a scuba diver.' (7:259.21)

Yes, I thought you were a scuba diver.'

(14) a. **She was** writing down bits and pieces (b:1105.0)

b. **He was** in Aberdeen (f:100.16)

(15) a. There was one nicht **we were** lyin at anchor. (g:875.32)

There was one night we were lying at anchor.

b. We played on at beach til **we was** tired, sailin boaties, bilin whelks(b:254.15)

We played on that beach until we were tired, sailing boats, boiling whelks.

(16) a. He was aie away when **you ones were**... (8:4.37)

He was always away when you were

b. So **you were all- you were all** just bairns (8:13.48)

So you were all, you were all just children

c. **You ones was** a wee bitty older and you treated her just a right (4:513.45)

You (plural) were a little older and you treated her just fine.

(17) a. **They were** na really conscious of it either lyke you ken (4:521.36)

They weren't really conscious of it either, like, you know.

b. **They were** wild as anything (!:606.5)

Here we see variability in the use of *was* vs *were* with first person plural, and with second person singular and plural. There is no variability in the other person/number combinations. First and third singular always give *was*, and third plural always gives *were*. We therefore have the following paradigm of variable/categorical forms:[5]

(18)

	singular	plural
1st	was	was/were
2nd	was/were	was/were
3rd	was	were

A simple Variation in Exponence approach here is untenable: if we were to say that [*u*singular:–] had two surface variants (*was* and *were*), we'd overgenerate **they was*; if we were to say that [*u*author:–] had these two surface variants, we would not capture *we was*. It follows that we would have to state rather specific forms as being variable as in (19):[6]

(19) a. [*u*singular:+] was

b. [*u*singular:+, *u*author:–] was/were

c. [*u*singular:–, *u*author:–] was/were

[5]There is also variability in the past tense of *be* with expletive constructions (although the variation here is minimal, with *there was* favoured in most contexts), and where the subject is a full plural DP rather than just a pronoun (this variability is robust). However, this latter kind of variability is not confined to the past tense, nor is it confined to just the verb *be*. We show in (Adger and Smith 2005) that it is a separate phenomenon with a different syntactic analysis (which appeals crucially to underspecification of number features of Ds, rather than person and number features on a verb).

[6]A referee notes that we could condense this slightly by appeal to some underspecification, and having *were* as the only form for [*u*singular:+, *u*author:–]; in any case the point remains.

d. [*u*singular:–, *u*author:+] was/were

e. [*u*singular:–, *u*participant:–] were

With no deeper syntactic explanation, we would then end up appealing to historical or functional factors for why it is these forms rather than others that are variable. I'll show below that this is not the way to go—a far more satisfying explanation can be derived from the ways that the feature bundles underlying the forms of *was* and *were* combine.

5.2 Frequency Distribution

As well as the patterns of categoricality and variability in the paradigm, there are also differences in frequencies within the variable cases:

(20)

pronoun	percentage of *was*	N
second singular	69	161
first plural	67	368
second plural	10	10

Smith (2000) showed that the person and number features of the subject have a significant effect on whether the copula occurs in one variant or the other, while other grammatical factors such as whether the sentence is negative or positive, whether the subject precedes or follows the verb, or whether the verb is a copula or an auxiliary, have no statistically significant impact on the frequencies of occurrence of these forms (or it is not possible to determine from the amount of data available whether they have an impact). The first plural and second singular pattern similarly here, with about two thirds *was* and one third *were*, while the second plural goes in the other direction. Unfortunately the second plural occurs in very small numbers, so conclusions about its behaviour must be tentative. However, informant judgment tests show that both *you (ones) were* and *you (ones) was* for plural second person are possible, so the fact that the frequencies go the opposite direction here from the behaviour of singular you is likely to be significant.

5.3 Historical Explanation

One approach one might take to the current patterns of copula use in Buckie English is a historical one. This is what Smith suggests in her thesis. She points out that the forms of the past tense copula in Northern varieties of Middle English look as in table 2.

Table 2: Survey of the pronominal forms be in Middle English (Forsström, 1948)

	Northeast	West Midland	Northern
1st, 3rd sg	was	was, wes	was (wes)
2nd sg	was, wore (ware)	was [north-west],were , were,	was
Plural	wore(n), ware/n/	were(n)	war(e) (were/e/)

In these varieties, we see you[singular] triggering *was*-type agreement in all three varieties, with *was/were* variability in two. Smith suggests that the *was/were* variability in second person singular forms in Buckie is a retention from the historical record. Similarly, she suggests that the categorical zero rate for *they was*, contrasting with high rates of *was* with plural NP subjects (see footnote 5, and also Smith 2000, Adger and Smith 2005 for details) is a retention of the Northern Subject Rule reported by Murray 1873. This rule essentially allows singular agreement to appear with plural subjects when the subject is an NP but not when it is a pronoun.

This explanation, however, is a little problematic. Firstly, it does not explain why *was* is possible for plural first and second person pronouns at all, since there is no *was* in the plural in the historical record. One might try to develop an account of the presence of variability in the plural based on the idea that the second person singular's occurrence with *was* spreads to the second plural, and thence to the first person plural. However, if some such process of analogy were relevant, levelling the agreement to *was* throughout, then one might also expect to see *was* in the third plural, which is simply ungrammatical in this variety, as we have seen. Appeal to the Northern Subject Rule to rule this out is just a stipulation which does not drive the explanation any deeper.

6 Formal Explanation

I now turn to showing how the theoretical tools we put in place earlier provide us with an explanation of how the variability arises, why the patterns of variability and categoriality are the way they are, why the pattern is stable, and why the frequencies of the variable cases are as seen.

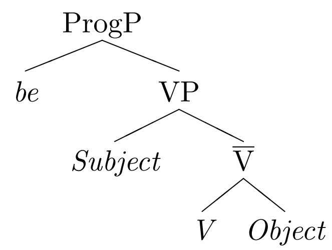
6.1 Features for pronouns and for auxiliaries

Let us assume the features we discussed above:

(21) [singular:±]; [participant:±]; [author:±]

A pronominal subject will be Merged in the specifier of the verb phrase. The auxiliary will be Merged higher. Movement processes will eventually site the subject to the left of the auxiliary (I abstract away from this here, as well as from further details of structure, such as any articulated shell or functional structure within the VP):

(22)



The question is: what are the features of the auxiliary? In the present tense we need to distinguish at least the following forms:

(23) am, are, is

For these we require a distinction between singular and plural (*he is/they are*), a distinction between author and addressee (*I am/you are*) and a distinction between participant and non-participant (*I am/she is*). The latter is required since both *you*[singular] and *she* are singular and not authors.

Consider the syntactic representation available to the child acquiring this system. We assume she will have in place the feature specification for the pronouns, which are all determined on the basis of interpretation correlating with form. The child will be exposed to data consisting of pronoun plus verb-form. The task is to determine what features are on the verb form.

When the system is set up like this there are a number of possibilities. The verb forms could be fully specified. In such a case, the choice of which version of the copula to introduce into the syntactic derivation will be wholly determined, since only one verb form will match the pronoun precisely. This means that, on the assumption that lexical items are bundles of phonological and syntactic features, we need to specify a fair amount of homonymy as follows:

(24)

$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{author:+} \end{bmatrix}$ am	$\begin{bmatrix} \text{singular:–} \\ \text{participant:+} \\ \text{author:+} \end{bmatrix}$ are
$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{author:–} \end{bmatrix}$ are	$\begin{bmatrix} \text{singular:–} \\ \text{participant:+} \\ \text{author:–} \end{bmatrix}$ are
$\begin{bmatrix} \text{singular:+} \\ \text{participant:–} \end{bmatrix}$ is	$\begin{bmatrix} \text{singular:–} \\ \text{participant:–} \end{bmatrix}$ are

However, a paradigm of forms like this clearly misses morphological generalizations: all the plural forms are *are*, as are all the [author:–] forms. Such syncretisms across paradigms can be dealt with in a number of ways, most obviously by underspecifying the feature content of the verb (see, e.g. Bierwisch 1967, Lumsden 1992, Halle 1992, Williams 1994) . Maintaining the assumption that lexical items are bundles of phonological and syntactic features, we can implement an underspecification analysis by assuming the following lexical items:

We derive this by looking at just those features that maximally correlate with the forms.[7]

(25) a. [*u*singular:–] are

b. [*u*author:–] are

c. [*u*singular:+, *u*author:+] am

d. [*u*singular:+, *u*participant:–] is

From this specification, it follows that *I* matches with only (c); *you[sing]* matches with only (b); she matches with only (d); *we* matches with only (a); *you[pl]* matches with (a) and (b); *they* matches with just (a).[8]

Thinking of how this works in a syntactic derivation, when the auxiliary is Merged above the pronominal subject, there are four possible lexical items which can be chosen. In actuality, the choice is almost completely determined. Only *you*[pl] allows more than one lexical item, and both have the same pronunciation. The system captures the categorical nature of present agreement with the verb *be* in standard

[7]We can of course reduce this lexicon, eliminating the final case of homonymy, by use of an elsewhere rule; I leave this possibility aside here, since the argument is more straightforward when we look at direct correlations between forms and features.

[8]I ignore the possibility of using an elsewhere form here, since what we are interested in are the direct correlations between forms and features.

English (and in Buckie).

There is a further question to be addressed here: what is the provenance of the lexical items specified above? Given the three features we are using here, and allowing underspecification, there are actually 26 possible feature-value combinations and hence 26 possible lexical items: those containing just one feature-value combination (6), those containing two (a further 12), and those containing three (a further 8). How is the child to determine which are relevant?

The first thing to note is that 6 of these will be unusable in the syntactic system: these are the items which bear a [*u*participant:–] specification, along with an [*u*author] specification. If these are in an Agreement Chain with a [participant:–] pronoun, then their [*u*author] feature cannot be checked, given the Feature Cooccurrence Restriction defined in section 2. Conversely, if they are in an Agreement Chain with something bearing an [author] feature, then there will be a clash in the value of [participant].

This gives us 20 possible lexical items, and we have determined that 4 are sufficient. How do we select the correct subset?

One simple algorithm is to look for maximal generalizations (that is, the best natural classes). In a feature system with underspecification, maximal generalizations are those made by lexical items with the fewest features. The idea, then will be to generate one-feature lexical items, evaluate the outcome, and then, if that outcome is not satisfactory, generate two-feature lexical items and recurse.

How do we decide when a lexical item is unsatisfactory. I suggest that the system tries to reduce optionality, synonymy, and the size of the lexicon. So we have the following algorithm:

(26) Seek Maximal Generalization by

a. generating all n-feature LIs, where n=1

b. Matching them with forms

c. Rejecting all LIs where a feature cannot always be mapped to a single form (Reject Optionality): that is, an LI is kept if there is a form which that LI always matches.

d. If a feature can always be mapped to a single form, but this creates synonymy, eliminate LI's to reduce synonymy as much as possible (Reject Synonymy).

e. Recursing over n=n+1, with the proviso that if a form has been successfully analysed, in the n-1th step, LIs capturing it in the nth step will be rejected (Minimize Lexicon).

This algorithm is an empirical claim about the procedure which leads to the mental representation of the lexicon of a speaker of this variety. See below for some discussion.

Let us see how this works concretely by looking at the present tense of *be*. The algorithm first creates lexical items using just a single feature, that is, it generates a list of the 6 one-feature lexical items as follows:

(27) a. [*u*singular:+]

b. [*u*singular:–]

c. [*u*participant:+]

d. [*u*participant:–]

e. [*u*author:–]

f. [*u*author:+]

We then attempt to associate these with morphological forms, assuming that a single LI correlates with a single form, so we are looking for matches where a feature can always be mapped to a particular form: that is, if two forms are found for say, [singular:+], then a lexical item with this feature specification is rejected (this is Reject Optionality). If we apply this procedure to our LIs, we have:

(28)
a. *[*u*singular:+] am/are/is
b. [*u*singular:–] are
c. *[*u*participant:+] am/are
d. *[*u*participant:–] is/are
e. [*u*author:–] are
f. *[*u*author:+] am/are

It follows that only two lexical items are possible from this list, and together they successfully analyse *are* wherever it occurs, giving an analysis of *are* of maximal generality. However, this analysis is clearly not complete, as it doesn't provide LIs for *am* and *is*. The algorithm's second pass is then to generate the list of two-feature lexical items, and to apply the same procedure:

(29)
a. [*u*singular:+, *u*participant:+] am/are
b. [*u*singular:–, *u*participant:+] are
c. [*u*singular:+, *u*participant:–] is
d. [*u*singular:–, *u*participant:–] are
e. [*u*participant:+, *u*author:–] are

f. [*u*participant:+, *u*author:+] am/are

g. [*u*singular:+, *u* author:+] am

h. [*u*singular:–, *u* author:+] are

i. [*u*singular:+, *u* author:–] are

j. [*u*singular:–, *u* author:–] are

The algorithm rejects cases where the two-feature LI cannot always be mapped to a particular form (a and f). There are a number of two-feature LIs that have successful analyses, for example [singular:–, participant:+] will always cooccur in the primary linguistic data with *are*. However, such LIs do not add to the analytical power of the system, since there are already extant LIs which will correctly analyse the input, so the algorithm rejects these items (Minimize Lexicon).

The analyses of *is* and *am* are successfully provided by:

(30) a. [*u*singular:+, *u*participant:–] is

b. [*u*singular:+, *u*author:+] am

At this point, we have successfully analysed all the forms, and so the algorithm halts.

What we essentially have here is an evaluation metric which chooses between sets of lexical items that the learner is able to construct as being compatible with the primary linguistic data. This particular metric seeks maximal generalizations (hence, minimally specified lexical items), rejects cases of formal synonymy (see below), and reduces the size of the lexicon as much as possible. One can imagine alternative metrics, for example one that seeks only to minimize the number of lexical entries, rather than the featural specification of those entries. This would be the usual procedure in linguistics, seeking to minimize homonymy in the paradigm by using an elsewhere form and thereby reducing the number of lexical items as much as possible, giving:

(31) a. [*u*singular:+, *u*author:+] am

b. [*u*singular:+, *u*participant:–] is

c. elsewhere: are

Instead, our algorithm produces the following lexicon, which correctly captures the feature-form pairings as discussed above:

(32) a. [*u*singular:–] are

b. [*u*author:–] are

c. [*u*singular:+, *u*author:+] am

d. [*u*singular:+, *u*participant:–] is

As has long been known (see, e.g. Chomsky 1965), the correctness of such a metric is an empirical matter. We will see in the next section that the approach proposed here impacts upon the predictions of the system about the frequencies of variant forms.

6.2 The combinatorial analysis of variability

The underspecification approach outlined above comes into its own when we look at a variable, rather than categorical system. Recall that in Buckie we had the following situation:

(33)

I $\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{author:+} \end{bmatrix}$ was	We $\begin{bmatrix} \text{singular:–} \\ \text{participant:+} \\ \text{author:+} \end{bmatrix}$ was/were
You $\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{author:–} \end{bmatrix}$ was/were	You (ones) $\begin{bmatrix} \text{singular:–} \\ \text{participant:+} \\ \text{author:–} \end{bmatrix}$ was/were
He/she/it $\begin{bmatrix} \text{singular:+} \\ \text{participant:–} \end{bmatrix}$ was	They $\begin{bmatrix} \text{singular:–} \\ \text{participant:–} \end{bmatrix}$ were

Now, the learner of Buckie is faced with this paradigm as her input. Using the procedure we defined above, she first generates the following lexical items:

(34) a. [*u*singular:+]

b. [*u*singular:–]

c. [*u*participant:+]

d. *[*u*participant:–]

e. [*u*author:–]

f. [*u*author:–]

g. [*u*author:+]

Running the algorithm associates these with forms as follows:

(35) a. [*u*singular:+] was

b. [*u*singular:–] were

c. [*u*participant:+] was

d. *[*u*participant:–] was/were

e. [*u*author:–] was

f. [*u*author:–] were

g. [*u*author:+] was

Note that, for example, [*u*singular] always maps to *were*, so (b) is a well-formed pairing. The fact that [*u*singular] also maps sometimes to *was* is irrelevant, given the definition of the algorithm. The LI which is [*u*participant:–] has two different forms associated with it: *was* when it occurs with a third singular pronoun and *were* when it occurs with a third plural. It is therefore ruled out, by Reject Optionality. Since [*u*author] always maps to *was*, and also always maps to *were*, we have two lexical items. This gives us a case of formal synonymy, with a single underlying specification having two pronunciations (or alternatively, two surface forms with the same underlying specification)

This situation raises the following question: is this kind of 'surface' variation allowable in the system? If so, then we should keep both lexical items. However, the *was* form of *you*[sing] is captured by other lexical items ([*u*singular:+] *was* and [*u*participant:+] *was*), and the *was/were* variability of *you*[plural] is also captured by two other lexical items ([*u*singular:-] *were* and [*u*participant:+] *was*), so a more minimal system would eliminate [*u*author:–] *was*, giving:

(36) a. [*u*singular:+] was

b. [*u*singular:–] were

c. [*u*participant:+] was

d. [*u*author:–] were

e. [*u*author:+] was

It is not possible to eliminate the association of the feature [*u*author:–] with the form *were*, or we would predict that *you were* is not a possible form for singular second person, contrary to fact. The claim I make here, then, is that the algorithm wil reject [*u*author:–] *was* in this situation, as a specific reflex of Minimize Lexicon which we might term Reject Synonymy.

If it were impossible to remove one of these two synonymous lexical items without loss of coverage, then the algorithm would keep both. This would effectively give us a Variation in Exponence situation, creating doublets. If Kroch (1994) is correct, such a situation would be diachronically unstable and should lead to change.

This set of lexical items will correctly analyse the patterns of grammaticality in the data, as well as the patterns of variability, so the analysis is complete and there is no second pass of the algorithm. For example, if we have a third plural pronoun as subject, the only LI that will be able to combine with this, is (b) [*u*singular:–] *were*. If any of the other LIs in this set combine with *they*, their feature will not be in an agreement-chain, and will hence violate Full Interpretation, correctly predicting ungrammaticality. On the other hand, if we have the first person plural subject *we*, this will combine with any of (b) [*u*singular:–] *were*; (c) [*u*participant:+] *was* or (f) [*u*author:+] *was*, correctly predicting the possibility of *was/were* variability with this pronoun.

One other point should be made here: none of these lexical items is in competition with another, since there is no subset relation between them (trivially, since they are all separate features). Given this, we could implement this instance of Combinatorial Variation in a Distributed Morphology type system (Halle and Marantz 1993), with fully specified syntactic feature bundles, but underspecified Vocabulary Items, since the latter would never compete. However, not all instances of Combinatorial Variation are of this sort: some indeed seem to involve lexical items which are in a subset relation with each other (see the next section). Unless we relax the usually strict application of the Pāṇinian Elsewhere Principle in Distributed Morphology, we cannot capture this straightforwardly (but see Parrott 2006 for an argument that this would be the right way to go.)

Note further that this is not the *minimal* set of lexical items. It is possible to remove the lexical item [*u*author:+], and we still capture the correct distribution of categorical and variable forms. However, (36) is the set of lexical items that follows from the evaluation metric we proposed above: this metric was set up so as to derive the patterns of grammaticality and the syncretisms on the basis of search for the maximal generalizations, rather than the minimal number of lexical items. This now makes a prediction about the distribution of forms.

The way the system is set up, it is possible to have a number of routes to a final form. So, for example, for *you*[singular] there are two ways to get *was* and just one way to get *were*, while for *you*[plural], it's the other way around:

(37) [singular:+, participant:+, author:–] ... (a)was; (c) was; (d) were

(38) [singular:–, participant:+, author:–] ... (b)were; (c) was; (d) were

Recall that the frequencies of *was* for standard *were* went in this direction.

(39)

pronoun	percentage of *was*	N
second singular	69	161
first plural	67	368
second plural	10	10

In fact, for *you*[singular] almost exactly two thirds of the data were *was*. For *you*[plural], the numbers are really too small to tell, but they do at least go in the right direction. There appears to be variability (one example of *was* and nine of *were*), and it goes in the opposite direction from the second person singular case.

These frequencies follow on the assumption that there is a random choice of form (see sections 4 and 6.4). Of course, they may be perturbed by performance related effects (see below), but the crucial point is that the grammar gives us variation in form without a corresponding variation in meaning, and moreover it makes predictions about the frequencies of forms when (and if) performance factors can be controlled for.

Similarly, there are two means to the *was* form for *we*, but only one to the *were* form. Once again this is the correct result in frequency terms.

(40) [singular:–, participant:+, author:+] ... (b)were; (d) was; (e) was

Here is where the empirical claim about the evaluation metric is relevant. If we took the minimal set of lexical items that are needed to capture the categorical versus variable patterns, rather than taking the set of items derivable using just one feature, then we would still predict that there are two forms which occur with *we*, but that there would only be one way to get each form (since, by assumption, the lexical item [author:+] would be absent from our lexicon). This would now predict that we should find *was* and *were* about 50% each in the corpus, and this is incorrect. As expected, the choice of evaluation metric has particular empirical effects.

6.3 Acadian French

In this brief section I want to show that the algorithm just outlined predicts the correct results for another categorial/variable split in a paradigm. King (2005) reports the following paradigm for Acadian French:

(41)

	singular	plural
1	parle	parle/ parlons
2	parles	parlez
3	parle	parlent/parlont

Here we have variability in the first and third person plural. The phonological realizations of these forms are roughly as follows:

(42)

	singular	plural
1	0	0/ɔ̃
2	0	e
3	0	0/ɔ̃

Applying our Maximal Generalization Metric, we have the following lexicon on the first pass:

(43) a. [*u*singular:+] 0(because this feature can always be mapped onto 0)

b. *[*u*singular:–] 0/e/ɔ̃ (Reject Optionality)

c. *[*u*participant:+] 0/e/ɔ̃ (Reject Optionality)

d. [*u*participant:–] 0 (because this feature can always be mapped onto 0)

e. *[*u*author:–] 0/e (Reject Optionality)

f. [*u*author:+] 0 (because this feature can always be mapped onto 0)

This gives us three lexical items which analyse the zero exponent of agreement. However, we need to do a second pass, since we still don't have lexical items for 'e' or 'ɔ̃'. The second pass gives:

(44) a. [*u*singular:–, *u*participant:–] ɔ̃

b. [*u*singular:–, *u*author:–] e

c. [*u*singular:–, *u*author:+] ɔ̃

However, the resultant lexicon is rather different from the Buckie situation, as [*u*participant:–] and [*u*singular:–, *u*participant:–] are in competition with each other, as are [*u*author:+] and [*u*singular:–, *u*author:+]. That is, for Acadian French, we end up with two lexical items which compete. If we strictly applied the Pāṇinian Elsewhere Principle, we would predict no variation, since the more specific form would always be chosen. One way out of this dilemma, which still maintains the Elsewhere Principle, is that we force two separate grammatical systems in this situation (that is, we have competing grammars, in Kroch's sense). Following Kroch, we therefore predict that this will be a diachronically unstable situation, which will be moving towards a stable system. This is exactly what King reports.

6.4 Potential objections

A referee suggests that there is an assumption here which needs to be examined in more detail: the assumption is that other factors than ϕ-feature specification are not systematic enough to affect the distribution of *was/were* variability in Buckie, specifically sociolinguistic and processing factors must be taken into account. As acknowledged in section 4, this is a fair point; however, the particular nature of the speech community under consideration means that a number of these extra-grammatical factors are immediately ruled out.

As detailed in Smith (2000), *was/were* variability in Buckie is not affected by extra-linguistic factors in any clearly systematic way. Smith shows that each generation of speakers has a very similar statistical pattern for use of *was/were* in the different person/number combinations, although it is true that the older generation have more *was* in general. Looking at gender, Smith shows that it is only middle aged females who have a markedly more standard pattern (although even these speakers show the same basic pattern of categoriality versus variability, and in fact show a broadly similar pattern of frequency distribution). Finally, the close knit nature of the community, and the fact that the interviews were conducted between community insiders means that questions of class/identity or of the nature of the conversational context are unlikely to have had a major impact.

Smith has also confirmed to me (pc) that other potential factors are unlikely to be affecting the distribution. For example, occurrences of the past tense of *be* in the corpus are in general too isolated from other occurrences for recency of use (which would presumably impact on lexical access) to be a factor. Furthermore, an informal inspection of the tokens shows that very few have divergent stress patterns, so intonational factors don't seem to play a role.

The objection might be made that there are still other factors which impact on the choice of lexical item—but such an objection can be made about any analysis of empirical data. The only refutation of such an objection is an explicit theory which makes predictions, and this is what I have proposed here.

Related to this point is a broader one, also made by the referee: this approach does not leave room, within the model itself, for the variants to be associated with social meaning, in contrast to an alternative, like Bender (2001). In Bender's system, a linguistic object might include social meaning, allowing this kind of information to be part of the language users' linguistic competence.

This is an interesting position, but one which I wholly reject, mainly for broader reasons of modularity, dissociation of linguistic and social skills, etc. (see, e.g. Smith, Hermelin, and Tsimpli 2003). The particular instantiation of this idea that Bender gives is that a variable (e.g. presence versus absence of a copular verb in a nominal versus verbal predication) can be modeled by a choice of lexical items with 'social meaning' specified in them. For example, she suggests the following two lexical items to deal with (a subpart of) the grammar of copular variation in African American Vernacular English:

(45) a. $0.4 \begin{bmatrix} \text{copula-be} & \\ \text{COMPS} & \left\langle \text{NP} \right\rangle \\ \text{CTXT — SOCIAL} & \text{'educated'} \end{bmatrix}$

b. $0.6 \begin{bmatrix} \text{silent-cop-ph} & \\ \text{ARGS} & \left\langle \text{NP} \right\rangle \end{bmatrix}$

The numbers before each lexical item specify the resting probability of the lexical item, which is essentially the probability that this one will be chosen over the other. In any particular speech situation, (b) is more likely to be chosen than (a).

However, Bender suggests that a speaker can, in a particular speech situation, have a particular social intent (say, to sound educated). In such a situation, the speaker can override the resting activation, and select an overt copula (a), even though this lexical item has a lower resting activation. In this way, the variability in use is affected by the linguistic specification of the lexical item.

I have two problems with this kind of model. One is rather specific: imagine that we are interlocutors in a non educated social context. In this case, the lexical item in (a) will simply never 'match' the context, so we predict that there will be zero occurrence of this lexical item, and therefore 100% copula absence. Now, this prediction may very well be true for AAVE, but it is clearly false for the Buckie *was/were* variation under consideration here. If we were to try to mark the *were* variant with *we* and *you* as, say, 'educated', then we predict no variability in the 'non-standard' speech situtation of an insider sociolinguistic interview, contrary to fact.

The second problem is that this model stipulates the resting probabilities in terms of frequencies, but provides no reason why the frequencies should be what they are. This ultimately may be correct: perhaps all one can do is stipulate these facts, but it's certainly worthwhile to attempt to provide an actual explanation for them, and that is what I have done here. In my model, the input probabilities for the sets of lexical items that constitute variants are all equal, but I agree with Bender that certain social and psychological factors may very well alter these input probabilities. I disagree that there's any need to build these social factors into the linguistic information in lexical items, however.[9]

There is, however, a different issue that should not be ducked here: to what extent is it possible to extrapolate I-languages from data which I have discussed here. The extra assumption I am making is that each community member will have the same grammar, and that it is legitimate to collapse the data from a number of individuals into a single analysis. I think that this assumption is reasonably motivated by the fact that the general patterns across individuals hold, for the most part, within a single individual's data (for example, all individuals have a categorial/variable split exactly as described here), but it is true that there just isn't enough data to be sure that the detailed *frequency* effects discussed here actually hold for each individual. This is a shortcoming of the analysis which I am aware of, and it is the reason that I offer this analysis as a plausibility argument rather than as a detailed empirical study.

[9]One final comment here on the embedding of the competence model in a performance model, which I'm taking for granted here. A referee claims that this is at odds with mainstream Minimalist argumentation, citing Seuren (1982), but that's simply not the case. My assumption is standard in Chomskian generative grammar as a more or less random choice of references from my bookshelf shows (Chomsky 1965, p9, Chomsky 1980, p203, Chomsky 2000b, p124).

7 Conclusions and implications

The system I have set up here captures the patterns of categoricality and variability in Buckie by means of a simple feature checking system along the lines of Chomsky (1993) and Chomsky (1995). Allowing lexical items to be underspecified for the uninterpretable features they bear automatically predicts the possibility of variability (that is, non-deterministic variation in form with no corresponding variation in meaning). I also proposed an evaluation metric whereby the child selects a set of lexical items which bear uninterpretable features, seeking the maximal morphological generalizations given the interpretable features borne by the subject. The particular features chosen capture the patterns of variability and categoricality in the data. However, they do more than this: they also predict variation in the frequency of occurrence of the morphological form, in a way that is analogous to Mendelian genetic combination: the combination of discrete elements may lead to statistical differences in frequency of occurrence, which are detectable when other factors are controlled for, or by looking across a large enough sample.

The system has a number of implications for how we think about variability. Most importantly, I think that there is a clear argument that variability is not just of the Variation in Exponence type: that is differential realisation of an underlying category as a number of surface forms. There is also a kind of variability that arises from the combinatorics of the syntactic system itself. Moreover, this variability is captured here essentially by manipulations of the featural specifications of lexical items, rather than by assuming the existence of a number of potentially competing grammars (Yang 2002, Kroch 1989), or of multiple mutually exclusive parametric options in a single grammar (Henry 1995, Wilson and Henry 1998). Furthermore, the system does not build the probabilities or frequencies into the grammar itself as weighted rules or constraint hierarchies (contrary to the practice in Stochastic Optimality Theory (Boersma 1997, Bresnan, Dingare, and Manning 2001) or Probabalistic Grammars (Manning 2003). Rather, it attempts to predict the population of lexical items required from the primary linguistic data and an evaluation metric. These lexical items then syntactically combine to derive broad frequency effects. Perhaps the closest approach to CV in the literature is Anttila's partially ordered grammar lattices (Anttila 1997, Anttila and Cho 1998), which have been used in Optimality Theory to capture statistical effects of variability in phonology. Working out whether Anttila's model is able to capture these morphosyntactic paradigmatic effects, with patterns of variability and grammaticality, will have to be left for another time.

A further point is that in the system defended here, there is a single grammar with an inventory of lexical items bearing particular feature specifications. If choice of those lexical items is random, then we expect to see particular frequency distributions; these may be perturbed by performance factors, such as ease of lexical access, and perhaps sub-conscious choices about appropriateness to register etc.

Finally, the system predicts that at least some variation should be diachronically stable: the variability between *was* and *were* as analysed here is a stable, learnable system. The lexical items do not compete with each other (see Kroch 1994) so there is no reason to expect a change in the pattern. Of course, the model described here is compatible with a fairly standard approach to language change (e.g. Lightfoot 1999): if one of the variants (V1) produced by the grammar comes to be highly favoured over another (V2), perhaps because of socio-linguistic reasons or the interaction of grammatical and processing factors, then V2 will eventually end up having a very low frequency. Children acquiring such a system, who have to organise their featural repertoire so as to account for the patterns of grammaticality they are exposed to, may analyse V2's low frequency as non-occurrence. This will result in their having a different set of lexical items from their parents, resulting in change.

References

Adger, David. 2003. *Core Syntax*. Oxford: Oxford University Press.

Adger, David and Smith, Jennifer. 2005. Variation and the Minimalist Program. In Leonie Cornips and Karen Corrigan, eds., *Syntax and Variation: Reconciling the Biological and the Social*, 149–178, Amsterdam: John Benjamins.

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout, and W. Leo Wetzels, eds., *Variation, Change and Phonological Theory*, 35–68, Amsterdam: John Benjamins.

Anttila, Arto and Cho, Young-Mee You. 1998. Variation and change in Optimality Theory. *Lingua* 104:31–56.

Bender, Emily. 2001. Syntactic variation and linguistic competence. Ph.D. thesis, Stanford, CA.

Bierwisch, Manfred. 1967. Syntactic features in morphology: General problems of so-called pronominal inflection in German. In *To honor Roman Jakobson: essays on the occasion of his seventieth birthday*, 239–270, The Hague: Mouton.

Boersma, Paul. 1997. How we learn variation, optionality and probability. *Proceedings of the Institute of Phonetic Sciences* 21:43–58.

Bresnan, J., Dingare, S., and Manning, C. D. 2001. Soft constraints mirror hard constraints: Voice and person in english and lummi. In *Proceedings of the LFG 01 Conference*, Stanford: CSLI.

Brody, Michael. 1997. Perfect chains. In Liliane Haegeman, ed., *Elements of Grammar: Handbook of Generative Syntax*, 139–167, Dordrecht: Kluwer.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Kenneth Hale and Samuel Keyser, eds., *The view from Building 20*, 1–52, Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000a. Minimalist inquiries: the framework. In R. Martin, D. Michaels, and Juan Uriagereka, eds., *Step by Step: Essays on Minimalist syntax in honour of Howard Lasnik*, 89–115, Cambridge, MA: MIT Press.

Chomsky, Noam. 2000b. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz, ed., *Ken Hale: A Life in Language*, 1–52, Cambridge, MA: MIT Press.

Halle, Morris. 1992. The Latvian declension. In Geert Booij and Jaap van Marle, eds., *Yearbook of Morphology*, 33–48, Dordrecht: Kluwer.

Halle, Morris and Marantz, Alec. 1993. Distributed Morphology and the pieces of inflection. In Ken Hale and Samuel Jay Keyser, eds., *The View from Building 20*, 111–176, MIT Press.

Harley, Heidi and Ritter, Elizabeth. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.

Henry, Alison. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

King, Ruth. 2005. Morphosyntactic variation and theory: Subject verb agreement in Acadian French. In Leonie Cornips and Karen Corrigan, eds., *Syntax and Variation: Reconciling the Biological and the Social*, 198–229, Amsterdam: John Benjamins.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

Kroch, Anthony. 1994. Morphosyntactic variation. In K. Beals et. al., ed., *Papers from the 30th Regional Meeting of the Chicago Linguistics Society*, 190–201, Chicago Lingistics Society.

Labov, William. 1969. Contraction, deletion and inherent variability of the English copula. *Language* 48:773–818.

Labov, William. 1972. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Lightfoot, David. 1999. *The Development of Language*. Oxford: Blackwell.

Lumsden, John. 1992. Underspecification in grammatical and natural gender. *Linguistic Inquiry* 23:469–486.

Manning, Christopher. 2003. Probabalistic syntax. In Jennifer Hay Rens Bod and Stefanie Jannedy, eds., *Probabalistic Linguistics*, 298–341, Cambrige, MA: MIT Press.

Newmeyer, Frederick. 2003. Grammar is grammar and usage is usage. *Language* 79:682–707.

Parrott, Jeffrey. 2006. Distributed Morphological mechanisms of Smith Island *weren't* levelling, ms. Georgetown University.

Pesetsky, David and Torrego, Esther. 2001. T to C movement: causes and consequences. In Michael Kenstowicz, ed., *Ken Hale: A Life in Language*, 355–426, Cambridge, MA: MIT Press.

Seuren, Pieter A. M. 1982. Internal variability in competence. *Linguistische Berichte* 77:1–31.

Smith, Jennifer. 2000. Synchrony and diachrony in the evolution of English: Evidence from Scotland. Ph.D. thesis, University of York.

Smith, Neil, Hermelin, Beate, and Tsimpli, Ianthi. 2003. Dissociation of social affect and theory of mind in a case of asperger syndrome. *UCL Working Papers in Linguistics* 15:303–322.

Williams, Edwin. 1994. Remarks on lexical knowledge. *Lingua* 92:7–34.

Wilson, John and Henry, Alison. 1998. Parameter setting within a socially realistic linguistics. *Language in Society* 27:1–21.

Yang, Charles D. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.