Combinatorial Variability*

David Adger
Queen Mary, University of London
d.j.adger@qmul.ac.uk

28th February, 2005
COMMENTS WELCOME

1 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 the 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

*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 Tromso (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 Daniel Harbour and Gillian Ramchand for helping me think through some tricky bits, and of course to Jennifer Smith, without whose fantastic empirical work and socio-linguistically informed comments none of this would have been possible.

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 (i.e free) variation **variability**, and it is what this paper is mainly concerned with.

The same kind of rule based approach to variation and variability has 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 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 variation 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, deal with Variation in Exponence. 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 untenable and that the Chomskian conception of I-language (Chomsky 1986) is, without significant modification, incorrect. For this reason (and for other reasons), generativists, convinced by the explanatory efficacy of the notion of I-language, 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 do not make the claim here that CV is the only source of variation, but just the weaker claim that we at least need something like CV. The question then is whether we also need the VE. 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 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, metalinguistic or social judgments on the form etc. 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.

2 Some background assumptions

2.1 Lexical Items and Features

Current thinking within the Minimalist Program (Chomsky 1995, Chomsky 2000, 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 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).

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.

2.2 Pronouns

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

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

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 addressee) or not. The final feature [addressee:±] allows us to distinguish between speaker and addressee. Note that having a specification for [addressee:±] entails having a positive specification for [participant] (that is, it is not possible to be classified for whether one is an addressee 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 [addressee], at least in English, since there are no pronominal forms in English which do not distinguish between speaker and addressee. So we have the following general restriction:

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

2.3 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:

(4) An agreement-chain is a pair of lexical items, where the uninterpretable features of one LI are a (non-proper) subset of the interpretable features of the other.[1]

[1] I abstract away here from questions of representation and derivation (see, for example, Brody 1997). I

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

(5)

$$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{addressee:-} \end{bmatrix} \quad ... \quad \begin{bmatrix} u\text{singular:+} \\ u\text{participant:+} \\ u\text{addressee:-} \end{bmatrix}$$

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

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

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:

(7) *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 [singular:-]:

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

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

3 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. 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:

---

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. In current work with Jennifer Smith, we are developing the valuation approach to the phenomena discussed here, and extensions to other cases of variable agreement.

(9)

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

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

(10)

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

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 (with a higher probability) than z.

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. This is true for an individual speaker's use of language across a large enough corpus of utterances, and also for groups of speakers. Note that the proportions of variability predicted by the grammar depend on the idea that a speaker's choice of a lexical item is made at random. Of course, this need not, and probably usually is not, the case. The choice of lexical item is something that the performance systems are able to influence, predicting that there will be fluctuation from the proportions of variants produced from the competence grammar. The performance systems are extremely complex,

[2]In this sense we provide a definition of the Labovian notion of (socio)linguistic variable in terms of interpretable and uninterpretable features. See further Adger and Smith in press.

including processing effects, and potentially higher-level sociolinguistic effects. However, if we control for these by having enough data, and by monitoring the effect of these factors, then we should be able to see variable I-language at work.

Note that this approach allows us to connect the notions of individual grammars (I-languages) and community grammars. The I-language consists of computations over discrete units which produce a range of variants. The individual speaker has acquired their I-language on the basis of the primary linguistic data they were exposed to during the acquisition process. This data may contain more than information just about structure—it may also contain information about the probability that a certain item will be used in certain situations. Moreover, the probability that a particular lexical item will be used is also affected by constraints imposed by other systems of the mind on language use, most obviously ease of lexical access, which leads to priming and word-frequency effects. The latter are presumably part of a separate internal system tightly connected to the grammatical system. All of these interact in the production of an utterance, explaining why it is that there are regularities in variability across speech communities as well as regularities in variability within individuals.

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 in press for an earlier analysis of this data).

4 Combinatorial Variability in Action

4.1 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 were collected using standard sociolinguistic methodology, is highly vernacular in nature and amounts to approximately forty hours of tape-recorded casual conversations which were fully transcribed and 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).

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

Table 1: Speaker sample

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

below:

(11) 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.

(12) 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.'

(13) a. **She was** writing down bits and pieces (b:1105.0)
b. **He was** in Aberdeen (f:100.16)

(14) 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.

(15) a. **Was you ones** so bad, like? (l:262.7)
Were you (plural) so bad, like?
b. **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.

(16) 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:[3]

(17)

[3]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 kind of variability is not confined to the past tense, nor is it confined to just the verb *be*, and we assume it is a separate phenomenon. We address the analysis of this pattern in current work

	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 [singular:-] had two surface variants, we'd overgenerate **they was*; if we were to say that [addressee:+] had two surface variants, we would not capture *we was*. It follows that we would have to state rather specific forms as being variable, with no deeper syntactic explanation, and 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.

4.2 Frequency Distribution

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

(18)

pronoun	percentage	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 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). Note that 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.

4.3 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.

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 3, and also Smith 2000, Adger and Smith in press for details)

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/)

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 too 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.

5 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.

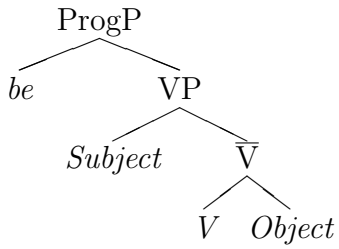
5.1 Features for pronouns and for auxiliaries

Let us assume the features we discussed above:

(19) [singular:±]; [participant:±]; [addressee:±]

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):

(20)



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

(21) am, are, is

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

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.[4]

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:

(22)

[4]Again, I am simplifying here, by assuming such a straightforward link between feature specification and form. If we are to capture broader generalisations about agreement in Buckie, then we almost certainly need to morphologically (and syntactically) decompose the verb forms. In current work, Jennifer Smith and I adopt a system which assumes such syntactic decomposition, feature valuing rather than checking, and Late Insertion of phonological material, as assumed in Distributed Morphology (Halle and Marantz 1993). None of these assumptions is materially relevant to the basic idea of combinatorial variability, so I implement the analysis here with simple checking theory, no syntactic decomposition, and with the assumption that lexical items are bundles of syntactic and phonological features when they enter the syntax (that is, roughly the system in Chomsky 1995).

$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{addressee:-} \end{bmatrix}$ am	$\begin{bmatrix} \text{singular:-} \\ \text{participant:+} \\ \text{addressee:-} \end{bmatrix}$ are
$\begin{bmatrix} \text{singular:+} \\ \text{participant:+} \\ \text{addressee+} \end{bmatrix}$ are	$\begin{bmatrix} \text{singular:-} \\ \text{participant:+} \\ \text{addressee:+} \end{bmatrix}$ 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 [addressee:+] 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:

(23) a. [*u*singular:-] are
b. [*u*addressee:+] are
c. [*u*singular:+, *u*addressee:-] 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).

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 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, together with the assumption that a bundle cannot be specified for [participant:-] and [addressee], and allowing underspecification, there are actually 22 possible feature-value combinations and hence 22 possible lexical items: those containing just one feature-value combination (6), those containing two (a further 10), and those containing three (a further 6). How is the child to determine which are relevant?

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. Let us assume then that the child first tries

to create lexical items using just a single feature, that is, she generates a list of lexical items that look as follows:

(24) a. [*u*singular:+]
b. [*u*singular:-]
c. [*u*participant:+]
d. [*u*participant:-]
e. [*u*addressee:+]
f. [*u*addressee:-]

The learner then attempts to associate these with morphological forms, assuming that a single LI correlates with a single form: that is, if two forms are found for say, [singular:+], then a lexical item with this feature specification is rejected. If we apply this procedure to our LIs, we have:

(25) a. *[*u*singular:+] am/are/is
b. [*u*singular:-] are
c. *[*u*participant:+] am/are
d. *[*u*participant:-] is/are
e. [*u*addressee:+] are
f. *[*u*addressee:-] am/are

The reason we reject cases where a single feature correlates with more than one form is that such cases will immediately overgenerate: for example, if [*u*singular:+] were to be retained as a lexical item, then we'd predict *I is*, and *he am*. In this sense the rejection of formal synonymy (where we have a single feature with more than one morphological form) connects with the predictive coherence of the grammatical system.

It follows that only two lexical items are possible from this list, and together they successfully analyse *are* wherever it occurs. The learner then generates the list of two-feature lexical items, and applies the same procedure, searching for analyses of *is* and *am*. These are successfully provided by:

(26) a. [*u*singular:+, *u*participant:-] is
b. [*u*singular:+, *u*addressee:-] am

There are other 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. Let us assume, then, that the learner's aalgorithm only expands the lexicon to handle cases which do not already have an analysis.

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, and bars superfluous additions to the system. One can imagine alternative metrics, for example one that minimizes the number of lexical entries, rather than the featural specification of those entries. 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.

5.2 The combinatorial analysis of variability

This underspecification approach outlines above, which is motivated as a means to explain why the syncretisms in the paradigm are where they are, comes into its own when we look at a variable, rather than categorical system. Recall that in Buckie we had the following situation:

(27)

I [singular:+ participant:+ addressee:-] was	We [singular:- participant:+ addressee:-] was/were
You [singular:+ participant:+ addressee+] was/were	You (ones) [singular:- participant:+ addressee:+] was/were
He/she/it [singular:+ participant:-] was	They [singular:- participant:-] were

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

(28)
a. [*u*singular:+] was
b. [*u*singular:-] were
c. [*u*participant:+] was
d. [*u*participant:-] was/were
e. [*u*addressee:+] was/were
f. [*u*addressee:-] was

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, as it overgenerates *they was* and *he were*.

At first glance, this also seems to be the case for [*u*addressee:+]; however, this case is different, because of the variability of the surface forms. Both singular and plural forms of *you* (that is, all forms characterised by [addressee:+]) allow both *was* and *were*. It follows that we can remove the association of the feature [*u*addressee:+] with the form *was*, hence removing the formal synonymy, without changing the predictions that the system makes about grammaticality. That is, if we have the following set of LIs, we still predict that it is possible to have both *you was* and *you were* for both singular and plural second person:

(29) a. [*u*singular:+] was
b. [*u*singular:-] were
c. [*u*participant:+] was
d. [*u*addressee:+] were
e. [*u*addressee:-] was

It is not possible to eliminate the association of the feature [*u*addressee:+] with the form *were*, or we would predict that *you were* is not a possible form for singular second person, contrary to fact.

This set of lexical items will correctly analyse the patterns of grammaticality in the data, as well as the patterns of variability. 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*addressee:-] *was*, correctly predicting the possibility of *was/were* variability with this pronoun.

Note that this is not the *minimal* set of lexical items. It is possible to remove the lexical item [*u*addressee:-], and we still capture the correct distribution of categorical and variable forms. However, (29) 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:

(30) [singular:+, participant:+, addressee:+] ... (a)was; (c) was; (d) were

(31) [singular:-, participant:+, addressee:+] ... (b)were; (c) was; (d) were

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

(32)

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

These frequencies follow on the assumption that there is a random choice of form. Although it is perhaps not the case that any individual use is random (although of course they very well may be), looking across a large enough sample of data we should see the broad patterns emerge, as we appear to here.

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.

(33) [singular:-, participant:+, addressee:-] ... (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 items, 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 [addressee:-] 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 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 are a number of implications that this system has 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). I do not claim this as an advantage, and it may be necessary to admit multiple grammars, but I do want to draw attention to the fact that there is a way of analysing at least some variability without appealing to multiple grammars. 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.

Furthermore, 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 (as in, e.g. 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 and Smith, Jennifer. in press. Variation and the Minimalist Program. In Leonie Cornips and Karen Corrigan, eds., *Syntax and Variation: Reconciling the Biological and the Social*, Amsterdam: John Benjamins.

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.

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. 1986. *Knowledge of Language*. New York: Praeger Publications.

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. 2000. 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. 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.

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

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

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

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.

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

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.