Variability and Grammatical Architecture[1]

March 2013

David Adger

1. Introduction

This paper is an attempt to connect formal minimalist syntactic theorizing with empirical problems raised by sociolinguistic investigation. This work is an extension of previous collaborations with sociolinguists over the past eight years (Adger and Smith 2005, Adger 2006, Adger 2007, Adger and Smith 2010, Cheshire, Adger and Fox 2013). In current sociolinguistic work there is only a small interest in generative syntactic models for sociolinguistic phenomena. The syntactic models generative linguists provide are considered by many sociolinguists to be orthogonal (perhaps even inimical) to their interests. Similarly, there has been scant interest in generative syntax in modeling structured variability in morphosyntactic data (with a number of notable exceptions, some of which are mentioned below). In some of the work I've done over the past eight years or so, I've been interested in exploring how generative models of syntax, devised with quite different goals in mind, might be useful for the analytical purposes of sociolinguistics. This paper reviews the basic ideas of this research programme, raises some problems for it, and sketches some ways forward. What will be crucial here is not the sociolinguistic side of relevant data, but rather the question of how to theoretically model community-internal variability, and more specifically person-internal variability.

There are a number of models in the literature which have attempted to tackle this issue: Tony Kroch has proposed that there are competing, multiple grammars (Kroch 1994) each categorically generating a variant, so that individuals who control multiple grammars have access to multiple variants, giving rise to intra-personal variability. More recently, Nevins and Parrott (2010) have developed a system involving a probabilistic version of the Distributed Morphology operation of Impoverishment, resurrecting the basic proposals of Labov (1969), in a modern theoretical setting. A more standard route, within generative syntax, pursues a view originally developed by Henry (1995), which ties the mechanisms of variation to optional movement.

The line of attack that I'm going to pursue here is different from these. It connects with issues addressed elsewhere in this volume. Cedric Boeckx, in his contribution, raised questions of 'memorization' or 'idiomaticity', and how those concepts relate to non-compositionality. Sjef Barbiers, in his contribution, raises the same issue but discusses it in terms of 'conventionalization', connecting this to questions of frequency. This will also

[1] Many thanks to the audience at the Barcelona Workshop on Linguistic variation in the minimalist framework, January 2010, and to Terje Lohndal and Peter Svenonius for comments on an earlier draft. Thanks too to Jenny Cheshire, Naomi Nagy, Sali Tagliamonte and Jennifer Smith for much needed sociolinguistic advice.

be a concern for me. Of course, these are live issues in the debate between generative grammar and construction grammar (see Goldberg 2003, Lidz and Williams 2009, Goldberg 2013, Adger 2013a and Adger 2013b). Proponents of the latter believe that grammar effectively emerges from the conventionalization of frequent collocations of items. The approach I'll pursue is effectively 'Construction Grammar in Reverse': grammar is a constraining factor in acquisition, providing the representational skeleton upon which the form meaning relationship takes flesh (cf. Borer 2003). The actual morphosyntactic properties of a particular language (especially the form and properties of morphemes) are a conventionalization of distributional occurrences, for which frequency may be an important factor, but the nature of these is severely constrained by the operation of universal grammar in a way that I will specify below. The eventual aim of the general research programme is to connect specifically minimalist theories of syntactic representations to empirical issues of interest to sociolinguists.

2. Internal Factors

The particular empirical domain I'll focus on is what sociolinguists call internal factors (Labov 1994). These are individual-internal factors connected to grammatical features and structures (or possibly to the processing of grammatical features and structures) that impact upon the choice of a particular variant in a particular context. I'll focus on two examples. The first is from Buckie, a small fishing village in the North East of Scotland. The example comes from joint work with Jennifer Smith, a native of Buckie and a sociolinguist who has spent some time working there (Smith 2000). As an example of the phenomenon I am interested in, consider (1):

(1a) Buckie boats were a' bonny graint.

'Buckie boats were all nicely grained'

(b) The mothers was roaring at ye comin' in

'The mothers were shouting at you to come in'

All speakers in this particular speech community allow this kind of variation (Adger and Smith 2010). In both examples there are plural subjects but the form of the verb is variable; in (1a) there is an apparently agreeing auxiliary *were* and in (1b) we find the apparently non-agreeing form *was*. We find this pattern also with certain pronouns. I've given, in (2) below, a table with the frequencies found in the Buckie corpus of 39 speakers (19 male, 20 female, spanning three generations), all born and raised in the community (see Smith 2000 for details of this corpus).

(2) Distribution of *was* with different Subject Types. N=4904

Subject	percentage of *was*	N
1st singular	100	691
2nd singular	69	161
3rd singular	100	2290
1st plural	67	368
2nd plural	10	10
3rd plural	0	435
Singular NPs	100	762
Plural NPs	56	187

For 2nd person singular and 1st person plural, we see that two thirds of the time the form is *was*, whereas in Standard English we expect *were*. All speakers allow this variation throughout the paradigm reported here, although we don't have sufficient numbers in this data to show that all speakers have this pattern of frequency distribution. Following standard practice in sociolinguistic variationist research, the data is presented for the whole community, embodying the idea that the individual's grammar is reflected in the aggregate at community level (e.g. Labov 1972). Similarly, for plural noun phrases (examples like 1a and 1b), we find 56 per cent *was*. Again all speakers allow this variability. Any analysis of the general pattern that proposes that, in Buckie, the patterns can be accounted for by simply substituting *was* in place of Standard English *were* is untenable, since, as can be seen in the table, the 3rd plural pronoun *they* has only the *were* form, apparently displaying obligatory agreement.

Variable agreement with plural NP subjects appears also with main verbs in the present tense, as well as with present tense *be*:

(3a) The money that that lads *is* making is nae canny compared to what I made.

"The money that those boys are making isn't good compared to what I made"

(b) Aye, and some of them *are* drawing dole on three, four different names.

"Yes, and some of them are receiving unemployment benefit using three or four different names"

(4a) What bairns walk any distance?

"What children walk any distance?"

(b) When they go back, the teachers asks them to write something and they send them till's

"When they go back, the teachers ask them to write something and they send them back to us)"

In (3) and (4), we find a plural subject in both (a) and (b) examples, but variation in whether the verb is marked with the *–s*.

Furthermore, with the pronominal *they*, the present tense main verb, just like the past and present auxiliary, appears with plural agreement (i.e. no *–s* is possible). There is just 0.6% appearance of *–s* with *they* in present tense (N=1271). Adger and Smith (2010) show that these are all narrative presents with the verbs *say*/*tell*, and they argue that *–s* here is performing a distinct function (that is, that there is a special morpheme *–s* that variably marks narrative present irrespective of the form of the subject). The conclusion that the grammar of Buckie rules out *–s* when the subject is *they* is backed up by judgment data we carried out on the individuals who were recorded for the initial corpus gathering phase of the research. We took cases where *–s* appeared with a plural subject and swapped in *they* in the same context. This led to uniform rejection by our speakers:

(5) *When they comes up to see us there Buckie judgment

(6) *When they gets home Buckie judgment

(7) *They does na do it yet Buckie judgment

(8) *They does na get what that cats get Buckie judgment

(9) *They is up in Elgin Buckie judgment

This phenomenon of variable agreement which becomes obligatory in the presence of a pronoun has been tackled in the literature by various people: Henry (1995) developed a proposal about the manifestation of this phenomenon in Belfast English which related it to differential possibilities in where pronominal vs non-pronominal subjects could be syntactically placed and how this syntactic placement impacts on agreement morphology. Tortora and Den Dikken (2010) extend this idea to Appalachian English, although more recently Bernstein and Zanuttini (2010) have proposed a featural perspective on this close to what Jennifer Smith and I argued for. As mentioned previously, Nevins and Parrot (2010) developed a proposal that ties the pattern to variable application of impoverishment rules, ultimately tied to markedness effects.

Adger and Smith (2005) proposed a basic architecture for such variable phenomena that was further developed in Adger (2006, 2007) and Adger and Smith (2010). What we argued is that the explanation of the variability is to be tied only to the featural composition of the agreeing finite element, and it has nothing to do with the syntactic position of the relata. Further, the general approach developed a learning procedure

which generates underspecified lexical items whose random combination leads to predictions of differential proportions of variants, rather than by embedding probabilities into morphological spellout rules, like the Nevins and Parrott proposal. Finally, the system allows individuals to have a single grammar, distinguishing variability from bilingualism. In fact, Adger 2006 points out that if UG is minimal, containing just the Merge operation plus constraints on the mapping of syntactic representations to semantic and morphophonological interpretations, there can't be multiple grammars in any meaningful sense: rather we have multiple individual lexical items which share interpretations and which constitute the basis of the variability and it is these items that are chosen in different speech contexts, rather than whole grammars. Of course, since these items are typically functional items, choosing one version over another will potentially have grammatical effects. Adger and Smith (2010) provide a number of arguments that this is the right way to approach the Buckie data outlined above.

An important consequence of the model developed in this work is that no radical change is required in the architecture of current minimalist syntactic theory. We already have all the technology at our disposal to tackle this issue: the featural representation of functional categories and the matching of morphological feature matrices to syntactic ones (as in, for example, vocabulary items in Distributed Morphology).

To see how this theory works schematically, take some functional category X, specified with features F, G and H, and take three vocabulary items, each of which matches one of these features:

(10a) X[F, G, H]

(bi) [F] < -- > i

(bii) [G] < -- > j

(biii) [H] < -- > j

The syntax builds X, and then X needs to be 'vocabularized'. Any of (bi-iii) will match, and so any can vocabularize X. On the assumption that the choice of vocabulary item is random, then we expect each of (bi-iii) to be inserted 33% of the time. However, if there is homonymy in the vocabulary list, as in (10b) above, then the surface forms will be unevenly distributed in terms of their frequency. In the schematic example given here, we will see 66% j to 33% i, approximately.

The idea, then, is that the variability arises because of an underspecification of the relation between the syntactic feature bundles and their morphological exponents; variability in the frequency of the forms emerges because of homonymy in the set of relevant exponents, and I proposed an algorithm (Adger 2006) that generates the set of

exponents from the input, and termed this approach to variation *Combinatory Variability*.[2]

The system is different in one important respect to standard models of the morphology syntax interface that involve underspecification, though: it does not incorporate a Subset Principle for resolving competition between lexical entries, at least not as a grammatical principle. The algorithm in Adger (2006) does reduce a certain amount of homophony, and therefore incorporates the intuition that a morphological learner generalizes beyond its input data (see also Pertsova 2007), but the algorithm produces what Adger (2007) calls a Pool of Variants, any of which can in principle be inserted. I showed that this was the right result in a number of cases.

The categoricity of agreement with pronominal *they* arises because of parametric variation in the usual sense: Buckie, like Welsh, but unlike standard English, bundles a pronominal feature, [+/–pronominal], in with person and number features in its finite agreement head (see Adger and Smith 2010 for details). The checking of this pronominal feature is what leads to the impossibility of *they* with *–s*.

However, this basic story is not quite satisfying. It lacks a place for the internal factors mentioned above in (at least) two important ways. The first is frequency effects: it is well known in the variationist literature that the frequency of a particular lexical item can impact on how robust the variability is: more frequent verbs, such as auxiliaries, appear more frequently with *–s* (see the discussion in section 4 on frequent verbs and grammatical change). In fact, it is generally true that variation is more robust for items that are more frequent (this is strikingly attested in historical change, see for example Ellegard 1953 on *do*; see also Erker and Guy 2012 for a nuancing of this idea in a synchronic corpus). For example, in Buckie frequent verbs like *be* or *do* (even just focusing on their main verb uses), when compared with less frequent verbs like *kick* or *interrupt*, have a higher proportion of *-s* with plural subjects (again, see section 4 below for the data). There is no obvious way in the combinatorial variability system as developed so far to capture that.

The second issue is how grammatical properties can influence the choice of a variant. As first noted in the Labovian tradition (e.g. Labov 1969), it turns out that there are grammatical effects on how frequent variants are. Grammatical features don't merely determine whether a structure is grammatical or ungrammatical; for variable cases, they rather impact on the probability that one variant rather than another will appear. That is, beyond the social effects on variability, which are well known (Labov 2001), there are also grammatical effects on variability. In general these effects are more 'global' than

---

[2] Adger 2006 implements the proposal in a lexicalist fashion, allowing the lexical items underlying the morphological exponents to be underspecified. However, as noted in that paper, the proposal can also be expressed in terms of underspecification between the syntactic features on a lexical item and their morphological exponents, which is the implementation offered in Adger and Smith 2010.

simply the features of a lexical item. This is the problem of internal factors mentioned above.

For example, take verbal *–s* (as the appearance of *–s* on verbs with plural subjects is often known): it turns out that in Buckie it is more likely that verbal *–s* will appear in subject relative clause constructions than it is that it will appear in non relatives (chi squared = 197.568 with one degree of freedom, two-tailed p-value less than 0.0001).

(11)

Subject relatives vs. other			
Subject relatives		other	
%	N	%	N
42	53	**34**	327

It appears that the syntactic context matters in this case. The interesting thing about this kind of effect is that the syntactic context impacts not on the grammaticality or ungrammaticality of a particular form, but rather on the probability of that form. All the context does is push the observed frequencies in one direction or the other.

Again the combinatorial variability model as it stands apparently has no obvious way of doing this. What the model states is just this: the agreeing verb form is just V plus some set of agreement and tense features. But we seem to need more information: specifically whether the clause that the verb appears in is a relative clause. The question is how we get that information into the system.

This internal factors question then raises a further issue about the empirical data: are there constraints or limits on the nature of the internal factors and if so, what characterizes these? Are they derivative of functional or surface relevant properties of the data, or are they constrained by the nature of the syntactic representation?

These are the issues in the theory of variation that I want to tackle: variability, frequency and internal-factor effects.

3. Connecting Syntactic Structure to Variable Form

What I want to do now looks extremely disconnected from this. I'm going to briefly explore the standard theory of how Merge works, and develop a particular perspective on how information flows in a syntactic representation which is reliant on the common distinction between extended projections and specifiers (Grimshaw 1991). In the system I sketch here (see further Adger 2013), I'll show that the information high up in an extended projection can have an effect on what is vocabularized low in that projection. The way that spellout, or morphological vocabularization to be more exact, is executed in this theory will give us a handle on frequency effects as well as on the limits of the impact of internal factors.

Merge is standardly taken to be binary, I have given below a (slightly simplified) definition from Collins and Stabler (2011), who formalize certain aspects of minimalism, that clarifies the stipulations inherent in Merge:

(12) Let W be a workspace and let A, B be syntactic objects where A, B belong to W and A and B are distinct. Then, external merge of A, B is equal to the set {A,B}.

Note that this definition (and all others I'm aware of that are set theoretic), has a crucial stipulation in it: that A and B are distinct. One might ask why that stipulation is there. If we remove it then we have the Merge of A and A being simply the set {A}. In fact, the idea of removing this stipulation and allowing *Self Merge* has been proposed already in the literature. Guimaraes (2000) suggested it and developed a system that constrained it by ruling certain applications of Self Merge out. Kayne (2010) uses it as a way of distinguishing nominal from verbal extended projections. Both Guimaraes and Kayne strictly limit the application of Self Merge. For example, Kayne says in a footnote in his paper, "alongside {x} there seems to be no need for {{x}}, which would be unavailable in principle if every merge operation must directly involve a head."

In Adger (2013), I propose jettisoning that assumption (that every Merge operation must involve a head), simplifying the theory of phrase structure by removing both the distinctness condition on Merge and the requirement that Merge always involves a head. That proposal then creates a system that generates unary branching structures of the kind that the classical bare phrase structure system standardly rules out (Chomsky 1995). In (13), the root Self-Merges to give a set of cardinality 1, whose member is root. We can then Self-Merge that set, giving a new set, of cardinality 1, whose member is {root}. Continuing, we generate a potentially infinite sequence of elements (the set containing that root, the set containing the set containing that root, and so on). We can represent this data structure in a more familiar form as a unary branching tree, where each x stands in for the position of a new pair of set brackets:

(13)
x
|
x
|
x
|
root

The immediate issue to address then is that of the label of these various constituents. The absence of labels might be one way to rule out iterated self-merge: there is nothing to give the output of self-merge a label (cf. the related idea in the footnote from Kayne 2010 discussed above). In a system like that of Chomsky 1995, where labels of higher structure

are always predicted from the set of labels in the contained structure, however, that proposal would still allow a structure where the label of the root is repeated at every level, requiring a further principle (perhaps an economy principle, as suggested by an anonymous referee) to rule that out. Alternatively, in a label free system like that of Collins 2002, one might allow labelless representations like (13) but take them to be ruled out by the principles that map syntactic structures to the interfaces. However, Collins 2002 is actually an argument not for the elimination of labels, but rather for the redundancy of triggered Merge and labels. Accepting the logical force of the argument, but taking the opposite line to Collins, then, I suggested in Adger 2013 that we jettison triggered Merge (following Chomsky 2004) and instead have an exocentric labeling system.

In such a system, labels are given exogenously, by something outside the core computational system: for example, one might take a language-external cognitive schema to apply to the fractionated structure that is given by the syntax, thereby labeling that structure with what are known commonly as functional categories (cf. Chomsky 2007, 14). More concretely, assume an antecedently given conceptual distinction between kind, sortal and substance concepts (see, for example, Xu 2010, Li, Dunham & Carey 2009). These cognitive distinctions are represented as linguistic distinctions via labels of pieces of structure (e.g the label Ki(nd) or Cl(assifier) from Svenonius 2008, following Zamparelli 2000 and Borer 2005). The conceptual relationships between the cognitive distinctions are represented as syntactic relations between labeled structures (e.g dominance, identity). Syntax provides a bare structure, which is labeled by cognitive schemata to give structures rooted by a lexical item dominated by a series of functional category labels: since these schemata are finite in size, the structures labeled by them are also finite, so syntax using just self-Merge provides a finite set of labeled structures (effectively, fragments of unary branching extended projections).

However, I take it to be false that the functional structure is identically represented in all languages for every root, or we would expect no variation at all. Rather language acquirers make use of what is universal (the finite set of structures generated by self-Merge and labeling) during the process of language development and associate the structures UG allows them to generate with distributional regularities in the data. The relevant regularities include morphophonological regularities (that is, what morphemes the label, or series of labels, are associated with) and morphosyntactic regularities (properties of these structures relevant to their cooccurrence with other structures). The outcome of this process is, then, a set of structures that are effectively annotated by morphophonological and morphosyntactic properties (syntactic features). I take the morphosyntactic annotations to be what bear the core 'parametric' properties discussed by Rizzi in his contribution to this volume: syntactic content and instructions for syntactic actions (see also Adger and Svenonius's 2011 distinction between first order and second order features).

The process of acquisition is then, in this model, a process that matches the possible unary extended projections with systematic regularities in the primary linguistic data (morphemes, for example), and associates these with both morphophonological and syntactic properties. Once acquisition reaches a more or less steady state, the acquirer

has, in effect, compiled a lexicon the elements of which contain fairly rich morphophonological and syntactic information. The outcome of the process of lexicon compilation is one that associates fragments of unary extended projections with both morphological and syntactic properties. These pre-compiled lexical items are available for language use and correspond to fairly classical morphemes.

A simple example of language variability to help clarify the working of these concepts is the difference between the English morpheme *child* and the Scottish Gaelic morpheme *clann*. The English morpheme is a singular count noun, while the Gaelic morpheme is what Acquaviva (2008) calls a 'lexical plural': *clann* is used to refer to a group of children, but is grammatically singular, occurring with, for example, a singular article (*a' clann*, the.fem.sing children; **na clann*, the.fem.pl children). In order to refer to a single child, a paraphrase *daoine cloinne*, person children.gen.sing, literally 'person of children' is used. In the system developed here, English *child* is the spellout of the structure that results from a series of self-Merge operations that add count semantics to a root (following Borer 2005). Its plural, *children*, is the spellout of a structure that additionally has had a plural label added to it (call this Pl). This label is, when present, interpreted as adding a plural semantics, but, in addition, it carries a plural grammatical feature (e.g. [-singular], following Noyer 1992, or Harbour 2004), a result of distributional regularities in the data available to an English child that effectively activates the syntactic relevance of the functional category label. *Clann*, on the other hand, is a root that has no spellout in the absence of a plural label. *Clann* is the spellout of the whole unary extended projection (the root, the functional structure that adds count semantics, and the functional structure that adds plural semantics. More concretely, there is a spellout for Pl-Cl-N-child, but no spellout for Cl-N-Child). The idea that a morpheme can 'span' a sequence of functional categories has its antecedents in Williams' 2003 analysis of auxiliaries. However, for *clann*, there is no grammatical (syntactic) feature [-singular] on this label because the distributional regularities in the data do not associate *clann* with this property. Gaelic does possess a grammatically plural label, which is associated with a variety of spellouts depending on noun class, gender, etc.; for example, the plural of *sùil*, 'eye' is the affixed form *sùilean*. In such cases the plural functional category Pl bears a plural syntactic feature [-singular]. Of course, the fact that *clann* already encompasses the plural functional category means that it's impossible to pluralize *clann*, in this way. Gaelic *clann* differs from English *children*, in that the latter includes while the former lacks, a precompiled association of morphophonological information (effectively the suffix *-ren* and its impact on the root) and morphosyntactic information (the annotation of [-singular] on the Pl functional category label).

Another aspect of variation that this system allows is a version of Brody's 2000 proposal that the syntactic action associated with a functional label might be an instruction to the spellout systems to pronounce the whole sequence in a particular position (Brody annotates this position with a diacritic which is interpreted by the spellout systems as an instruction to pronounce the whole word at that position in the structure, see Adger and Svenonius 2011, Adger 2013 and especially Svenonius 2012 for recent discussion). For example, in a V2 language, where the finite verb is spelled out in C, we can take the label for declarative C to bear an instruction that the phonology of the entire finite extended projection is to be linearized at C (see below for more detail and an example). That is, the

C[+finite]-T fragment of the projection line is associated with an annotation that requires the morphophonology of the whole projection line to be at C (that is, the spellout of the verb, its aspectual markers and tense and agreement markers are all spelled out in the structural position C). This is importantly different from the proposal made by Starke (this volume) who takes variation between systems to involve the spellout of whole tree structures as a single morpheme. In that system, the generalization that all finite verbs appear in C is not trivially implementable.[3]

In addition to this kind of variation between and within languages, dependent on the presence or absence of a syntactically active feature, on the 'span' of functional categories lexicalised by a morpheme, and on the position at which the morpheme is pronounced, languages also have labels associated with particular syntactic requirements, for example, the requirement that a particular label have a specifier: for T this would give rise to EPP effects, while, for a V2 language, it would apply to the C which is the spellout point for finite extended projections discussed immediately above. Again, this involves associating fragments of extended projections with distributional regularities. It is notable that the available distributional regularities seem to be extremely restricted, confined to the shape of the morpheme, its position, and its relationship to its specifier.

Self-Merge, then, generates an infinite set of sequences which are rooted. Some (finite) subset of this contains sequences whose parts are labeled; a further subset of this is extracted during acquisition and associated with morphophonological and morphosyntactic information. Some (perhaps most) morphemes will lexicalise more than one label in a unary extended projection, giving the effect of insertion of morphophonological material at non-terminal points of a structure.[4] Let us call the acquired set of such pairings between labeled unary extended projections and morphosyntactic and morphophonological representations the Compiled Vocabulary, adapting the Distributed Morphology term for this (Halle and Marantz 1994).

In addition to self-Merge, the system also has standard (binary) set-Merge in both its internal and external manifestations. Because of this, the system generates an infinite set (because of recursion) of tree structures whose nodes may branch (in which case we have a specifier structure) or not. This system, although derivational, is very similar in its yield to Brody's (2000) Mirror Theory system of phrase structure. It is a consequence of both

---

[3] This observation about the architecture and the argument that V2 is problematic for a Nanosyntactic approach to spellouts of non-terminals is die to Peter Svenonius.

[4] This architecture shares, with Nanosyntax (Starke 2009, Caha 2009) the idea that morphemes are associated with non-terminal structures, but with Mirror Theory, the idea that morphological structures are associated with unary branching head-head sequences. See Ramchand (2008) for a slightly different implementation of a similar intuition involving an autosegmental style association line architecture, and Adger, Harbour and Watkins (2009), Bye and Svenonius (2012) and Adger (2013) for further developments and applications of the idea.

systems that the syntactic complement relation is reserved for extended projections (see Adger 2013 for argument that this consequence is beneficial).

To see the system in action, applied to the data that concerns us here, take, for example, a fragment of the Buckie sentence in (4b):

(14) The teachers asks them

The verb root *ask* here self-Merges giving a syntactic object that contains just the root. This object is then labeled as a V (presumably there is a small set of choices as to how roots are labeled, Baker 2003). If Borer (2005) is correct, then the choice of label here is available to a speaker throughout his/her life, so that roots can be ambiguously nominal or verbal.

The label V bears some feature that provides a well formed interpretation when a specifier is Merged as an object (or further Self Merge leads to an aspectual category, as in Ramchand (2008), that bears such a feature), allowing Merge of *them* followed by further Self Merge which is labeled v, which allows Merge of the subject, as is standard. Further Self Merge builds T, Internal Merge raises the subject, and further Self Merge builds matrix C. So much is fairly standard, with Self Merge and Labelling taking the place of external merge of heads. The final structure is a 'telescoped' structure, similar to Brody's:

(15)

C

T

The teachers v

< The teachers > V

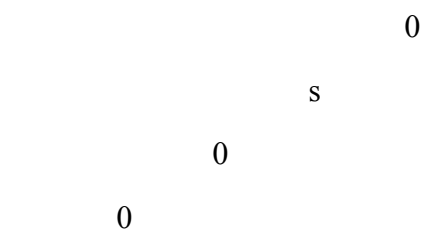
them ask

With this in place, let us turn in more detail to how these structures are spelled out. In Brody's system, a core principle is Mirror, which states that:

(16) The syntactic relation "X complement of Y" is identical to an inverse-order morphological relation "X specifier of Y" (Brody 2000, 42)

Brody's system gives us an agglutinative interpretation of the morphology, so that the morphological structure associated with (15) is:

(17)



The second part of Brody's system is the diacritic mark, mentioned above, that specifies at which point of the extended projection (what Brody calls the extended word) the morphological form is pronounced. If v bears this mark, then the word *asks* will be spelled out 'in' v: that is, it will be spelled out after the subject and before the object.

I adopt this proposal here, combining it with the idea, already discussed, that overt, apparently atomic, morphemes may 'span' more than one category label. For example, the relevant compiled vocabulary item is [be V v T[past] C] <--> *was*, capturing the suppletion. Note however, that in addition to this, in the absence of any further stipulation, we also have to have [be V v T[past] C] vocabularized as be-0-ed-0, since each of these labels, at least in principle, and probably in this case in actuality, has a vocabularization independently of the others. The root [be-V] needs to be associated with *be*, [v] presumably can have a zero realization, as can [C], and T[past] has an *ed* realization with regular verbs.

Standardly, one would appeal to a blocking principle (Kiparsky 1973) to rule out *beed* as the past tense of *be*. However, given the framework outlined above for dealing with variation, blocking cannot be a principle of grammar, it must rather be a fact about the routinization of structures that is emergent from the learning process, so both *was* and *beed* have to coexist as possible vocabularizations of this syntactic structure (where by vocabularization, I just mean the morphophonological properties associated with the built syntactic object). As well as the item *was* in the Compiled Vocabulary, the grammar allows the construction of a set of alternative forms, including the regular, but non-occurring, *beed*. In terms of the Combinatorial Variability model, they are both in the Pool of Variants, however *beed* is so low in that Pool of Variants (that is, its probability is miniscule compared to *was* because its frequency in the input during acquisition is so low) that although it matches the category label sequence, it is never (or very rarely) going to be put to use.

During development the child is using all of the syntactic resources given to her by UG to build up syntactic structures and to linearize them into morphological chunks. However, as the child's development progresses, variation emerges as a result of the tension between the child's generated structures and the evidence from the input. This variation then reduces as the child begins to attempt to match the frequency of her caregivers' variants (Smith et al 2007). The evidence for the non-routinized version *beed* is very low,

so the variation vanishes from the child's production (although not from the child's I-language). Blocking is then simply a side effect of this process: the routinized version is associated with such a high probability that the regular version is judged as unacceptable.

Given this, a non-suppletive structure like *asks* will also exist in (at least) two forms, a routinized version (or perhaps more than one of these, see below) and the version that is the algorithmic output of the linearization of the syntax (the equivalent of *beed*).

In situations where the child's input is variable, more than one vocabularization of a single extended projection segment is possible: one of these will be the output of the morphological linearization of the structure on a label by label basis, but the child's input may consist of only routinized suppletive forms (for example, *was* and *were*), and so the child will have these forms matched to sequences of syntactic categories and keep them both in the pool of variants.

With this framework in mind, consider again the variable agreement in Buckie. Maintaining the approach sketched above, we have two vocabularizations for the root *ask* in the context where it has a plural NP subject: *ask* and *asks*. Adger and Smith (2010) show that an application of Adger's (2006) algorithm to the Buckie data results in a system that allows T with a NP[-singular] subject to bear, via agreement, a feature [-singular] associated with a zero phonology and a feature [-pronominal] associated with the phonology *–s*. This results in a choice of which is spelled out (see Adger and Smith 2010 for details), giving rise to the variable effect.

The latter of these forms itself has both a routinized and a generated source. The more frequent the verb *ask* is in the input, the more routinized its forms are. It is always possible to generate the *–s* form, (as indeed it is for the *–0* form) by associating the relevant feature with the relevant morphophonological signal in essentially an agglutinative fashion, but there may be no routinized form. In the absence of such a routinized form, or if such a form is barely routinized, then there will only be one way to get to the *–s* form: linearize the syntax into an appropriate morphology. But a compiled Vocabulary Item is, of course, more immediately accessible, creating a downward pressure on the use of the syntactically computed version. In this way, the system captures the generalization that the variability between the forms is more robust for frequent verbs than for infrequent ones.

The approach also gives us a handle on the problem of internal factors. Indeed it begins to give shape to a possible theory of internal factors. Note that in the structures generated by the theory, C is actually part of the possible vocabularization of the verb. Although the verb is pronounced 'in' v, C is part of the projection line, and hence a vocabularization that results in a verbal *–s* form may span C, and hence the relative/non-relative specification of C will be relevant to that vocabularization.

Thinking about the higher proportion of *–s* in relatives in Buckie, we can see this as combinatorial variability applied not to a feature bundle at the head (i.e X0) level of structure, as in the theory presented in Adger (2006), but rather to a sequence of labels, where certain vocabulary items span the C that appears in a relative clause (C[rel]), while

others don't. Putting aside the featural details covered in Adger (2006) and Adger and Smith (2010), we would have something like the following:[5]

(18a) [V v T C[rel]] < -- > -s

(b)[V v T] < -- > -s

(c) [V v T] < -- > -0

Now in a non-relative, only (b) and (c) match the syntactic context, predicting a rough equality in the frequency of each variant. However, in a relative clause, all of (18a-c) are available for insertion, which predicts a higher frequency of the surface form *–s*. Of course, this is a toy example, as numerous other factors will be relevant (the featural content of the subject and hence T, possible aspectual information contributed by v (see, e.g. Poplack and Tagliamonte 1989) or an Asp projection, the frequency of the verb root itself, etc.). However, even in this toy case we can see that the system constrains the set of potentially relevant internal factors to be those that are encoded in the projection line of the verb. We predict, then, that this kind of agreement variation cannot be affected by, for example, the number of the subject of a clause embedded by the agreeing verb, or the subject of a sentential subject of the verb. To my knowledge this has never been tested because it was thought to be so unlikely. The theory here makes a robust prediction that it is unlikely.

4. Language Change and Variability

Let me now turn to some broader issues connected to questions of acquisition and change. In terms of acquisition, of course, the approach outlined here allows us to connect fairly well with some of the work that has been done by proponents of Construction Grammar and their predecessors (e.g. Bybee 2006, and for work on variable agreement systems see, especially, Pietsch 2005), who have pointed out that there are frequency effects in the process of development that potentially have an impact on frequencies in corpora and on ease of processing. However, the Combinatorial Variability system allows us to do this while keeping the notion of frequency or probability firmly outside the grammar: it is relevant only for questions of lexical access during production. The grammar generates a restricted range of variants depending on the syntax (the Pool of Variants) and then some function will apply to the Pool of Variants to select a single one in any occasion of use. The structure of the Pool of Variants, how homonyms are distributed within it and whether there are both generated and routinized forms, has a direct impact on the frequency distribution of the surface forms. In addition, the Pool of Variants may be externally structured by the processing systems that effect lexical access: some items are generally highly routinized and easily accessible, so that when they appear within the Pool of Variants this property itself has an effect on whether that item is chosen.

---

[5] Actually the data we have shows this effect for subject relatives (we don't have the relevant data for object relatives), so the relevant features will have to include the phi-features on T shared with the subject, if, indeed the effect is restricted to just subjects.

On the question of language change there are two interesting phenomena I want to briefly mention here. The first is related to the loss of variability and the emergence of a possible parametric setting, while the second is related to the emergence of grammatical structuring in previously unstructured variation.

In Buckie there is a change happening in variable agreement. Older speakers use more *–s* forms with plural NP subjects than younger speakers as is seen in the following table, which includes all kinds of apparently singular agreement with plural subjects (*was*, *is*, other auxiliaries and main verbs) (chi squared value is 64.674, with two degrees of freedom and a two-tailed p value of less that 0.0001):[6]

(19)

Distribution *–s* in plural NP contexts by age. N=380

old		middle		young	
%	N	%	N	%	N
73	59	**24**	136	**32**	185

There's an interesting dip in the middle aged speakers which appears in many of the variables which have been investigated for this corpus (see Smith 2000). Smith argues convincingly that this is due to a particular sociolinguistic factor connected to language standardization for middle aged females which depresses the numbers.

What is interesting in the context of the approach developed here is that this change is affecting auxiliaries more than it is affecting main verbs:

[6] This discussion reports joint empirical work with Jennifer Smith.

(20)

Distribution *–s* with plural NP by age and verb type. N = 375						
	old (chi squared = 0.632, p=0.4267)		middle (chi squared = 18.657, p less than 0.0001)		young (chi squared = 31.391, p less than 0.0001)	
	%	N	%	N	%	N
auxiliaries and copulae	**56**	32	**69**	92	**71**	130
main verbs	**44**	25	**31**	42	**29**	54

In the older generation, there is no significant effect of whether the *-s* appears on an auxiliary or on a main verb. In the middle and younger generations a marked shift appears and there is a highly significant effect of auxiliary versus main verbs. Given the theory of intra-personal variability sketched above, we can attribute this change to the frequency difference alluded to above (and in fact to the further effect of the suppletive forms of the auxiliaries, which implies that they must span the sequence of labels). Given that children track the frequencies of their caregivers with some exactitude (Smith, Durham and Fortune 2007), the existence of change in the frequencies of variables seems odd. However, at some point in development, children alter their frequency of use of particular variables. This is possibly connected to changes in how the developing child fixes its linguistic affiliation, changing from its family to its peer group, or at least adding to its linguistic repertoire. Assume that children alter this frequency uniformly, depressing their use of a particular variant by a stable amount. In such a case, the more frequent verbs will maintain a higher proportion of use of the disfavoured variant, an effect heightened by the upward pressure exerted by the accessibility of these compiled forms. However, the grammatical system is still the same: the grammar still produces plural NP subjects with *–s* and we would predict that novel verbs, for example, would be grammatical in both forms. Low frequency verbs that are variable would arise just via application of the grammatical principles involving the linearization and vocabularization of category labels. However, given the frequency of auxiliaries, we then see a bifurcation between auxiliaries and main verbs that depends precisely on the fact that auxiliaries are in almost every sentence, whereas the different main verbs are much rarer. Now, when these speakers bring up the next generation, after iteration of the process, the variable agreement, depending on its sociolinguistic status, may become yet more disfavoured until at some point the bifurcation between auxiliary and non-auxiliary verbs becomes so sharp that acquirers attribute what was a variable distinction to a categorical property of grammar (for example the auxiliary/main split). This would then correspond to the emergence of a parametric distinction between varieties that develops from the loss of variation interacting with frequency effects.

The other side of language change is explored in Cheshire, Adger and Fox (2013). In that paper we examine the appearance of a new condition on variation in the speech of a group of London teenagers. The study shows that, for both the teenagers and the older generation, both *who* and *that* can be used as subject relativizers with animate antecedents of the relative clause, just as in Standard English:

(21) apparently a chav is like someone that wears like big gold chains

(22) I'm the only one who's gone to college

As in standard English, *who* does not appear with inanimates. However, although the frequency of *who* as a subject relativizer with animates is the same in each generation, there are very different conditions on the use of *who*. The younger generation are statistically more likely to use *who* than *that* when the antecedent of the relative is an ongoing topic of the discourse. No such effect is found for the older generation, so what we see is the emergence of a structuring of the variation.[7]

We capture this behaviour within the kind of system developed here by specifying the relativizers as follows:

(23) a. [C, relativiser:+] <--> *that*

b. [C, relativiser:+, animate:+] <--> *who*

c. [C, relativiser:+, animate:+, topic:+] <--> *who*

Any of these feature bundles can occur with an animate topical antecedent. However, when the antecedent is non-topical the vocabularizations in (23) are not all possible: non-topical antecedents can only appear with (a) (pronounced as *that*) and (b) (pronounced as *who*). Schematically, we have the following situation:

(24) a. NP[animate:+, topic:+] (a) *that*, (b) *who*, (c) *who*

b. NP[animate:+, topic:-] (a) *that*, (b) *who*

This captures the higher proportion of relativizers pronounced as *who* that occur when the antecedent is topical: since both (b) and (c) can be used with a topic, but only (b) with a non-topic, topics are expected to occur with *who* more often than with *that*, while there is no such expectation for non-topics.

Cheshire, Adger and Fox argue that the availability of the topic feature in the specification of the relative complementizer (we take *who* just to be a realization of C) arises because of the multilingual environment the speakers develop their linguistic competence within, an environment which includes many languages which have

---

[7] In fact, this effect only holds for the teenage group from multilingual inner city London. Another group that was investigated, from a non-multilingual part of London, do not show this effect.

grammatical marking for topicality and other information packaging features (see the paper for discussion). However, what is crucial for the discussion here is that the co-opting of this feature creates a differentiation in the Pool of Variants available to speakers in topical vs non-topical syntactic environments, and hence a difference in the frequencies with which the variants are produced in the different environments. That imbalance may then lead to a change in the grammatical system, in this case actuated by the multilingual environment rather than the attenuation of grammatical contexts due to sociolinguistic pressure to differentiate generational usage patterns.

5. Conclusion

At a general architectural level, the system developed here and in the cited papers allows us to maintain a strictly modular grammatical architecture, which keeps the theory of syntax fairly pristine, both in terms of the mechanisms and operations involved, and in terms of the theoretical vocabulary over which they operate. Moreover it provides the wherewithal for theoretical linguists to begin to address questions of individual variability, frequency and the way that internal factors structure variability, topics that have not been addressed in generative grammar for many years. At the same time, this system connects well with current sociolinguistic work on language variability and change, and provides a viable grammatical architecture for sociolinguists interested in linking their work with linguistic theory.

6. References

Acquaviva, Paolo. 2008. Lexical Plurals: a morphosemantic approach. Oxford: Oxford University Press.

Adger, David. 2006. Combinatorial variability. Journal of Linguistics 42:503-530.

Adger, David. 2007. Variability and modularity: A response to Hudson. Journal of Linguistics 43:695-700.

Adger, David. 2012. Constructions are not explanations, lingbuzz 001675.

Adger, David. 2013a. Constructions and grammatical explanation. Mind and Language, to appear.

Adger, David. 2013b. A Syntax of Substance. Cambridge, MA: MIT 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.

Adger, David and Smith, Jennifer. 2010. 'Variation in agreement: A lexical feature-based approach', Lingua 120. 1109–1134

Adger, David and Svenonius, Peter. 2011. Features in minimalist syntax. In Cedric

Boeckx, ed., Oxford Handbook of Linguistic Minimalism, 27-51, Oxford, UK: OUP.

Baker, Mark. 2003. Lexical Categories. Cambridge, U.K.: Cambridge University Press.

Bernstein, Judy and Raffaella Zanuttini. 2010. 'What is verbal -s in English?' Ms. Yale.

Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In John Moore and Maria Polinsky, eds., The Nature of Explanation in Linguistic Theory, 31-67, Stanford, CA.: CSLI Publications.

Borer, Hagit. 2005. In Name Only: Structuring Sense. Oxford: Oxford University Press.

Brody, Michael. 2000. 'Mirror theory: syntactic representation in perfect syntax'. Linguistic Inquiry 31:29-56.

Bybee, Joan. 2006. 'From usage to grammar: The mind's response to repetition' Language. 82.4, 711-733

Bye, Patrick and Svenonius, Peter. 2012. Non-concatenative morphology as an epiphenomenon. In Jochen Trommer, ed., The Morphology and Phonology of Exponence, Oxford, UK: Oxford University Press.

Caha, Pavel. 2009. The Nanosyntax of Case, PhD, University of Tromso.

Cheshire, Jenny, Adger, David and Fox, Sue. 2013. 'Relative who and the actuation problem', Lingua 126: 51-77.

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

Chomsky, Noam. 2004. 'Beyond Explanatory Adequacy' in Adriana Belletti, ed, Structures and Beyond, OUP, New York.

Chomsky, Noam. 2007. 'Approaching UG from below'. In Hans-Martin Gaertner and Uli Sauerland, eds., Interfaces + Recursion = Language? , Berlin and New York: Mouton de Gruyter.

Collins, Chris. 2002. 'Eliminating labels'. In Sam Epstein and Daniel Seeley, eds., Derivation and Explanation in the Minimalist Program, 42-64, Blackwell Publishing.

Collins, Chris and Stabler, Edward P. 2011. A formalization of minimalist syntax, lingbuzz/001691.

Ellegard, Alvar. 1953. The Auxiliary 'Do': The Establishment and Regulation of Its Use in English. Almqvist & Wiksell, Stockholm.

Erker, Daniel and Guy, Gregory. 2012. The role of frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. Language 88:526-557.

Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. Trends

in Cognitive Sciences 7:219-224.

Goldberg, Adele E. 2013. Explanation and constructions. Mind and Language, to appear.

Grimshaw, Jane. 1991. 'Extended projections', ms, Brandeis University.

Guimarães, Maximillian. 2000. 'In Defense of Vacuous Projections in Bare Phrase Structure,' in M. Guimarães, L. Meroni, C. Rodrigues & I. San Martin (eds.) University of Maryland Working Papers in Linguistics, 9, 90-115.

Halle, Morris and Marantz, Alec. 1994. Some key features of Distributed Morphology. In Andrew Carnie, Heidi Harley, and Tony Bures, eds., Papers on Phonology and Morphology, 275-288, MITWPL 21.

Harbour, Daniel. 2004. On the unity of `number' in semantics and morphology, Lingbuzz 000089.

Henry, Alison. 1995. Belfast English and Standard English: Dialectal variation and parameter setting. New York and Oxford: Oxford University Press.

Kayne, Richard S. 2010. Antisymmetry and the lexicon. In Anna Maria Di Sciullo and Cedric Boeckx, eds., The Biolinguistic Enterprise, 329-353, Oxford, UK: Oxford University Press.

Kiparsky, Paul. 1973. 'Elsewhere in Phonology', S. Anderson and P. Kiparsky, eds, A festschrift for Morris Halle, 93-106. New York: Holt, Rinehard and Winston.

Kroch, Anthony. 1994. 'Morphosyntactic variation'. In Beals, K., J. Denton, R. Knippen, L. Melnar, H. Suzuki and E. Zeinfeld, eds., Papers from the 30th regional meeting of the Chicago Linguistic Society, volume 2: the parasession on variation in linguistic theory. Chicago: Chicago Linguistics Society, 180-201.

Labov, William. 1969. 'Contraction, deletion, and inherent variability of the English Copula', Language 45. 715-762.

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

Labov, William. 1994. Principles of Linguistic Change. Volume 1: Internal Factors. Oxford: Blackwell.

Labov, William. 2001. Principles of Linguistics Change. Volume 2: Social Factors. Oxford: Blackwell.

Li, Peggy, Dunham, Yarrow & Carey, Susan. 2009. Of Substance: The Nature of Language Effects on Entity Construal. Cognitive Psychology 58, 487-524

Lidz, Jeff and Williams, Alexander. 2009. Constructions on holiday. Cognitive

Linguistics 20:177-189.

Nevins, Andrew, Parrott, Jeffrey.K. 2010. 'Variable rules meet impoverishment theory'. Lingua 120, 1135–1159.

Noyer, Rolf. 1992. Features, positions and affixes in autonomous morphological structure. Ph.D. thesis, MIT.

Pertsova, Katya. 2007. Learning form-meaning mappings in the presence of homonymy. Ph.D. thesis, UCLA.

Pietsch, Lukas. 2005 Variable Grammars: Verbal Agreement in Northern Dialects of English. Max Niemeyer Verlag, Tuebingen.

Poplack, Shana, & Tagliamonte, Sali A. 1989 'There's no tense like the present: Verbal s inflection in early Black English'. Language Variation and Change 1:47–84.

Ramchand, Gillian. 2008 Verb Meaning and the Lexicon: A First Phase Syntax. Cambridge: Cambridge University Press.

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

Smith, Jennifer, Durham, Mercedes and Fortune Liane. 2007. '"Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of variation in a Scottish dialect'. Language Variation and Change, 19.63–99.

Starke, Michal (2009) Nanosyntax: A Short Primer to a new approach to language, lingbuzz/001230

Svenonius, Peter. 2008 'The position of adjectives and other phrasal modifiers in the decomposition of DP' in Louise McNally and Chris Kennedy (eds) Adjectives and Adverbs: Syntax, Semantics, and Discourse, Oxford University Press, 16-42.

Svenonius, Peter. 2012. Spanning, lingbuzz/001501.

Tortora, Christina, and Marcel den Dikken. 2010. 'Subject agreement variation. Support for the configurational approach'. Lingua 1089–1108.

Williams, Edwin. 2003. Representation Theory, MIT Press, Cambridge, Mass.

Xu, Fei. 2010. Count nouns, sortal concepts, and the nature of early words. J. Pelletier (ed.), Kinds, things, and stuff: New Directions in Cognitive Science , Vol. 13 (pp. 191-206). Oxford University Press.

Zamparelli, Roberto. 2000. Layers in the Determiner Phrase Garland, New York.