ENGLISH DIALECT SYNTAX IN WORD GRAMMAR

ABSTRACT

The paper focuses on inherent variability in syntax and the challenge that it presents for theories of language structure, using illustrative data from the Scottish town of Buckie (Smith, 2000). Inherent variability challenges a linguistic theory at three levels of theoretical adequacy: structural (Does the theory distinguish the relevant structures?), contextual (Does it allow structures to be related directly to their social context?) and behavioural (Does it allow an explanation for the observed frequencies?). The paper summarises the relevant claims of Word Grammar and shows (1) that it has at least as much structural adequacy as any other theory, (2) that it has more contextual adequacy than other theories because it formalises the theory of Acts of Identity, and (3) that it at least provides a theoretical foundation for future advances towards behavioural adequacy. The paper also argues against the minimalist analysis of the *was*/*were* alternation in Buckie (Adger, 2006).

§1 INHERENT VARIABILITY AS A TEST OF THEORETICAL ADEQUACY

What do English dialects have to do with linguistic theory? A number of connections are possible, but the most interesting one concerns the notion of **inherent variability** – the fact that dialect study always turns up cases where 'variability' of form is 'inherent' in the dialect concerned. Inherent variability is common in phonology (e.g. rhyme /r/ may be present or absent in words such as *farm*) and in morphology (e.g. the {ing} suffix may end in either a velar or an alveolar nasal consonant, as in *hunting* or *huntin'*), but it is also found in syntax (Cornips and Corrigan, 2005; Hudson, 1997; Labov, 1972a). For example, some dialects of English allow variation between *was* and *were* where Standard English allows only one or the other form – a matter of syntax rather than mere morphology because the choice of verb-form is influenced by the choice of subject. We shall return to this example in some detail below.

But what do English dialects have to do with inherent variability? The link lies in the history of our subject rather than in the subject-matter itself, because inherent variability pervades language and is by no means unique to 'dialects', whatever we may mean by this term. The term *inherent variability* refers to the existence of alternative expressions for the same meaning, a notion that is already very familiar under the term *synonymy*. However, synonymy tends to be interpreted more narrowly, as the relation between distinct words that share the same meaning. Inherent variability includes these, but also a wide range of other patterns:

- syntactic variability (e.g. the position of the particle in phrases such as *look up the word* versus *look the word up*).
- morphological variability (e.g. the choice between *was* and *were*).
- phonological variability (e.g. the presence or absence of /r/ in words such as *farm*).
- phonetic variability (e.g. the details of how /r/ is pronounced).

All these kinds of variation are found in standard English, so the link between inherent variability and 'dialects' lies not in the facts but rather in the history of our discipline. One link is a matter of pure methodology: non-standard dialect studies usually use a collection of texts rather than self-report, so quantitative analysis is possible. Another is the historical fact that it was in quantitative dialectology, as applied to non-standard vernacular speech, that William Labov developed the first methods for studying variability. In short, non-standard dialects have turned out to be a particularly good test-bed for general theories of the inherent variability which is in fact found throughout language.

Inherent variability is a serious test for any linguistic theory because it requires three different kinds of linguistic analysis:

- **structural**: the variants (the competing expressions) convey the same meaning but otherwise they have different structures.
- **contextual:** the variants are associated with different social and linguistic contexts.
- **statistical**: the variants have different probabilities of use.

This list helps with the analysis of inherent variability by defining its component parts, but in doing so it reminds us that each of these parts applies outside the realm of what we normally understand by inherent variability.

- Structure: Expressions may have the same meaning in completely different languages (e.g. French *homme* and English *man* are synonyms), but when these alternate we speak of code-mixing rather than inherent variability.
- Context: Contextual associations may apply to expressions that cannot be involved in inherent variability because there is no synonymous alternative; to take a simple example, the term *illocutionary* is associated with linguists and philosophers, but so far as I know it has no ordinary-language synonym.
- Statistics: Every expression has a measurable probability, regardless of whether it alternates with a synonym. This probability is mentally real – frequency effects pervade all our knowledge, and are particularly well-researched in language; for

example, frequency is a major determinant of how easily we recognise a word (Harley, 1995:146-8).

These three mental phenomena are independent of one another, and inherent variability is simply one point where they happen to intersect.

If this is true, then there will be nothing left to say about inherent variability once we have a complete account of each of the three phenomena. This hypothetical state lies in the very distant future, but meanwhile we can map out a rough map of the terrain to be covered in terms of a hierarchy of theoretical adequacy comparable with Chomsky's observational, descriptive and explanatory adequacy:

- **Structural adequacy**: A theory is structurally adequate if it explains the intra-linguistic structures of a language (e.g. the relations between words and meanings).
- **Contextual adequacy**: A theory is contextually adequate if it explains the extra-linguistic structures of a language (e.g. the relations between words and their typical speakers).
- **Behavioural adequacy**: A theory is behaviourally adequate if it explains performance in relation to (among other things) the speakers' competence, including the relative frequencies of variants in inherent variability.

This is the terrain that we and our theoretical descendants have to cover before we can claim to understand inherent variability; and by the time we understand that particular part of language, we shall probably be near to understanding the whole of language.

I described the list of adequacies as a hierarchy of levels because structural adequacy seems to be the easiest goal and behavioural adequacy the hardest.

- Structural adequacy: We have a great many competing theories of language structure, all of which explain the structural part of inherent variability, shared meaning.
- Contextual adequacy: Very few existing theories can explain situational constraints, but some can.
- Behavioural adequacy: This is barely in sight for any theory at the present time. For example, I see no immediate prospect of even the beginnings of an explanation for the very regular tendency for nouns (including pronouns) to account for 37% of word tokens in formal written English (Hudson,1994), or for the fact that the percentage of nouns increases with age through childhood (Hudson, 2004).

The rest of this paper is an attempt to apply these ideas about inherent variability in syntax and about levels of adequacy to one particular theory, Word Grammar (**WG**); consequently, the next section explains the most relevant tenets of this theory. In the following sections I consider how WG fares in relation to the three levels of adequacy, so for the sake of concreteness I tie the discussion to one small dataset, the figures for the was/were variable in the dialect of a town called Buckie which I present in section 3. I shall argue in section 4 that WG achieves structural adequacy as easily as every other theory of language structure, but section 5 will claim that WG does better than other theories on contextual adequacy because it, and perhaps it alone, allows linguistic items to be linked directly to situational categories. Unfortunately section 6 will admit that WG fails on behavioural adequacy, though it may offer a better basis for progress than most other theories thanks to the fact that it includes the theory of processing introduced in section 2.3.

§2 SOME TENETS OF WORD GRAMMAR

WG was first named in 1984 (Hudson, 1984), when it evolved out of Daughter-Dependency Theory (Hudson, 1976) which in turn had evolved out of Systemic Grammar (Hudson, 1971). During its evolution it has adopted features which are found in other linguistic theories, so some of the tenets listed below can be found elsewhere, but the total theoretical package is unique. Even the theory called WG has changed a great deal since 1984, and the version described here is the one presented in detail in Hudson (2007).

Perhaps the most important and distinctive fact about WG for inherent variability is that it includes sub-theories of learning and processing as well as language structure. The following sub-sections present these sub-theories.

§2.1 Language structure

As far as **language structure** is concerned, the main claim is that language is a **cognitive network** – a vast collection of concepts connected to each other by links of various kinds. The concepts concerned include not only meanings, but also words, morphs and sounds; and the links include not only the 'isa' relation of classification (as in 'He is a linguist', 'A robin is a bird' or '*cat* is a noun'), but also a host of more or less specific relations such as 'subject', 'realisation' and 'sense'. This idea is by no means unique to WG. Apart from the general structuralist view of language as a 'system where everything hangs together', the main source for WG is Stratificational Grammar (Lamb, 1966; Lamb, 1999), but the network view can also be found in other cognitive-linguistic theories such as Construction Grammar and Cognitive Grammar (Goldberg, 1995; Langacker, 2000;

for a survey of such theories see Gonzálvez-García and Butler, 2007). Other theories might accept that parts of language, such as the lexicon, are a network, but not that everything in a language could be reduced to a single network.

A number of important consequences for inherent variability follow from the idea that language is a cognitive network:

- If the network is cognitive, it belongs to an individual rather than to a community, so one aim of work on inherent variability (apart from collecting data and analysing it) is to understand how individuals handle inherent variability in their language knowledge and also in their own behaviour.
- A cognitive network for language must be related in some way to the other cognitive networks which are part of standard psychological theory (Reisberg, 1997: 256-303), not least the whole of long-term memory, i.e. general encyclopedic knowledge. The easiest assumption – and one for which there is good evidence – is that the networks are all integrated into a single super-network. This view is diametrically opposed to the idea that language is a module with boundaries separating it from the rest of knowledge (Chomsky, 1986). For inherent variability, this means that individual items (such as words or sounds) in the language sub-network may easily be related to non-linguistic concepts; for example, a word may be related to a particular kind of person or situation.
- If language is nothing but a network (as WG claims), then a concept is nothing but a node in the network, so its 'content' consists of nothing but its relations to other nodes. The content is not held in some other format – there is no other format – and less still in the label, because labels are redundant. One consequence for inherent variability is

that social concepts are defined in the same way as linguistic ones; for example, the concept 'Londoner' which might be related to some linguistic expression is also related to the concepts 'person' and 'London'. These other links provide a definition of the concept and explain its cognitive status – including, of course, its imprecision. In Chomskyan terminology, the linguistic concepts of 'I-language' are related to social concepts of what we might call 'I-society', the individual's internal representation of social structure. Of course, I am not suggesting that a dialect syntactician is obliged to provide a complete analysis of I-society for every social category invoked; but what I am suggesting is that social concepts should be presented as nodes in a network which could, in principle, be expanded. This is an important difference between WG and the two other theories that I know which allow 'social' information in the linguistic analysis: Head-driven Phrase Structure Grammar (Pollard and Sag, 1994: 94) and Systemic Functional Grammar (Halliday, 1994); in both of these theories, social features are simply undefined labels which could not, even in principle, be linked to non-linguistic categories. Other cognitive-linguistic theories could generalise to social structure in the same way, but although this idea was introduced in Kemmer and Israel (1994), so far as I know it has not been taken any further since.

- If language is a network, the structures which we assign to tokens of language in speaking and listening must also be networks. In a network, relations are primitive in the sense that if two nodes are related directly, this must be shown by a direct link between them. For syntax (including dialect syntax), this means that the dependencies that all theories recognise must be basic, and not merely derived from phrase structure; and since phrases can be derived from dependencies, they are redundant. In short, network theory requires dependency structure rather than phrase structure – a conclusion that has always been fundamental to WG, but which has not yet been drawn by other cognitive linguists. Fortunately, dependency structure is much easier and more revealing than phrase structure for most purposes, including the kinds of analysis that are of interest in dialect syntax. This will emerge from the Buckie data, where we shall need to relate the choice between *was* and *were* to the choice of subject pronoun. In phrase structure the relation between *was* and (say) *we* is very indirect thanks to the various intervening phrasal nodes, but in dependency analysis the pronoun *we* (without any phrasal node) is directly related to *were*.

- If language is a network, it is completely static and declarative, like a map rather than a set of directions. This rules out all procedures, including those sometimes invoked in morphology – rules of affixation, vowel change and the like. Instead of processes we have relations, including the very general relation 'realisation' which links a word to its morphological structure (and then links the latter to a phonological structure). For inherent variability this means that variants are alternative realisations of some more abstract entity, with equal status; in contrast, procedural models such as 'variable rules' (Labov, 1972b) encourage or even force the analyst to take one variant as basic and to derive the other from it.
- The realisation relation forces a clear architecture in terms of 'levels of analysis' such that realisation applies only between levels. WG recognises three distinct levels where some other theories recognise only two:
    - syntax, whose units are words
    - form, whose units are morphs and morph-combinations

  - phonology, whose units are sound-segments.

This is important for inherent variability because variation between syntax and form is different from that between form and phonology; for example, the choice between the morphs {was} and {were} is independent of the phonological options for realising either of these morphs.

- If language is a network, there is no boundary between 'grammar' and 'lexicon', but instead there is a clear hierarchy of more or less general categories ranging from 'word' through general classes such as 'verb' and 'auxiliary verb' to specific lexemes such as the verb BE, sub-lexemes such as the BE which is used with progressive participles and even specific examples of sub-lexemes. For dialect syntax, this means that we cannot define the field as the study of general (i.e. non-lexical) features; the only distinctive characteristic is that it involves two or more syntactically related words. The hierarchical view also means that we are free to invoke categories that are as specific as we may find helpful; for example, when dealing with the verb BE in Buckie we can distinguish very specific uses of *was* and *were* as part of the grammar.
- Moreover, if language is a network, there is no boundary between the stored types of language and the temporary tokens of ongoing experience; tokens inherit from types, and types are induced from tokens. One consequence is that stored types may include situational features of tokens such as the typical speaker or time of utterance – categories which are crucial for inherent variability.

These are not the only general principles that follow from the basic theoretical premise that language is a cognitive network, but they are probably the ones which are most directly relevant to inherent variability and dialect syntax.

§2.2 Learning

As in other cognitive-linguistic theories, language is learned rather than innate, and the learning is 'usage-based' (Barlow and Kemmer, 2000). Rather obviously, this means that we learn language from our experience of 'usage' – i.e. what other people say and write; but less obviously, it also means that the knowledge that we learn in this way reflects the usage on which it is based. In outline, we first store particular tokens (the 'exemplars' of Goldberg, 2006) and then we induce generalisations across tokens to give increasingly general categories and patterns; and since we continue to experience usage throughout our lives, our language continues to change too. This very dynamic view of the language network has a number of consequences for inherent variability in syntax.

- There is no innate 'universal grammar' to provide the analytical categories that a child applies to its experience. Instead, every category must be induced from experience, and different dialects – indeed, different speakers – may induce different categories or have categories with slightly different defining characteristics. Seen from this perspective, what is striking (and in need of explanation) is the uniformity of actual speakers, but even so we may expect some variation at all levels of analysis, including the social categories of dialect syntax.

- In any kind of learning, frequency has an effect on the learning. If language is learned by general-purpose learning mechanisms (as WG claims), then we should expect the frequency patterns of usage to have an effect on a learner's stored language – as indeed they do. For example, as mentioned earlier, the frequency of a word has a strong affect on its accessibility (Reisberg, 1997:45). It would be strange if frequency had no effect in inherent variability, where two alternatives that have different frequencies are a priori likely to have different probabilities of being selected. One interpretation of the frequency effect is in terms of 'resting activation' as I explain in section 2.3.
- Children exposed to inconsistent input from adults can be expected to reproduce this in their own speech. There is evidence from Buckie that this is true of children as young as age two (Smith, Durham and Fortune, 2007).
- Conversely, adults may continue to change their grammars in step with changes in their experience, so age may affect inherent variability in complex ways (such as the behaviour of middle-aged women in Buckie, who may have reacted in middle-age against the behaviour of young men – see section 3). This possibility means that it may in fact be wrong to assume that older speakers are still speaking as they did when younger. This is important in inherent variability, because age differences are often taken as evidence for historical change.

§2.3 Processing

Processing includes both speaking (or writing) and listening (or reading), which are usually assumed to need separate theories; but in WG a single theory can cover both directions because the directionality is determined by the purpose of the processing – to find a sound-pattern (speaking) or to find a meaning (listening). Most of the details of this theory are irrelevant here, but inherent variability involves choice and frequency, so we do need some elements of the theory of processing as applied to speaking. The main idea in the theory is that the network is active, in the sense that individual nodes and links have activity levels which (at least in the ultimate analysis) correspond to neuro-chemical properties of neurons. (There is no assumption that the nodes and links of the model correspond in any simple way to neurons; indeed, it is unlikely that there is any such simple correspondence.) Nodes have a 'resting' activation level which reflects the frequency (and, perhaps, the recency) of their use; and when they receive activation from neighbouring nodes which brings them up to their threshold level, they fire, distributing their surplus activation equally among their neighbours. This blind process is called 'spreading activation', and is the basis of many theories of processing, including the WG one. Rather surprisingly, even cognitive-linguistic theories typically pay very little attention to spreading activation (with the honourable exception of Stratificational Grammar – Lamb, 1999).

The main role of spreading activation in speech processing is in selecting a single target node from among the millions of nodes in the network. By a very simple process, the parameters of the search are defined by activating a schematic 'target' node and at least one 'clue' node; for example, in searching for the word that means 'octopus' the clue node is the concept 'octopus' and the target node is a specially created one which hardly has any properties other than 'word-form'. The aim is to enrich the target node by finding a rich 'source' node (or nodes) from which it can inherit properties. The source node is the one which receives the most activation spreading from both the clue and the target nodes. This rather crude process works well most of the time, and generally chooses the candidate that makes the best fit with the properties of the cue; but sometimes it fails (as in speech errors). Indeed, speech errors are one of our best pieces of evidence for blindly spreading activation, along with priming experiments (where one word makes another one more accessible if the two are related, cf. Reisberg, 1997:265-7).

This theory has a number of consequences for inherent variability:

- Inherent variability is the result of choices between alternative realisations for the same word or form, so it can only be understood if we understand how choices are made. If the winner in any competition is the most active variant, the question is how activity levels are determined, and the theory outlined here, based on spreading activation, has two main attractions which I expand below: it explains why choices are variable rather than categorical, and it explains why multiple influences combine in complex ways.
- One particularly important consequence of the theory is that (contrary to Adger, 2006) the stronger of two candidates does not always win. Suppose a node A is linked to a very active node X, whereas A's rival B is linked to a less active node Y. How does B ever get a chance to win? This is because of the mechanism of spreading activation whereby a node only spreads its activation ('fires') when it reaches a threshold level. Once X has fired (and activated A), it returns to its resting level of activation, giving the less active Y a chance to fire and activate the rival B. In this way, B could win (say) one competition in ten rather than being permanently unsuccessful.
- The theory allows stable long-term effects such as frequency to interact with temporary effects such as the immediate social situation: the former determine the

node's resting activation level (which is always below threshold level), while the latter modify this temporarily, possibly bringing it up to the threshold for firing.

- The fact that the network includes non-linguistic as well as linguistic concepts explains why both kinds of concept can influence a linguistic choice. In this kind of network, a linguistic node will receive activation from all its neighbours, whether they are related to it linguistically (e.g. through syntactic relations) or in terms of relations such as 'speaker'.

So far as I know, these ideas about spreading activation have never been applied to variable sociolinguistic data. Even the general idea of relating variable output to a competence based on usage has hardly been explored at all (but see Hudson, 1997 and Kemmer and Israel, 1994). The next section introduces a typical data-set from variationist sociolinguistics which the remaining sections discuss in terms of WG theory.

§3 SOME DATA FROM BUCKIE

Buckie is a small fishing town in the north of Scotland, 60 miles north of Aberdeen, and the home town of Jennifer Smith, who carried out the research reported here (Adger and Smith, 2005; Smith, 2000; Smith et al, 2007;). The project was an exploration of syntactic patterns in the speech of 39 working-class Buckie residents from three age groups (16 aged 22-31, 14 aged 50-60 and 9 aged over 80); these groups are divided equally into male and female. The data consist of forty hours of tape-recorded casual conversation, and all 300,000 words of the data have been transcribed. This database is more than adequate for the study of common syntactic patterns.

The data from this project are convenient for my purposes partly because they are so well presented but partly because they have already been given a theoretical interpretation in terms of the Minimalist Program (Adger, 2006), to which I return below. I shall focus on the variable which receives most attention in this theoretical discussion, the *was/were* variable. This variable involves variation between *was* and *were* only in one syntactic context: when the subject is one of the pronouns *you* (or the plural *you ones*) and *we*. (There is no variation with other pronouns, and the variation found with full NPs and with *there* involves a more general pattern which is irrelevant here.) Adger illustrates the variation with the following examples (where a gloss is sometimes needed):

(1) He says, 'I thocht **you were** a diver or somethin'.

He says, 'I thought you were a diver or something'.

(2) Aye, I thocht **you was** a scuba diver.

Yes, I thought you were a scuba diver.

(3) There was one nicht **we were** lyin at anchor.

(4) We played on 'at beach til **we was** tired, sailin boaties, bilin whelks.

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

(5) He was aie away when **you ones were**.

He was always away when you (plural) were.

(6) So **you were** all- **you were** all just bairns.

So you were all, you were all just children.

(7) **You ones was** a wee bitty older and you treated her just a right.

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

As can be seen, the choice between *was* and *were* is variable whether the subject pronoun is *we* or *you*, and whether *you* is singular or plural; but the data included only 10 tokens of plural *you* with *was/were* (compared with 161 of singular *you* and 368 of *we*) so I shall concentrate on singular *you* and (plural) *we*.

The raw data for *was/were* are shown in

Table 1 (from Smith, 2000:61). For both subject types, the age differences are very highly significant (p = 0.000 in both cases), and the effect of the subject type is significant (p < 0.05) for both the older age groups and almost significant (p = 0.062) for the young speakers. The figures are presented more accessibly in Figure 1, which shows for each pronoun the number of *was* tokens as a percentage of the total of *was* plus *were*.

[Table 1 here]

[Figure 1 here]

The most striking feature of Figure 1 is the clear effect of age when the subject is *you*, with a general decline in the use of *was*. However, *we* produces a very different age effect, with a reverse of the decline between middle-aged and young speakers. It turns out that this is mainly due to the female middle-aged speakers shunning *we was*, in contrast with young males who increasingly revel in it (Smith, 2000:64); this complex pattern is shown in Figure 2. Without real-time data it is not possible to know what changes have actually taken place to produce this effect, but one possibility is that middle-aged women have changed (considerably) within their own lifetime while young men have changed (slightly) in comparison with older men; and equally speculatively, it is possible that these changes were motivated by a desire to distance themselves from one another – the more often young men said *we was*, the more often middle-aged women, who disapproved of this usage, said *we were*.

[Figure 2 here]

We need not pursue the analysis of the Buckie *was*/*were* variable further here. The main point is that even this small amount of data is enough to show that a speaker's choice of variant on any particular occasion is affected by both linguistic and non-linguistic influences – in this case, the linguistic influence of the subject pronoun, and the non-linguistic influence of the speaker's sex and age. Our ultimate aim is to explain these effects by answering questions such as:

- Why do speakers treat *we was* differently from *we were*?
- Why has the use of *was* increased among men?
- Why do middle-aged women avoid *we was*?

We can already take some steps along this road (such as those taken in Smith 2000), but it will be difficult to make serious progress until we can accommodate the facts in a general theory of linguistic and social structure such as WG.

One of the contributions of inherent variability to general linguistic theory is to remind us that the relation between linguistic expressions and their semantic or social function is essentially arbitrary. The linguistic and social distribution of *was* and *were* in

Buckie cannot follow from any general principles because the same two variants alternate in other places, and where they do alternate, the facts may be quite different. Take, for example, the *was/were* pattern found in a school ('Midlan' School) in Lancashire (Moore, 2003). Whereas *was/were* variation with pronoun subjects is found in Buckie only after *you* and *we*, in Midlan School it happens only after *I* and *he, she* and *it*. Furthermore, the choice between *was* and *were* in Midlan School turned out to be heavily dependent on the pupils' self-selected social groupings (what Moore calls 'communities of practice'), whereas the only social influences reported in Buckie were age and sex.

Arbitrary or not, the data from Buckie present the following challenges for linguistic theory:

- structural adequacy: how to explain why *was* and *were* are synonymous and both available after *you* and *we*.
- contextual adequacy: how to explain the effect of the speaker's age and sex on the choice between *was* and *were*.
- behavioural adequacy: how to explain the observed frequencies in the data, e.g. that female speakers vary between 90% and 9% in their use of *we was.*

The remainder of the paper will consider how WG responds to these challenges, and where relevant, how other theories do.

§4 STRUCTURAL ADEQUACY: EXPLAINING SYNONYMY

The first question is how to analyse the part of the grammar of Buckie English shown in the verb paradigm of Table 2 (based on Adger, 2006).

[Table 2 here]

The challenge here is rather small, and trivial for any theory of grammar which allows one word to be realized in more than one way. What we need is an analysis which shows on the one hand what *was* and *were* have in common, and on the other what distinguishes them. What unites them is that they both belong to the lexeme BE and they both have past tense, so we need an abstract word which we can call BE:past. This classification is shown in Figure 3, where the little triangles indicate that the relation concerned is an example of 'isa', the relation which is the basis for all classification (e.g. BE:past isa BE, i.e. it is an example of BE and inherits all the properties of BE). This figure shows just three nodes from the network for Buckie English, but later figures will introduce other nodes (and other kinds of relations); and it is important to remember that, according to WG, the language contains nothing but what is shown in the network. The network is like a static map of the available structures, and requires the theory of processing outlined in section 2.3 as an 'instruction manual' which explains how to use it.

[Figure 3 here]

Recognising the abstract word BE:past may seem a trivial achievement, but it is important in dealing with variation because it gives us a common point which is notationally and conceptually different from the variants *was* and *were.* It allows us to ask the more interesting question of how the unifying BE:past is linked to *was* and *were*, which is strictly speaking not a matter of syntax at all, but of morphology: BE:past is realised either by the form {was} or by the form {were}. Given the three-level architecture of WG (section 0), words are realised by morphs, so where just one realisation is possible, it is as shown in Figure 4 for the past tense of the lexeme TAKE.

[Figure 4 here]

Inherent variability is different because it allows more than one realisation for the same word, so the immediate question is how to allow BE:past to be realised either by {was} or by {were}. This requires a mechanism for handling alternatives, and raises surprisingly deep theoretical issues. If knowledge of language is a mental network in which each concept is represented by a node and defined by its connections to other nodes, any choice is a choice between distinct concepts, each requiring a different node. How, then, can we represent an 'either … or' relation in a network? The WG answer is to represent each of the alternatives as a separate sub-case; so in the case of *was/were*, we can distinguish two sub-cases of BE:past, one of which is realised by {was} and the other by {were}, as in Figure 5. What we call these sub-cases is unimportant, so we might as well call them BE:past$_{was}$ and BE:past$_{were}$. In this analysis, each of these sub-cases automatically inherits all the properties of BE:past, but also combines it with a realisation. Of course, the whole point of the Buckie data is that *was* and *were* are different in syntax as well as in morphology, and I return to these differences below; but for the time being we shall ignore them.

[Figure 5 here]

The analysis so far is roughly equivalent to a range of different treatments in other theoretical frameworks, each of which has its weakness:

- a phonological rule which converts an underlying *was* into *were* (or the other way round), perhaps with variable weightings added as in Labov's variable rules (Labov, 1969; Sankoff and Labov, 1979). Weakness: the analysis stretches the notion 'phonology' too far.
- a parameter (or microparameter) which defines two distinct grammars, one with {was} and the other with {were}, which coexist in a single competence (Black and Motapanyane, 1996; Henry, 1995; Kroch, 1994,). Weakness: the analysis stretches the notion 'grammar' too far.
- distinct lexical entries for BE:past, each with a distinct realisation (Adger, 2006). Weakness: the analysis stretches the notion 'lexical entry' too far, as I explain below, and it also does not explain the observed data, as I explain in section 6.

The last analysis is the most similar to the one that I am proposing, in that it assumes that the different realisations are each attached to a distinct item stored in the grammar/lexicon. However, Adger's two-entry analysis also ignores a crucial problem: BE is a remarkably versatile verb with many different complement types, each of which presumably requires a different lexical entry. The following examples merely hint at a much wider range of possibilities.

(8) He was tall.

(9) He was David.

(10) He was walking.

(11) He was interviewed.

(12) He was to give the next lecture.

The choice between {was} and {were} is logically independent of the choice between these complement types, and in Buckie it is statistically irrelevant (Smith, 2000:59), in contrast with Midlan School where structures such as *He was stood by the door* favoured {was} (Moore, 2003:88). If the *was*/*were* contrast really is independent of the complement type, and if each complement type has a separate entry, it follows that there must be a separate entry for each combination of {was} or {were} with a complement type – one entry for BE with an adjective complement and realised as {was}, another for BE with an adjective complement and realised as {were}, another for BE with progresive complement and realised as {was}, and so on and on. The objection to this analysis is not so much the effect on the size of the lexicon but rather that it misses the simple generalisation that any past-tense BE may be realised in either way.

How does WG avoid the same problem? The solution in WG is multiple inheritance: a single word may inherit from two or more word categories. We have already seen examples of this in TAKE:past and BE:past, each of which inherits from both a lexeme and an inflectional category; and the same principle allows a word to inherit both from some complement-type of BE and also from some realisation-type of BE. This is shown in Figure 6, where $BE_{progressive}$ and $BE_{passive}$ are the cases of BE which have progressive and passive complements; in a real grammar, of course, there would be far more such categories.

[Figure 6 here]

Of course, the network structure in Figure 6 on its own tells us nothing about how categories combine. Commonsense might suggest that sub-categories are mutually exclusive – i.e. nothing can have an 'isa' relation to two sisters (or near-sisters) in a classification hierarchy such as this. However, WG does not restrict classification in this way because both inside and outside language, sisters often are compatible. For example, a gerund 'isa' both noun and verb (Hudson 2003), and a person may 'isa' both a pedestrian and a citizen. Instead of a formal restriction, we rely on the theory of processing which was introduced in section 2.3 to ensure that spreading activation will activate only the relevant super-categories sufficiently for selection.

We now turn from morphology to syntax in order to consider the syntactic constraints on the two variants, *was* and *were*. Inherent variability usually involves a combination of structural – in this case, syntactic – and social constraints, so the syntactic constraints need a clear place in the total analysis. According to the analysis so far, *was* and *were* are simply in free variation, but this is wrong for Buckie (and, so far as I know, every other dialect of English). Table 2 showed that in Buckie, *was* and *were* are divided as in Standard English between the pronouns *I, he/she/it* and *they*, and it is only after *we* and *you* that they are both possible. In order to show these restrictions we recognise even more specific sub-cases of BE, each distinguished by the identity of its subject. Figure 7 shows the analysis, which extends the one in Figure 6 downwards. This analysis links *I* exclusively to *was* and *they* to *were*, but either realisation is compatible with *we* (and with *you*, which is not shown).

[Figure 7 here]

The example shows how easily syntactic constraints can be imposed on morphological realisations when syntactic relations are expressed directly as dependencies rather than indirectly via phrase structure. The relevant syntactic constraints could not be defined more easily than by a dependency relation such as 'subject'; and, indeed, it follows from the general theory of networks that constraints may only be imposed on one word by other words which are directly related to it by a single dependency link. (It would be much harder even to formulate this principle in a phrase-structure analysis, let alone to explain it.)

§5 SPEAKER TYPES AND THE CHOICE OF VARIANTS

One thing that emerges clearly from work on inherent variability is that speakers associate particular expressions with particular types of context. The clearest expression of this idea is the theory of 'acts of identity' according to which every act of speaking expresses a complex identity that the speaker chooses to present at that time (LePage and Tabouret-Keller, 1985; Trudgill 1983; Hudson 1996). The way we speak is not simply a copy of the way those around us speak, and could not be so because those around us speak in different ways. Rather, the way we speak is a matter of choice in which we choose how we want to sound in terms of a complex of social contrasts such as age, sex, region, educational background and so on; in short, we 'locate ourselves in a multi-dimensional social space' (LePage and Tabouret-Keller, 1985).

This view of social variation locates it firmly in the individual speaker, so it fits well in a cognitive theory of language and language use. The alternative is a theory which locates variation in ‘the community’ (Labov 1972b). It is true that other people are our models for this variation, just as they are for every other aspect of language, but community-based theories are not a serious alternative to a theory of how individuals vary. For one thing, it is very hard to define ‘the community’ in any meaningful sense, and especially so in the complex societies of cities where so much of ‘urban dialectology’ has taken place. And for another, the only mechanism for explaining individual behaviour is individual cognition, so patterns in the collective behaviour of a community are only relevant to the extent that they are part of individual cognition.

According to this view, our cognition includes both a multi-dimensional model of social space and a set of mappings between elements of this model and individual elements of language; for example, Buckie people seem to associate *you were* with youth (Figure 1) and *we were* with middle-aged femaleness (Figure 2). This emerges from their behaviour, but at least some of them are sufficiently aware of it to put it into words. For example, one Buckie middle-aged female who never used *we was* but used *you was* a great deal showed considerable self-awareness: 'Well, maybe I would say *you was* but I would never say *we was*. I ken that's just wrong...' (Jennifer Smith, p.c.) Of course, like most of our knowledge about most things - and perhaps especially about language – this kind of knowledge tends to be below the level of consciousness; but it always emerges in behaviour, and sometimes in self-report.

Examples like this show that our social knowledge about language can apply to very specific elements of the language rather than globally to entire ‘dialects’ or ‘languages’, but they also show that the connections we know link these elements to social categories which may be harder to self-report. For this particular Buckie speaker, the categories are 'I' and 'wrong'. In terms of the theory of acts of identity, these refer to two different locations in the multi-dimensional social space. Presumably 'I' is not only the speaker, but also people similar to her, whereas 'wrong' evokes speakers of whom she disapproves – she knows that some people do in fact say *we was*, but dissociates herself from them. Given the clear difference between middle-class women and most men in Figure 2, it would be reasonable to assume that these 'wrong' speakers are men. This is clearly not the end of the social analysis – a strong link to the other sex is not enough in itself to justify the description 'wrong' – and a fuller analysis would certainly have to include education in order to explain it. The main theoretical point is that a model of linguistic competence must be able to link specific linguistic elements to specific social categories.

This kind of linkage is possible in WG because the language network is fully integrated into the same general cognition that includes the social network, so there is no boundary to be crossed between linguistic categories such as BE:$_{\text{was/were,we\_,was}}$ (the *was* of *we was*) and social categories such as 'Buckie man'. Moreover, the relations needed for making such links are precisely the situational categories such as 'speaker' and 'time' that can be induced from tokens (section 0). The typical middle-class female speaker in Buckie may therefore have a competence which includes the network in Figure 8. This simple example merely hints at more complex analyses which would include other parts of the multidimensional social space.

[Figure 8 here]

Supposing this analysis is roughly correct, how does it help us to understand the choice between *we was* and *we were* in speaking? According to the theory of acts of identity, the choices we make depend to at least some extent on how we want to 'identify' ourselves to other people (and to ourselves). In the Buckie example, whether a woman chooses *was* or *were* after *we* depends on how much she can identify with the male stereotype which (she thinks) typically uses *was*. This is generally a matter of degree rather than categorical, so a full answer must wait for the quantitative discussion which I start in section 6, but it is obvious that a female speaker is likely to prefer forms that are compatible with her own female identity. Perhaps somewhat less obviously, a male speaker could have exactly the same grammar, with the same social links, as a female speaker, but could still produce a completely different pattern of choices thanks to his different self-classification. Paradoxically, therefore, by allowing the grammar to include links to social categories we allow a clear separation of the language-internal patterns from social links.

To summarise this section, I have argued that a model of I-language should be able to include very specific linguistic categories such as 'the *was* of *we was*' and to link these linguistic categories to particular kinds of speakers defined in terms of whatever social categories are available in the speaker's social cognition ('I-society'). I have also shown how this 'contextual adequacy' can be achieved in WG.

§6 BEHAVIOURAL ADEQUACY: EXPLAINING THE NUMBERS

A behaviourally adequate analysis is one which explains the statistics of behaviour – for example, why the female speakers in Figure 2 chose to use *was* after *we* on 90%, 9% and

64% of occasions depending on their age. The first question, of course, is whether there really is anything to explain in the observed figures; could the differences be attributed to chance? This part, at least, is easy; statistical tests show that these differences are very highly significant (p = 0.000), so we can reject the idea that they are due to chance. But this leaves much more difficult questions such as the following:

- Why these figures rather than others? For example, why do older females use *was* on 90% of occasions rather than, say, 60%?
- Why does age have this particular quantitative effect when the subject is *we* but not when the subject is *you*?

The questions are easy but answers are much harder to find, or even to imagine.

One possibility that is worth exploring is that the figures in some way follow automatically from the structure of the grammar. This is the avenue that Adger explores (Adger, 2006) in relation to the same Buckie data that I have discussed. His theoretical framework is the Minimalist Program, but his idea depends very little on this framework except for the technical details. He suggests that the lexical entries which map features to morphology are defined in such a way that some verbs match two different entries both of which have the same realisation. In particular, his analysis allows BE to match two entries realised as *was* when its subject is either singular *you* or *we*, but only one entry realised as *were*; as he puts it, there are two routes to *was* but only one to *were*. This explains, he claims, why *you was* is twice as frequent as *you were* in the averaged figures for Buckie shown in

Table 1, where *was* accounts for 69% of the tokens after *you* and 67% of those after *we* – remarkably close to the 66.6% that his analysis predicts. If all speakers tended towards the same 2:1 ratio of *was:were*, the explanation would be complete; but unfortunately they do not. As we saw in Figure 1 and Figure 2, these averages conceal a great deal of variation, and especially among women (notably the difference between 90% for old women and 9% for middle-aged women). It seems unlikely that a purely structural explanation can be found for differences such as these.

The Buckie research suggests another kind of explanation: activation levels. The relevant data come from a different project reported in Smith et al (2007), which studied variation in the speech of small children (aged 2;6 to 4;0) and their caregivers. For one variable (the vowel in words such as *house*), the variants are a monophthong and a diphthong, with monophthongs somewhat outnumbering diphthongs on average (43% monophthongs for adults and 37% for children). However, this average once again conceals a wide range of individual scores ranging from about 2% to 90%. This makes it all the more impressive that the individual children's choice of variants correlated closely ($p <0.05$) with that of their caregivers, and suggests that children can in some way record the relative frequency of events in their experience and reproduce these frequencies in their own behaviour – an idea which is very familiar in the literature on usage-based learning (see for example Bybee and Hopper, 2001; Goldberg, 2006). In the WG model of cognitive structure, frequencies affect the 'resting' activation level of each node or link, so (returning to the *was/were* example) we can imagine the node for the *was* in *we was* having a resting level in the child's mind which corresponds roughly to the number of times the child has heard the adult using it, and similarly for *we were*. Thanks to the way in which spreading activation operates, it raises these levels until they reach the required threshold for firing, so the higher the resting activation, the more easily (and the more often) the node will fire. Using activation levels in this way is similar to assuming that we know probabilities, but avoids the problem of deciding exactly what it means to 'know' a probability, and how this knowledge might be represented mentally.

But how can activation levels show the effects of context? Once again we can illustrate the problem from Buckie, using data for the *house* vowel in children's speech which varies dramatically with context. Both adults and children use the monophthong far more in 'casual' contexts such as play than in more 'formal' contexts such as teaching or discipline, and once again the figures for adults and children are very similar; for example, the averages for both are between 70% and 80% monophthongs in play compared with 10-20% in teaching (Smith et al, 2007). Clearly, the child does not have a single activation based on a global experience of adult behaviour but distinguishes one context from another.

How context exerts this effect is a matter for future research, but at least two possibilities deserve exploration: very specific recording of frequency and very general recording. It may turn out that both will be needed in some as yet unforeseeable combination.

- Very specific recording would require even more specific nodes than we have considered so far; for example, a Buckie speaker might have distinct nodes for *we was* according to whether it was spoken in a formal or an informal setting: BE:past$_{\text{we_,_adjective,was,formal}}$ and BE:past$_{\text{we_,_adjective,was,informal}}$, each with a different resting activation level. These nodes are located in the 'isa' hierarchy beneath the node for BE, so they automatically inherit all the properties of BE, but they also show a number of effects which might be called contextual: the effect of past tense, of the subject *we*, of the complement type '_adjective' (section 4) and of the formality of the situation. These nodes may seem implausibly specific, but there is massive evidence for specific entries elsewhere. Take, for example, the Buckie research on the *house* vowel. This provides spectacular evidence for specific lexical effects: the monophthong percentage for children tracks that of the adults closely across a range of lexical items with scores ranging from close to zero for *how* to nearly 100% for *about*; for example, both children and adults produced 80% monophthongs for *house*. Given this degree of detail it is easy to believe that specific contextual effects could also be recorded in the same way.

- Very general recording attaches the frequency effect to some general concept such as 'formal', from which it may be inherited by any particular example of formal language. This model is plausible, and phonological variation seems to confirm that generalisations are available for rare or new lexical items. But even if research confirms the need for contextual generalisation, a host of other research problems remain. Can learners really induce generalisations about activation levels? And can users really inherit such generalisations? At present I have no idea what answers may emerge for such questions.

In short, we can be sure that very specific nodes record very specific frequency effects which combine the influence of both linguistic and non-linguistic context. What is less clear is whether, and how, more general contextual effects are exerted.

What can WG contribute to research in this area? The main contribution is an overarching theoretical framework which provides all the main elements – a network model of I-language embedded in a network which includes I-society, very specific sub-lexemes, specific links to both linguistic and non-linguistic contexts, a usage-based theory of learning and a theory of activation in use. In terms of this theory (in contrast, say, with the Minimalist Program), quantitative contextual effects are to be expected. It also provides a formal network notation for modelling all the relations concerned. But what WG cannot yet offer is a formal theory of activation or a formal notation for it. Such a dynamic object cannot really be modelled experimentally in the same way as a static language network, and the computer modelling which it requires does not yet exist (although a start has been made).[1]

§7 CONCLUSIONS

WG is a very general theory which applies equally to all languages, so English dialects as such are of no special interest or relevance. The real challenge comes from the inherent variability which has loomed large in recent research on English dialects (and also, of course, on many other languages). The findings of these studies challenge any linguistic theory by testing its adequacy at three increasingly demanding levels: structural adequacy, contextual adequacy and behavioural adequacy.

What I have argued in this paper is that cognitive-linguistic theories in general are well suited for this work because they allow language to relate directly to non-linguistic categories. However, I have also claimed that WG has a number of distinctive features which make it particularly promising:

- It assumes a dependency analysis of syntax.
- It recognises a level of form between syntax and phonology.
- It includes a theory of processing based on spreading activation.
- The theory of processing explains why choices vary even when one variant is dominant.

It would also be fair to add, I believe, that the logic and notation of WG are especially well developed; but for these attractions I must refer the reader to Hudson (2007).

[1] Two ESRC-funded projects produced a partially functional computer model called WGNet++ (http://www.phon.ucl.ac.uk/home/dick/ell2-wg.htm) and Mark P Line has started to develop a very general-purpose network manipulator called Babbage which could accommodate WG networks (http://www.babbagenet.org/).

Author's details

Richard A. Hudson

Department of Linguistics and Phonetics

University College London

Gower Street

London

United Kingdom

WC1E 6BT

dick@ling.ucl.ac.uk

REFERENCES

Adger, D. (2006). Combinatorial variability. Journal of Linguistics 42: 503-30.

Adger, D. & J. Smith. (2005). Variation and the Minimalist Program. In Cornips, L. & K.P. Corrigan (eds.) Syntax and variation: reconciling the biological and the social. Amsterdam: John Benjamins. 149-78.

Barlow, M. & S. Kemmer (eds.) (2000). Usage based models of language. Stanford: CSLI.

Black, J. & V. Motapanyane. (eds.) (1996). Microparametric syntax and dialect variation. Amsterdam: John Benjamins.

Bybee, J. & P. Hopper. (eds.) (2001). Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins.

Chomsky, N. (1986). Knowledge of language: its nature, origin and use. New York: Praeger.

Cornips, L., and K.P. Corrigan. (2005). Convergence and divergence in grammar. In Auer, P., F. Hinskens, & P. Kerswill (eds.) Dialect change: convergence and divergence in European languages. Cambridge: Cambridge University Press. 96-134

Goldberg, A. (1995). Constructions: a Construction Grammar approach to argument structure. Chicago: University of Chicago Press.

Goldberg, A. (2006). Constructions at work: the nature of generalization in language. Oxford: Oxford University Press.

Gonzálvez-García, F. & C. Butler. (2007). Mapping functional-cognitive space. Annual Review of Cognitive Linguistics 4: 39-96.

Halliday, M.A.K. (1994). An introduction to functional grammar. 2nd edition. London: Arnold.

Harley, T. (1995). The psychology of language. Hove: Psychology Press.

Henry, A. (1995). Belfast English and Standard English: dialect variation and parameter setting. Oxford: Oxford University Press.

Hudson, R.A. (1971). English complex sentences. an introduction to systemic grammar. Amsterdam: North Holland.

Hudson, R.A. (1976). Arguments for a non-transformational grammar. Chicago: Chicago University Press.

Hudson, R.A. (1984). Word Grammar. Oxford: Blackwell.

Hudson, R.A. (1994). About 37% of word-tokens are nouns. Language 70: 331-9.

Hudson, R.A. (1996). Sociolinguistics. 2nd edition. Cambridge: Cambridge University Press.

Hudson, R.A. (1997). Inherent variability and linguistic theory. Cognitive Linguistics 8: 73-108.

Hudson, R.A. (2003). Gerunds without phrase structure. Natural Language & Linguistic Theory 21: 579-615.

Hudson, R.A. (2004). Grammar for writing. http://www.ioe.ac.uk/schools/ecpe/ReconceptualisingWriting5-16/seminar2.html.

Hudson, R.A. (2007). Language networks: the new Word Grammar. Oxford: Oxford University Press.

Kemmer, S. & M. Israel. (1994). Variation and the usage-based model. In Beals, K., J. Denton, R. Knippen, L. Melnar, H. Suzuki & E. Zeinfeld (eds.) Papers from the 30th regional meeting of the Chicago Linguistics Society: the parasession on variation in linguistic theory. Chicago: Chicago Linguistics Society. 165-79.

Kroch, A. (1994). Morphosyntactic variation. In Beals, K., J. Denton, R. Knippen, L. Melnar, H. Suzuki & E. Zeinfeld (eds.) Papers from the 30th regional meeting of the Chicago Linguistics Society: the parasession on variation in linguistic theory. Chicago: Chicago Linguistics Society. 190-201.

Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. Language 45: 715-62.

Labov, W. (1972a). Negative contraction and negative concord in English grammar. Language 48: 773-818.

Labov, W. (1972b). Sociolinguistic patterns. Oxford: Blackwell.

Lamb, S. (1966). Outline of Stratificational Grammar. Washington, DC: Georgetown University Press.

Lamb, S. (1999). Pathways of the brain: the neurocognitive basis of language. Amsterdam: John Benjamins.

Langacker, R. (2000). A dynamic usage-based model. In Barlow, M. & S. Kemmer (eds.) (2000). Usage based models of language. Stanford: CSLI. 1-63.

LePage, R. & A. Tabouret-Keller. (1985). Acts of identity: creole-based approaches to language and ethnicity. Cambridge: Cambridge University Press.

Moore, E. (2003). Learning style and identity: a sociolinguistic analysis of a Bolton high school. Unpublished PhD thesis, University of Manchester.

Pollard, C. & I. Sag. (1994). Head-Driven Phrase Structure Grammar. Chicago: Chicago University Press.

Reisberg, D. (1997). Cognition: exploring the science of the mind. New York: Norton.

Sankoff, D. & W. Labov. (1979). On the uses of variable rules. Language and Society 8: 189-222.

Smith, J. (2000). Synchrony and diachrony in the evolution of English: evidence from Scotland. Unpublished PhD thesis, University of York.

Smith, J., M. Durham & L. Fortune. (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.

Trudgill, P. (1983). Acts of conflicting identity. In Trudgill, P. (ed.) On dialect. Oxford: Blackwell. 141-60.

Table 1: *Was* or *were* after *you* and *we* for three age groups in Buckie

Speaker age	You (singular)		We	
	You was	*You were*	*We was*	*We were*
Old	45	5	113	36
Middle	23	12	32	41
Young	43	33	101	45
All	111 (= 69%)	50	246 (= 67%)	122

Figure 1: *Was* as a percentage of *was/were* after *you* and *we* for three age groups in Buckie

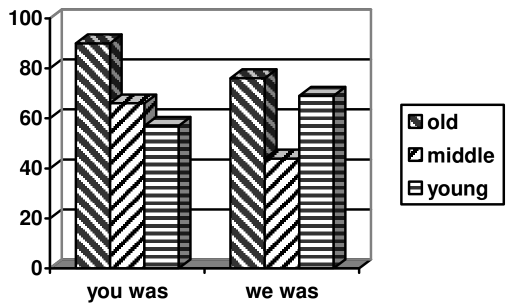

Figure 2: ***We was*** **as a percentage of** ***we was/were*** **for three age groups and two sexes in Buckie**

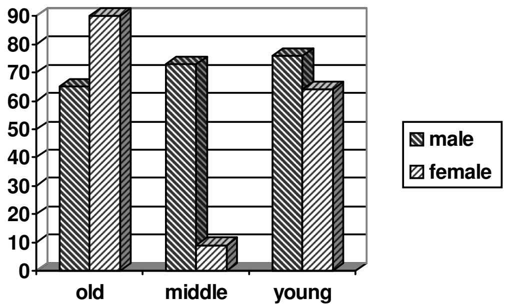

Table 2: *Was* or *were* with pronoun subjects in Buckie

I was	We was/were
You was/were	You was/were
He/she/it was	They were

Figure 3: BE:past isa both BE and Past

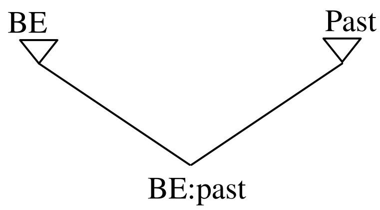

Figure 4: TAKE:past is realised by the form {took}

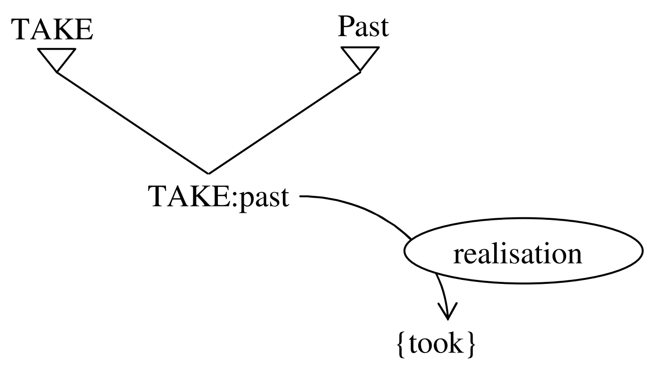

Figure 5: BE:past may be realised by either {was} or {were}

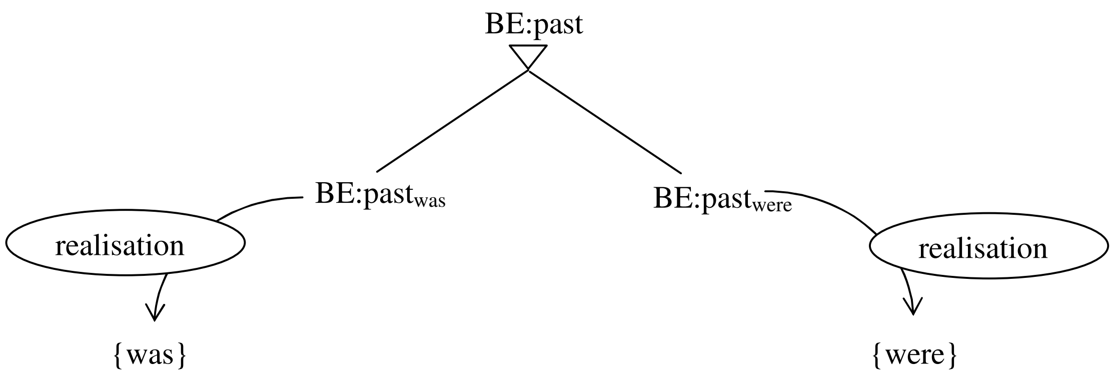

Figure 6: The *was/were* contrast is independent of other lexical subdivisions of BE.

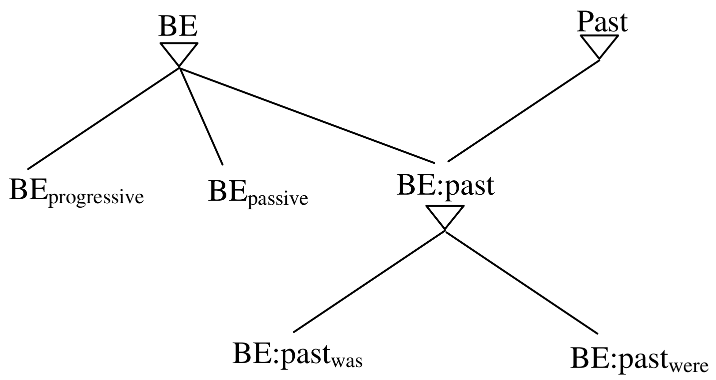

Figure 7: The relation between *was/were* and pronoun type in Buckie.

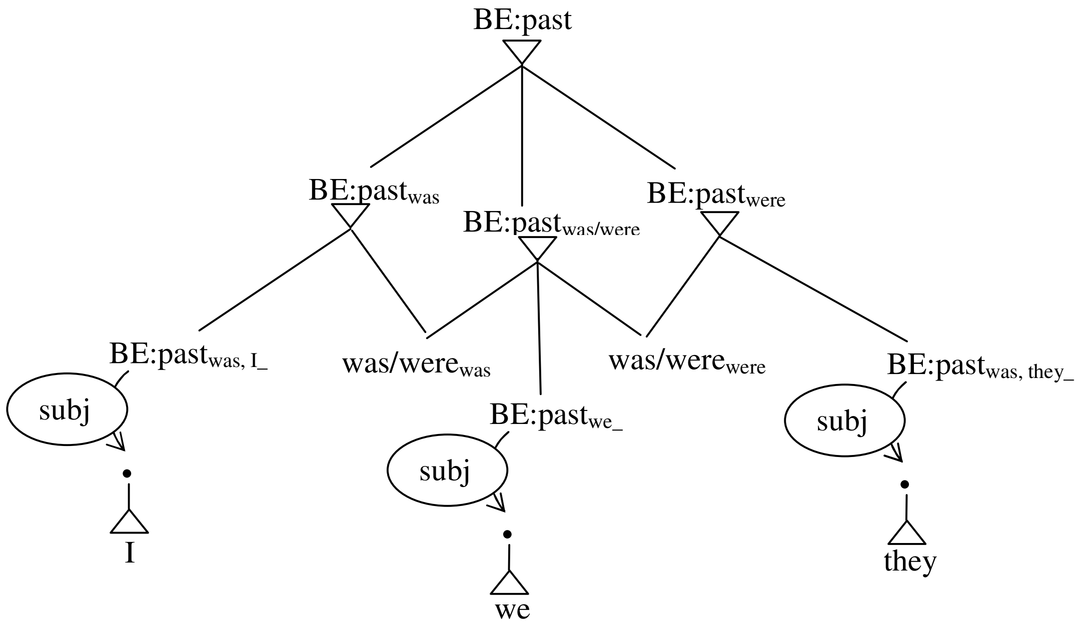

Figure 8: In Buckie, *We was* is typically used by men.