Inherent variability and linguistic theory

Richard Hudson, University College London

Abstract

Recent work by Guy and Kroch has used statistical data on variation in performance as evidence for specific theories of language structure, namely Lexical Phonology and Principles-and-Parameters syntax. I accept the validity of their data and of their statistical interpretations - Guy's `exponential model' and Kroch's `constant rate effect'; but I question their interpretations of these data in terms of language structure. Instead I argue that their data support a radically different model of language structure based on networks of prototypes, such as either Cognitive Grammar or Word Grammar.

1. Inherent variability as evidence for competence theories[1]

One of the most important discoveries in modern linguistics is surely the existence of `inherent variability' (Labov 1969), the coexistence of alternative `ways of saying the same thing' within the speech of a single speaker who alternates between them in a statistically regular way. The study of inherent variability has turned into a major area of linguistic research and greatly increased our understanding of variation in both place and time. But most of this work has fallen clearly within the sphere of sociolinguistics, with its special focus on the relationships between linguistic and social structures; very little could be described as the study of language structure as such, and even less has had any influence on (synchronic) theories of language structure. Indeed, it is hard to think of a single example (until very recently) where statistical data on inherent variability has been used as evidence in discussions of language structure. It is true that there have been a few sophisticated marriages between quantitative data and existing theories of language structure, the obvious example being Labov's idea of associating transformations or phonological rules with probabilities to give `variable rules' (Labov 1972:216ff, Cedergren and Sankoff 1974); but none of this work really challenged, or even influenced, the (then) current views on language structure, and in any case `the variable rule as a part of linguistic theory has quietly been abandoned' even in sociolinguistic studies of variation (Fasold 1990:256).

After Labov's early work there was a long period of separation between the work on inherent variability and work on language structure. Although the units that varied were parts of language - words, sounds, morphemes, constructions - inherent variability was left to the sociolinguists on the grounds that this variation had nothing to do with anything that `theoretical linguists' were interested in. A typical view is expressed by Smith (1989:180):

> To be of interest to a linguistic theorist it is not sufficient that the talk be of words and such like, rather the talk has to have implications of some kind for the theory concerned, by supporting or contradicting one of the claims derivable from it. ... Any social parameter whatsoever may be the locus for some linguistic difference. Unfortunately nothing of interest to linguistic theory follows from this, so quantifying the difference is irrelevant to linguistics even though it may be of interest to the sociologist if it gives him or her a recognition criterion for some socially relevant variable.

It is easy to criticise this comment for assuming in advance that `linguistic theory' cannot contain any claims which make contact with variability; but there is no denying that at the time when it was written it contained an important grain of truth. Even if variability had potential for illuminating theories of language structure, this potential had not been realised. It is true that some of us had asked what the existence of variability implied for theories of language structure (Romaine 1982, Hudson 1985, 1986, in press), but none of us managed to get beyond the stage of producing general statements of principles or programmes for future research.

The stimulus for the present article is an important collection of papers (Beals et al 1994) presented as a parasession to the Chicago Linguistic Society's 30th regional meeting, with the title `Variation in Linguistic Theory'. Some of these papers are serious attempts to use variable data as evidence for particular theoretical positions on the structure of language. They show beyond any reasonable doubt that the data are not only robust, but also relevant; but (unsurprisingly) different authors draw different theoretical conclusions, with two main groups of contenders. Roughly speaking, we can distinguish the `classical theories' from the `prototype-based theories'. The former are represented by Lexical Phonology (Guy) and Principles-and-Parameters Theory (Kroch), and the latter by Cognitive Grammar (Kemmer and Israel) and Pierrehumbert's view on phonology. The former articles (by Guy and Kroch) use statistical data very skillfully to support their theoretical claims, but the prototype-based theories are presented without statistical support.

This article is an attempt to contribute to the debate started in that collection. My theoretical sympathies are with the

[1] Many thanks for help and encouragement from David Denison and for comments and criticisms from those who attended seminars at UCL and Oxford in which I presented part of the material.

prototype-based theories, and my strategy will be to recycle statistical data quoted in support of classical theories as evidence for a prototype-based alternative. I shall consider the two specific cases which Guy and Kroch discuss, one in English phonology (so-called `t/d deletion') and one in English historical syntax (`do-support'). The argument will involve some fundamental questions about the nature of language-structure, not least of which is the question whether phonology and syntax are really different (as claimed most explicitly by Halle and Bromberger, 1989). I shall try to show that both cases allow the same kind of reanalysis (in terms of prototypes), so to that extent the argument will count as support for the view that all language structure is formally homogeneous, as well as being formally similar to structures found in general cognition.

2. Phonology: t/d loss as evidence for lexical phonology

The first case is the alternation between the presence or absence of a final alveolar stop following a consonant at the end of a word, so-called t/d deletion. For example, mist and raised can both be pronounced with or without the final stop, and according to the available data on free speech, both stops are more likely to be pronounced than to be omitted, though they are both omitted on at least some occasions. The data are summarised by Gregory Guy (1994), who has also been responsible for collecting and analysing most of it over the last decade or so (Guy 1980, 1991a, 1991b, 1993).

His very careful statistical analyses have shown beyond reasonable doubt that this alternation is not a simple phonetic matter (though of course it is phonetically motivated by the universal tendency to reduce consonant clusters). The most interesting conclusion (for present purposes) is that the survival chances of the stop (which we shall now call simply `t/d') depend on the morphological structure of the word containing it. If t/d is the past-tense (or participle) suffix, it is much less likely to be omitted than if it is simply the last segment in the word's stem. For example, mist and hold are much more likely than missed and holed to lose their t/d; and for past-tense verbs like left and felt where t/d is a suffix but not the sole inflectional marker the figure is intermediate. These differences between the three types of word are easily understood in functional terms: when t/d is a suffix it carries far more information than when it is just the last segment of the stem. But a functional explanation cannot be the whole story because suffixes are often omitted regardless of the loss of information (Labov 1994: Chapter 19). In any case, the functional explanation does not help at all in understanding the most interesting fact of all, to which we now turn.

Guy has discovered a regular mathematical relationship among the t/d omission figures for the three word-classes. He calls the morphological word-classes `monomorphemes', `irregular past' and `regular past', so we can let M, I and R stand for the probability of t/d being pronounced in each of these types of word. The following formulae define the relationship among these figures:

The Exponential Model

$$I = R^2$$
$$M = R^3$$

That is, the chance of t/d surviving in an irregular past (e.g. left) is the square of its chance of surviving in a regular past (e.g. missed); and its survival probability in a monomorpheme (e.g. mist) is the cube of this figure.

Guy calls this pair of formulae the `exponential model'. If it were based on just one set of data it might be explained as the result of chance; but it is in fact based on three different collections of spoken data, involving 7 + 45 + 16 = 68 speakers in total, and for each collection the model predicts almost exactly the figures which were observed (with differences well below statistical significance). The relevant figures are shown in Table 1.

[Table 1 about here]

At present there is no reason to doubt Guy's conclusion that the exponential pattern is `a real phenomenon', so I propose to accept it and move on to the question of explanation. Guy points out that the exponential pattern allows a very simple and intuitively satisfying explanation if we assume that t/d may be exposed to the possibility of being omitted on more

than one occasion, but that on each such occasion t/d has just the same chance of survival. If we call the probability of retention `r', then after two exposures to omission a given t/d's chance of survival would be r^2 (the chance of surviving the first round times the chance of surviving the second round). For example, if 4 out of 5 t/d's survive the first attempt to omit them (r = 0.8), then each of the surviving 4 has another 4/5 chance of surviving the second attempt, giving a total probability of 0.8 x 0.8 = 0.64. A third exposure will reduce this number still further, but again by the same amount (0.8), giving a final survival chance of $0.8^3 = 0.512$.

According to this interpretation of the figures, then, we can explain the exponential pattern in the observed data if we can find a linguistic model that allows the same rule or process to apply once to regular past verbs like missed, twice to irregular past verbs like left and three times to monomorphemic words like mist. This is a very exciting challenge for theoretical linguistics which has never before been confronted in this way with quantitative data.

What model of word-structure makes precisely these predictions? It is easy to think of models which fail to make them - for example, a model in which t/d-loss is handled by a rule which is blind to morphological structure and applies just at one level. It is very much harder to find a plausible candidate.

Guy's suggested linguistic model is a version of Lexical Phonology, which seems to have just the required characteristics. It distinguishes three cycles of processing for each word - before lexical insertion a word is subject to two cycles within the lexicon, and after insertion another `post-lexical' cycle applies. The difference between lexical and post-lexical applications means that the effect of the following word is only relevant to the latter. These three cycles are sensitive to morphological structure so they promise to distinguish our three word-classes. Furthermore, any rule can apply on any cycle to which it is relevant (subject to general constraints), so the omission of t/d may be due to a single rule of t/d-deletion which applies (optionally) on any level, with a constant probability of application. If mist, left and missed can be given appropriate structures, then t/d-deletion will be able to apply on all three cycles to mist, just on the last two to left and just on the last cycle to missed. According to Guy it is possible to justify structures of precisely this type for the three kinds of word, and the explanation is complete.

The overall structure of the argument is impossible to fault, but I have reservations about some of the supporting assumptions.

■ The analysis rests on a variety of theoretical assumptions which are controversial even within Lexical Phonology, to judge by Booij and Rubach 1992, McMahon 1994 and Roca 1994.

- ■ `In the classic version of Lexical Phonology, phonological rules are .. assigned to a specific level.' (Booij and Rubach 1992:329); but Guy assumes that t/d deletion is free to apply at any level.

- ■ Furthermore, `most linguists would admit that we must distinguish word-level (postcyclic) rules from cyclic rules which are subject to the Strict Cycle Condition' (ibid); but t/d deletion applies cyclically without being subject to the Strict Cycle Condition. The SCC is designed to prevent cyclic rules from changing structure (e.g. by deletion) within the underlying representation (Roca 1994:53), but this is precisely what t/d deletion does when applied to mist. Guy recognises this problem (142), but argues that variable rules should be exempted from the SCC (a considerable weakening of the theory unless variable rules are formally distinct from categorial rules; and every rule must be variable during the period when it is either entering the language or leaving it).

- ■ `Only postlexical rules may apply between as well as within words.' (McMahon 1994:2159) But t/d deletion clearly applies between words (because it is sensitive to the initial segment of the next word). How can this be reconciled with cyclic application?

Guy's assumptions may in fact be defensible, and it may even be that the statistical data will be able to contribute to theoretical debates about Lexical Phonology, but it should at least be recognised that his analysis is not a completely straightforward

application of a standard version of Lexical Phonology.

■ Another objection is that Guy's own earlier work (Guy and Boyd 1990) shows that the status of irregular past tense forms like left changes as speakers mature. For many years they are stored as unanalysable units, so during those years they have just the same status as monomorphemes like mist. Only in later life is the final /t/ recognised as a separate morpheme. But Guy's current model assumes that the form retrieved from the lexicon must be something like [li:v], which then combines with the Past suffix to form first [[li:v]ed], and then [left] (138). Even if we assume that all Guy's speakers have recognised /t/ as a separate affix, this implies that they have forgotten (i.e. `un-stored') the complete form /left/ so that they can no longer retrieve it, ready-made, from their lexicon. So far as I know there is no evidence for this kind of `un-learning', and it seems unlikely that it exists. It is much more likely that the past tense of the lexeme LEAVE is retrieved ready-made from the lexicon, and similarly for all the other irregular past-tense forms.

■ The probability of pronouncing d/t is heavily influenced by the first sound of the following word: retention is favoured by a vowel and omission by a consonant. Guy documents this variation in detail, and points out (especially in Guy 1991a) that according to Lexical Phonology the following word is irrelevant at both the lexical levels (since these take place within the lexicon, and not in the sentence). This means that two of the three deletion-opportunities in words like mist are blind to the following word, in contrast with missed where there is only the one, post-lexical, opportunity. We can therefore predict that the effect of the following word should be sharper for missed than for mist, where it is masked by the deletions at two other levels which were unaffected. For example, suppose the retention-rate was 50% before a consonant and 90% before a vowel, with 70% as the mean. This would be the predicted distribution for regular pasts like missed, so we should expect a difference of 40% between the two environments. For monomorphemes like mist, on the other hand, the effect of the two environments applies only to the figure of 49% (70% of 70%), which is the output of the first two cycles (unaffected by the following context). In this case the expected difference is just 20% (50% of 49% = 24.5%, compared with 90% of 49% = 44.1%).

The figures reported by Guy (1991b:235f) conflict with this prediction. For one corpus (Wolfram 1969) monomorphemic and regular past words show the same size of effect, and for another (Guy 1991b) the following word has a great deal more effect on t/d loss in monomorphemes than in regular pasts - the reverse of the predicted trend.

■ As Guy recognises (142), if t/d deletion is a simple deletion rule, it should apply to stem morphemes within words, even allowing impossible forms like *tessing (for testing). His suggested solution is to decompose the rule into two separate rules, a variable rule which optionally detaches the t/d from its place in syllable structure, and a late rule which erases this stray segment unless it has been `rescued' by a vowel-initial suffix. In other cases, a segment may be rescued by a vowel in the next word (e.g. /l/ may be clear in sequences like feel active), so why not here too, giving impossible sequences like at leas.t eight?

■ Lexical Phonology crucially assumes ordered rewrite rules, which are very controversial. As Pierrehumbert notes (1994:240), many phonological theories use declarative constraints rather than rewrite rules, so Guy's analysis is no stronger than the evidence for rewrite rules and against constraints. It is true that the relationship can be reversed, so that the exponential data are seen as evidence for ordered rewrite rules; but what if the data can be explained in terms of some alternative theory (as I hope to show below)?

3. Prototype-networks in linguistic competence

Having accepted Guy's statistical interpretation of the data on t/d-loss (the exponential model), I have questioned his explanation in terms of Lexical Phonology. What alternative should we consider? I shall now follow Kemmer and Israel (1994) in arguing for an analysis in terms of a network of prototypes. (Pierrehumbert also concludes in favour of prototypes (1994:251), but unlike Kemmer and Israel she does not discuss t/d-loss explicitly; prototypes are a keystone of cognitive linguistics, whose tenets are well summarised in Taylor 1989.) A prototype is a concept which is defined by its `typical'

features, any of which may be overridden in exceptional cases. For example, a human can typically speak, but a baby is still a human, albeit an untypical one; moreover, the exceptional nature of babies is a regular feature of the world, so some exceptional cases are rule-governed. The relevance to t/d loss is that the pronunciations in which t/d is present are typical, and those in which it is absent are rule-governed exceptions, like infant humans. I shall show that the structure of word-prototypes provides three opportunities for t/d loss in mist, two in left and one in missed, as required for an explanation of Guy's exponential model.

In the following discussion I shall outline the relevant parts of Word Grammar (**WG**; Hudson 1984, 1990), but apart from terminology and diagramming conventions the same ideas can be found in Cognitive Grammar (**CG**; Langacker 1987, 1990, 1995), the theoretical framework within which Kemmer and Israel work. Wherever possible I shall translate my terms into CG, but I hope it will be easy for CG experts to provide their own translations for my WG diagrams. The foundation of the whole system is **default inheritance** in an inheritance, or `isa', hierarchy. The basic idea is commonplace and hardly controversial, although details are debatable (e.g. Fraser and Hudson 1992). A robin is a bird (CG: `bird' is schematic for `robin') and therefore inherits by default all the properties of the typical bird (CG: `robin' is an instance of `bird'); i.e. as soon as we know that a robin is a bird, we assume it has all the typical bird-properties, as well as more specific ones (such as the robin's distinctive red breast). The properties of any concept apply just to the prototype, the abstract `typical' member, so any typical property may be overridden by an exceptional one. For example, a penguin is a bird although it has some exceptional properties (CG: `bird' is schematic for `penguin', but `penguin' is an extension of `bird').

Fig 1 shows how these relationships can be displayed graphically. The `isa' relation is shown by a triangle whose base is on the more general category, and whose apex is connected by lines to any concept which `isa' that category. (The triangle is meant to indicate that the general category includes the more specific ones.) The properties can be shown by labelled arrows, which (in this example at least) are not to be taken too seriously. It is easy to see how a serious analysis would produce a **network** of related concepts.

[Figure 1 here]

Default inheritance means that `robin' automatically inherits all the facts about `bird', but `penguin' inherits only those which are compatible with what is indicated specifically for `penguin'.

One of the attractions of default inheritance is precisely that we know, with some confidence, that it applies to general knowledge (e.g. about birds), so the same should be true of each special area of knowledge, including language. (An extreme version of modularity which predicted zero formal similarity between language and other knowledge would be absurd.) To see the relevance to language take the word mist. This has a number of characteristics including a meaning (which we can call simply `mist') and a form (/mist/), and it is a noun.

[Figure 2 here]

Fig. 2 shows what is presumably the typical pairing of form and meaning, but deviations are possible in both directions. The sense can be extended metaphorically, and the form can be changed for all sorts of reasons, one of which is the application of a variable rule such as our t/d-loss. Suppose we have a learner who has heard the word mist used on two occasions; arbitrarily we can call the first token of mist `word 305' and the second `word 507'. On each occasion the word was used as a noun, meant `mist' and had roughly the same pronunciation, but their exact pronunciations were different: /mist/ for word 305 and /mis/, without the /t/, for word 507. This would be a problem if categories were defined by necessary and sufficient conditions, with a word's normal pronunciation as one of its necessary conditions, but a prototype-based theory accommodates it easily. Word 507 can be analysed as an exceptional member of the same prototype as word 305, with /mis/ as an unusual realisation of the normal pronunciation /mist/; and the mismatch can even be stored in memory for future use - an essential prerequisite for eventually learning the t/d-loss rule. The resulting structure is shown in Fig. 3.

[Figure 3 here]

Kemmer and Israel call their model `**usage-based**' because the permanent knowledge structures are induced from usage (i.e. experience), and are shaped by the statistical patterns in the experience. Suppose our learner hears a drunk pronounce mist as [mi ʃ]. This is close enough to the target to be recognised, but let's suppose the learner notes its odd

pronunciation, and remembers it (i.e. stores it along with words 305 and 507). What effect will this isolated experience have on the learner's competence? Virtually none, except as evidence for how drunks speak. But suppose over the years the learner hears /mis/ 10 times, and /mist/ 20 times. Each such occasion increases the **entrenchment** (the CG term) of the pattern concerned. The more often we experience a pattern, the more firmly embedded into our competence it becomes; so the links between mist and its two pronunciations, /mist/ and /mis/, will achieve different degrees of entrenchment. We don't know, of course, how this works either in terms of neurology or in terms of mathematical models, and we can be sure that the mechanics involve a great many more influences than brute frequency; but the general idea is clear and plausible. Even more importantly, it promises to lead us (in the long run) to an explanation for inherent variability.

We can now make our example slightly more realistic. The alternation between /mist/ and /mis/ is part of a much larger pattern which applies to all those words which are subject to t/d-loss. Having observed the same alternation in a number of words, our learner can induce a generalisation: any word that ends in /t/ or /d/ following another consonant has two alternative forms. There is no need to formulate this relationship as a deletion rule, because any number of the shorter forms may already be stored; indeed, it would be impossible to induce the generalisation until a large number of them really are stored. The pairs of forms can be related in a number of ways, but the easiest way is probably to stipulate that a form without /t/ is in the `isa' relationship to one with /t/. Fig. 4 gives the relevant part of our learner's language system after the generalisation has been made, with just two words shown: mist and want. Notice how default inheritance allows the generalisation to apply freely to any form, while also allowing individual cases to be recorded in full.

[Figure 4 here]

The analysis shown makes no pretence of phonological sophistication - in fact it doesn't even generalise from /t/ to /d/, let alone paying attention to the known influence of preceding and following contexts. We shall improve it a little below, but for the time being this crude analysis will do instead of a proper analysis. The main point is that there is no `deletion rule', just a pair of related structures of which one contains t/d and the other lacks it. If the first structure (/...Ct/) matches a word's form, then the word should have an alternative form; otherwise, not. It does match /wont/, so we know the latter's alternative form is /won/; but it doesn't match, for example, /kæt/ or /tri:/, so these have no alternative form.

As already noted, each part of this network structure has some degree of `entrenchment' which reflects the experiences of the person concerned. In the case of t/d loss, each example of the normal pattern supports the form with t/d, and exceptional examples reinforce the exceptional pattern without t/d. For present purposes we can represent the degree of entrenchment of a concept simply as a probability of that concept being preferred to any relevant alternatives, without any commitment as to how this probability is reflected neurologically. If, for example, I have heard 4 examples of /mist/ to every one of /mis/ (as a pronunciation of mist), this imbalance will be represented in my mind by the probabilities 0.8 and 0.2 attached to these two pronunciations; and if the same is true more generally for words that can end in t/d, the more general concepts will carry the same probabilities, as shown in Fig. 5. Of course the figures for the individual cases (e.g. for mist and for want in our example) need not be the same; on the contrary, cases of lexical diffusion would seem to suggest the contrary (e.g. Hudson 1980:168ff). Presumably the entrenchment value for the general rule in such cases could be different from all the individual values.

[Figure 5 here]

We now have a declarative, non-procedural analysis of t/d-loss which requires just two elementary operations: pattern-matching and default inheritance. Speakers and hearers need to know that alternative forms can be used instead of the basic form, and of course in reality the choice between them is influenced by the linguistic and social context in ways that we must leave unexplored for the time being. Fig. 6 just hints at how these extra variables could be introduced.

[Figure 6 here]

All we need to assume is that these influences are constant whenever the choice is made; this is equivalent to Guy's assumption that t/d-loss is a single variable rule which always has the same statistical parameters.

It is important to emphasise that the proposed model is itself a model of competence, of linguistic knowledge. I assume that linguistic knowledge is closely integrated with other relevant knowledge (indeed there is good evidence in favour of this assumption - see Hudson 1990: Chapter 4 - and I know of very little evidence against it), so there is no need to distinguish linguistic knowledge from general knowledge. And it is a commonplace that general knowledge varies in entrenchment:

`factual assumptions are entertained with greater or lesser confidence' (Sperber and Wilson 1986:75); the more frequently a word is used, the easier it is to retrieve (Aitchison 1987:180); and so on. These two facts support a model of knowledge in which linguistic concepts, such as words and pronunciations, are closely linked to non-linguistic concepts and carry quantitatively different entrenchment values. However, none of this means that the proposed model is a model of performance. On the contrary, in order to give a complete causal explanation for Guy's observed performance figures the competence model needs to be embedded in a model of how speakers exploit their competence. Some basic characteristics of such a model are clear - in particular, many different mental operations must take place in parallel, and information must be used as soon as it is both available and relevant. This is not the place to try to develop a performance model, but it is important to stress that the competence model suggested here cannot explain the performance data fully without one.

4. Phonology: t/d loss as evidence for prototype networks

The critical element in the competence model is the generalisation which allows any postconsonantal (and syllable-final) t/d to be omitted. This provides the foundation on which we can build an explanation of Guy's exponential model. The next step is to show that it has more opportunities to influence the form of a monomorphemic word than an inflected word. The solution requires a more sophisticated understanding of the word word. What exactly is `the word mist', for example? Does it include mists as well as the singular mist? We could answer either yes or no, according to whether we are thinking of the lexeme MIST or of its singular, which we can call a `paradigm-member' of MIST. To avoid confusion we can give them different names: MIST (capitalised) versus $MIST_{sing}$. $MIST_{sing}$ and $MIST_{plur}$ are both paradigm-members of (`isa') MIST. Suppose we read (or hear) sentence (1).

(1) The autumn mists are famous around here.

The third word is a unique utterance, so it needs a unique name; for simplicity we can call it simply `word 3'. This uttered word is an example of both the lexeme MIST and also its plural, $MIST_{plur}$; and in the next sentence word 3 `isa' MIST and $MIST_{sing}$.

(2) The autumn mist hid the sun from view.

Now suppose the word written mist in the last example was pronounced /mis/; how could we explain this? There are two places where the stipulated inheritance relation could have applied: between the form of MIST and that of $MIST_{sing}$, or between the latter and the form of word 3. The two structures are shown in Fig. 7, with the relevant relationships highlighted. In the first diagram, $MIST_{sing}$ inherits the full form of MIST, complete with /t/, but the /t/ is not inherited from it by word 3; in the second, it is $MIST_{sing}$ itself that inherits the reduced variant from MIST.

[Figure 7 here]

The main point of this diagram is to show how the loss of /t/ at the higher level - between MIST and $MIST_{sing}$ - `bleeds' the rule at the lower level, in exactly the same way as in Guy's t/d-deletion rule. If /mist/ has a constant probability n of being realised as /mis/ at the next level down, the chance of /mist/ surviving at each level is 1-n, so the chances of survival at the bottom level are $(1-n)^2$.

We almost have an explanation for the exponential model, but first we must show why the figure for monomorphemes such as MIST is actually $(1-n)^3$, while regular and irregular past tenses have a different figure. To complete the analysis of monomorphemes we need a third level of structure, distinct from both the lexeme and the paradigm-member. One attractive candidate for this level is the abstract morpheme which may be shared by a number of lexemes. For example, take the set of words containing -ceive - CONCEIVE, DECEIVE, PERCEIVE and RECEIVE. They are all verbs, and they share at least one word-formation pattern (nominalization in -ception), so we may well link them together in our minds by abstracting the morpheme -ceive. And yet this is not the whole root of any word, so it exists only as a word-part. If it is stored anywhere, it

must be in a list of (meaningless) morphemes, bricks out of which words may be built. Similarly for the -stand in UNDERSTAND, which must be related somehow to the root of STAND although they have no meaning in common (Hudson 1984:44). These relations are shown (orthographically) in Fig. 8.

[Figure 8 here]

If some word-parts are stored in this way, why not all of them? In that case, the list of word-parts will provide the third level that we need: the form of word 3 `isa' the form of $MIST_{sing}$, which `isa' the form of MIST, which `isa' the word-part /mist/ (also found e.g. in MISTY). Each of these `isa' patterns allows the same mismatch between a target that has final /t/ and an instance that lacks it. The relevant patterns (where t/d loss is possible) are highlighted in Fig. 9.

[Figure 9 here]

We can now deal with the other two morphological classes, starting with irregular past-tense forms like left, for which we need two levels. For these the lexeme's form (e.g. /li:v/ for LEAVE) has no /t/, so it is irrelevant. However the abstract word-part list certainly is relevant, because many of these irregular past-tense forms clearly belong to a morphological prototype of the kind discussed by Bybee and Moder (1983). According to Quirk et al (1985:105f), there are 32 verbs which have a t/d suffix following another consonant in the past tense and which are irregular in their stem vowel (e.g. leave ~ left), in their stem-final consonant (e.g. bend ~ bent), or in the voicing of the t/d suffix (e.g. spell ~ spelt); but the vowel in the past-tense form is /ϵ/ in no fewer than 23 of the 32. Table 2 shows the relevant verbs classified by vowel in present and past. It is clear from this table that there is a strong tendency for past tenses that have an irregular /t/ suffix after a consonant to have the vowel /ϵ/ (regardless of the present-tense vowel).

[Table 2 here]

The conclusion that I draw from Table 2 is that irregular past-tense forms like left inherit from a general /.ϵCt/ prototype as well as from their specific stored form, as shown in Fig. 10.

[Figure 10 about here]

Each of the highlighted inheritance relationships is subject to the general t/d-loss pattern, giving just two chances for the suffix to be lost - exactly as required by the exponential model.

The third morphological class is the class of regular past-tense forms like missed. Here there is only one opportunity for t/d-loss, between the form of the past-tense of MISS and the uttered word (which we can continue to call `word 3'). The t/d-loss pattern applies only to a final /t/ which follows a consonant, but the /t/ of the affix and the /s/ at the end of the stem are not combined before the form of $MISS_{past}$. Consequently only one pattern is highlighted in the diagram:

[Figure 11 here]

To summarise, this discussion has outlined a radical alternative to Guy's Lexical Phonology analysis of t/d-loss, which explains the exponential retention figures without the weaknesses in the Lexical Phonology analysis. The key ideas in this alternative are:

- that the possibility of t/d loss is shown in competence as an example of rule-governed deviation in an inheritance hierarchy;
- that variable phenomena such as this have some degree of entrenchment in our competence, reflecting (inter alia) the statistical patterns in our experience; and
- each word-form to which t/d loss could apply has one, two or three distinct relevant models, depending on the word's morphological structure; each of these models is a point at which the t/d could be lost.

6. Syntax: DO-insertion as evidence for parameters

Another important contributor to the debate about variation in linguistic theory is Anthony Kroch. His paper in the CLS volume (1994) gives data on changes in the frequency of periphrastic do in early Modern English, which he uses to illustrate the statistical generalisation about syntactic change called the Constant Rate Effect. (From now on I shall refer to periphrastic do simply as DO.) He then offers an account of the rise of DO in terms of a Principles-and-Parameters analysis, so the critical question is how successful this account is, and whether it can be improved by changing the theoretical assumptions.

First, the data. He quotes the following examples to show how DO distinguishes modern English from early Middle

English.

(3) a How great tribulations suffered the Holy Appostels?
b How great tribulations did the Holy Apostles suffer?

(4) a .., whiche he perceiueth not.
b .., which he does not perceive.

(5) a Queen Ester looked never with swich an eye.
b Queen Ester never looked with such an eye.

In his numerical data which plot the course of this change he distinguishes five sentence-types, which are shown in Table 3 (with the Middle English examples modernised in all irrelevant respects).

[Table 3 here]

For each sentence-type, Kroch quotes both the total number of textual examples counted (N) and the percentage of these in which DO was used. His raw numbers (based on Ellegård 1953:159)[2] are repeated in Table 4.

[Table 4 here]

As can be seen from the second column, the data-points are not distributed evenly through time. It will be helpful to present the figures graphically, so we can reorganise them in equal periods of 25 years by collapsing the figures from 1526-1550, and by splitting those for 1426-1475 (with some averaging to smooth transitions where appropriate). The result is the rather complex graph in Fig. 12, which appears to confirm Kroch's Constant Rate Effect - that `the rate at which the newer option replaces the older one is the same in all contexts' (1994:181). According to Kroch, a statistical analysis in terms of logistic regressions shows that the rate of slope is the same in all contexts (ibid:182). More generally, he says that the same pattern, the Constant Rate Effect, `is found repeatedly in empirical investigations'.

[Figure 12 here]

Kroch takes this statistical linkage between the various changes shown on this graph as evidence for a fundamental linkage in the speakers' competence:

> We take its general validity to indicate that what changes in frequency in the course of time during a syntactic change is language users' overall tendency to choose one abstract grammatical option over another in their language production. ... The unity of the change is defined at the level of the grammar, not at the level of the surface contexts.

Like Guy, then, he finds a single statistical pattern behind the diversity for different linguistic contexts, and seeks an explanation for this unifying pattern in competence; but whereas Guy proposed a single variable rule responsible for all the alternations, Kroch rejects an explanation in terms of variable rules. Instead, he follows recent developments in Chomskyan syntactic theory in assuming that Middle English and modern English are distinguished by a single parameter which has many ramifications, so that the transition period covered by his data involved competition between two distinct grammars.

The parameter which changed between Middle and modern English is the presence or absence of verb-raising (or, in more recent terms, whether or not verb-raising is forced to apply before the structure is `spelled out' phonetically). Fig. 13 shows the basic sentence-structure which he assumes.

---

[2]According to Warner (1993:220), Ellegård's percentages are consistently too low for all the years before 1500, because he excluded any text that did not contain any examples of periphrastic DO at all. Warner has estimated the correct percentages, but unfortunately only two of his sentence-type categories correspond to those in Kroch's table so I have left the latter unchanged. I cannot tell whether the correction has any consequences for Kroch's statistical analysis.

[Figure 13 here]

The assumption is that the node `I' (for Inflection or INFL) is crucial for subject inversion and negation, so if a verb is raised to I both of these possibilities are open to it; and at the same time verb-raising allows the verb to precede the adverb (e.g. looked never in (5)a). Without verb-raising, neither inversion nor negation is possible, so DO must be supplied; and the verb must follow the adverb (never looked). (Auxiliary verbs are exempt from the loss of verb-raising, which is why they do not need DO; see p. 192.)

This analysis has the advantage of relating the rise of DO to the position of adverbs, which would appear at first glance to be an unrelated phenomenon. Adverb-preposing started somewhat earlier, but apparently its rate of increase has the same slope as the spread of DO. The graph in Fig. 14 includes Kroch's figures for the one adverb never, showing how often it is used before the verb. For convenience I have collapsed all the DO figures into two sets; as can be seen from the previous graph, the figures for negative questions and (affirmative) transitive questions are much higher than for the other three sentence-types, so I have lumped the former together as `high do' structures, leaving the remainder as the `low do' residue. The line between them is the mean DO figure for all sentence-types.

[Figure 14 here]

To judge by this graph, the high-do contexts had a growth-spurt around 1500 (about 25 years after the spurt for never + V), which was not matched by the low-do contexts until much later - in fact we can only assume the existence of a later spurt during the next century after the period covered by Kroch's data. We must obviously also assume that the spurt for the high-scorers levelled off during this period, as the figures approached 100%. In short, we seem to have evidence here for three separate S-shaped curves whose spurts cover about a century, but which (according to Kroch's statistical tests) have the same slope and therefore could illustrate different applications of the same change.

Most of Kroch's paper deals with the mechanism which drove our linguistic ancestors to sort out the conflict between their two grammars, called Grammar Competition. He suggests that:

> the historical evolution of competing variants in syntactic change is similar to the evolution of morphological doublets. In both cases, the coexistence of the variant forms is diachronically unstable: One form tends to drive the other out of use and so out of the language. (196)

He suggests that all syntactic change may be located in the requirements of individual formatives (184), and Grammar Competition eliminates alternative syntax just as it eliminates synonyms (except when these are supported by sociolinguistic differences). Unfortunately his general discussion of Grammar Competition does not deal directly with verb-raising, but he presumably sees this too as something which is triggered by the formative INFL, so the competition is between an INFL which triggers raising and one which does not.

Kroch's statistics do seem to support his claim that DO spread through the various sentence-types at the same speed, and that this spread could be linked to the change in adverb position. However there are serious weaknesses in his theoretical interpretation.

■ If the change involved nothing but the verb-raising requirements of I, why did it enter the different sentence-types at different times? Kroch's analysis predicts that DO and adverb-preposing should have grown not only at the same speed but also at the same time, with no difference at all across sentence-types. Kroch recognises this, but offers only a partial explanation:

> .. the approach taken here implies directly that the frequency differences in different contexts of a change must be due to factors orthogonal to the grammatical change itself .. and that these orthogonal factors are responsible for the differences in the .. parameter values in [our tables]. Such factors are not well-understood but must involve psycholinguistic and information processing preferences, which, in usage, favor one form or the other differentially in different linguistic contexts whenever a language, for any reason, happens to allow more than one option for

> expressing a given linguistic content. (183)

No doubt processing factors are relevant, and I myself shall invoke them below, but Kroch's account makes them bear a remarkably heavy burden in the explanation. In effect, processing effects have to be powerful enough to override the effects of the subject-raising parameter. For instance, by 1525 the order never + full verb is almost obligatory (89%), so at least 89% of utterances must have been based on a grammar without verb-raising; but at that time only 8% of negative declaratives had DO. If the verb-raising choice is independent of sentence-types, as Kroch claims, at least 89% of negative declarative sentences had un-raised verbs, so processing factors have to explain why 80% of un-raised verbs could still carry negation. Of course, an alternative interpretation would be that the processing factors influence the choice between the two grammars, so that negative declaratives tended strongly (in 1525) to favour the verb-raising grammar; but this would be to deny Kroch's claim that the choice of grammar-type is independent of the sentence-type in which it is used.

■ Warner (1993:219ff) discusses the spread of DO in the context of a wide range of other contemporaneous changes which Kroch does not mention, but which according to Warner are interconnected (as I shall explain below). One simple example is that CAN and MAY both lost their non-finite forms in the same period in which DO became normal in questions (Warner 1993:221). Presumably Kroch would have to deny a causal connection between the two changes, which may of course be historically accurate, but Warner makes a convincing case for a linkage, and presumably the more changes can be covered by a single explanation, the better.

■ The assumptions of Principles-and-Parameters theory force Kroch to consider the old and new patterns as distinct grammatical systems (183):

> The options in question .. are not alternating realizations within a single grammar, like extraposed versus non-extraposed constituents. Rather they seem always to involve opposed grammatical choices not consistent with the postulation of a single unitary analysis. In the present case, for example, contemporary accounts of verb-movement to INFL all agree that it is forced by the morphosyntactic contents of functional heads and cannot be optional. Because the variants in the syntactic changes we have studied are not susceptible of integration into a single grammatical analysis, the variation does not stabilize and join the ranks of a language's syntactic alternations. Instead, the languages always evolve further in such a way that one or the other variant becomes extinct.

Does this really mean that Shakespeare alternated between two distinct `grammatical systems' according to whether or not he used DO? And if we could find evidence for another similar but unrelated change would that mean he alternated among four distinct systems? How could Kroch's interpretation be distinguished from one in which the verb-raising feature of INFL has two alternative values, both within the same grammatical system? If the distinction cannot be supported empirically, we are left without evidence for Kroch's fundamental distinction between `alternating realizations within a single grammar' and `grammar competition'.

■ As mentioned earlier, Kroch himself recognises auxiliary verbs (HAVE, BE and the modals) as verbs which continue to be raisable, so how can verb-raising be controlled only by INFL?

■ Near the end of the paper (195) Kroch points out that verbs and nouns, unlike adpositions, never seem to have idiosyncratic positions, and suggests that the reason for this may be that verbs and nouns are always raised to a higher functional head before surface structure. This suggestion clearly conflicts with the claim that modern English verbs do not raise (to INFL or anywhere else) at surface structure.

■ Kroch suggests that all syntactic variation is located in functional head nodes. In particular, he claims that all word-order

variation is determined by heads rather than by dependents. This seems to imply that the position of an attributive adjective relative to its head noun can never vary from adjective to adjective - but this is precisely what we do find, at least on the surface, in the Romance languages (where most adjectives follow the noun but a few are allowed to precede it). Similarly, in English sentence-adverbs vary lexically in their preferred positions relative to their heads; for example, also typically precedes its head, whereas its synonyms too and as well have to follow it:

(6) a John also/*too/*as well snores.
b John snores also/too/as well.

Indeed, the adverb never which Kroch himself discusses is a clear example of idiosyncratic positioning, because unlike most other adverbs it cannot occur after the verb:

(7) a John often/sometimes/never snores.
b John snores often/sometimes/*never.

■ The general thesis that syntactic variation is always resolved definitively in favour of one pattern does not seem to be borne out by some well-known cases. He quotes Taylor's (forthcoming) work on changes in Greek word-order between Homer and the New Testament, which does indeed show an increasing tendency to prefer SVO rather than the earlier SOV; but the outcome was only a trend, and quite unlike the relatively rigid SOV order found in (say) German subordinate clauses. This flexibility continues into modern Greek, two millenia later, where SVO is still dominant but all five other orders of S, V and O are still found (Tzanidaki 1995; Lascaratou 1989).

■ Kroch, like many other linguists, assumes that `verb-raising' is no longer found in English. According to Warner (1993:66) subject inversion was still common with some full verbs (e.g. MEAN, SAY, THINK) through the eighteenth century, so to the extent that subject inversion depends on verb-raising this would suggest a much more gradual elimination than Kroch's estimate (`.. 1575, the date at which V-to-I movement is definitively lost' 181). Moreover, there is one particularly important and problematic auxiliary which is usually overlooked: the possessive verb HAVE. In many UK dialects the following sentences are all grammatical (Trudgill 1978:14f).

(8) a He has a car.
b Has he a car?
c He hasn't a car.

We have already seen that Kroch accepts that auxiliaries still raise, but for most of them it is presumably possible that they start in some position other than the ordinary `V' node, which would explain why ordinary verbs do not raise. But this HAVE is transitive, so it must be the V which case-marks the object; but it allows inversion and negation. Admittedly this odd auxiliary appears to be on the retreat, under threat from the UK alternative HAVE GOT and the USA possessive HAVE which is just an ordinary verb, like GET; and it seems reasonable to see this shift as part of the same historical process that Kroch documents. But it is now four centuries since the supposed end of verb-raising (which as we have seen Kroch dates at 1575), so it is very strange that this residual raising verb is still lurking in our vocabularies.

■ The theoretical framework that Kroch assumes is itself highly controversial, so the validy of his argument is tied to that of the theory. This means that the argument is weakened for those of us who question the Chomskyan family of theories for other reasons (Hudson 1976, 1984, 1990).

5. Syntax: DO-insertion as evidence for prototype-networks

Kroch's discussion has led us to much the same conclusion as the one we reached over Guy's paper: the data are interesting, apparently robust and challenging, but the theoretical explanation has serious weaknesses. Once again I shall offer an alternative in which a network of prototypes plays a major role. In doing so I shall be largely following Warner (1993), a careful and well-documented study of modern English auxiliary verbs and their history, supplemented by relevant sections of Denison 1993, a general history of English syntax. Like Warner and Denison, I believe the major change that affected English in the early modern period involved the distribution of grammatical characteristics between the two subclasses of verbs, the auxiliary and full verbs. (The term `auxiliary' verb has to be taken with caution, bearing in mind that they do not necessarily `help' another verb; even in modern English we have a transitive auxiliary verb - the possessive HAVE discussed above - and a pure copula auxiliary verb - BE; in Old English and Middle English auxiliary verbs were even more diverse, as we shall see.)

Warner's account of developments in this part of English grammar is very different from Kroch's. The unifying factor, which could explain the Constant Rate Hypothesis, is provided by the two word-classes concerned (auxiliary and full verbs), whose membership stayed roughly the same. What changed so dramatically was the range of characteristics which distinguished the classes, but the increase in the fifteenth and sixteenth centuries was part of a much longer development which extended far beyond the period covered by Kroch. In a nutshell, we have moved from the Middle English list of differences in Table 5 to the modern list in Table 6. (The numbers after the examples refer to the pages where Warner quotes them.)

[Table 5 here]

[Table 6 here]

The most striking difference between these two tables is in their length. Since Old English, the class of auxiliary verbs has grown up. As a `baby' category it had very few distinctive characteristics, and a somewhat fluid membership at the points where these characteristics disagreed; but in its modern form it is a perfect example of a well-developed prototype, distinguished by a large number of characteristics which, by and large, all apply to the same range of members, but with a penumbra of exceptional forms.

The fourth column of Table 6 shows roughly when the new form or restriction entered the grammar. The dates show that the development of the auxiliary class is spread over a thousand years, with Kroch's changes as just one segment of the development. It's true that the sixteenth century seems to have been a particularly important period in this history, but as I pointed out earlier, not all the changes happening at this time can be explained by the loss of verb-raising. Warner (192) explains this `apparently coherent long-term development of an auxiliary group' in terms of cognitive principles, namely Rosch's (1978:28) `principle of cognitive economy':

> The task of category systems is to provide maximum information with the least cognitive effort.

This principle favours categorization systems in which distinct characteristics are highly correlated (as in the case of our modern auxiliary-verb class). Warner suggests two causal links between this principle and the observed change, but the one which I should like to highlight is that the classification of human behaviour (such as language) affects the material classified, unlike our classification of everything else. In the case of language, Rosch's principle provides a feed-back loop: the more closely linked we think two features are in other people's usage, the more we link them in our own usage, which in turn encourages our hearers to link them even more closely in their usage, providing us with even better evidence for their linkage, and so on.

The last column of Table 6 shows whether it was the auxiliary class or the full-verb class that had acquired the new characteristic. For example, the `+ adverb' characteristic of auxiliaries is simply a continuation of the ability to occur before an adverb such as never which was previously shared by all verbs, but which became distinctive for auxiliaries when full verbs lost it; this is shown as a change to full verbs. In contrast, the `reducible to clitic' characteristic was an innovation which only affected auxiliary verbs. It is therefore misleading to refer to this historical change as the development of the auxiliary class; it could just as accurately be described as the development of the full-verb class. According to the cognitive account, it is simply

the separation of two classes, neither of which has any particular priority. Some of the characteristics of OE verbs have been inherited by modern full verbs, and others by modern auxiliaries; but both are still verbs.

We still have to explain Kroch's constant rate hypothesis - why did DO change from impossible to obligatory at the same speed in all different sentence-types? A crucial element in the explanation will be the assumption that each change involved lexical diffusion through the same range of individual verbs, so it is important to be clear that the growth of the auxiliary and full-verb classes only involved their defining characteristics, and not their membership. More precisely, as prototype theory would make us expect, the two classes have always had untypical members, and partly because of this their membership has always been slightly unstable, but this variation is only minor. The core membership has always been the class of modal verbs, defined by the characteristics in Table 7 (which also reached roughly the modern state of mutual predictiveness by the end of the sixteenth century - Warner ibid:200ff):

[Table 7 here]

However, the membership of this class has varied:

(9) Old English modals (Warner 153's tentative list)

CAN, MAY, SHALL, WILL, MOT (`may, must'), OWE, þARF (`need'), UTON ("let's"), DARE;
marginal members lacking one characteristic: BE, WEORðAN (`become')

(10) Modern English modals

MAY, SHALL, WILL, MUST, OUGHT (past of OWE);
marginal members (see Table 7): CAN, DO, BE + TO
currently turning into full verbs? OUGHT, USED, DARE, NEED

In addition to the modal verbs, the class of auxiliary verbs has always included BE, and since Middle English, $HAVE_{perf}$ (perfect HAVE, as in have seen; Warner 117). The two systems are diagrammed in Fig. 15. The main change in late Middle English was the rise of DO (periphrastic do) in the thirteenth century (Warner 220, Denison 1993:264), which I have located in the modern class of modals because it has to be finite and takes a plain infinitive complement. It is the spread of this verb that we now have to explain.

[Figure 15 here]

This discussion of verb classes provides the essential background to an explanation of Kroch's changes. Once a change has become associated with members of one or the other of these classes, we should expect it to spread at the same rate because it has the same route to follow: first, through the same range of verbs, and second, through the same range of speakers. We can imagine it starting as an idiosyncratic characteristic of a handful of verbs, which eventually generalises to whichever class these verbs belong to. But this is just the second half of the story. The first half is about why these particular changes took place - in other words, why did auxiliary verbs acquire a special association with questions and negation, which eventually turned into a monopoly? Any explanation must be speculative, and it may well be that multiple causes were in fact responsible for the events, but the essential point to remember is that the events to be explained are quite minor - just a slight tendency for one structure to be preferred to another for one particular sentence-type. Because of the feed-back mechanism noted above any tendency in in one person's speech may influence other people's speech, thereby reinforcing the initial tendency. Just as the beating of a butterfly's wing could be the ultimate trigger for a thunderstorm, so a single choice by one speaker could ultimately trigger a major grammatical change.

We saw in Fig. 12 that the first two sentence-types to adopt DO were transitive and negative questions. Why should these have been affected before ordinary affirmative or intransitive questions? Kroch's suggestion (1989) is that the use of DO made it easier to distinguish the subject and the object: in V + NP + NP there might be some ambiguity or uncertainty which is absent from DO + NP + V + NP, where the first NP must be the subject. This advantage would clearly not apply to intransitive

sentences, where there is only one NP. As for negation, Denison (1993:467) reports a small but significant statistical link between auxiliary verbs and not even in the late fifteenth century, before Kroch's major change started. This trend would presumably have been reinforced in questions by the functional link between negation and questioning which makes us interpret a negative yes/no question as conducive. For example, Can't you swim? is a special way of finding out whether you can swim, and not just the negation of Can you swim?. Processing may be slightly easier if the word which signals both the question and the conduciveness is distinct from the one that signals lexical content, as would be the case if an auxiliary (e.g. DO) was used: its position before the subject signals the question, while the following not signals the conduciveness - hence (perhaps) the preference for `DO + subject + not + V' over the syntactically simpler `V + subject + not'.

It would probably be easy to multiply these speculative functional explanations for the early link between questions and transitivity or negation, and each new functional pressure could reinforce the others. Let's assume that we have an explanation for the early stages of DO-support for full verbs. What about auxiliary verbs? Why do we not find DO-support for them? First, they are (typically) intransitive, so the pressures on transitive verbs do not apply. Secondly, they were already linked to negation via both morpho-syntax and semantics. Morphologically, as we have already seen in Tables 5 and 6 in Old English the negative marker ne was frequently cliticized to auxiliary verbs as a proclitic (e.g. nis for ne is), and in late Middle English its successor, not was often attached as an enclitic (e.g. isn't); and for a syntactic link between auxiliaries and negation we have Denison's evidence mentioned above. Semantically, negation often interacts with the meanings of modal verbs (e.g. the negative of must is need not). These counter-pressures may well have been enough to protect auxiliary verbs from the tendency to use DO with full verbs; and of course once a difference is established between auxiliary and full verbs cognitive pressures are likely to magnify it in order to improve the informativeness of these classes.

The main point of this discussion is that Kroch's explanation for the changes in late Middle English may not be the best one. The alternative that I have offered (following Warner) has the advantage of explaining far more changes in English grammar and of predicting the gradual evolution that Kroch's figures indicate. To summarise this explanation, it has two parts, one cognitive and the other functional. The cognitive part involves the idea that word-classes are prototypes, mental categories which link characteristics in bundles. These categories become more useful as the bundle of linked characteristics grows; but we only link characteristics mentally when our experience provides some justification (even if the justification is only partial). But language, unlike other kinds of experience, is under the control of the experiencer, so these cognitive pressures can influence the very experience on which we base our prototypes, with the result that the prototypes tend to get more and more distinct. The functional part of the explanation plays a rather minor role, providing an initial trigger for any change in behaviour. Fortunately for us, it is fairly easy to study the categories and their characteristics which are relevant to the cognitive account. What is much harder is to get beyond mere speculation in finding functional explanations for the birth of the distinctinctive characteristics.

6. Conclusion

This discussion has raised some very general questions which go well beyond the details of t/d loss and periphrastic DO, questions about the principles of linguistics and the nature of human cognition. I should like to finish by sketching some consequences for these two questions.

As far as linguistics is concerned, I agree entirely with both Guy and Kroch that linguistic theory should be confronted with the statistical data of variation studies. The two papers reviewed here both have the merit of having identified data which is clearly relevant to theoretical issues, though I disagree with the theoretical conclusions. At the very least the debate has shown beyond doubt that statistical data can `have implications of some kind for the theory concerned, by supporting or contradicting one of the claims derivable from it', contrary to the views quoted at the start of the paper. Maybe we can at last hope for a fruitful meeting between variation data and theories of language structure.

The question of human cognition is partly about the nature of categories (classical Aristotelian categories or prototypes?) but also about the place of language in cognition (as a separate module or integrated?) and the source of knowledge (innate or learned?). Guy and Kroch both assume a `classical' theory of language structure which is modular and

partly innate, and the fundamental differences between Lexical Phonology and Principles-and-Parameters theory imply a theory of language structure in which phonology and syntax are basically very different from each other as well as from the rest of cognition. In contrast, my `cognitive' explanation for their data emphasises the similarities both between language and other areas of cognition and also between phonology and syntax. Both parts of language, like large areas of general cognition, are assumed to be organised in terms of prototypes organised in `isa' hierarchies which allow default inheritance. Each of these opposed views is shared by a large number of people, but the difference between them is fundamental and hard to bridge. It may be that variation statistics will prove to be a meeting-point where the merits of the two views can fruitfully be compared.

REFERENCES

Aitchison, J. 1987. Words in the Mind. An introduction to the mental lexicon. Blackwell.

Bayley, R. 1993. Variation in Tejano English: evidence for variable lexical phonology. Paper presented at the Conference on Language Variety in the South, Auburn University.

Beals, K., Denton, J, Knippen, R., Melnar, L., Suzuki, H. and Zeinfeld, E. (eds.) 1994. CLS 30. Papers from the 30th Regional Meeting of the Chicago Linguistic Society, Vol 2: The Parasession on Variation in Linguistic Theory. Chicago Linguistic Society.

Booij, E. and Rubach, J. 1992. Lexical phonology. In Bright, W. (ed.) International Encyclopedia of Linguistics, Vol 2. Oxford University Press, 327-30.

Bybee, J. and Moder, C. 1983. Morphological classes as natural categories. Language 59, 251-70.

Cedergren, H. and Sankoff, D. 1974. Variable rules: performance as a statistical reflection of competence. Language 50, 333-55.

Denison, D. 1993. English Historical Syntax. Longman.

Ellegård, A. 1953. The auxiliary `Do': the establishment and regulation of its use in English. Stockholm: Almkvist and Wiksell.

Fasold, R. 1990. The Sociolinguistics of Language. Blackwell.

Fraser, N. and Hudson, R. 1992. Inheritance in Word Grammar. Computational Linguistics 18, 133-58.

Guy, G. 1980. Variation in group and the individual: the case of final stop deletion. In Labov, W. (ed.) Locating Language in Time and Space. Academic Press, 1-36.

Guy, G. 1991a. Explanation in variable phonology: an exponential model of morphological constraints. Language Variation and Change 3, 1-22.

Guy, G. 1991b. Contextual conditioning in variable lexical phonology. Language Variation and Change 3, 223-39.

Guy, G. 1993. Lexical phonology and the problem of variation. In Guenter, J., Kaiser, B. and Zoll, C. (eds.) BLS 19, Proceedings of the 19th Annual Meeting of the Berkeley Linguistics Society, Berkeley Linguistics Society, 171-81.

Guy, G. 1994. The phonology of variation. In Beals et al (eds.), 133-49.

Guy, G. and Boyd, S. 1990. The development of a morphological class. Language Variation and Change 2, 1-18.

Halle, M. and Bromberger, S. 1989. Why phonology is different. Linguistic Inquiry 20, 51-70.

Hudson, R. 1976. Arguments for a Non-transformational Grammar. Chicago University Press.

Hudson, R. 1980. Sociolinguistics. Cambridge University Press.

Hudson, R. 1984. Word Grammar. Blackwell.

Hudson, R. 1985. A psychologically and socially plausible theory of language structure. In Schiffrin D. (ed.) Meaning, Form and Use in Context: Linguistic Applications, Georgetown University Press, 150-9.

Hudson, R. 1986. Sociolinguistics and the theory of grammar. Linguistics 24, 1053-78.

Hudson, R. 1990. English Word Grammar. Blackwell.

Hudson, R. in press. Syntax and sociolinguistics. In Jacobs, W. et al. (eds.), Syntax. An International Handbook of Contemporary Research, Vol 2, de Gruyter, 1514-1528.

Kemmer, S. and Israel, M. 1994. Variation and the usage-based model. In Beals et al. (eds.), 165-79.

Kroch, A. 1989. Function and grammar in the history of English: Periphrastic Do. In Fasold, R. and Schiffrin, D. (eds.) Language Change and Variation. Benjamins, 132-72.

Kroch, A. 1994. Morphosyntactic variation. In Beals et al. (eds.), 180-201.

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

Labov, W. 1972. Sociolinguistic Patterns. University of Pennsylvania Press.

Langacker, R. 1987. Foundations of Cognitive Grammar Vol 1. Theoretical prerequisites. Stanford University Press.

Langacker, R. 1990. Concept, Image and Symbol. The cognitive basis of grammar. Mouton de Gruyter.

Langacker, R. 1995. Raising and transparency. Language 71, 1-62.

Lascaratou, C. 1989. A Functional Approach to Constituent Order with Particular Reference to Modern Greek. Implications for Language Learning and Language Teaching. Athens.

McMahon, A. 1994. Lexical phonology and morphology. In Asher, R. (ed.) Encyclopedia of Language and Linguistics. Pergamon Press, 2155-60.

Pierrehumbert, J. 1994. Knowledge of variation. In Beals et al. (eds.), 232-56.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. A Comprehensive Grammar of the English Language. Longman.

Roca, I. 1994. Generative Phonology. Routledge.

Romaine, S. 1982. Socio-historical Linguistics. Cambridge University Press.

Rosch, E. 1978. Principles of categorization. In Rosch, E. and Lloyd, B. (eds.) Cognition and Categorization. Erlbaum, 27-48.

Santa Ana, O. 1992. Locating the linguistic cycle in vernacular speech: Chicano English and the Exponential Hypothesis. In Denton, J. et al. (eds.) CLS 28. Papers from the 28th Regional Meeting of the Chicago Linguistics Society, Vol 2: The Cycle in Linguistic Theory, 277-87.

Smith, N. 1989. The Twitter Machine. Reflections on language. Blackwell.

Sperber, D. and Wilson, D. 1986. Relevance: Communication and cognition. Blackwell.

Taylor, A. forthcoming. The change from SOV to SVO in Ancient Greek. Language Variation and Change 6,

Taylor, J. 1989. Linguistic Categorization. An essay in cognitive linguistics. Oxford University Press.

Trudgill, P. (ed.) 1978. Sociolinguistic Patterns in British English. Arnold.

Tzanidaki, D. 1995. Greek word order: towards a new approach. UCL Working Papers in Linguistics 7.

Warner, A. 1993. English Auxiliaries. Structure and history. Cambridge University Press.

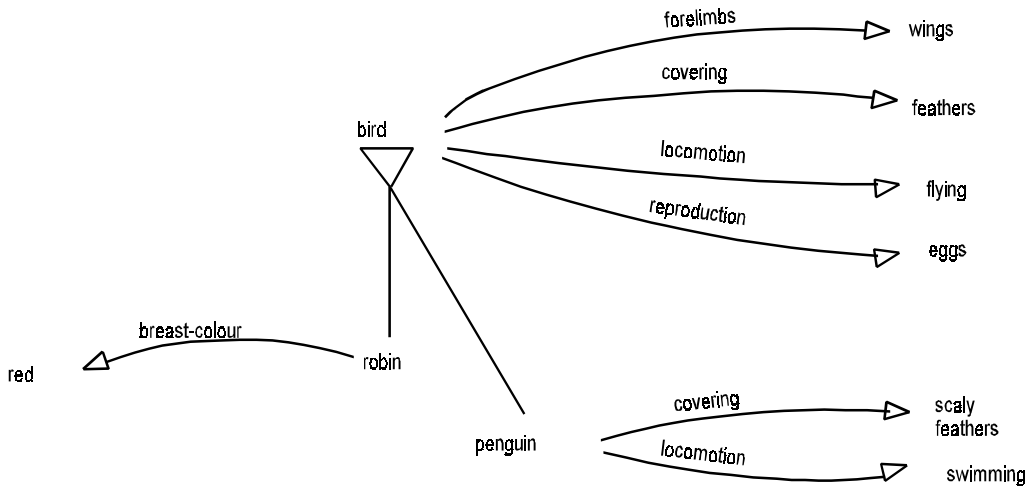
forelimbs
wings
covering
feathers
bird
locomotion
flying
reproduction
eggs
breast-colour
red
robin
penguin
covering
scaly
feathers
locomotion
swimming

Fig. 1

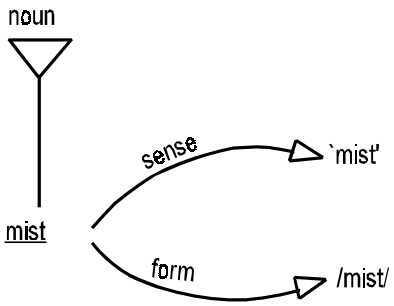
noun
sense
`mist'
mist
form
/mist/

Fig. 2

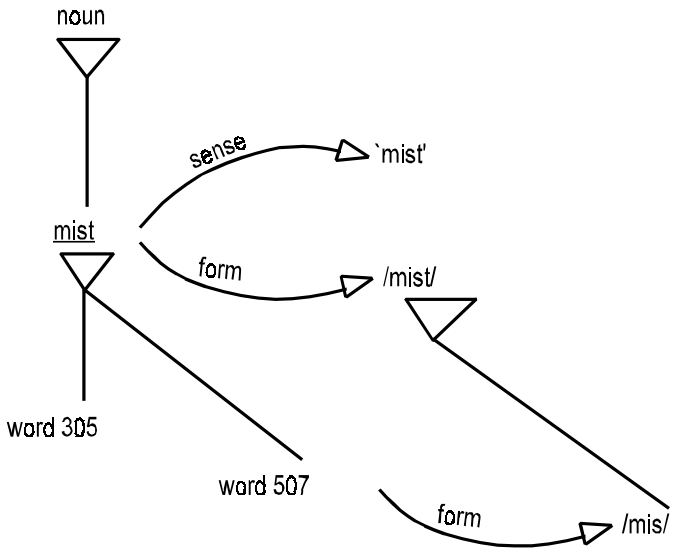
noun
sense
`mist'
mist
form
/mist/
word 305
word 507
form
/mis/

Fig. 3

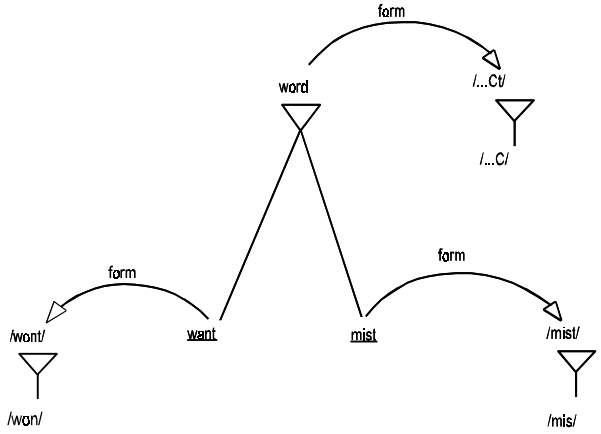
form
word
/...Ct/
/...C/
form
form
/wont/
want
mist
/mist/
/won/
/mis/

Fig. 4

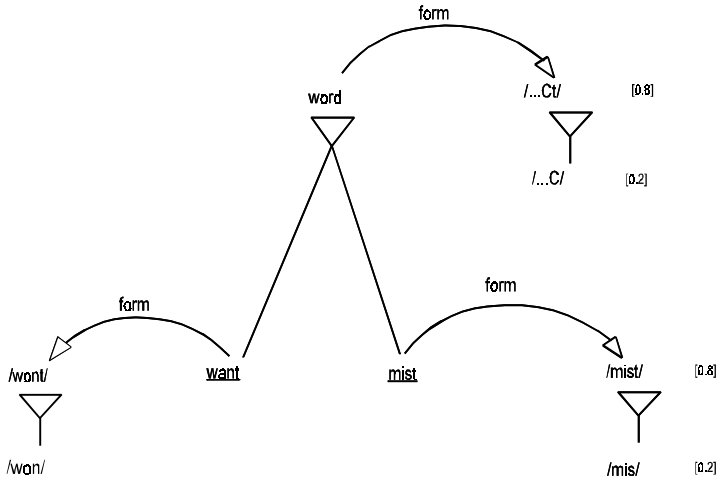
form
word
/...Ct/
[0.8]
/...C/
[0.2]
form
form
/wont/
want
mist
/mist/
[0.8]
/won/
/mis/
[0.2]

Fig. 5

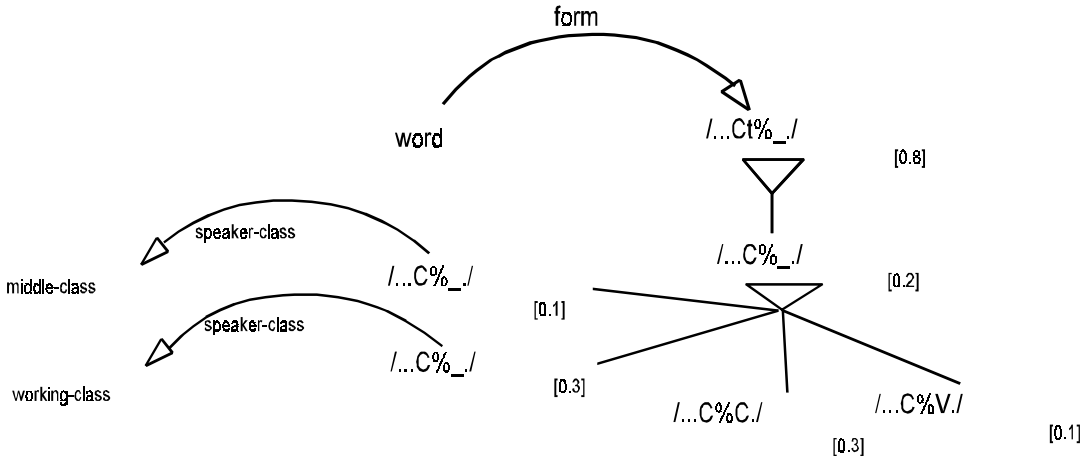
form
word
/...Ct%_./
[0.8]
speaker-class
/...C%_./
middle-class
/...C%_./
[0.2]
[0.1]
speaker-class
/...C%_./
working-class
[0.3]
/...C%C./
/...C%V./
[0.3]
[0.1]

Fig. 6

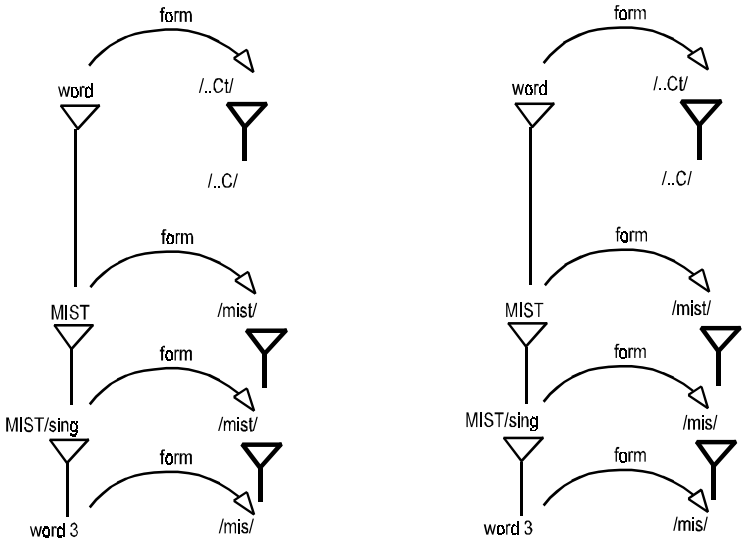
form
word
/..Ct/
/..C/
form
MIST
/mist/
form
MIST/sing
/mist/
form
word 3
/mis/
form
word
/..Ct/
/..C/
form
MIST
/mist/
form
MIST/sing
/mis/
form
word 3
/mis/

Fig. 7

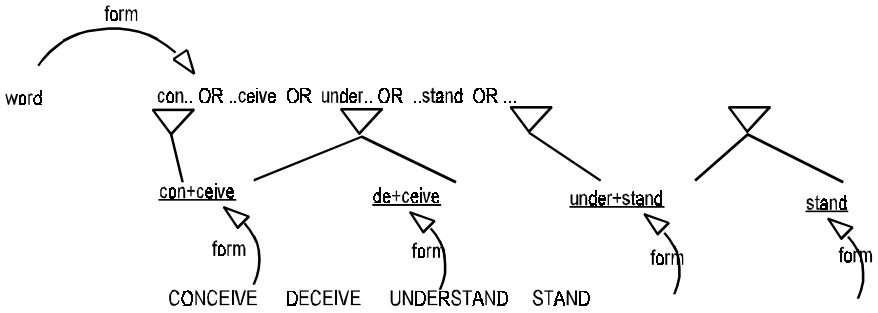
form
word
con.. OR ..ceive OR under.. OR ..stand OR ...
con+ceive
de+ceive
under+stand
stand
form
form
form
form
CONCEIVE
DECEIVE
UNDERSTAND
STAND

Fig. 8

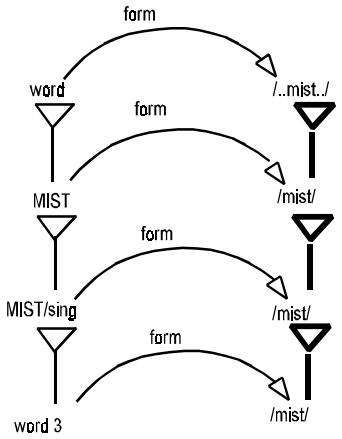
form
word
/..mist../
form
MIST
/mist/
form
MIST/sing
/mist/
form
word 3
/mist/

Fig. 9

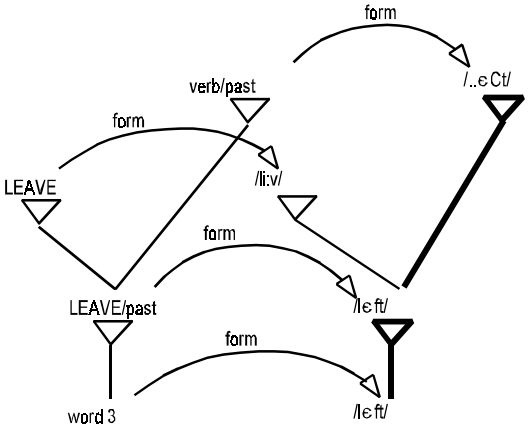
form
/..ɛCt/
verb/past
form
/li:v/
LEAVE
form
LEAVE/past
/lɛft/
form
word 3
/lɛft/

Fig. 10

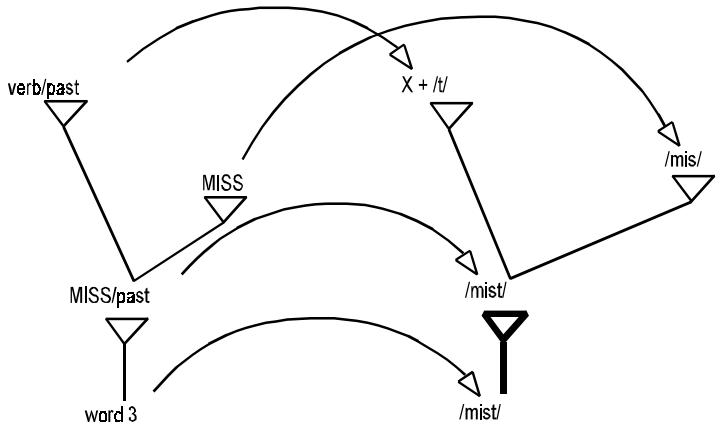
verb/past
X + /t/
/mis/
MISS
MISS/past
/mist/
word 3
/mist/

Fig. 11


% do
110
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
Neg quest
Aff trans quest
Aff intr quest
Neg decl
Aff wh-obj quest

Fig. 12.

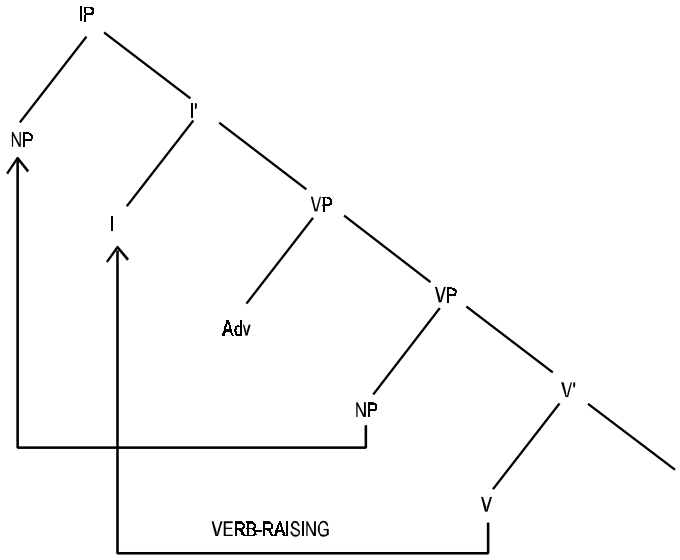
IP
I'
NP
VP
I
VP
Adv
V'
NP
V
VERB-RAISING

Fig. 13

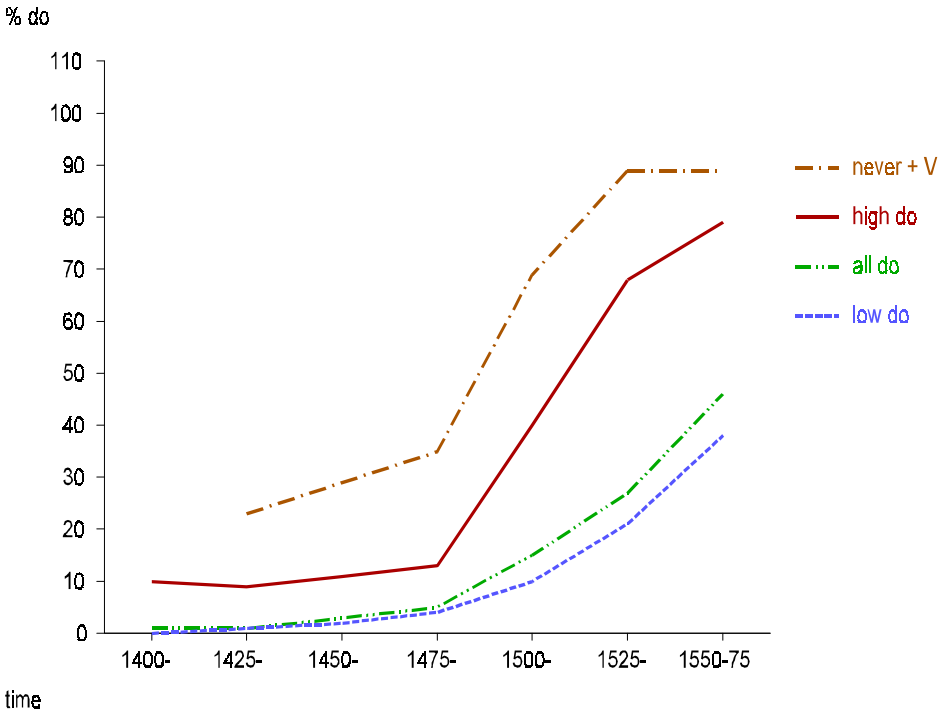
% do
110
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
never + V
high do
all do
low do

Fig. 14

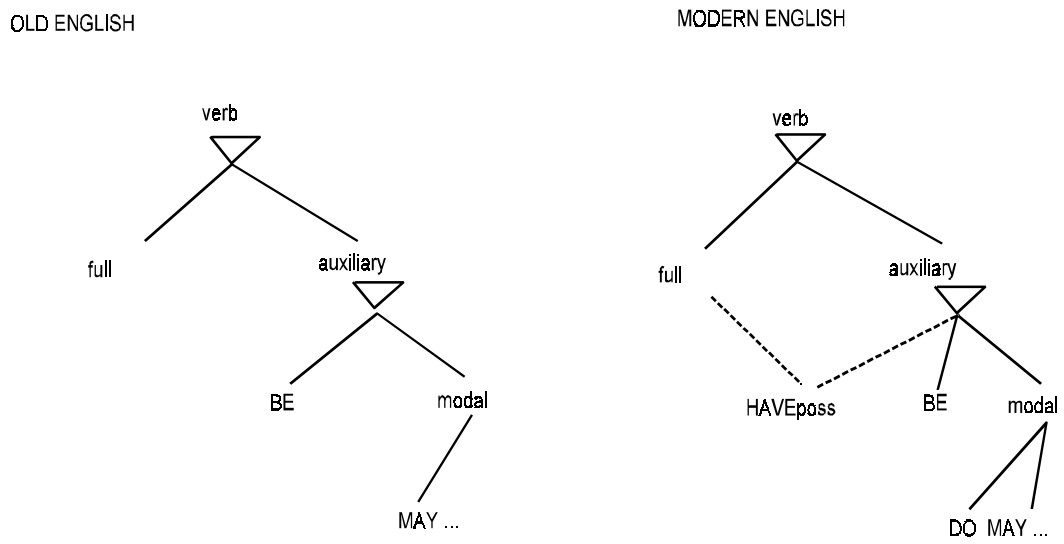

Fig. 15

corpus	word-type	observed		predicted
		N	% t/d retained	% t/d retained
Guy 1991a (7 speakers)	monomorpheme	658	61.9	61.4
	irreg past	56	66.1	72.3
	regular past	181	84.0	85.0
Santa Ana 1992 (45 speakers)	monomorpheme	3724	42.1	42.2
	irreg past	297	59.3	56.3
	regular past	836	74.3	75.0
Bayley 1993 (16 speakers)	monomorpheme	2065	43.3	43.9
	regular past	541	76.2	76.0

Table 1. t/d retention rates in three corpora (Guy 1994)

stem vowel		
present	past	
/ϵ/	/ϵ/	DWELL, SMELL, SPELL, BEND, LEND, REND, SEND, SPEND
/i:/	/ϵ/	BEREAVE, CLEAVE, CREEP, DEAL, DREAM, FEEL, KEEP, KNEEL, LEAN, LEAP, LEAVE, MEAN, SLEEP, SWEEP, WEEP
other	same	BURN, EARN, LEARN, SPILL, SPOIL
other	different	LOSE, SELL, TELL, (HEAR)

Table 2. Irregular suffixed verbs classified by stem-vowel

Sentence-type	Old	New
Negative declarative	He went not.	He did not go.
Negative question	Went he not?	Didn't he go?
Affirmative question transitive	Saw he the dragon?	Did he see the dragon?
Aff. question intransitive	Went he?	Did he go?
Aff. wh-object question	What saw he?	What did he see?

Table 3. Middle English and modern English treatments of 5 sentence types.

Time		Negatives				Affirmative questions					
Dates	yrs	declarative		question		transitive		intransitive		wh-object	
		%	N	%	N	%	N	%	N	%	N
1400-1425	25	0	177	11.7	17	0	3	0	7	0	1
1426-1475	50	1.2	903	8.0	25	10.7	56	0	86	0	27
1476-1500	25	4.8	693	11.1	27	13.5	74	0	68	2.0	51
1501-1525	25	7.8	605	59.0	78	24.2	91	21.1	90	11.3	62
1526-1535	10	13.7	651	60.7	56	69.2	26	19.7	76	9.5	63
1536-1550	15	27.9	735	75.0	84	61.5	91	31.9	116	11.0	73
1551-1575	25	38.0	313	85.4	48	73.7	57	42.3	71	36.0	75

Table 4. Frequency of DO by sentence-type.

property	examples
VP ellipsis and pseudo-gapping	deofol us wile ofslean gif he mot. `The devil will kill us if he can' (112)
transparent/raising (also: ONGINNAN, AGINNAN, `begin', WEORðAN, `become' and others?)	hine sceal on domes dæg gesceamian beforan gode. `Him (acc) shall at Doomsday be-ashamed before God' (123)
negative form in n- (also: WITAN, `know', HABBAN, `have')	nylle `don't want', nam `am not', næbbe, `have not' (151)

Table 5. Distinctive characteristics of Old English auxiliary verbs

Distinctive characteristic of auxiliary verbs	exceptions	example	date	page	cha-n-ged
VP ellipsis and pseudo-gapping		.. It would _ me _.	OE	111	-
+ adverb		*ran never	15c	206	full
subject inversion		*ran you?	15c	220	full
+ not		*ran not	15c	215	full
incompatible with DO		*do be	16c	209	aux
reducible to clitic	OUGHT, DO, USED, DARE, NEED	's going	16c	207	aux
tag question		.., is he?	16c	207	aux
VP as complement	$HAVE_{poss}$, BE	*will coffee	17c	202	aux
not -> n't		isn't	17c	208	aux
individual treatment of inflected forms[3]		have been to, *is to	19c	63	aux

Table 6. Distinctive characteristics of Modern English auxiliaries.

characteristic of modal verbs	modern exceptions

[3]Chapter 2 of Warner considers the idiosyncratic restrictions on the use of specific inflected forms of individual auxiliaries (e.g. *is having left, *will must, *is to London), which he uses as evidence for the theoretical claim that each form of an auxiliary, unlike full verbs, is stored individually.

Only finite	many dialects still allow multiple modals, e.g. will can
Preterite-present morphology - i.e. no -s with singular subject	BE + TO, DO (except non-standard He don't)
Plain infinitive complement	OUGHT, USED
Irregular meaning of past tense	CAN, DO, BE + TO
`Modal meaning' and subcategorization for VP	

Table 7. Distinctive characteristics of modern English modal verbs