The rise of auxiliary DO: verb-non-raising or category-strengthening?[1]

Richard Hudson
UCL

Abstract

The paper contrasts two explanations for Ellegård's statistical data on the rise of auxiliary DO during the 15th and 16th centuries. One is Kroch's explanation in terms of a change on the parameter of verb-raising, which is shown to have a number of serious weaknesses. The other is Warner's explanation in terms of the gradual development of the distinction between auxiliary and full verbs. Though Kroch quotes Ellegård's figures in support of the Principles-and-Parameters analysis, they actually support Warner's view much better. The paper also considers developments in the auxiliary system since the 16th century and offers a mixture of cognitive and functional explanations for the changes since the 13th century.

1. A qualitative history of the auxiliary DO

One of the most striking characteristics of Modern English is the importance of the auxiliary DO. In the absence of any other auxiliary, it is used in subject-inversion, in negation, in emphatic polarity, in `VP anaphora' and in tag questions, as shown by the following pairs of examples where a sentence containing DO is contrasted with one containing another auxiliary, HAVE.

(1) a Have they finished?
b Did they finish?

(2) a They haven't finished.
b They didn't finish.

(3) a They HAVE finished!
b They DID finish!

(4) a They have.
b They did.

(5) a They have finished, have(n't) they?
b They finished, did(n't) they?

Auxiliary DO is often called `periphrastic' DO because it has no meaning independent of the meaning of the construction concerned; the only reason for using auxiliary DO in Modern English is because the syntax requires an auxiliary and no other auxiliary is needed by the sentence's meaning. DO fills the gaps where non-auxiliary verbs are not allowed and where other auxiliaries are not needed.

In Middle English, in contrast, DO had no special role because auxiliary and non-auxiliary verbs could be used in much the same ways, as shown in Table 1 (where Middle English word-forms have been modernised). The choice of sentence-types to illustrate the change will be explained below.

[Table 1 about here]

This complementarity between auxiliary DO and other verbs means that any history of DO must also be a history of the whole system of auxiliary and non-auxiliary verbs. In a nutshell, what has happened over the last thousand years is that a range of constructions have developed in which only an auxiliary verb is allowed. The main constructions are exemplified in (1)-(5). As we shall see, there are other constructions in which auxiliary and non-auxiliary verbs are treated differently, but the ones just mentioned are special because they occur so often in any text, and correspondingly in the experience of any language user or language learner. Any change in these constructions has a high profile in the history of the language. We also know a great deal about the history of DO in these constructions, and can locate the change quite precisely in the fifteenth and sixteenth centuries. During that period, as non-auxiliary verbs waned, so DO waxed; and it makes no sense to try even to describe the history of DO on its own, let alone to explain it. This paper therefore has to present the changes in the use of DO as part of a much larger pattern of change which has affected all verbs.

This much is uncontroversial. The point of contention is the nature of these large-scale changes. I shall compare two very different views which are exemplified by the work of Kroch (e.g. 1989, 1994) and of Warner (especially 1993). I shall try to show that the numerical data which Kroch quotes in support of his account actually support Warner's account much better. In both cases the rise of DO turns out to be just a small corner of a much larger picture of grammatical reorganisation, but the overall pictures are very different. In Kroch's picture the details of change all follow from the resetting of a single underlying `parameter', the verb-raising parameter, during the 15th and 16th centuries. Before the change, English allowed all verbs to raise to the position of the abstract inflectional node in a Barriers-type sentence structure (Chomsky 1986, Pollock 1989); but since the change verb-raising is no longer allowed. In principle this parameter can only be set in one direction, but an individual may have two different grammars, each with a different setting for this parameter. Warner's picture shows a much more gradual development which had nothing to do with verb-raising; in fact, verb-raising is not recognised as a reality at any time in the history of English. Instead, what changed is the meaning of being an auxiliary or non-auxiliary verb. From modest beginnings, this distinction came to be relevant to a larger and larger range of constructions (and other patterns), a development which could extend over many centuries beyond the 15th and 16th centuries - a strikingly different view from Kroch's.

Whereas Kroch locates his explanation firmly in the transformational paradigm, Warner's belongs explicitly to the cognitive tradition with its stress on structured categories (such as word-classes). The debate involves serious questions about the nature of grammars and sentence structure; in particular, is the verb-raising transformation real in any languages? Indeed, are any transformations real? If Kroch's account is successful then it must count as serious support for the transformational view of sentence structure. If Warner is right, on the other hand, the diachronic facts are equally compatible with monostratal theories of syntactic structure such as Head-driven Phrase Structure Grammar (Pollard and Sag 1994), Lexical Functional Grammar (Bresnan 1982) or Word Grammar (Hudson 1990).

The facts for which these two views are alternative explanations are not in dispute. As mentioned earlier, Warner recognises a broader range of facts as relevant, but we can start by focussing on the patterns which Kroch discusses. These are the five constructions shown in Table 1 - negative and interrogative sentences of various kinds. These are the most relevant to Kroch's account because they are the ones where (on his analysis) verb-raising once applied but no longer does. However they are also relevant because we have useful statistical data which we shall consider in the next section. We shall consider Warner's broader view in section 4.

2. A quantitative history

The historical facts relevant to negative and interrogative sentences can be summarised as follows (I follow the account in

Denison (1993:255f)). The `periphrastic' (semantically empty) DO appeared first in the 13th century, but at that time it had no special connections with negation or inversion; nor, indeed, were these restricted by rule to auxiliary verbs. The link between auxiliary-hood and these sentence-types appeared first in the 15th century, and had almost reached its modern solidity (with some differences of detail) by the end of the 16th. The changes shown in Table 1 were completed within two centuries, and our challenge is to explain not only why they happened at all, but also the way they spread through both the grammar and the population.

First, the data. For each of the sentence-types shown in Table 1, Kroch quotes both the total number of textual examples counted in which DO might have occurred (N) and the percentage of these in which DO was used. His raw numbers (based on Ellegård 1953:159)[2] are repeated in Table 2.

[Table 2 about 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. 1, 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. He offers an explanation for the change which we shall consider in the next section.

[Fig 1 about here]

One of the most striking features of Fig. 1 is that the five lines appear to form two clusters. The lines for negative questions and transitive (positive) questions are consistently higher than the other three, and by the mid-16th century the two clusters are very clearly separated. I shall call the two clusters the `high-do' and `low-do' constructions. How seriously should we take the apparent difference shown in this graph? One way to test its significance is by testing the statistical significance of differences between the constructions at each point in time. For example, in 1400 11.7% of negative questions contained DO, whereas 0% of transitive questions did. This percentage difference is enough to separate the two lines in Fig. 1, but when we look at the raw figures we find that they are based on a total of only 17 negative questions and three transitive questions - i.e. 2/17 negative questions had DO, compared with 0/3 transitive questions. It is easy to test the statistical significance of this difference (by the Chi-square test). It is not at all significant - the probability (p) is over 0.5, meaning that such a difference is likely to be found purely by chance in over 50% of random samples. It is normal to set a threshold of 5% ($p < 0.05$) for statistical significance, so the question is which of the differences shown in Fig. 1 are significant in this technical sense.

In a nutshell, the answer is that (with a handful of exceptions which we shall consider below) the only significant differences are those between our high-do and low-do groups. More precisely, at least one high-do construction (transitive question) is almost always significantly different from at least one low-do construction (intransitive question); the two high-do constructions are almost never significantly different from each other; nor are the three low-do constructions. Fig. 2 shows the much simpler picture which results if we ignore all the differences which are statistically below the threshold of significance. Clearly, we are dealing with two groups of concomitant changes, rather than with a single one; so any explanation must be able to explain why the changes are grouped in this way.

[Fig. 2 about here]

As mentioned above, Fig. 2 is actually too simple. The statistical analysis reveals two exceptions.

■ In 1500-25, transitive and intransitive questions fall together, each leaving the group that it belongs to at other periods - i.e. DO is significantly rarer in transitive questions than in the other high-do construction, and significantly more frequent in intransitive questions than in the other low-do constructions.

■ In 1525-50, Wh-object questions become significantly rarer than the other low-do constructions.

These two aberrations are shown in Fig. 3. If we use statistical analysis at all, we must respect it in all cases. A complete explanation for the historical changes that led eventually to the relatively simple two-way split in 1550-75 will have to explain these details, but I admit immediately that I have no such explanation.

[Fig. 3 about here]

To judge by these graphs, the high-do contexts had a growth-spurt around 1500 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 two 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. The challenge, therefore, is to explain first, what this change consisted in, and second, to explain why it affected negative and transitive questions before all the other sentence-types.

3. An explanation in terms of the verb-raising parameter

Kroch takes the statistical linkage between the various changes 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.

This unity he finds in a single Chomskian parameter which has many ramifications, so that the transition period covered by his data involved competition between two distinct grammars.

The parameter in question 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. 4 shows the basic sentence-structure which he assumes.

[Fig. 4 about 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. Without verb-raising, neither inversion nor negation is possible, so DO (which can occur at I) must be supplied, whereas (according to Kroch (ibid:192)), auxiliary verbs are exempt from the loss of verb-raising, which is why they do not need DO.

One of the attractions of the verb-raising analysis is that it also explains a change in the position of adverbs. Kroch quotes the following example to illustrate the change:

(6) a (Middle English) Queen Ester looked never with swich an eye.

b (Modern English) Queen Ester never looked with such an eye.

If we assume that adverbs such as never are generated as adjuncts on the left edge of the VP, then verb-raising would move looked past never; so the loss of verb-raising would explain why never cannot follow the verb in Modern English. Kroch quotes further figures from Ellegård (1953) which confirm that this change took place at roughly the same time as the rise of auxiliary DO, and that it spread at roughly the same rate as the other changes. More precisely, the pre-verbal position of never arrived about a generation before the use of DO in high-do contexts as can be seen from Fig. 5.

[Fig. 5 about here]

Most of Kroch's paper deals with the mechanism, called Grammar Competition, which drove our linguistic ancestors to sort out the conflict between their two grammars. 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 does trigger 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 INFL, why did it affect 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; and similarly for the various sentence-types which were affected. Kroch recognises this problem, 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. The problem is as follows. Whatever the influence of processing may be, this influence is presumably constant across time; so the only possible source of variation through time is the number of speakers who have the alternative grammars. Accordingly, the model predicts that the relationships among the sentence-types should stay constant through time; but this is not what we find.

For example, take the relationship between preposed adverbs and negative declaratives with DO. In 1525 the respective figures are 89% and 8%, but 25 years later it was about 90% and 43% - a very different relationship. Putting this another way, let's assume that the order adverb-verb presupposes a non-raising grammar; therefore anyone who uses this order must already have such a grammar. Given the figure of 89% adverbs preposed in 1525, at least this proportion of the population must have had non-raising grammars by that date. Therefore the much lower figure of 8% for negative declaratives with DO must be due to some processing consideration which makes verb-raising in this sentence-type about ten times as likely as adverb-preposing. But if this processing consideration is constant it should still have roughly the same effect 25 years later. At this time the number of people with a non-raising grammar is roughly similar to the previous period (when almost everyone already had one); therefore the proportion of negative declaratives with DO should also be roughly the same. But it is not - it is five times greater.

The same problem arises with any pair of syntactic constructions. If their differences are due to processing factors, they should stay constant; but they don't. What we observe is much more suggestive of a change which spreads through the grammar construction by construction, with one group of constructions setting the pace and other constructions following after. As with lexical diffusion, the change spreads through both the grammar and the population at the same time, so the figures that emerge from a corpus of texts produced by different people reflect both these changes.

■ 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.

In other words, at least some of the variability is due to code-switching. But the codes concerned are grammars which are identical except for one parameter, the presence or absence of verb-raising. How can we distinguish empirically between the predictions of this claim and of one in which the trigger of verb-raising is optional? And if we can distinguish the two, how plausible is Kroch's view? At present its only support is its compatibility with Principles-and-Parameters theory, so those of us who question that theory need more empirical evidence.

■ The explanation depends crucially on the the verb-raising analysis, but the empirical and theoretical underpinnings for this analysis are weak. Some of these weaknesses have been pointed out by others (notably Kim and Sag 1996), so I shall not try to repeat the exercise. The devil lies in the detail of such analyses, and one of the problems of verb-raising lies in dealing with all the details. Kroch discusses a lot of details in the paper quoted (1994), but there appear to be some important inconsistencies. First, if auxiliary verbs (HAVE, BE and the modals) are verbs which continue to be raisable, how can verb-raising be controlled only by INFL (as Kroch claims)? According to this analysis, INFL either raises all verbs, or it raises none of them, so there is no way to accommodate lexical exceptions. And second, Kroch suggests (195) that verbs and nouns may always be 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.

■ According to Kroch verb-raising had disappeared from English by the end of the 16th century (`.. 1575, the date at which V-to-I movement is definitively lost' - ibid:181). This assumption flies in the face of a variety of facts, not least the figures which he himself quotes showing that most low-do constructions still managed without DO in 1575. Furthermore, according to Warner (1993:66) subject inversion was still common with some full verbs (e.g. MEAN, SAY, THINK) through the eighteenth century, so if subject inversion did involve verb-raising this would suggest a much more gradual elimination than Kroch's estimate. Even in some varieties of Modern British English we still have the possessive auxiliary HAVE as in Have you a car?, which seems to call for a raising analysis although it is also a transitive verb. The problem is well known (Pollock 1989:407), but still lacks a convincing solution (as argued by Kim and Sag 1996).

■ Kroch explains the drive to eliminate verb-raising by claiming that syntactic variation is always resolved (by `Grammar Competition') definitively in favour of one pattern. This general claim does not seem to be borne out by some well-known cases. For example, he quotes Taylor's (1994) 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 1996; Lascaratou 1989).

In conclusion, Kroch's explanation leaves enough important ends dangling to suggest that we should look for an alternative explanation. A common thread runs through all these criticisms, namely the need to locate changes within the details of the grammar, in relation to individual constructions or lexical items, rather than to aim at a single global variable. The next section will develop this theme.

4. An explanation in terms of word-classes

An alternative way to explain Kroch's data is in terms of the word-classes `auxiliary verb'[3] and `full verb'. In this explanation I shall be largely following Warner (1993), supplemented by relevant sections of Denison 1993. Like Warner and Denison, I believe the major change that affected English in the early modern period involved the distribution of grammatical characteristics between these two subclasses of verbs.

Warner's account of developments in this part of English grammar is very different from Kroch's. The unifying factor, which may 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 3 to the modern list in Table 4. (The numbers after the examples refer to the pages where Warner quotes them; the sign ≤ means `precedes'.)

[Table 3 about here]

[Table 4 about 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 4 shows roughly when the new form or restriction entered the grammar. The dates show that the development of the auxiliary class is spread over at least three centuries, with Kroch's changes as just one segment of the development. It is 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 way we classify human behaviour such as language, unlike our classification of everything else, is not mere classification, but affects the material classified. 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 4 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 (e.g. We have never eaten pears) which was previously shared by all verbs (We ate never pears), 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 collection of historical changes as the development of the auxiliary class; it could just as accurately be described as the development of the full-verb class. Indeed, this would be a more accurate description because it would reflect an interesting historical generalisation that emerges from Table 4, namely that full verbs changed before auxiliaries did. All three changes covered by Kroch's discussion involved the relations between full verbs and some other word in the sentence (an adverb, the subject and the negative marker); but all the subsequent changes involved the characteristics of auxiliary verbs. As far as the earlier changes are concerned, it is auxiliaries rather than full verbs that are in direct line of descent from OE verbs; but the reverse is true for the later changes. Consequently, rather than seeing the change as the rise of the auxiliary class, it would be better to see it as 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 can now consider some of the questions that are raised by the historical data discussed earlier, starting with Kroch's constant rate hypothesis - why did DO change from optional to obligatory at the same speed in all different sentence-types, and at the same speed that adverbs changed position? 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, 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 5 (which also reached roughly the modern state of mutual predictiveness by the end of the 16th century - Warner ibid:200ff):

[Table 5 about here]

However, the membership of this class has varied:

(7) Old English modals (a tentative list following Warner 153)
CAN, MAY, SHALL, WILL, MOT (`may, must'), OWE, þARF (`need'), UTON ("let's"), DARE;
marginal members lacking one characteristic: BE, WEORðAN (`become')

(8) Modern English modals
MAY, SHALL, WILL, MUST
marginal members (see Table 5): CAN, DO, BE + TO
currently changing (back) into full verbs? OUGHT, USED, DARE, NEED

(We shall discuss the marginal and changing members of the modern class in the last section.) 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. 6. The main change in late Middle English was the rise of auxiliary DO in the 13th 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.

[Fig. 6 about here]

We can now suggest a reason for the constant rate of the various changes. Once a change has become associated with members of one or the other of these word-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 a similar range of speakers. We can imagine it starting as an idiosyncratic characteristic of a handful of verbs, used by a handful of speakers, which eventually generalises to whichever class these verbs belong to, in the usage of all speakers. Of course it could be objected that it is easy to imagine different changes progressing at different rates even if their routes take them eventually through the same territory; for example they could start with different social groups whose prestige drives the changes at different speeds. However, the forces that drive change are so complex that variation in one area is likely to counterbalance variation in other areas, resulting in roughly the same speed of change overall. If this is so, then we might expect to find less constancy when we consider the details - and this is in fact precisely what we did find in Fig. 3. For example, in the period 1525-50 DO did not become any more common in Wh-object questions than it had been during the previous period, though it was increasing steadily in all the other sentence-types during that period and though it caught up in Wh-object questions in the following period. We must bear this erratic behaviour at the `micro' level in mind when trying to explain the generally systematic macro-level progress of changes.

Another major challenge is to explain why the change passed through the sentence-types in the order that we have observed. Starting with the earliest change, what did the position of adverbs have to do with the other changes? A functional link is obvious between adverb-placement and subject-verb inversion. Suppose that the function of subject-verb inversion is to make the finite verb, as bearer of tense and polarity, into the first word in the sentence (Halliday 1985). What if the finite verb is modified by an adverb? In the earlier period, when adverbs followed verbs, the presence of an adverb was irrelevant to subject-verb inversion, but after the change in adverb position, the adverb will prevent a full verb from occupying the first position:

(9) a They never eat pears.

b Never eat they pears?

In contrast, an auxiliary allows the adverb to remain attached to the second verb:

(10) a They do never eat pears.
b Do they never eat pears?

The presence of DO allows the finite verb to be in first position while still allowing the adverb to precede the verb it modifies. At the time when DO was simply optional, and made no difference to the meaning, it provided the perfect solution to this problem. The problem may not have arisen very often, but even occasional examples like these would be enough to establish at least a statistical linkage between DO and inversion.

Now we turn to the sentence-types without adverbs. Why were negative and transitive questions affected first, almost a generation before the remaining sentence-types? The 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 tendencies which could have had quite minor beginnings - 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 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.

Let us start with the link between inversion and transitivity. In this case I agree with Kroch, who offers a functional explanation (1989). His suggestion is that when `periphrastic' DO became available, its use helped the hearer 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; specifically, at that time auxiliary verbs were more likely than full verbs to be accompanied by *not*. This trend may 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 conducive 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 + *not* + subject + V' over the syntactically simpler `V + *not* + subject'.

These explanations are functional; they invoke the benefits to the speaker and/or hearer of preferring one of the available options to the other. However they presuppose the existence of a meaningless and optional auxiliary verb (DO), which is the result of a previous structural change. Moreover it is important to consider the changes from the point of view of their effect on the grammar, which is also a matter of structure. We can now distinguish two stages in the history of this part of English grammar. At stage A, inversion and negation were equally possible for all verbs. Then our functional pressures were grammaticised, which led to restrictions on the inversion of full verbs, but made no change to auxiliary verbs at all. The grammars contained the following rules:

(11) Stage A (as in early Middle English)

a. You may invert any verb, whether auxiliary or full.
b. You may negate any verb, whether auxiliary or full.

(12) Stage B.
a. You may invert any verb, whether auxiliary or full; but:
a'. Do not invert a full verb if it is negated or has an object.
b. You may negate any verb, whether auxiliary or full

The next change generalises the constraints on full verbs by preventing all inversion of a full verb. The easiest way to do this is to replace the general rule plus restriction by a single less general rule applying just to auxiliary verbs rather than to verbs in general. This change is not driven by function, but by cognition - the desire for a simpler grammar. Similarly, the ban on inverted full verbs which prevents them from being negated is generalised to all full verbs, and re-expressed as a negation rule which only applies to auxiliary verbs. The result is stage C.

(13) Stage C (as in Modern English)
a. You may invert any auxiliary verb.
b. You may negate any auxiliary verb.

Stage B may therefore be seen as a cognitively complex grammar driven by functional pressures which leads from the cognitively simpler stage A to the cognitively simpler stage C. Stages A and C both have the same range of rules for inversion and negation, but they apply to different word-classes (`verb' in stage A, `auxiliary verb' in stage C).

If this account is right, then the Constant Rate hypothesis becomes even more interesting. When we first distinguished between `high-do' and `low-do' sentence-types, we found that, although they started to favour DO at different dates, the overall change probably took place at the same speed in both types of sentence. But what I have just suggested is that it was functional pressures that drove the introduction of DO in high-do contexts but cognitive pressures that drove it in the low-do contexts. If we put all these facts, claims and ideas together, we may be able to conclude (very tentatively) that functional and cognitive pressures are roughly equal in strength.

5. The recent history of auxiliary verbs

The story of auxiliary DO does not end with the 16th century. As can be seen from Table 4, Warner lists four[4] other distinctive characteristics which developed in the 16th or 17th centuries:

(14) a cliticization (is ~ 's, will ~ 'll, etc.)
except: OUGHT, USED, DARE, NEED
b tag questions (..., isn't it?, etc.)
c exclusively VP complements
except: $\text{HAVE}_{\text{possess}}$, BE
d not reduced to suffix n't (isn't, etc.)

Why should these extra characteristics have developed? The verb-raising analysis does not help at all, but a cognitive perspective does suggest a very tentative explanation. As Table 3 showed, the word-class `auxiliary verb' was distinguished in Old English only by a small handful of characteristics:

(15) a VP ellipsis and pseudo-gapping (e.g. modern ...but it would me after It didn't worry her)

b transparency/raising (e.g. hine (acc) sceal ..gesceamian, `He shall be-ashamed')

c uninflected 3rd singular (e.g. he sceal, `he shall')

d negative in n- (e.g. nylle, `don't want')

These characteristics were only rather loosely related to one another, as at least one (b) was also true of some non-auxiliary verbs, and another (c) was not true of all auxiliaries (the exception being BE). The last characteristic was particularly undistinctive, as proclitic ne was the regular negative marker even with full verbs, and it did not contract to n- before all auxiliaries (Denison 1993:449). In any case, this characteristic disappeared altogether at the end of the Middle English period when ne was abandoned in favour of post-verbal not (ibid:450). In short, the category `auxiliary' did not predict many characteristics in late Middle English. In terms of Rosch's principle of cognitive economy, it was a rather uneconomical category. In fact, it would be quite easy to write a grammar of Old or Middle English in which the category `auxiliary verb' played no part at all, because all the characteristics in (15) could easily be treated as properties of the individual lexical items concerned.

The developments in late Middle English changed the picture completely. According to the analysis which I suggested in the last section, by stage B the distinction between auxiliary and full verbs played a crucial role in the grammar as the basis for deciding whether or not to use auxiliary DO in a question. This role was crucial because the new stage-B rules blocked subject inversion with any full verb that was negated, a general constraint which could not conceivably be replaced by a collection of rules referring to specific lexical items. These constraints on inversion (given as (a') in (12)) greatly `strengthened' the category `full verb' and also, (by contrast), `auxiliary verb'. At stage C this strengthening remained, though transferred to `auxiliary', and if anything it was increased by the greater generality of the rules concerned.

It is clear (and presumably uncontentious) that the subsequent history of the auxiliary verbs has made `auxiliary' verb a more effective category, in Rosch's terms. Each further characteristic that was added provided another feature that could be predicted from this category; and the more harmoniously the various features defined the same membership, the better the category was. I have already mentioned five new features that have been added, but we should also recognise various changes that are still taking place, and which all promise to increase the harmony among the features by removing exceptions. The following list is probably incomplete.

■ The famous possessive HAVE is still an ordinary auxiliary for many speakers in the UK (Trudgill 1978:14f), but it is exceptional in its valency (taking an object rather than a VP complement). This usage persists in most of the UK giving forms like the following:

(16) a They've a car.

b Have they a car?

c They haven't a car.

Meanwhile the USA and the south of England have generally replaced the `possessive auxiliary' either by an ordinary full verb, as in (17), or by HAVE GOT, in which the auxiliary-hood of HAVE is separated from the transitivity of GOT as in (18):

(17) a They have a car.

b Do they have a car?

c They don't have a car.

(18) a They've got a car.

b Have they got a car?
c They haven't got a car.

The alternatives make the future of possessive auxiliary HAVE uncertain. The loss of this transitive auxiliary can be seen as the last battle in the war against transitive auxiliaries which has been taking place since the 17th century (Warner ibid:202), when WILL stopped allowing an object (e.g. I will an apple).

■ USED and OUGHT are both exceptional in taking TO, rather than a bare infinitive; and neither of them has an abbreviated form. They appear to have become full verbs for many speakers.

(19) a Did they used/ought to sell it?
b They didn't used/ought to sell it.

■ DARE and NEED are exceptional in allowing their subject to have a semantic role (unlike the usual `raising' semantics of auxiliary verbs); and again, neither has an abbreviated form. They both have synonymous and homonymous full-verb counterparts which take TO, and they are both severely restricted in distribution to the same `affective' contexts as negative polarity items (which may be a generalisation from the inverted and negated contexts peculiar to auxiliaries). Furthermore auxiliary NEED has no past tense (compare He needn't with *He neededn't). In short, these auxiliaries are severely restricted and becoming more so, and their eventual replacements are already in use.

(20) a *She dare/need jump.
b Dare/need she jump?
c She daren't/needn/t jump.

■ As usual, non-standard dialects have taken the tidying-up process further than standard English. Most auxiliary verbs are modal verbs, which have present-tense forms without -s even when used with third-person singular subjects; so we might expect this characteristic to be generalised to the non-modal auxiliaries. In standard English BE, DO and HAVE maintain the -s (is, does, has), though it is noticeable that they are almost the only verbs in the language (alongside says) whose -s-form is irregular. However, in some non-standard dialects the present-tense paradigm has been simplified on the model of the modal verbs. This is especially true in the negated forms, where ain't doubles up for both isn't and hasn't, as well as for aren't and haven't as in the following example produced by a speaker from Reading, England (Freeborn et al 1986:148):

(21) ..she ain't got to bother have she?

Invariant auxiliary don't is used even more widely in the UK, including London, and indeed until this century it was standard in the casual speech of upper-class and middle-class society; Denison (forthcoming) quotes the following two sentences from a letter by Queen Victoria:

(22) Alix don't like her. ...... Alix does not sleep well.

Most of the changes reviewed in this section increase the harmony among the various characteristics of auxiliary verbs. Put another way, they all have the effect of reducing the number of exceptions to any rule which expresses one of these characteristics as a generalisation about auxiliary verbs. This is simply a matter of fact, though there is room for disagreement about the details of the changes. We are on less certain ground when we ask why the various harmonizing changes have happened, but I think most people would agree that the explanation must be a cognitive one - the changes make the grammar simpler by removing exceptions and increasing harmony (which is the same thing). However the cognitive pressure towards harmony must interact with functional pressures to increase (or maintain) expressive power: functional pressures keep us busy in the workshop while cognitive pressures make sure we tidy up the mess that we make. The following speculations explore this interaction.

We started this section with four changes that have affected auxiliary verbs since the 15th century. One of these (the loss of transitive auxiliaries) can be seen as a tidying-up operation to increase harmony. Another (the use of tag questions) presumably had some kind of functional motivation since it increases the range of expressible meanings; but it was cognitively tidy to recycle the auxiliary verbs as bearers of inversion and negation.

This leaves two changes to explain, both involving phonological reduction - the reduction of the auxiliary itself, and that of the negator not to n't. As speakers we presumably minimize our effort, which means that we favour phonological reduction, but we also minimize misunderstandings that might arise from too much reduction; these functional pressures are always with us, so the question is why reduction triumphed just in the case of auxiliary verbs, and only in the last few centuries. Why, for example, do we reduce had to 'd when it is an auxiliary but never (at least in Standard English) when it is a full verb?

(23) a I had/I'd finished.
b I had/*I'd a bath every Friday.

From one functional point of view, reduction is still costly. Why do we put up with the homonymy produced by reducing is and has to 's, or had, should and would to 'd? Why do we demote the sole signal of negation from a full syllable to a mere syllabic consonant, and put up with the awful problem of distinguishing can and can't? However this is only one part of the functional view, and these costs only arise on those rare occasions where the context allows ambiguity. In other respects the reduction has two important functional benefits which outweigh the danger of misunderstanding. Reduction signals two important bits of information, one grammatical and the other sociolinguistic, which may explain the timing and restrictions on the rise of reduction.

Grammatically, reduction signals an auxiliary verb with all the other characteristics that this word-class entails. The grammatical benefit of being able to identify an auxiliary verb by its phonological form is proportional to the amount of grammatical information that this identification allows us to infer, so reduction becomes really profitable only in the 16th century, after the major changes of the previous two centuries; and of course the benefits of this signal are also proportional to its reliability, which is why it is reserved so exclusively for auxiliary verbs. (In this case the functional and the cognitive are in complete harmony because the functional pay-off depends on the effectiveness of the cognitive tidying-up which made all those auxiliary characteristics predictable.) Another functional benefit of reducing not to n't has been that two syntactic words have been collapsed into a single one which can jointly signal both interrogative inversion and negation as in Haven't you finished? as opposed to Have you not finished? Interestingly, these functional pressures are precisely the same as the ones which produced the first high-do context, negative questions.

The second factor that may have encouraged phonological reduction is the sociolinguistic need to signal casual style. Reduced forms are very clearly linked in Modern English to casual style, though (as always) the linkage is probabilistic rather than categorial (Labov 1969). Impressionistically, it seems that the use of reduced forms ranges between two extremes: zero use in the most formal and public written documents, and almost obligatory use in the most casual and intimate speech. It is important to remember that the new reduced forms in no way threaten the older unreduced forms, all of which survive as alternatives; since they are exact synonyms, Kroch's Grammar Competition principle would predict that competition would by now have eliminated one of them unless they each have a distinctive social role, as they clearly do. Their present-day sociolinguistic content is clearly crucial, so we must ask what part it may have played in the rise of reduced forms. Why were auxiliary verbs selected for this role, and why did the role develop only in the 16th century? My guess is as follows.

By the end of the 16th century auxiliary verbs had become more common, as tokens in speech, than they had ever been before because of the ban on inversion and negation of full verbs; in other words, their previous numbers were boosted by DO. The more frequently a linguistic variable occurs, the better it is for signalling subtle sociolinguistic information. Moreover, to judge by Japanese and Basque (Hudson 1996:131), the root verb of a sentence seems to be a particularly good place for locating such information; and the positions where auxiliaries are now required are especially likely to be the sentence's root. These two facts may explain why auxiliaries were chosen, but why is it so important to distinguish casual and formal style? Let's start with the present day. To use unreduced auxiliaries in ordinary conversation is to be accused of sounding pedantic or bookish, which most of us try hard to avoid. In the modern world the distinction between reduced and unreduced auxiliaries has a very clear social function of distinguishing symbolically between different `domains' of life. It is tempting to speculate that this social distinction only started to have widespread significance with the rise of a large middle-class society involved in both private and public domains - a very different kind of functional pressure.

6. Summary

We have considered a number of well-documented changes to the English verbal system which spread over many centuries, and which are still in progress. For these changes I have offered a structural analysis in terms of changes in the grammar, and specifically changes to the `content' of the word-classes `auxiliary verb' and `full verb'. According to my analysis, these two classes have become far more clearly differentiated than they used to be: various rules that used to apply to all verbs now apply only to auxiliary verbs, a number of new patterns have developed which also apply only to auxiliary verbs, and DO is now available only with full verbs. This gives a series of structural changes, but that was not the end of the story; it continued into speculations about the possible reasons for the changes. I suggested two kinds of cause: functional pressures to make communication easier and more efficient, and cognitive pressures to increase `harmony' and `tidiness' in the system. In some cases various pressures of both kinds supported each other, but in other cases they were in conflict.

The following summarises the story:

(24) a Auxiliary DO is introduced, allowing the option of using an auxiliary without changing the meaning.

b Adverb-preposing make subject-verb inversion awkward for verbs modified by adverbs, so auxiliary DO comes to the rescue.

c Further functional pressures exploit auxiliary DO to help speakers to avoid ambiguities in questions that contain an object, and to put the markers of questioning and negation near to each other. These pressures are grammaticised as constraints on full verbs in some questions, giving stage B of (12).

d Cognitive pressures for simplicity generalise these constraints to all full verbs, and re-express them as positive rules referring to auxiliary verbs, giving stage C of (13).

e Cognitive and functional pressures (including sociolinguistic pressures) combine to make this newly-enriched category more easily recognisable by allowing auxiliaries alone to be reduced to clitics and to take reduced n't.

f Cognitive pressures for simplicity and harmony have removed some exceptions, and are still removing others, thus tidying up the effects of earlier changes.

Dept of Phonetics and Linguistics,
University College London,
Gower Street,
London, WC1E 6BT

REFERENCES

Bresnan, Joan (ed.), 1982. The Mental Representation of Grammatical Relations. Cambridge, MA: MIT Press.

Chomsky, Noam, 1986. Barriers. Cambridge, MA: MIT Press.

Denison, David, 1993. English Historical Syntax. London: Longman.

Denison, David, forthcoming. Syntax. In Suzanne Romaine (ed.). The Cambridge History of the English Language, IV. 1776 - Present Day. Cambridge: Cambridge University Press.

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

Freeborn, Dennis, 1986. Varieties of English. An introduction to the study of language. London: Macmillan.

Halliday, Michael, 1985. An Introduction to Functional Grammar. London: Arnold.

Hudson, Richard, 1990. English Word Grammar. Oxford: Blackwell.

Hudson, Richard, 1996. Sociolinguistics. Second edition. Cambridge: Cambridge University Press.

Hudson, Richard, forthcoming. Inherent variability and linguistic theory. Cognitive Linguistics.

Kim, Jong-Bok and Sag, Ivan, 1996. French and English negation: a lexicalist alternative to head-movement. mimeo.

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

Kroch, Anthony, 1994. Morphosyntactic variation. Chicago Linguistics Society Parasession 30, 180-201.

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

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

Pollard, Carl and Sag, Ivan, 1994. Head-driven Phrase Structure Grammar. Chicago: University of Chicago Press.

Pollock, Jean-Yves, 1989. Verb-movement, universal grammar and the structure of IP. Linguistic Inquiry 20, 365-424.

Taylor, Ann, 1994. The change from SOV to SVO in Ancient Greek. Language Variation and Change 6, 1-38.

Trudgill, Peter (ed.), 1978. Sociolinguistic Patterns in British English. London: Arnold.

Tzanidaki, Dimitra, 1996. The Syntax and Pragmatics of Subject and Object Position in Modern Greek. London University PhD.

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

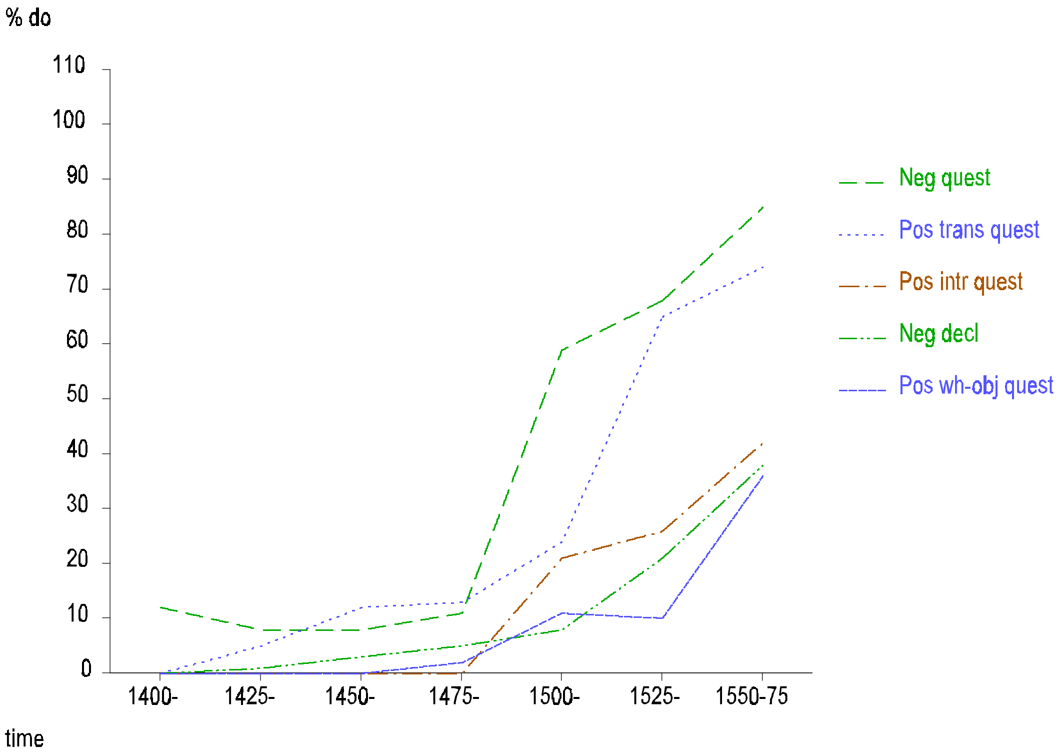
% do
110
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
Neg quest
Pos trans quest
Pos intr quest
Neg decl
Pos wh-obj quest

Fig. 1.

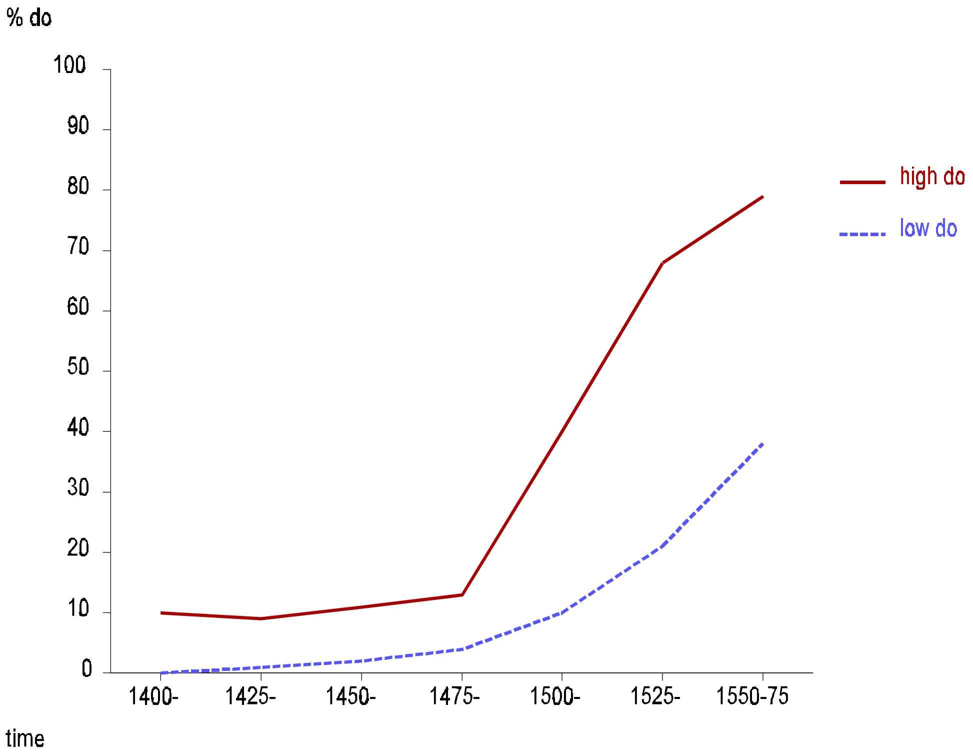
% do
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
high do
low do

Fig. 2

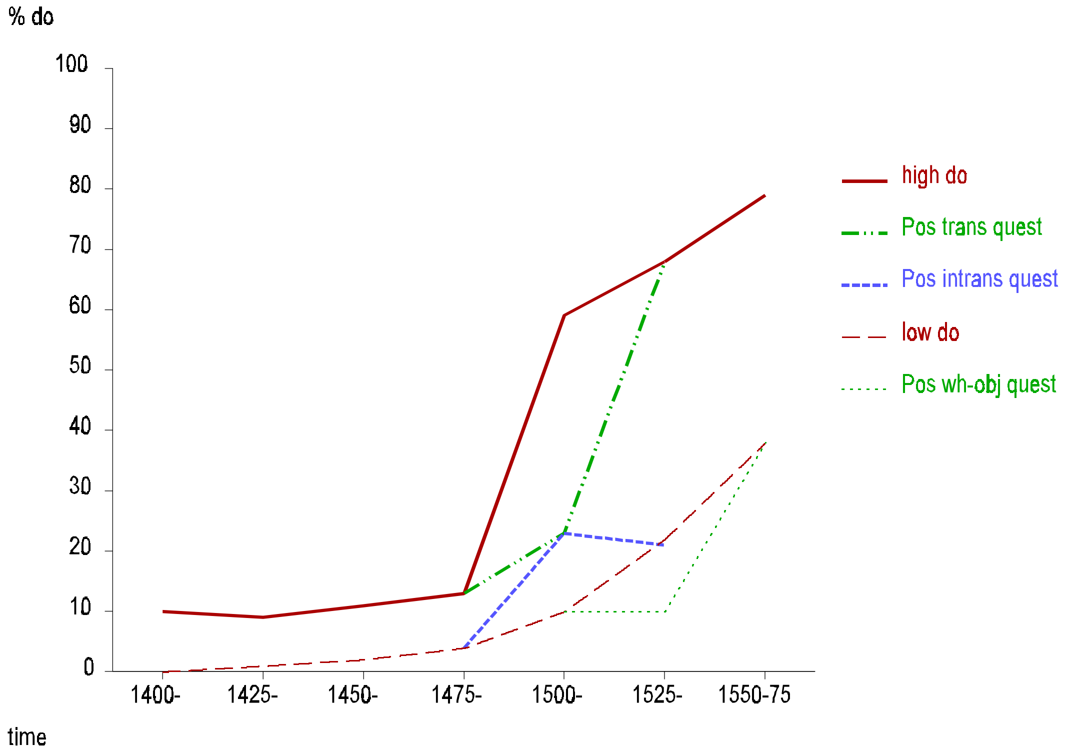
% do
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
high do
Pos trans quest
Pos intrans quest
low do
Pos wh-obj quest

Fig. 3

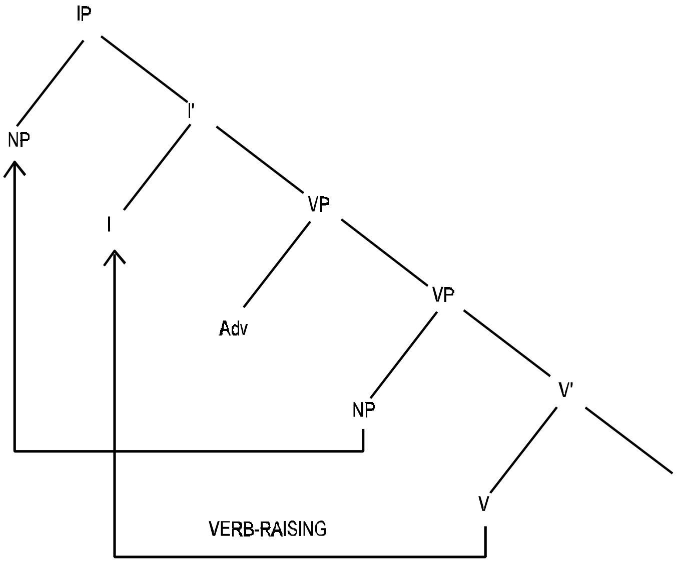

Fig. 4

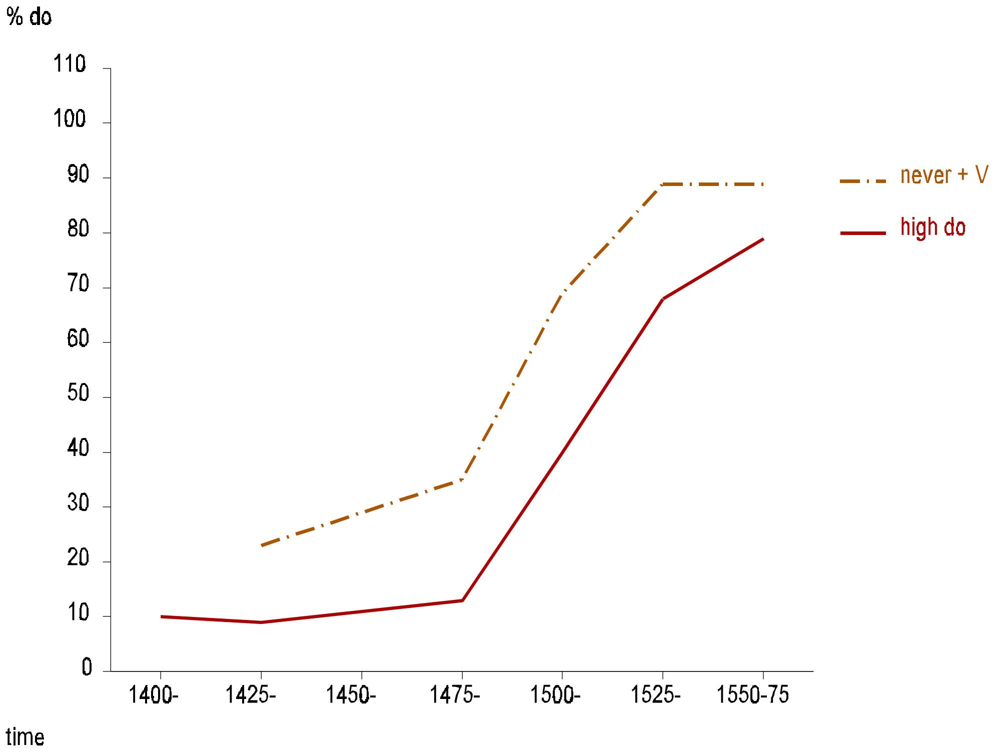
% do
110
100
90
80
70
60
50
40
30
20
10
0
1400-
1425-
1450-
1475-
1500-
1525-
1550-75
time
never + V
high do

Fig. 5

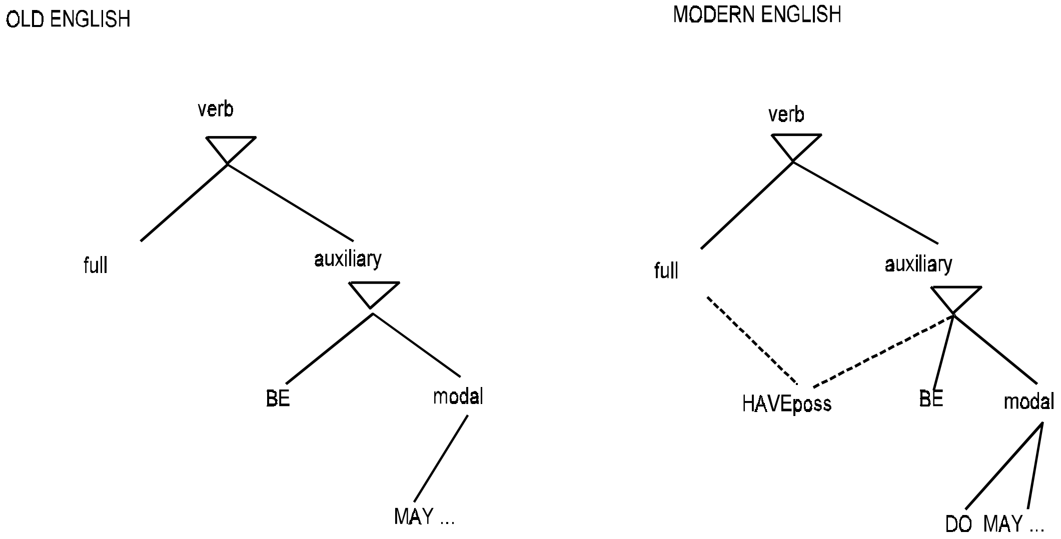
OLD ENGLISH
verb
full
auxiliary
BE
modal
MAY ...
MODERN ENGLISH
verb
full
auxiliary
HAVEposs
BE
modal
DO
MAY ...

Fig. 6

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

Table 1.

Time		Negatives				Positive 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 2.

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)
uninflected 3rd singular	[see previous example]
negative form in n- (also: WITAN, `know', HABBAN, `have')	nylle `don't want', nam `am not', næbbe, `have not' (151)

Table 3.

Distinctive characteristic of auxiliary verbs	exceptions	example	date	page	changed
V allows VP ellipsis and pseudo-gapping		.. It would _ me _.	OE	111	-
V < adverb		*ran never	15c	206	full
V < subject		*ran you?	15c	220	full
V < not		*ran not	15c	215	full
V reducible to clitic	OUGHT, USED, DARE, NEED	's going	16c	207	aux
V in tag question		.., is he?	16c	207	aux
V allows VP as complement	HAVE$_{poss}$, BE	*will coffee	17c	202	aux
V not ~ V-n't		isn't	17c	208	aux

Table 4.

characteristic of modal verbs	modern exceptions
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 5.

1. The main ideas in this paper were presented publicly at seminars at UCL and Oxford University, and at the workshop `English Historical Syntax: What now?' in Manchester in May 1996. I should like to acknowledge comments from the participants of all these meetings, and in particular from Anthony Warner, David Denison, Robert Stockwell and an anonymous referee who all commented on an earlier written version. The same ideas and some of the same material are included in a more wide-ranging discussion in Hudson (forthcoming), which argues more specifically for Word Grammar as a suitable theory of language structure for explaining both synchronic and diachronic variation.

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.

3. The term `auxiliary verb' should be treated with care; I myself prefer the name `polarity verb' to avoid implying that auxiliary verbs by definition `help' other verbs. A verb can be an auxiliary verb even if it is the only verb, overt or understood, in the sentence. If inversion is possible only for auxiliary verbs, then is must be an auxiliary verb in Is Pat here?, and similarly for has in Has Pat a car?. Nevertheless, I recognise the strength of tradition, so I shall follow tradition in calling verbs like BE and WILL `auxiliary verbs'. I draw the line at the term `main verb', though, so I shall contrast auxiliary verbs with `full verbs'.

4. Warner actually lists a fifth characteristic, the development of lexical idiosyncrasies by individual inflected forms of auxiliary verbs; for example, he points out that the perfect participle of BE, unlike any other form of this lexeme, can mean `go' (as in We have been to Paris). However I am not convinced that this is a peculiarity of auxiliary verbs. Potential counterexamples are the GOT of the possessive HAVE GOT, the imperative and infinitive of BEWARE (which are the only possible forms) and the various verbs that are reputed to occur only as passives (such as REPUTE).