OG 2

**KonDoc - Bestellschein** **KOPIE**

**Bestellart:** **Lieferweg: EMAIL**

-

Eingang: 14.10.2013 - 17:00

Lieferung bis: **17.10.2013 - 16:00**

**Subito-BestNr.: KonDoc:2013101459272**

Kundennummer: K000011867

Bestellnummer: **E000154732**

Benutzernr.: **03/037078**

E-Mail:
frans.plank@uni-konstanz.de

PROF. Frans Plank

Sprachwiss.

**78464 Konstanz** **de**

| | |
|---|---|
| Signatur: | **spr 2/s81** |
| Titel: | Studies in language : internat. journal sponsored by the foundation "Foundations of Lang |
| Jahrgang/Heft:: | 8 |
| Erscheinungsjahr: | 1984 |
| Seiten: | 305-364 |
| Autor: | Plank |
| Artikel: | The modals story retold |

Weitere Angaben:

Bemerkung zum Dokument:

Kundennummer:


Bestellnummer

*Studies in Language* 8.3. 305-364 (1984). 

# THE MODALS STORY RETOLD

FRANS PLANK
*Fachgruppe Sprachwissenschaft*
*Universität Konstanz*

The following statement might well have deserved inclusion in Richard Hudson's (1981) meritorious recent collection of issues on which linguists can agree: "A definition of a possible grammar will provide the upper limits to the way in which a given grammar may change historically, in so far as it cannot change into something which is not a possible grammar." On the other hand, as Hudson says he is convinced that every one of the eighty-three agreements he decided to include in his list "has implications for some area of practical life" (1981:335) — of the world at large, but presumably of linguists too —, this particular omission perhaps should not be attributed entirely to his or his consultants' forgetfulness: it might have been intentional, motivated by the feeling that truisms of this sort are unlikely to have much practical, or other, importance.

Be that as it may, this statement is a quote from David Lightfoot's Foreword (p. viii) to his *Principles of Diachronic Syntax*, and Lightfoot, on the contrary, proceeds as if this were no truism at all but an insight of enormous practical and theoretical importance for diachronic syntax and grammatical theory alike. And he has his reasons. He maintains that a theory of grammar is available which is sufficiently restrictive to be of some diachronic relevance — so restrictive in fact that certain historical changes of individual grammars must be interpreted as attempts to steer away from the limits which the theory of grammar is setting individual grammars, once these limits have been approached or even infringed on account of prior changes. Not surprisingly, if one is familiar with Lightfoot's writings in the period following his generative-semantics euphoria, it is the Extended Standard Theory in the then current version which is claimed to be the best, i.e. most restrictive, theory of grammar available. It may seem unfortunate that Lightfoot's book

happened to be assembled while Chomsky and his Extended Standard disciples were once more about to change the rules of their game rather drastically. Consequently, some of Lightfoot's theoretical convictions and analytical suggestions already rang slightly démodé when his book appeared; and some developments (e.g. the 'modular' approach, the separation of sentence and discourse grammar, and of core grammar and peripheral rules) apparently were so brand-new that Lightfoot, himself in the van of this revisionary movement, did not manage or want to readjust the whole book for their sake, contenting himself with a rather cursory treatment instead. Ultimately, however, I think it would be wrong to conclude that adverse historical circumstances conspired to diminish the significance of Lightfoot's book as a contribution to diachronic syntax. I tend to believe rather that the diachronic significance of any transformational-generative foray into the field of language and grammar change, irrespective of whether the paraphernalia date from before or after Pisa, is bound to be negligible when it comes to matters of explanation: the kinds of restrictions on particular grammars in the investigation of which Chomsky Grammarians, of whatever vintage, have so far distinguished themselves are in my opinion most unlikely to be profitably invoked as reasons or causes, directly or indirectly, of particular diachronic developments. Having worked through Lightfoot's book, I see no reason to alter this opinion; on the contrary, I feel I can hold it with greater conviction than before. I hasten to concede, though, that this particular book strikes me as a most unfortunate choice if someone with similar persuasions really wanted to become a convert.

What Lightfoot sets out to demonstrate is this. There are changes in the grammars of individual languages which are entirely accidental; hence there is nothing about such chance developments which linguistic theory could possibly be required to explain. (Monod is a handy name to drop here, of course.) Other changes, however, happen of necessity, and these are to be explained, ultimately, by the theory of grammar (perhaps in conjunction with perceptual theory and other components of a general theory of mind) which predicts that the grammars of individual languages must change 'catastrophically' (in the sense of — guess who!) whenever they are, or have gotten into, particular states. There is no need or possibility of predicting anything about the ways and means of such necessary changes: in principle, anything goes in this respect. (Alas, Lightfoot has to miss this opportunity of dropping Feyerabend's name: Popper and especially Lakatos are his avowed philosophical favourites, at least in theory — in practice, in his manner of handling

empirical evidence, his inclinations seem to me definitely Feyerabendesque.) What needs to, and can, be predicted is merely that under these circumstances grammars somehow change into less offending states, the language resulting from such grammar changes remaining intelligible to speakers operating with a less advanced grammar. Specifying the states where particular grammars must be improved is obviously crucial, and this is the task of the theory of grammar. Quite simply, these offending states obtain when derivations are not minimally complex and when deep structures are not fairly close to their corresponding surface structures (approx. quote, again from the Foreword, p. viii). We are promised a 'Transparency Principle', which constrains derivations, and thus individual grammars, accordingly, in co-operation with the various other restrictions on the form and functioning of rules which are also part of the theory of grammar. Thus, Lightfoot seems to agree, interestingly, that the usual generative-transformational constraints per se very often are not restrictive enough to have much bearing on diachrony. But if so much explanatory weight is being attributed to a single additional constraint, one eagerly expects to learn more about it, its initial characterization in terms of minimal complexity of derivations and fairly close fits of deep and surface structures obviously being a little vague. And indeed, promises abound in the first third of the book that a Transparency Principle shall be "gradually developed" (p. viii), that it shall be given "some fairly precise content ... in subsequent chapters" (p. 121), that it "will need to quantify the amount of exceptional behaviour needed in order to bring about a re-structuring of the grammar" (p. 129), that the author shall "then seek to formulate a Transparency Principle which prescribes the tolerance level for opacity" (p. 135), and that his proposal of such a principle is "evidently falsifiable in principle" or "in essence" (p. 125), but "hitherto not actually falsified", in spite of being "a 'bold' or 'risky' hypothesis" (p. 125).

To be sure, there are linguistic historians who would be less modest than Lightfoot: rather than focusing exclusively on the problem of constraints, they would also seek explanations, from whatever source, concerning what have been called the problems of transition, embedding, evaluation, and actuation (cf. Weinreich/Labov/Herzog 1968); and they might also recognize as explananda changes which are not strictly necessary, without being entirely random — a category which is outside the purview of Lightfoot's explanatory aspirations (cf. p. 124). But instead of pondering whether Lightfoot's relatively agnostic overall policy about diachronic matters deserving some explanation constitutes a realistic assessment of our current explanatory

possibilities or a considerable step backwards theoretically as well as empirically (I lean towards the latter view), let us go along with him as far as his impoverished conception of a theory of change is concerned. Let us also refrain from debating at this point the value of pronouncements such as "communicability must be preserved between generations" and "grammars [or languages, or speakers — you choose; FP] practise therapy rather than prophylaxis" (p. 149), which almost exhaust the content of Lightfoot's theory of change, having been taken over, without much argument or elaboration, from linguistic folklore (presumably via Morris Halle and Elizabeth Traugott) and neogrammarian ideology, respectively. It is evidently the Transparency Principle upon which the success or failure of Lightfoot's endeavour is predicated. Considering how much Lightfoot appears to bank on this principle, the treatment it receives in the book is curious, to say the least.

The development of the English modals (Chapter 2, but summarized repeatedly later) serves as Lightfoot's paradigm case where the Transparency Principle can be seen in operation. I think one can only agree with Lightfoot that this particular well-documented and often-investigated development in the history of English is indeed highly instructive for anyone interested in how the immutable bears upon the transitory in grammar and lexicon, and in what the observation of the latter reveals about the former. The development of the English modals is a paradigm case of grammaticization, showing in an exemplary manner how more or less ordinary lexical items are appropriated for the grammatical system, with the linguistic forms involved being gradually adjusted to the functions that transparently motivate them (cf. Plank 1981, 1983). Considering its generally recognized significance for theories about grammaticization, the story Lightfoot tells about this development is again curious, to say the least. Oddly enough, this story has occasionally been mistaken for history, and has been accepted as a reliable descriptive account of the historical chain of events even by people more or less radically disagreeing with Lightfoot about its explanation (including Steele et al. 1981: 275-8, Lehmann 1982: 29-30). Moreover, even those reviewers who were not impressed by Lightfoot's Transparency Principle usually did not go to any lengths to question the verisimilitude of the story of its alleged prime manifestation. (Warner (1983: 194-200), which appeared after the present article had been virtually completed, did note a number of its major factual flaws, but managed to be impressed by Lightfoot's theoretical conclusions.) Nor, interestingly, did the modals come up for discussion in Lightfoot's (1981a) own valiant attempt to lay the phantom of Transpar-

ency. To facilitate future, and hopefully more responsible, theorizing about the development of the English modals in particular, and about grammaticization in general, the bulk of the present article is, therefore, devoted to chronicling the set of changes that demand an explanation. With the modals story retold, it will become evident that there still is such a demand: what Lightfoot claims to constitute an explanation at least is easily seen to fail on all counts.

There has been much controversy recently about the proper analysis of auxiliaries in Modern English, and also about the status of the notion 'auxiliary' itself in grammatical theory. Lightfoot does not really advance this discussion in any important way. Without going into details and without presenting full analyses, he assumes that the etymological ancestors of today's modals (called the premodals) were ordinary, complement-taking main verbs in Old and Middle English, requiring no special syntactic (initial or surface) category of their own. As to ModE, he argues, against Ross and others, that auxiliaries, and specifically modals, are not main verbs underlyingly, outside the sentence they end up in on the surface, but he does not address many of the controversial issues about what goes on between initial and surface structures in an optimal grammar of ModE. He assumes, at any rate, that an optimal grammar of ModE has to include an initial-structure category 'Modal' distinct from the category 'Verb' (and outside the Verb Phrase) to account for the no longer quite verb-like behaviour of the modals. It follows, according to Lightfoot, that this category 'Modal' was introduced between ME and ModE or in early ModE (to be more precise, in the 16th century, as Lightfoot then argues). In fact, 'Modal' is not the only initial-structure category newly introduced in early ModE times, in his view: the entire category 'Aux' supposedly dates from this period, a new phrase-structure rule expanding it into 'Tense' (not yet a separate initial-structure constituent in OE/ME) and, optionally, 'Modal'. Although Lightfoot's synchronic analysis of OE/ME/ModE, or lack thereof, would deserve closer scrutiny, let me merely register my basic agreement with the claim of a categorial change of the modals. I am not convinced that this categorial change is best interpreted primarily or exclusively as one of initial-structure categories, although I agree that there were concomitant morphosyntactic changes in the grammar of English going well beyond mere changes in the subclassification of certain verbs. The main focus of my critical attention will be on the overall scenario and on the explanation of the development of the modals suggested by Lightfoot, and I believe my criticisms would be compatible with a more surface-

structure-oriented view of the relevant changes.

Now, Lightfoot lists twelve changes affecting, or pertaining to, the verbs to become 'Modals'; where appropriate, I take the liberty of commenting on, and revising, his story. For one thing, I only count nine changes where he counts twelve.

(1) During the ME period the premodals generally lost the ability to take direct objects and finite-clause complements. Only *can* can still be found with ordinary direct objects as late as 1649 (*yet can I Musick too* (Lovelace)), not counting even later examples, which are, however, likely to be conscious archaisms. Says Lightfoot. But *can*, though being somewhat unusual in its subcategorization vis-à-vis other premodals (note, for example, that of the 22 occurrences of *cunnan* in Beowulf only 6 take an infinitive complement, whereas of the 83 occurrences of *magan* in Beowulf only 4 are without infinitive complement; similarly in other OE texts, according to counts in Standop 1957: 66), is not really uniquely exceptional in this respect: at least *will*, *would*, and *could*, and presumably *may* too, also continue to take direct objects in the 17th century, contradicting Lightfoot's claim that this particular change affecting the subcategorization of all premodals was completed in ME.

Moreover, it is misleading to imply that this development, allegedly affecting all premodals indiscriminately, took place regardless of the meanings with which they were used. In fact, nothing at all happened to premodals when they were used epistemically and perhaps also deontically, because the possibility of occurring with a direct object and no intervening infinitive had been limited to dynamic uses of premodals at all times. (Like for instance Palmer (1979), I prefer this three-way distinction of modalities, adapted from von Wright (1951), to the customary transformational two-way distinction of root and epistemic modalities. As to genuinely alethic modality, its linguistic significance seems negligible. An additional desiderative kind of modality, corresponding to the traditional 'will-mood', could be recognized, but I doubt that it would merit a status entirely separate from the other kinds, especially the deontic one.) Lightfoot, who is a believer in the autonomy of syntax, and accordingly maintains that the syntax of the English (pre-)modals changed quite independently of their semantics, does not attempt to make his grammar fragment responsive to the semantics of premodal - object combinations to begin with. Given a premodal and its object nominal, it used to be possible to infer particular verbal concepts, which in later English are

required to be present on the surface: 'play' or 'produce' rather than, say, 'hear' in the above Lovelace example, 'possess' or 'have' rather than, say, 'destroy' in *Wultu kastles and kinedomes?* (1225 Ancr. R.); 'do' or 'have' rather than 'forget' in *If it had beene the pleasure of him who may all things* (1597 Morley); 'have' rather than 'refuse' or 'kill' in *I will no reconcilement* and *I would no other company* (both from Shakespeare); 'read' or 'understand' rather than 'memorize', 'recite backwards' or 'carry' in *She could the Bible in the holy tongue* (1632 Ben Jonson). Note again how limited the range of modal meanings is in such infinitive-less constructions: even where non-dynamic meanings would be available if there were an explicit infinitive accompanying the premodal, no such epistemic or deontic interpretation is possible here (cf. Plank 1981). Independently of how such semantic facts are eventually handled in a grammar, I doubt that the loss of premodal-object constructions can be made sense of when seen in isolation from the semantic development of the (pre-)modals. The ability to take plain direct objects, without intervening main verb, would definitely seem to correlate with the presence of what is usually called 'notional' meaning, and consequently ought to disappear when a verb loses this kind of meaning, exchanging its lexical status for a grammatical one.

(2) With the exception of *willan*, itself an 'anomalous' verb always lacking the regular 3rd person singular fricative desinence, the premodals have always been members of a particular inflectional class, the preterite-presents, whose past tense (or rather perfective) forms had acquired present tense meaning long ago, without ever acquiring, however, the regular 3rd person singular present tense desinence. Significantly, this class grew much smaller during the ME and ModE periods. Some of the original preterite-presents joined other, more regular inflectional classes (e.g. *witan* 'know, understand, learn, be aware of', *agan* 'possess, obtain, have to pay, have to do', partly also **dugan* 'avail, be of value, be capable of'), others simply disappeared from the language altogether (e.g. **(be/ge-)nugan* 'suffice, have at one's disposal', *unnan* 'grant, allow, desire', *þurfan* 'need, be required, be under an obligation to, have occasion to', *(ge-)munan* 'remember, think about, consider, intend', later (in ME) also 'be about to, have to', and ultimately also *witan* (except in *to wit*)). These developments certainly were not uniform across all English dialects; most notably, *duȝen/dowen/dow*, *thurfen* (with the shortened stem forms *thar-/ther-/thor-/thur-*), and *munen/mun* survive as modality expressions in northern (especially Scots) and partly in Midland

dialects, or became obsolescent in these dialects much later than in the standard language. Notwithstanding such divergencies, the morphological class of the preterite-presents shrank gradually, and — what is perhaps more remarkable — the shrinkage was not entirely random: the verbs uniquely identifiable by their lack of a 3rd person singular agreement marker in the present tense indicative also represented a semantically more natural and coherent class in late ME/early ModE than in former periods. (For similar developments in other Germanic languages see Hammerich (1960).) In opposition to Lightfoot's syntactic autonomy ideology, I would again prefer not to dissociate these developments entirely from semantic changes: smaller, semantically more coherent and well-structured classes would in principle seem to be more characteristic of elements with grammatical meaning or function than of lexical elements with purely notional meaning. And as well-structured systems of grammatical elements do not normally admit of extensive synonymy, the grammaticization of the (pre-)modals into such a system might in fact help explain why particular preterite-presents dropped out of this morphosyntactic-semantic class or of the language: note that *witan* was a near synonym of *cunnan*, both verbs referring to intellectual capability; that **dugan* and *magan* both had dynamic-modal meaning, referring to physical capability; that *þurfan* had a deontic-modal meaning much like that of *need*; that *munen* in ME had become a synonym of *shall*, sharing its deontic and futural meanings; and that *unnan* also had deontic ('allow') and dynamic ('desire') meanings resembling those of some modals. None of these semantic facts is noted by Lightfoot, who consequently regards the various reductions of the class of the preterite-presents as entirely accidental.

(3) Having originally been past-tense (perfective) forms themselves, the preterite-presents developed a new past tense in Germanic times, using a dental formative. Presupposing that present-past tense form oppositions such as *(ic) cann - cuþe, sceal - sceolde, mæg - meahte, mot - moste* (and *wille - wolde*, to include the non-preterite-present premodal) were essentially, or even exclusively, a matter of present and past time reference in OE, the relationship between such forms soon ceased to be of a purely temporal nature, except in the sequence of tenses required by reported speech, where *could, should, would, might* still are the regular past-tense forms of *can, shall, will*, and *may*. This lack of a systematic and regular tense distinction, increasingly noticeable during the ME period, further separated the modals-to-be from standard verbs. Here, too, the individual developments ought to

be seen in the context of the integration of these verbs into the grammatical system of modality — and indeed, Lightfoot himself must concede that genuinely semantic changes are at issue when original past-tense forms (*could*, *should*, *might*, *would*, *must*) acquire new modal meanings outside straightforward tense oppositions.

(4) Following the word-order change to (underlying) SVO in early ME, some mechanism had to be innovated, Lightfoot suggests, in order to get epistemic premodals, underlyingly still main verbs (whether one-place or two-place Lightfoot is not so sure), in front of main-verb infinitives instead of leaving them in their underlying sentence-final position, or of having them appear in extraposition-like constructions. I doubt that this supposed innovation can claim historical reality; my guess is it is an artifact due to some defect in Lightfoot's analysis.

(5) The introduction of *to* as a semantically empty and automatic marker of infinitives, deriving from the directional preposition still homonymous today, was a long drawn-out and complicated process beginning in OE, or rather West Germanic, and not yet completed in the 16th century, where *to* infinitives (and incidentally also infinitives with *till*, *at*, *for*, these variants surviving in dialects) and bare infinitives are still found to compete in constructions where their distribution is strictly regulated today. What is remarkable, however, is that the conjunction *to* (or its variants) almost never marks infinitives occurring immediately after a premodal (except of course *ought*, which Lightfoot overlooks). Examples such as these are exceedingly rare and most often involve *will*, perhaps due to confusion of the anomalous modal-to-be (OE *willan*) with regular weak verbs of similar meaning and form (OE *willian*, *wilnan*), which more regularly took *to* infinitives: *Thei ... wele to conquere this londe be force* (c 1450 Merlin); *þat we not wille forto hoolde ... eny þing which is his* (c 1445 Pecock); *Nor is it thoght he will to stay onely to get y^t saylers aboard* (1648 Kem, Letter); *neiþer he schal mowe to studie, to preche, to speke myche, neither to singe* (1443 Pecock). Infinitives are accompanied by a conjunction more often, though by no means as a rule, when they occur with a (pre-)modal (in particular *will/would*, *shall/should* and *may/might*) without immediately following this (pre-)modal in surface structure (cf. Visser 1969: 1829-35). It is true, bare infinitives continue to occur with a few other main verbs as well (*let*, *make*, and those of perception when used in the active voice); the evolving distribution of the infinitival

conjunction, nevertheless, served to further identify the premodals as a morphosyntactic class (almost) of their own.

(6) To some extent the non-finite paradigms of the preterite-presents, and especially of the premodals, appear to have been defective all the time; and since most preterite-presents were high-frequency items, this absence of some non-finite forms from the extant texts is presumably not entirely accidental. Lightfoot disregards such early indications of paradigmatic defectiveness, and maintains that four independent changes took place simultaneously in the mid-16th century or perhaps a little earlier:

(6a) Modals cease to occur as infinitives.
(6b) They cease to occur with the suffix *-ing*, i.e. as present participles, gerunds, and nominalizations.
(6c) Two or more modals cease to occur in sequence accompanying one verb, in the standard dialects.
(6d) Modals cease to occur as past participles with *have*.

I disagree with Lightfoot in two respects. Firstly, I doubt that these changes took place all of a sudden in the 16th century, dramatically depriving the (pre-)modals of grammatical privileges all of them had previously been able to enjoy without limitations. The non-existence or at least sparsity of a number of non-finite forms of preterite-presents has already been mentioned. Furthermore, premodals when used epistemically in general do not seem to have occurred non-finitely in OE and ME in the first place. (Note that an identical finiteness requirement characterizes epistemic modals also in other Germanic languages where there can be no question of modals not being verbs; in fact, a requirement to this effect can presumably claim general rather than language particular validity.) Thus, the overall loss was in this respect not as substantial as Lightfoot implies, and what was lost was not lost at one fell swoop. Secondly, I doubt that Lightfoot's four changes were really independent, logically or empirically: not only is his (6c) nothing but a special case of his (6a), but all of his changes manifest the single generalization that the premodals ceased entirely to occur in contexts where non-finite forms were called for. This is not to say that they could no longer occur in these syntactic contexts ***because*** they lacked the respective paradigmatic forms — if this were the sole reason for the modals' syntactic limitations, one would need to explain why it never happened that the paradigmatic gaps were filled again, relying on morphological regularities observable in the

paradigms of ordinary verbs. (See Pullum & Wilson (1977:769) and Baker (1981:315f.) for attempts to explain — not very convincingly, in my opinion — this persistence of allegedly incomplete paradigms, blaming it on the very irregularity of the existing paradigmatic forms.) Certainly Lightfoot is right to point out that it is imaginable for a verb to lack only individual but not all non-finite forms. But regardless of whether such a situation is likely to be very stable diachronically, it seems to me mistaken to interpret the development at issue along such lines. What happened was not that one or the other paradigmatic form dropped out without affecting the categorial structure of the paradigm: the categorial distinction finite/non-finite as such was given up for these (formerly verbal) words, and any separation of the paradigmatic and the syntagmatic nature of these categories would here be entirely artificial.

I can only mention in passing two further developments which are no doubt related, more or less directly, to the disappearance of non-finite modals. The first has to do with infinitival (pre-)modals preceded by one of the futural (pre-)modals, as in *I shall mowe/cunne come*. One might expect that the most natural continuation of such constructions after the demise of non-finite modals would have been to use syntactically more versatile modality expressions instead, as in *I shall have to/be able to come*. But what actually happened, in ME and increasingly in early ModE, was that the futural modals were simply dropped, the deontic or dynamic modals themselves no longer requiring an additional indicator of futurity: *I must/can come* (see Visser 1972:687-9). The second, apparently much later development was the innovation of 'Modal Raising' in the *can't seem to* construction (cf. Langendoen 1970), where the modal semantically belongs in the complement clause (*It seems that I can't understand Kant/I seem not to be able to understand Kant*), but is barred from surfacing there on account of the finiteness requirement (**I seem not to can understand Kant*), and manages to sneak into the finite main clause along with the negative (*I can't seem to understand Kant*).

In view of Lightfoot's later theoretical claims it is significant to note that there are dialects which exhibit non-finite morphosyntax of the modals well after its supposedly radical abandonment. There are the notorious double modals (such as *might could, might would, might should, might oughta, had oughta, usta could*) in American dialects, especially in the South, where, however, a good case could be made, it seems to me, for analyzing the first of these modal elements (in particular *might*) as an adverb on a par with *maybe* rather than as a verb or an auxiliary. More worrying are the non-finite

modals, specifically *can* and *could*, in Scots and other regional (especially northern) British varieties (cf. Ščur 1968). Lightfoot only admits that they are exceptions to (6c), but in fact they also contradict his (6a), (6b), and (6d):

(6a) *Nobody seems to can understand it*; *He's the man to can do it*; *Shay'd use to couldn't sit nur stan'* (all examples from the Leicester area); *I can't do it now but I used to could* (Scots)

(6b) *Wi' hym noa cannin' fynd them* (Scots); Ščur (1968) also mentions *cannan'* as the separate form of the present participle in Scots

(6c) *We'll can agree fine*; *He wouldna could dae't* (Scots, with negative enclitic *-na*)

(6d) *If wey had cuid cum* (Scots); *I haven't could get across the doors*

That the 'exceptionality' of some modals in these varieties with respect to the finiteness requirement is complete rather than specific to individual non-finite categories would seem to support my contention that the four allegedly independent changes (6a-d) occurring in the standard dialects are not really independent at all.

Now, Lightfoot (p. 110) says he does not know whether these dialectal non-finite modals are innovative or conservative forms, and he does not care either. I find this attitude surprising. After all Lightfoot has chosen the modals as his paradigm case, and a few pages later (p. 113) he even asserts that in this case "there is no need to idealize the data, in a way that one would ordinarily expect to have to do, abstracting away from different dialects, literary styles, etc."! Well, perhaps we better should abstract away from Scots and other dialects in the ordinary manner; or we could assume that these are not properly dialects of English any longer; or in sum: we best ignore as a matter of methodological principle anything in the way of synchronic variation as utterly irrelevant to a theory of change. Ironically, Lightfoot warns us (especially those of us who are foolhardy enough to concern themselves with syntactic reconstruction and typology) repeatedly that practically only a handful of languages are suitable for investigations of their syntactic history: not only must their documented history be rather long, the extant documents also must be substantial and cover a sufficient amount of stylistic and dialectal variety. But if all these criteria are met, as in the case of English, Lightfoot abstracts away from, or simply disregards, crucial data.

Returning to the non-finite modals in Scots and elsewhere in Britain, the infinitive (*can*), present participle (*cannan'*), and gerund (*canning*) are certainly formal innovations in so far as the corresponding OE/ME non-finite

forms had a different stem vowel (OE *u*, ME *u/o*); and the past-tense infinitival use of *could/cuid* (cf. last example at (6a) above) also seems to be a rather late innovation. But the occurrence of modals in contexts where non-finite forms are called for as such does not seem to be an innovation of some British dialects effectively re-introducing kinds of constructions that had previously been lost. In particular the construction with a modal following one of the future 'tense' modals (as in *Cryseyde shal not conne knowe me* (c 1374 Chaucer, Troylus) or *before my letters shall may come unto your grace's hands* (1532 Cranmer) and the like) seems to have remained in use quite continuously in the north of Britain. If it is nevertheless rather poorly attested in the literature of the crucial period, this is not very surprising in particular in the case of Scots: vernacular Scots, although the national tongue, was not a literary, i.e. written and later printed language (printing in Scotland beginning in the late 1500s). At any rate, if Lightfoot had decided not to abstract away from, or ignore, the dialectal data, he would be in trouble either way. His 'prediction' that a re-analysis of the premodals was necessary at a certain point in the history of English (see below) would flounder if he accepted that such non-finite modal constructions persisted unchanged in certain dialects. And if these constructions could be shown (as I think they cannot) to have been innovated again once they had been given up because they were in conflict with the Transparency Principle, how can he uphold his claim that "to undergo these changes backwards would clearly entail massive ***increases*** in opacity" (p. 152), and would therefore be impossible?

Incidentally, one may have wondered why no change (6e), analogous to (6d), was taken into consideration:

(6e) Modals cease to occur as past participles with passive *be* (or earlier *beon/wesan*).

If the premodals really were completely normal verbs which could take direct object nominals as well as non-finite and, many of them at least, finite complements, they might be expected to have been passivizable in OE and ME. But they were not, at least when occurring with an infinitival complement. In instances such as *þæt heora eagum se weg wære up to heofenum cuþ to locienne* (Blickl. Hom.) 'that to their eyes the way up to heaven was clear to see', *Hwæt is cuðost mannum to witanne?* (Salomon & Saturn) 'what is surest/best known for people to know?', *cuþ*, though formally the past participle of *cunnan*, functions as an adjective in active constructions and definitely not in passive constructions regularly related to corresponding actives.

With regard to passivization, the premodals thus seem to have been somewhat exceptional from the beginning. Lightfoot's analysis indeed can account for the absence of passives: he suggests that the complements of the premodals are underlyingly S's and superficially presumably VP's, and never NP's, which is the category he thinks is appropriate only for *to* infinitives in OE and ME. But since premodals can take nominal direct objects, finite-clause complements, and also *to* infinitives (when these are not placed immediately after the premodal), I am not convinced that this is an adequate solution of the problem why many verbs with infinitival complements passivize whereas the premodals don't (not even in non-standard dialects, apparently).

(7) Once the category Modal had been established as distinct from the category Verb, some rules formerly involving verbs were re-formulated. Negative Placement is one such rule, which originally, according to Lightfoot, inserted 'neg' after the first verb, yielding constructions like 'Beowulf can not come' if this verb is a premodal. Continuing with this rule, one would end up with 'Beowulf can come not' because the modal is now no longer the first verb. But the difference between the former and the latter surface pattern was "of a kind to strain communicability across generations to breaking-point" (p. 406), and "therefore some re-formulation was necessitated by the theory of change in order to prevent a breakdown of communication" (p. 407) — suggests Lightfoot, apparently quite seriously. I do not want to take issue with Lightfoot's account of negation (his OE Neg Placement rule is clearly wrong anyway, because finiteness rather than verbhood per se was crucial for the position of *ne*); conjuring up the danger of a breakdown of communication surely does not explain anything. Of course this is not to to deny that the grammar of negation, which since the end of ME, incidentally, also involves a new unmarked neg. element, viz. *noht/nat/not*, was modified considerably in the 16th and especially the 17th century, in particular in connection with the systematization of *do* Support.

(8) Subject-Verb Inversion is the other transformation mentioned by Lightfoot as requiring some re-formulation: initially it referred to verbs in general, but later only to auxiliaries including the new modal ones (and, incidentally, also the copula *be* and, less consistently, with the well-known differences between British and American English, the possessive *have*), without any breakdown of communication between the Subject-Verb inverting generation and the avant-garde Subject-Auxiliary inverters. Superficial

though Lightfoot's analysis of the intricate regularities of inversion at the different periods is, it suffices to indicate that the modals and other auxiliaries indeed diverged from verbs with respect to these particular syntactic regularities.

Obviously the systematization of the uses of 'empty' *do* plays a key role in the evolution of the syntactic regularities under which auxiliaries and verbs turned out to behave differently — and these regularities include not only negation and inversion, but also emphatic affirmation (*Beowulf dóes come* vs. *Beowulf cán come*) and the choice of pro-forms, in tag questions and elsewhere (e.g. *Beowulf comes and so does Grendel* vs. *Beowulf can come and so can/*does Grendel*). Since Lightfoot of course recognizes that his rule re-formulations and the development of empty *do* are inextricably intertwined, it is somewhat surprising that he insists on the rule modifications being necessary consequences exclusively of the introduction of the modal category. But the more substantial difficulty perhaps is that the distribution of 'vicarious' *do* in particular appears to be sensitive to the distinction between lexical verbs and auxiliaries (or verbs with the grammatical functions of today's auxiliaries) before the category auxiliary was established according to Lightfoot's scenario. *Do* commonly occurs as a vicarious form for standard verbs in OE, as in these examples: *Reced weardode unrim eorla, swa hie oft ær dydon* (Beow.) 'countless warriors guarded the mansion, as they often did (had done) before'; *Se wilnode ðæs westdæles, swa se oðer dyde ðæs eastdæles* (Ælfred, Oros.) 'he wished for the west part, as the other did for the east part'. There are also a few early as well as later examples of vicarious *do* with periphrastic aspect and voice forms: *þonne beo we sittende be þæm wege, sua se blinda dyde* (Blickl. Hom.) 'then should we be sitting at the road, as the blind man did/was'; *And if they be scattred among the church and the dyuers sectes ... so long as they so dooe, ... so long be they a vnknowen secte* (1532 St Th. More). But these are exceedingly rare, and may simply be solecisms: normally *do* is not used as a vicarious form for *be* in periphrastic and copula functions. Moreover, *do* appears not to have been used vicariously for the premodals either, or at least not very often — unless a premodal was the only verb in a clause, unaccompanied by an overt infinitive, as in *Wenaþ þæt heo moten to þære mæran byrig up to englum swa oðre dydon* (Christ & Satan) 'they expect that they are allowed (to come) to that splendid city up to the angels, as others were'. When a premodal governs an overt infinitive and there is a pro-form *do*, this *do* is usually vicarious for the infinitival verb rather than for the finite premodal: *Se man nolde gan, swa swa oðre men*

*dydon* (Ælfric, Saints' Lives) 'the man did not want to go, as others did/had done'; *Als dos þe fader to þe sun, he can þam teche and lere* (14c Cursor M.). (Even in the example from Christ & Satan just quoted, *dydon* may be vicarious for the implicit infinitive 'to come' rather than for the finite *moten*.) Pending a more thorough investigation of verbal pro-forms in OE and ME, I tentatively suggest that the distribution of vicarious *do* was always sensitive, at least statistically, to a distinction of lexical verbs with a notional meaning and verbs with certain grammatical functions (copula, aspect/tense/voice periphrasis) including the premodals. When other syntactic rules which likewise have to do with empty *do* turn out to distinguish auxiliaries, including the modals, from verbs, the pattern of the distinction as such, therefore, ought not be regarded as a radically new ModE innovation without any historical precedent.

(9) Finally, there was a set of 'quasi-modals' to complement the modals, viz. *be going to*, *be able to*, *have to*; and these have important morphosyntactic properties which had been lost by the modals: they can occur in contexts where non-finite verb forms are required, and they have regular tense oppositions. To Lightfoot, this is "perhaps the most remarkable change of all" (p. 112), and he also finds the dates of the alleged innovation striking: these three "new" verbs "came into the language", he claims (ibid.), immediately after the premodals had been re-analysed as non-verbal modals, in order to fill the morphosyntactic vacuum created by this re-analysis. What is really striking here is, not so much what happened historically, but what Lightfoot fancies to have happened. Nobody would be foolish enough to deny any connection between the fate of the premodals and the development of other modality expressions. Nevertheless, the dates of the allegedly first pertinent modal attestations which Lightfoot himself supplies from the OED (*have to* 1579, *be going to* 1482, *be able to* 1440) do not strike me as particularly uniform, to begin with; and the first of these attestations, referring to *be able to* as semantically equivalent to *can* (which, it will be recalled from (1), is rather a late developer otherwise), moreover considerably antedates the step of *can* and its kind from verbhood to modalhood in Lightfoot's own schedule. But matters really are much worse for this schedule and the whole neat scenario.

Although *have to* may be predominantly possessive in OE, it early acquires a modal sense in addition, and is found with a purely modal sense of intention/futurity and/or obligation already in OE and increasingly in ME,

as examples like the following would seem to demonstrate: *þone calic þe ic to drincenne hæbbe* (OE Gospel St Matthew) 'the chalice which I have to/will drink'; *Hæfst ðu æceras to erigenne?* (Ælfric, Grammar) 'do you have fields to plough/have to plough fields?'; *hwæðer he hæbbe hine to fullfremmenne* (OE Gospel St Luke) 'whether he had to perfect himself'; *Gramaire ferste hath forto teche To speke upon congruite* (c 1390 Gower, C.A.). Admittedly, *gan/gangan* in OE as a rule refers to movement, but when accompanied by an infinitive or gerund of purpose, its meaning already approaches the futural modal one of today's *be going to* (cp. *Ic geongo to cunnanne ða ilca* (Lindisf. Gosp.) 'I go to know/find out that'), and this non-local force is gaining ground in ME (e.g. in uses such as *ȝif þu gest herof to disputinge* (1250 Owl & Night.) 'if you hereof proceed/turn/are going to dispute'). *Beon (h)able* seems to have been employed with some modal sense almost as soon as it had been borrowed from French. The OED in fact has examples dating from before 1440 (such as *Thyne was the action, and I nought but abyl for to suffre* (1413 Lydgate)), and so does Visser (1969:1749ff.), that invaluable treasury of information which Lightfoot duly celebrates, but utilizes only very selectively *(cf. feithful men, which schuln be also able to teche other men*; *the hauene was not able to dwelle in wynter* (both c 1380 Wyclif); *Ther ben of suche ... That ben noght able as of hemselve To get love* (c 1390 Gower, C.A.); *It is a feble leche þat can not helpe þat is able to be holpen* (c 1400 Lanfranc)). One wonders how Lightfoot managed to disregard all this, and of course also how he could choose to consider only have to, be able to, and be going to. Not only semantically *be to* fits in perfectly as well; its modal meanings again are no ModE innovations but may be found already in OE and ME (see Klöpzig 1922 for details): cp. *se þearfa ... þe mid þe is to cumenne to engla gebeorscipe* (Anglo-Saxon Homilies) 'the poor one who is with you to come to the feast of the angels'; *þe mid þe is to onfonne þæs undeadlican gegyrlan on neorxna wange* (Anglo-Saxon Homilies) 'who is with you to receive the unperishable garment in paradise'; *þatt irre þatt to cumenn iss* (c 1200 Orm); *he wist what he was to do* (c 1380 Wyclif). What are more recent innovations are the demise of the impersonal construction of *be to* and the ban on all its non-finite occurrences. The following typical examples suggest that this latter innovation must have been fairly recent indeed: *A King of Egypt, being to erect The image of Osiris* (1622 Massinger); *The Merchant wil be to seeke for Money* (1625 Bacon); *I must be to be depended upon* (1798 Malthus); *reading it with the knowledge that he had been so soon to die* (1933 Vera Brittain); *But they may be yet to come* (1936 Innes). In fact, there were

numerous further predicates in OE and ME with approximately the same meanings as the premodals; here is a selection from those most current in ME: *birrþ, let, yemen, neden, haven nede to, ben yhalde to, besemen, him were bet(ere) to* (supplanted by *he were better to*, and eventually by *he had better*), *haven lewer, haven lyking, kepen, chosen, be aboute to, þolen, shapen to, bidden, coweiten, desiren, haven desir, purposen, taken purpos, daynen, graunten, enforcen, think, demen, holden, lewen, wenen, trowen* (see Wilde 1939/40 or Visser 1969: 1749ff., 1812f., 1827 for details).

Thus, if one does not want to exclude arbitrarily a great number of modality predicates continuously complementing the hard-core (pre-)modals, one must conclude that the picture of a sudden appearance out of the blue of three 'new' verbs to express modality in the manner of the original premodals is certainly an overdramatic misconstrual of history. And as to the claim of these new or not-so-new modality expressions filling a morphosyntactic vacuum, none of the relevant predicates was, or is, exclusively employed non-finitely; most of them were, and are, usually finite, at least more often so than non-finite; and a few are exclusively finite (*is to*), and even may have been exclusively finite all the time (*were/had better*). The situation is essentially the same as far as past-time reference is concerned.

This concludes my annotated version of Lightfoot's modals story. Of course the story is not really Lightfoot's, but more or less common knowledge as far as the basic facts are concerned. Lightfoot's own contribution, apart from abstracting away from and idealizing some of these basic facts, is essentially confined to less pedestrian matters of explanation, and to these I shall turn somewhat later. But first let me add fifteen or so further changes not dealt with by Lightfoot, the last two of which seem particularly important. With all these addenda it ought to be more difficult for certain critics (such as J. Anderson 1976: 34-42) simply to deny that there is a story worth telling, regardless of whether the present account can claim to be the full unexpurgated version of the modals story. Judging by the recognized auxiliary criteria of ModE as conveniently assembled in Huddleston (1980), it would seem relatively comprehensive, not necessarily in descriptive detail, but in the principal matters of theoretical interest.

(10) Impersonal constructions are well known to have disappeared in late ME and early ModE. In the chapter where he deals with this development (§ 5.1), Lightfoot indeed also notes that a few verbs apparently not construed impersonally in OE began to allow this construction in ME as an alternative

to the personal one, although it was doomed to become obsolete not much later. Lightfoot is not interested in the semantics of the opposition between personal and impersonal, and hence does not care whether these impersonal new-comers are an arbitrary or semantically natural set. Obviously the latter is the case, and it ought to be mentioned here that verbs which are to become modals make up a significant part of this set. Let me exemplify these impersonal uses not met with in OE (cf. van der Gaaf 1904; Visser 1970: 26-9): *Us oghte ... have patience* (c 1386 Chaucer, C.T.); *Ne þarrf ȝuw nohht nu follȝen me* (c 1200 Orm); *Us moste putte oure good in aventure* (c 1386 Chaucer, C.T.); *Me nedeth not to bost* (1534 St Th. More); *Hym hadde lever have ben at home* (c 1300 Coer de Lyon); *Hem had better be at Rome* (14th c Beves of H.). No doubt the usual explanation of such occasional impersonalizations referring to analogy is basically correct: there were verbs with similar meaning, in particular from areas close to deontic and also dynamic modality, which had been impersonal all the time, eventually attracting the premodals towards this syntactic pattern. Although increasing the variety of constructions available to these verbs, these analogical developments need not necessarily be regarded as complications or as indicative of utter confusion in this area of syntax. They may well have been attempts at regularization, providing the full range of morphosyntactic means appropriate to the expression of certain relational semantic distinctions.

(11) Since the end of the 15th century a separate verb *con/cun*, which later became largely obsolete, was formally and semantically differentiated from the (pre-)modal *can*. This new verb, based on existing forms of the old preterite-present, but in addition innovating a number of paradigmatic forms characteristic of regular weak verbs, continued the meanings 'get to know, learn, study' of the old preterite-present in its typical main-verb usages: *Tunes, Measures ... als' hee kons* (1607 Sylvester), *They had cand their lesson* (1587 Fleming), *Not to cun by heart, nor to write out* (1587 Golding), *Patiently conning the page again and again* (1838 Dickens).

An analogous differentiation seems to have occurred with *could* in the 16th century. Note that the current spelling of *could* reflects an orthographic innovation: an unetymological *l* began to be inserted about 1525, presumably on analogy of *should* and *would*, where the etymological *l* had become silent. When used with the meaning 'know' (intellectual capacity), however, the etymological spelling *couth/coud(e)* was retained longer (as in *So well his leere he Couth* (1652 C. Stapylton)), which I interpret as revealing a tendency

to split up one lexical item into two: one with the lexical function of a full verb, the other with the grammatical function of a modal auxiliary. (Recall, incidentally, that other independent lexical items also dissociated themselves from the premodal *can*, specifically from its non-finite paradigm: *cunning*, *(un-)couth*.)

Insofar as *shall* continued throughout ME to be used in its original meaning 'to owe', accordingly requiring an object noun phrase referring to tribute (and perhaps another one for the creditor), one could assume a similar lexical split once *shall*, with infinitive complements, had established itself with modal and temporal meanings. The differentiation in this case, though, would be a matter of meaning and subcategorization, without attendant formal distinctions as in the two previous cases.

Comparing the auxiliary *will* and the standard weak verb *to will*, which has all of the characteristically verbal properties the modals today are lacking (cf. *God wills that man should be happy* (1, 2); *God has willed it so* (1, 6d); *You just don't will yourself to keep awake* (1, 5, 6c, 7)), one might be tempted to assume that a lexical split of a similar kind has occurred with the non-preterite-present premodal as well. However, in this case there were two (or rather three) different lexical items even in OE: the irregular verb *willan* which then turned into a modal auxiliary on the one hand, and the perfectly regular weak verb *willian* (and *wilnan*) on the other. Due to the fact that the forms and meanings of these two or three lexical items continuously overlapped to a considerable extent, the distinction between them has tended to be blurred in some respects, and may even have been re-drawn occasionally (cf. (5) above); but these were not strictly speaking developments where one lexical item split up into two in historical times.

Some time later, in the 19th century, one encounters double forms of the 3rd person singular present tense of *do*, viz. *doth* [dʌθ] and *doeth* [du:ɪθ]; and if their distribution in examples such as *Man doeth this and doeth that, but he knows not to what ends his sense doth prompt him* (quoted from Jespersen 1942:21) is not accidental, it could be indicative of a dissociation of an auxiliary *do* from the full verb *do* that continues to be inflectionally regular. The precursor of *do* as a periphrastic auxiliary, ME *gan/con*, has rather more obviously dissociated itself from its full-verb source *ginnan* 'to begin', as is demonstrated by its occurrence as a present tense form despite its origin from a past tense form, also taking the 2nd person singular ending *-es* appropriate to the present, but not the past, tense (cf. Tajima 1975).

(12) Discussing the development of the infinitival conjunction out of the preposition *to*, Lightfoot claims that "as long as *to* was interpreted as a preposition, one could predict that it would never occur after the pre-modals, because they conveyed no notion of 'direction towards'" (p. 108). A considerable amount of idealizing and abstracting away from raw data must have been necessary to arrive at this 'prediction'! As a matter of fact the (pre-)modals occurred quite commonly with directional adverbs, including prepositional phrases with *to*, in OE, ME, and early ModE: *þin fæder sceal mid me to mynstre* (Ælfric, Saints' Lives) 'your father shall/must come/go with me to the minster'; *Ic to sæ wille* (Beowulf) 'I will/must go back to the sea'; *no þy ær fram meahte* (Beowulf) 'he could get away none the faster'; *hie ... wiston hwider hie sceoldon* (Ælfred, Orosius) 'they knew where they had to go'; *Adoun he moste* (1250 Fox & Wolf); *Thou shalt with me to helle to-night* (c 1386 Chaucer, C.T.); *The Sarezynes myght neyther in ne oute* (14th c. Coer de Lyon); *Thow mon to Paris to the King* (c 1475 Rauf Coil$_3$ear); *I'll to him*; *thou shalt to prison*; *That I may backe to Athens by day-light* (all from Shakespeare); *And I can not away* (c 1620 Ben Jonson); *I must to Prospero* (1649 D'Avenant). Further verbs which were not hard-core modals but not entirely outside the sphere of modality either (see (9) above), likewise used to construe with directional adverbs: *With othir men that intended to the same place* (c 1425 Found. St Bartholomew's); *The Erle of Oxenford ... is purposed into Skotlond* (1473 Paston Letters); *Come let's to dinner*; *Desire them all to my pavilion* (both from Shakespeare); *With shyppes .xii. to Italy had they mente* (c 1470 Harding, Chron.). Surely the modal expressions together with the adverbs or prepositional phrases did convey the notion 'direction towards (or from)', there being always the possibility of inferring a semantically neutral verb of movement in such contexts. And of course such movement verbs could actually surface in such contexts just as well: e.g. *Hwyder wilt ðu gangan? Min Drihten, ic wille gangan to Rome* (Blickl. Hom.) 'Where will you/do you intend to go? My Lord, I will go to Rome'. In the 17th century directional adverbs and prepositional phrases largely cease to occur with the modals, but also with the modality expressions which definitely retain verbal status, and an infinitival verb of movement now has to be employed to convey directional notions. There are later instances of the original infinitive-less construction, but most of them are consciously archaic and formulaic (cf. *I intend for England this spring* (1817 Byron); *Thou shalt with me to Iona* (1828 Scott); *I must down to the seas again* (1902 Masefield); proverbs such as *Murder will out*; not stylistically marked, however, is this late example from

the OED: *in these days when everything must to the papers* (1889 Macm. Mag.)), or they are Scotticisms (such as *I'll away out*).

Having dealt with the increasing inability of the (pre-)modals to co-occur, at surface structure, with directional adverbs (12) and with direct objects and finite-clause complements (1), it is appropriate to mention a further reduction of their superficial co-occurrence potential — if in parenthesis, because the abandoned construction at issue seems to have been much more peripheral. At least in OE, certain premodals could be employed in almost copula-like manner, followed by predicative adjectives without intervening overt copula. As indicated in the translation of one pertinent example from the 'Battle of Maldon', an overt copula later became obligatory in this construction: *Hige sceal þe heardra, heorte þe cenre, mod sceal þe mare, þe ure mægen lytlað* 'Courage must be the firmer, heart the bolder, spirit must be the greater, the more our strength wanes.'

(13) Matters of inflection have already been dealt with in connection with the continuing lack of a 3rd person singular desinence in the indicative present (2), the erosion of a regular present-past tense opposition (3), and the developments of the non-finite categories (6). The inflectional morphology of the OE preterite-present premodals, however, involved some further categories:

(13a) There were distinct forms for the second person singular indicative, utilizing the desinence *-t* (*scealt*, *meaht*, *þearft*) or *-(e)st* (*canst*, *dearst*, *-manst*/*-munst*, *ahst*, *most*, and regularly in the preterite).

(13b) Desinences (*-on* 1/2/3pl) and ablaut patterns signalled the singular-plural distinction in the present tense indicative (*deah* - *dugon*, *can(n)* - *cunnon*, *þearf* - *þurfon*, *dear(r)* - *durron*, *sceal* - *sculon*, *man* - *munon*; and with stem-vowel alternation which is not of the ablaut type, *mæg* - *magon*), whereas desinences alone encoded the number opposition with a few other premodals (*agon* , *moton* 1/2/3pl) and generally in the preterite indicative (*-on* 1/2/3pl) and in the present and preterite subjunctive (*-e* sg - *-en* pl).

(13c) Desinences and sometimes stem vowels distinguished indicative and subjunctive moods, more consistently in the present than in the preterite (where 1/3sg indicative is as a rule homonymous with 1/2/3sg subjunctive).

(Genuine imperatives of preterite-presents were not common; subjunctive forms were usually employed in this function.) In all these respects the pret-

erite-present premodals were much like normal verbs: the desinences and stem-vowel patterns may occasionally differ, the categorial distinctions encoded are essentially the same. And the fate of these inflectional distinctions was, in the long run, also much like that of normal verbs: they ultimately disappeared, as victims of sound change (affecting the desinences *-on*, *-en*, *-e*), of analogical levelling (affecting primarily the stem-vowel patterning, as when the singular stems oust the plural stems), and of morphological replacement (affecting the 2nd person sg., which is ousted by the 2nd person pl. when the originally plural pronouns *ye/you* take over from the old singular *thou*, first as pronouns of 'respect', but later without this pragmatic restriction) — not without some non-reductive interludes, to be sure (such as the extension of *-st* 2sg at the expense of *-t* in ME *may(e)st*). The schedule of these disappearances, culminating towards the end of ME and in early ModE, moreover, was essentially the same for normal verbs and the (pre-)modals. This implies, for instance, that the replacement of the old 2nd person singular, perhaps the last of the changes at issue to come to an end (of sorts), was not completed until the 17th century, although it had been initiated, under pragmatic conditions, at least as early as the 13th century. (But note that mood distinctions were also maintained quite long, as in *List if thou can heare the tread* (1596 Shakespeare), where the Quarto edition, employing *canst*, has already levelled the distinction between the two moods at the expense of the subjunctive.) In Shakespeare the old 2nd person sg. desinences abound, although occasionally they are replaced by uninflected forms in some editions: *Thou wilt keep ...*; contracted *That's a deed thou't dye for*; *Thou shouldst know it*; *Wouldst thou aught with me?*; contracted *Thou'ldst shun ...*; *Canst thou bring me to the party?*; *Thou ... shalt be shown in Rome*. But the frequency of *thou* and, concomitantly, of this old desinence decreased soon — not in non-standard dialects, one should add again, where old 2nd person singular forms were retained much longer. Indeed, judging from the last attestations given in the OED, one might conclude that the old 2nd person singulars of modals as well as normal verbs survived much longer even in the standard language. Examples such as *In sacred vestments may'st thou stand* (1717 Pope); *Shouldst thou point out to me* (1820 Scott); *Wilt thou then straightway bring him in?* (1849 M. Arnold), however, are likely to be dismissed as consciously archaic and certainly non-colloquial. They remain in colloquial use, though, if only in one particular register: the language of prayer. Taking into account such stylistic variants, one must admit that the modals, just like any normal verb, have retained an inflectional feature: the

distinction between an unmarked and a marked (*-t/-st*) 2nd person singular agreement form.

Disregarding this stylistically marked 2nd person singular as well as the relationship between present and past tense forms in reported speech, whose inflectional nature might be contested, the modals in the standard dialects are the perfect manifestation of the Sapirean drift towards the invariable word: they have lost all of their finite and non-finite morphology. It seems to me, though, that the net result of this overall development can be interpreted in different ways.

The first interpretation would be to assume that the present-day modals are obligatorily finite, lacking all non-finite forms; that they lack regular, non-suppletive past-tense forms, not to mention mood distinctions; and that their exponent of 3rd person singular agreement happens to be zero. In an alternative interpretation the modals would belong to a word-class for which all these verbal paradigmatic categories as such are irrelevant in the first place, rendering the concepts of paradigmatic gaps and zero exponents strictly speaking meaningless. This latter option, which seems to me preferable, and aspects of which will be elaborated later, would entail a reconsideration of the notion of finiteness. The modals themselves, including the assertive/interrogative/negative auxiliary *do/don't*, accordingly would have to be conceived of as finiteness markers of their clauses or verb phrases, with a function, and distribution, which is complementary to that of *to* and perhaps other conjunctions marking clauses or verb phrases as non-finite. (If *to* in fact should turn out to be more appropriately categorized as a non-finite auxiliary itself, rather than as a genuine hypotactic conjunction as has recently been suggested by Pullum (1982), this would not materially affect the present interpretation; in fact, it would seem more compatible with it than with its alternative, where *to* would reinforce the ranks of the elements with defective finiteness paradigms.) Finiteness, in ModE and perhaps more generally, would thus not, or not exclusively, be a matter of tense marking and/or person and number agreement: modality, grammaticized as mood, might be the feature, or one of the features, determining the finiteness of a clause or verb phrase. I cannot go into the implications of this interpretation for generative-transformational accounts; it is obvious, however, that this conception of finiteness would be incompatible with the idea that modals are obligatory tense carriers and recipients of zero agreement marking. (In spirit, Baker's (1981:306ff.) suggestion that modals in ModE are phrasal heads of finite predicate phrases, Evers & Scholten's (1980) hypothesis that

non-finite verbs are minus-modality verbs by universal definition, and Chomsky's (1981: 140 fn. 28) off-hand remark that perhaps the modals also appear within INFL together with [±Tense] and the agreement features, may be similar to the interpretation proposed here. And of course there are also Akmajian/Steele/Wasow (1979), Steele et al. (1981, with noteworthy speculations on finiteness in chapter 4), and other advocates of the category AUX who would not take the modals for verbs with defective paradigms.)

(14) Under certain circumstances negative elements tend to be closely associated with other words in the Germanic languages (and no doubt elsewhere). In OE, for instance, the negative *ne/ni* could be contracted with following verbs, adverbs, and pronouns if these began with a vowel, with *w*, or with *h* plus vowel; and it is with verbs in various grammatical, rather than purely lexical, functions including the premodals that such contractions were employed most frequently (e.g. *nolde/ne wolde*; *nat/ne wat*, *nyste/ne wiste*; *nyllan/ne willan*; *nabban/ne habban*, *nis/ne is*). Two later developments more specifically reveal closer ties between negative elements and the modals and other auxiliaries than between negatives and ordinary lexical verbs.

(14a) First consider Subject-Verb/Auxiliary Inversion in ME and early ModE, with *not* already serving as the negative element and with *do* Support not yet fully automatized. Before *not* was reducible to *n't*, there developed two possibilities of placing the negative element in clauses with inversion (e.g. in negative interrogatives): it could be inverted together with the verb or auxiliary to precede the subject (A), or it could stay behind the subject, inversion fronting the verb or auxiliary alone (B).

| | | |
|---|---|---|
| (A) | Vb/Aux-neg-Subj | 'Comes not Beowulf?', |
| | | 'Can not Beowulf come?' |
| (B) | Vb/Aux-Subj-neg | 'Comes Beowulf not?', |
| | | 'Can Beowulf not come?' |

One factor relevant to the choice between patterns (A) and (B) was whether the subject was pronominal or not; there was a clear preference for pattern (B) with pronominal subjects (cp. *Wol nat oure lord yet leve his vanytee? Wol he nat wedde?* (c 1386 Chaucer, C.T.)). But it seems to me that the identity of the verbal element was also relevant, at least statistically. Whereas pattern (A) (e.g. *Appereth not he still impenitent?* (1528 St Th. More); *Sounds not that better?* (1616 Ben Jonson)) is apparently much less frequent than pattern (B) (e.g. *Herys thou not this ugly noyse?* (c 1300 Harrowing of Hell);

*Came he not home to-night?* (1590 Shakesp.)) with lexical finite verbs, the overall frequency of pattern (A) increases considerably when the finite verb (or finiteness marker) is an auxiliary element, i.e. a (pre-)modal or *do*, an aspectual or passive auxiliary, or the copula and perhaps also possessive *have* (cp. (A): *why may not he be there again?*; *Had not you come* (both from Shakesp.); *art not thou pryamus sone ...?* (Caxton) — (B): *Shall I not be slayn?* (c 1460 Towneley Pl.); *Have I not forbid her my house?*; *art thou not horribly afraid?* (both from Shakesp.)).

(14b) The difference in the strength of the ties between the negative and auxiliaries on the one hand and full verbs on the other is categorial rather than statistical in the case of a subsequent innovation in the expression of negation. *Not* began to be contracted to *n't* around 1600, contracted forms making their appearance in the written language about 1660 (see Jespersen 1917: ch. 11). Significantly, *not* could, and still can, only be contracted after an auxiliary (*can't, mustn't, isn't, doesn't, hasn't* etc., also *needn't, daren't, use(d)n't*, when these are used as auxiliaries) and after the semi-grammaticized copula *be* and possessive *have*, or, more specifically, only after a leftmost auxiliary */be/have* (cf. *Beowulf couldn't have/*could haven't found Grendel*; **We used to haven't much money*), and never ever after a full verb (cf. *We haven't/*possessn't much money*; *He wasn't/*seemedn't very intelligent*;**He spoken't a word*; **I thinkn't*; **I hopen't*; *go fiddle, I care not/*caren't what you do* (1697 Vanbrugh)). Exceptionally, *may not* resists contraction today, although *mayn't* used to be found quite commonly earlier (cf. *Mayn't my cousin stay with me?* (1697 Congreve); *Why mayn't I say to Sam that I'll marry him? Why mayn't I?* (1894 Th. Hardy); *you mayn't know it but ...* (1886 Hughes)). In a further step of this development, several negated modals and other auxiliaries gained some degree of independence vis-à-vis their positive counterparts. Negative forms such as *won't, shan't, can't, don't, ain't* (and some less common ones, such as *han't* for *haven't, in't* for *isn't, wan't* for *wasn't* — see Jespersen, loc. cit.) cannot at any rate be related to their corresponding positive forms accompanied by uncontracted *not* by regular phonological or morphonological rules (such as the rule of interconsonantal *t*-deletion which is operative in *mustn't*, or the rule of negative contraction itself). On account of the recurring enclitic *n't* (if it is an enclitic, which is contested by Zwicky & Pullum (1983), who argue that it is an inflectional suffix) one could probably analyze such contracted forms as still bi-morphemic (regardless of whether one prefers to categorize them as clitic groups

or as inflected words). But on the other hand, in view of their formal opacity one might also be justified in taking them for (almost) autonomous negative auxiliaries, which after all are a phenomenon familiar from other languages (more familiar, it would seem, than genuine negative inflections).

(15) Turning from the formal to the semantic side of negation, it is necessary in sentences with modals to distinguish negation affecting modality from negation affecting the main verb (so-called proposition or event negation; see Palmer 1979: 24ff., *passim*). This distinction is not only a matter of interpretively associating the negative *not/-n't* with the modal or the main verb in a given sentence; different modals in fact may be required in positive and negative sentences depending on the kind of negation. Thus, whereas event negation with a deontic modal as in *You must take this down* requires only the introduction of the negative element, *You mustn't take this down*, proposition negation with an epistemic modal as in *He must be working in his office* in addition requires an exchange of the modal, *He can't be working in his office*. In this sense, epistemic *must* and *can* could therefore be regarded as suppletive variants with respect to proposition negation. Similarly, different modals turn out to be suppletive variants in the scope of modality negation. Thus, *can*, *may*, and *could* correspond to positive epistemic *may*, *must*, and *might* respectively under modality negation (cf. *He may be working in his office - He can't be working in his office*; *He must be in his office - He may not be in his office*; *Bloggs might be a Bavarian - No, he couldn't*), and *need* to positive deontic *must* (e.g. *You must take this down - You needn't take this down*). Suppletive relationships among the (pre-)modals as such are certainly no ME or ModE innovation, but have been in existence as long as the respective verbs have been employed as expressions of modality. What has changed to some extent historically, in accordance with the semantic development of individual (pre-)modals, is which particular (pre-)modals contract suppletive relationships to which others with respect to the two kinds of negation. For instance, as long as *must/motan* was not used epistemically (that is, in OE times), it could not be the positive counterpart of modality-negated *may* or of proposition-negated *can* as in the above examples. But clearly the establishment and abandonment of particular suppletive pairings are no autonomous developments to be accounted for independently of everything else in the grammar and lexicon: they are automatic consequences of changes in the meaning of the (pre-)modals concerned.

(16) It is commonly assumed that (pre-)modals, when governing a non-finite verb, have always governed only (bare) infinitives. As a matter of fact, however, the (pre-)modals, until the 16th and 17th century in the standard dialects, were also able to govern what appear to be past participles (cf. Visser 1973:203ff.). These past participles could be interpreted as passives (16a) or, more commonly, as active preterites or perhaps perfectively (16b), although there were no finite auxiliaries which one could usually expect with these interpretations.

(16a) *Scip sceal genægled, scyld gebunden* (Maxims) 'the ship shall (be) nailed, the shield bound'; *in the sovereynes substaunces ... jugement is more clear, and wil nat I-corrumped* (c 1374 Chaucer, Boece); *no grene, false and deseveable tymbre or haillard wherethorgh the byer therof may deceyved* (1439 Red Bk. Bristol); *He ... that mynystrith wel shal haue grete rewarde & crowned in heuen; thenne he must haue his seculer clothes ayen, and put out [sc. of the monastery] for euer* (both c 1490 Rule St Benet, The Caxton Abstract)

(16b) *swa swa manige trywe men sædan þe hit geseon sceoldan* (Peterb. Chron.) 'as many faithful people reported who should/ought to (have) seen it'; *For many one there dyde for colde That warmythe of howese savyd wolde* (c 1475 Siege of Rouen); *If they coude amended it* (1523-5 Lord Berners, Froiss.); *Where would you had remain'd until this time?* (1590 Shakesp.); *hee might easily, and according to the manner of men, occupied the monarchy* (1633 John Donne); *And but for you I might despair'd of* (1790 R. Burns)

The construction exemplified in (16b) might be explained historically as the result of a gradual reduction of preterite or perfective *have* after a modal, intermediate steps of which process are attested plentifully even in the written language (cf. *The Jewes wolde a stoned him* (c 1400 Mandeville); *She might ha' been a grandam ere she died; So would I a done by yonder sunne* (both from Shakesp.)). On the other hand, there is clear evidence for the innovation of preterite infinitives since the 13th and 14th centuries (cf. *He assigned Harald to Ingland, to had it in fee* (c 1338 Rob. of Brunne); *had it not been much better To kept thy promise than be thus surpris'd* (1592 Marlowe); *to loved* 'Preter tense', *to had loved* 'Preter pluperfectum' (1586 Bullokar, Bref Grammar for English)); and it is therefore conceivable that some of the apparent participles in constructions like (16b) in fact are such preterite

infinitives. Of course, our assumptions of what exactly changed when constructions such as (16a) and (16b) disappeared depend upon how they are accounted for synchronically: the disappearance of (16b) could be due to the loss of preterite, i.e. tensed, infinitives; or the disappearance of both (16a) and (16b) could be due to the loss of a syntactic rule, or rules, deleting (or of interpretive-semantic rules inferring) passive and preterite/perfective auxiliaries after a modal; or it could be due to the modals' losing the ability to govern past participles, if this is a real alternative to saying that the modals lost the ability to govern the syntactic or semantic rules just alluded to. At any rate, the net result of these developments is a further decrease in the modals' ability to govern anything — except morphologically unmarked infinitival verb forms (the only admissible morphological specifications of which may be aspectual and/or passival), if it is legitimate to call this absence of any overt marking government.

(17) There must have been fairly recent changes in the distribution of pro-forms admissible in the environment of a modal when the verb phrase or parts of it are in anaphoric relationships.

(17a) Today it is often considered impossible to gap a main verb, with a non-verbal verb-phrase constituent ending up adjacent to a modal (or other auxiliary): **Smith will buy the tomatoes, and Jones may buy the potatoes*; **Beowulf lived in the hall Heorot, and Grendel did/may live / has lived in the moor*. It has been suggested (by Hudson 1976: 543) that such ellipses are ungrammatical because Gapping requires that the gap include the first verb of the conjunct concerned (thus: *Smith will buy the tomatoes, and Jones will buy the potatoes*); but I doubt that this account is correct, simply because it seems to me mistaken to assume that the modals, in ModE, are verbs. To say that Gapping cannot take place with unlike auxiliaries in the two (or more) conjuncts (cf. Jackendoff 1971), is descriptively more adequate, but of course still fails to explain this particular restriction. Maybe an explanation can ultimately be derived from what has been implied about the status of the modals (and other auxiliaries) in connection with the developments (1), (12), and (16): the ungrammatical gapped constructions are perhaps another instance of the modals' increasing resistance to combinations with anything else but morphologically unmarked infinitival verbs. Their incompatibility with adjacent noun or prepositional phrases even if an infinitive is easily recoverable from the preceding conjunct would then require that entire verb-

phrases co-occurring with modals (minus eventual aspect and passive-voice marking infinitival auxiliaries) be deleted, if anything is to be deleted anaphorically at all (see 17b). Disregarding further details of an explanation along these lines, and also the possibilities that the rule of Gapping itself has been modified diachronically or that a different rule applied in certain types of elliptical co-ordinate constructions in the translation equivalents of which Gapping would apply today, it is noteworthy that the constraint against surface adjacency of modals and objects or certain adverbials in Gapping constructions has emerged later than the constraint(s) barring this kind of adjacency when it does not result from identity deletions. Such gaps were evidently not offensive until quite recently, that is, some time after the developments (1) and (12) had been essentially completed: *lloren ich haue Josep, þat ich louede so swiþe, & nou ich ssall Beniamin* (c 1250 Jacob & Josep); *Lay down thy swerd, and I will myn alswa* (c 1386 Chaucer, C.T.); *She has deceived her father, and may thee* (1604 Shakesp.); *which at that time relieved my Body from a Distemper, and will my Mind for ever from a Folly* (1711 Spectator); *Even if my father forgives my birth, he will not my religion* (1859 Meredith); *I have not spoken to the king One word, and one I must: Farewell* (1876 Tennyson).

As a matter of fact, there is a danger of exaggerating the difference between the state of affairs exemplified by elliptical sentences like those quoted last and that obtaining presently. Although often considered impossible, gappings yielding auxiliaries and adjacent verb-phrase internal constituents (such as direct objects) indeed are not categorically rejected by all speakers under all circumstances in present-day English: at any rate, there definitely are dialects where they are commonly used in informal speech and under certain grammatical conditions (including the presence of a contrastive verb-phrase internal constituent following the gap; see Levin 1980:§ 2 for a detailed analysis of these conditions on what she calls 'Pseudo-gapping'). Thus, whether or not the grammatical conditions on Pseudo-gapping have been tightened in recent times, the continuing availability of this kind of ellipsis (to illustrate it once more, with examples where the relevant conditions are met almost perfectly: *Does that annoy you? It would ánnóy me*; *You can't cut off that branch but you could cút óff these two*) suggests that if something is changing it is changing gradually. This particular elliptical construction is receding from all styles or registers (recall the earlier, rather literary examples) to the most informal conversational ones, and that apparently at a different rate in different dialects.

(17b) Turning to the kind of ellipsis commonly described in terms of a rule of Verb-Phrase Deletion (as in *Smith must buy the potatoes, if he can, or Jones will*), we note that this has been a possibility since OE (cf. Visser 1969:1835ff.): *cwædon þæt hi þa burg werian woldon, ȝif þa wæpned men ne dorsten* (Ælfred, Oros.) 'they said that they would defend the town if the men wouldn't dare'; *You love I best, and shal, and oother noon* (c 1386 Chaucer, C.T.); *yf the Kyng of Romayns maye subdew Gaunte ... as ... wee thynke he schall* (1475-88 Cely Papers); *whoso will, may go in* (1551 Robinson); *Your great looks fright not me - But my deeds shall* (1633 Massinger); *I always loved you, and always will* (1739-40 Richardson, Pamela). What did change presumably was that Verb-Phrase Deletion now tends to spare existential or copula *be* after a modal whereas until about the 17th century these *be*'s could optionally be deleted just like aspectual and passive *be*'s, which continue to be deletable: *His lif beð bliðe, his ending sal* (c 1250 Gen. & Ex.); *The god was ever and ever shal* (c 1390 Gower, C.A.); *I am thy mortal fo & euer wylle to my deth daye* (1470-85 Malory, M. d'A.); *For I thy seruant am and shal* (1566 The Whole Boke of Psalmes). This emerging difference in the behaviour of copula/existential *be* on the one hand and passive and progressive *be* on the other under deletion, which now conforms to this pattern:

> *He is already dead, and you soon will be dead/will be/*will, if ...* vs.
> *He has already been killed, and you soon will be killed/will be/will, if ...*
> and *He is not working, but she may be working/may be/may*

perhaps ought to be interpreted as a concomitant of the categorial and constituent-structure elaboration of the distinction between auxiliaries and full verbs: as long as there was no more or less clear-cut distinction of this sort, there was little reason for the various *be*'s to behave differently depending on whether they combined with plain adjectives (or nouns), past participles, or present participles (both of which are arguably adjectival forms themselves). Surely, in the end copula/existential *be*, like other grammaticized or semi-grammaticized verbs (possessive *have*, for example), turned out to share several morphosyntactic features with the prototypical auxiliaries (as was mentioned in (7), (8), and (14) above), and passive and progressive *be* continued to differ in some respects among themselves as well as from the modal auxiliaries — and these various behavioural likenesses and unlikenesses have stimulated much controversy, not only among transformationalists. However, one fundamental factor clearly differentiates two types of *be*: copula/existential *be* is a main verb in so far as it may be the only verb in a full clause whereas

passive and progressive *be* are necessarily auxiliaries in the sense that they require a main verb (if in the non-finite form of adjective-like participles) to add up to a full clause. And this factor apparently became crucial for the purpose of anaphoric deletions in the verb-phrase, after the modals had been transformed from verbs or very verb-like elements into auxiliaries. The modals have thus been established as auxiliaries which minimally require an accompanying copula/existential *be* when the predicate part of copula/existential sentences is being deleted under identity.

(17c) Probably this restriction on anaphoric deletion is relatively minor in comparison with a development concerning a kind of anaphora which used to be an alternative to anaphoric deletion of the verb-phrase about until the end of the 18th century or even longer: viz. the use of nominal pro-forms such as *that* and *it* as objects of dynamic and deontic (though not of epistemic) (pre-)modals: *That shall I not, said sir Dynadan* (1470-85 Malory, M. d'A.); *I pray you good vncle, procede you farther in the processe of your matter. - That shall I, cosin, with good wille* (1534 St Th. More); *the Bisshops could not remoue him. - Yes, M. Horn, that they might* (1567 T. Stapleton); *His Mastership will do well to look to himself. - That he should, re-echoed Craigengelt* (1818 W. Scott). Functionally at least, this constraint ruling out pronominalizations such as **Smith must buy the potatoes, if he can it, or Jones will it - *That he will not* appears to be analogous to the one on Gapping (17a): nominals are prevented from co-occurring with modals, whether they contract this close relationship as a result of gapping the verb which semantically governs them or on account of their being pro-forms of constituents in the scope of a modal. Verb-Phrase Deletion (see 17b) and *Do it, Do so,* or simply *do* or *so* anaphora (cp. *Smith must buy the potatoes, if he can do it, or Jones will do so - That he will not do; Smith doesn't like potatoes, but Jones may (do); I'd like to see the potatoes - Well, so you shall one of these days*) are anaphoric strategies which allow one to circumvent this constraint on the surface combination *Modal - $NP_{pro}$. These alternative non-zero pro-forms themselves, incidentally, are no recent innovations but have always been in use with the (pre-)modals (and certain main verbs as well):

*do so:* *I wolde for my parte well agre them for heresies, but yet haue I hard som or this that wold not do so* (1528 St Th. More); *Yet oughte we most chiefly so to do, when ...* (1552 Bk. Common Prayer)

*do:* *ich him luuie & wulle do* (c 1200 St Juliana); *All jantyllmen ... ought of ryght to honoure sir Tristrams ... and shal do unto the Day of Dome*

(c 1470-85 Malory); *we have not yet found them all, ... nor ever shall do* (1644 Milton); *do* anaphora after auxiliaries may have fallen into disfavour in American English dialects (see Pullum & Wilson 1977: 761)

*so*: *Cwædon þæt heo rice ... agan woldan, and swa eaðe meahtan* (Genesis) 'they said that they wanted to have the power, and they easily could have it'; *Me thunketh myn herte breketh a tue, Suete God, whi shal it swo?* (1310 Lyric. P.); *In ȝouthe I maye both ryde and goo, when I ame alde I may nott so* (1400 Isumbras); *I will stand, and so shall Trinculo* (1610 Shakesp.)

And the historical availability of these pro-form variants, moreover, would seem to weaken, if not eliminate, the import of another factor one might invoke to explain the disappearance of nominal pro-forms with modals: that is, one might assume that pro-forms such as *that* or *it* went out of use because the non-finite complete constituents combinable with the modals ceased to be of a category (noun phrase or sentence) where nominal pro-forms would be appropriate. I do not wish to deny that there have been such categorial changes altering the character of modalized constituents; on the contrary, I regard such modifications as very likely (cf. my speculations at the end of (13)), and thus see no reason to reject entirely the consideration that the ban on nominal pro-forms is to some extent contingent on them. But I doubt that it alone suffices to account for the changes and continuities of the anaphoric processes admissible in the environment of the modals, making any reference to a constraint against (surface) combinations Modal - NP superfluous.

(17d) It ought to be noted that the constraint against Modal - NP combinations at the surface is not absolute even now, not even in non-informal, non-conversational speech: if this combination results from the process known as Comparative Deletion, it is as unobjectionable (cf. *Smith has eaten as many tomatoes as Jones has/will potatoes*; *Smith may buy more potatoes than Jones will/may/did tomatoes*) as it always was (e.g. *for that ech of hem sholde helpen oother In meschief, as a suster shal the brother* (c 1386 Chaucer, C.T.); *We follow'd then our Lord ... So should we you, if you should be our king* (1594 Shakesp.); *I love him more Than he can Gaveston* (1593 Marlowe); *I could no more stir the canoe than I could the other boat* (1719 Defoe, Rob. Crusoe); *Count its stones as you would jewels of a crown* (1840-71 Ruskin)). As far as I can see, there are no indications of an imminent prohibition of

such ellipses in comparative constructions, where the presence of markers of comparison may contribute to identify the ties between modals and superficially adjacent object noun phrases as very tenuous indeed.

(18) In (9) modality expressions were mentioned which utilized the originally possessive verb *have*, viz. *have to* and *had better*. Of these two, *had*, obligatorily accompanied by *better* or also *rather*, *sooner*, *best*, *as well*, is integrated into the morphosyntactic system of the modal auxiliaries rather better: it entirely lacks the finite and non-finite inflectional morphology characteristic of verbs and the corresponding syntactic possibilities (there had been present and preterite indicative forms, but the last attestation of such forms in the OED is from 1595: *Poesie ... like Venus ... hath rather be troubled in the net with Mars, then enioy the homely quiet of Vulcan* (Sidney)), and it negates like an auxiliary (*hadn't better/had better not*) etc. (see Palmer 1979: 9, 69f.). But *had better* etc., the originally past subjunctive form which replaced the modality expression *him/he were better* in ME times, was not the latest modal acquisition from the area of possession. *Have got*, originally the perfective of *get* ('have acquired'), in addition acquired simple possessive meaning ('possess/have') after the 16th century, and simultaneously or perhaps rather subsequently also epistemic, deontic, and dynamic modal meanings roughly similar to those of *have to* (cf. Visser 1973: 2202-6). In the mid-19th century at the latest, *have got to* was commonly used as a modality expression, and what is perhaps even more significant than the semantic change, it was adapted to the morphosyntax of the hard-core modal auxiliaries to a greater extent than *have to*: it is not employed non-finitely; it has a defective or at least not very productive tense opposition insofar as its past tense occurs very rarely (e.g. *Nevertheless it had got to be done* (1903 S. Butler)); and it negates and inverts without supporting *do*. Finally, the reduced variant of *have got to*, viz. *got to*, the earliest written instances of which likewise date from the mid-19th century, is even more like the modal auxiliaries in so far as it lacks finite morphology as well. The range of occurrence of *got to* seems more limited, though, in so far as it is avoided with negation (*You *gotn't to/*gottan't/??gotn't decide now*) and inversion (**Got we (to) decide now?*). (For a detailed analysis of *have got* in British and American English, without, however, paying much attention to its modal use per se, see Fodor & Smith 1978.) In one respect, *have to*, *have got to*, and *got to* appear to be equally unlike auxiliaries at first sight: they must all occur with the complementizer *to* rather than with bare infinitives.

(19a) However, *have to*, *have got to*, *got to* and a host of other elements taking *to* infinitives in fact allow the conjunction to be reduced and contracted with them in informal speech, by processes which are partly idiosyncratic (e.g. the reduction of the syllabic nucleus in *gonna* vis à vis *going to*) and partly more generally characteristic of fast speech (e.g. the reduction of [tuw] to [tə], the devoicing of final consonants (as in *hafta* vis à vis *have to*), the weakening of lengthened consonants to flaps (as in *oughta*), the deletion of coronal consonants (as in *wanna*, *usta*)). But significantly, these contracted forms appear where usual fast-speech forms would not: in not very fast informal speech and before pauses. Here is an extensive list of elements which commonly or not so commonly admit (and of some which don't) conjunction contraction (see Bolinger 1980): *wanna*, *gonna*, *oughta*, *usta*, *sposta* (vis à vis *supposed to*), *bounta*, *hafta/hasta/hadda*, *gotta* (**obligeta*), ?*specta* (for *expect to*), ?*liketa*, ?*careta*, ?*meanta* (**wishta*, **desireta*, **intendta*, **refusta*), ?*seemta* (**appearta*), ?*tryta* (**endeavourta*), ?*ableta* (**proneta*), ?*sureta* (**certainta*), ?*scaredta* (**frightenedta*), ?*loveta*, ?*gladta*, ?*planta*.

It ought to be obvious that this list is not a semantically arbitrary collection of main-clause predicates: the items which acquire the typical auxiliary property of not subordinating an infinitive with the conjunction *to* all have epistemic, deontic, or dynamic modal, or aspectual (e.g. the habitual *usta*), meanings not very different in principle from those of the auxiliaries of somewhat longer standing. Of course nearly all of these predicates continue to occur with meanings and in functions which are not like those most typical of modal (and other) auxiliaries; but note that under these circumstances they also resist the contraction of *to*: for example, *to* does not contract with *use* used in the instrumental meaning (*Poison was used to/*usta get rid of Jones*), nor with *got* with the meaning 'be privileged' (this reading of *I got to spend that money* disappears when *gotta* is used), nor with *going* if *to* is supposed to be the homonymous preposition (*We're going to/*gonna Bath* vs. *We're gonna bath Jones*), nor with any of these predicates when they are used with a nominal object in addition to the infinitive complement and this object in surface structure vacates its position after the predicate (thus, *Bloggs is the man I wanna succeed* can only be related to *I want to succeed Bloggs*, not to *I want Bloggs to succeed*; cp. also the above *used to/*usta* example). But I have also tried to indicate in the list above that not all predicates with suitable meaning and construction are automatically eligible for *to* contraction. In some of these cases, the continuing resistance to contraction may be accounted for by purely formal properties of the predicates. On the other

hand, I mentioned earlier (under (2)) that grammatical systems tend to be better structured than lexical ones and consequently ought not admit extensive synonymy; and since conjunction contraction is plausibly interpreted as one step towards grammaticization (auxiliarization, more specifically), it is not surprising that the number of predicates thus (semi-)grammaticized should be limited.

Understandably, the chronological details of this step towards grammaticization here are rather difficult to determine because it is not usually reflected in writing. No doubt conjunction contraction is a fairly recent development with many of these modal and aspectual expressions; with some, however, it may have been possible as early as the 17th and 18th centuries. *Usta* appears to be among the earliest, as is indicated by occasional spellings *use to* instead of *used to*: *I am not soe much at liesure as I use to be* (c 1652-4 Dorothy Osborne in a letter); *How dost thou pass thy time? - Well, as I use to do* (1672 G. Villiers); *You are not so fond of me, Jenny, as you use to be* (1728 Gay, Beggar's Opera).

(19b) Often the contractability of the conjunction *to* is not the only symptom of an ongoing process of grammaticization, but is part and parcel of a more general, if gradual exchange of morphosyntactic properties of main-clause predicates for those of auxiliaries. Thus, *used to/usta*, for example, developed a defective tense paradigm, being essentially restricted to past-tense uses since the 18th century; it never quite managed to pattern with full verbs in constructions where these require *do* Support; and non-finite forms have always been rare when *use to* was intended to have the habitual meaning. Similar tendencies, more often than categorical restrictions, are also characteristic of other modality predicates when they begin to exhibit un-predicatelike behaviour; and it seems that tense and non-finiteness are the morphological categories which tend to be affected first once a predicate is drawn into the drift towards auxiliariness. *To* contraction, at any rate, is clearly favoured when the predicates eligible are finite and in certain tenses (usually the present), and sometimes also in certain persons.

(20) The elements mentioned in (18) and (19) do not exhaust the fund of lexical items undergoing the process of grammaticization as modal auxiliaries, or semi-auxiliaries, in the history of English. It ought to be obvious by now that auxiliarization progresses gradually, obviating any neat categorical delimitation of the set of elements affected. The Black English *come*

of speaker indignation (as in *He come coming in here raising all kind of hell*), and to a lesser extent also the Black as well as non-Black American English *go* of disapproval (as in *Don't go going around ringing people's doorbells*), should be mentioned as further illustrations of this point (see Spears 1982 for details). At least the modal *come* (but partly also *go*) has come to be differentiated from the corresponding motion verb by a number of criteria that also characterize other no longer purely lexical modality expressions: when used as a modal, it is inflectionally defective and maybe uninflecting (the regular 3rd person singular and past tense forms are at any rate lacking: **comes*, **came*; cf. (2), (3) above), always immediately followed by a verbal *-ing* form (never by other subordinate constructions or by directional adverbials; cf. (5), (12) above), restricted in its co-occurrence with other auxiliaries, especially the aspectual ones (**He has come coming in here* ...; cf. (6) above), and often reduced to [həm?], [həm] or even [m̩] (cf. the reduced variants of other modals such as *'ll*, *'d*, *c'n*). It seems that none of the possible origins of the Black English *come* of indignation — models in non-Black dialects of English, survival from the creole past of Black English, or Black English innovation — can as yet be ruled out conclusively (cf. again Spears 1982:865-7), which precludes any precise dating of the dissociation of semi-auxiliary from motion verb *come*.

(21) I would like to mention one further set of changes pertaining to modality without being certain that they naturally go with the others. They have to do with implicit modality, that is, with sentences which require or invite a modal interpretation without containing a modality expression. The putative developments are, at any rate, not directly comparable to diachronic losses of auxiliaries of the kind reported, for instance, from Salish languages (cf. Thompson 1979:739f.), where positionally exposed, i.e. sentence initial, semantically rather unmarked auxiliaries tended to be omitted in rapid, informal speech and then disappeared altogether in some languages, such as Squamish.

(21a) There is, firstly, Visser's (1972:667) speculation that, beginning only in the 16th century, certain activity verbs in the present tense could be interpreted not only as referring to the present performance of the activities concerned but also, and in some contexts exclusively, to the ability of the subject referent to perform the activities — as in *A Grammarian is better liked, that speaketh true and good Latin, than he that speaketh false* (1551 T.

Wilson, Logike). In fact, such modalized readings of modalless sentences are not restricted to the present tense; under suitable circumstances past-tense and future sentences can be understood as reports and predictions/expectations about capabilities or dispositions as well: *Bloggs spoke English as well as his own language* ('could speak'); *I'm sure Kaspar will speak in a few months* ('will be able to speak'). Probably Visser is right that implicit-modality sentences of this kind were innovated rather late, and perhaps even roughly simultaneously with other modal developments. However, it is certainly no coincidence that the only type of modality ever relevant here is the dynamic one (apparently only dynamic possibility), and it is well known that there are close parallels between dynamic modality and what von Wright (1951) calls the existential mode, i.e. existential and universal quantification. Thus, one may presumably expect that in all languages, independently of any other morphosyntactic properties of modality or mood expressions currently utilized, certain quantified sentences without additional modal invite dynamic-modal interpretations (e.g. *Nobody lives twice/You only live once* ('can live'); *Everybody dies some day* ('must die')) or allow their quantifiers to be replaced by modals without change of meaning (e.g. *Some moles are dangerous/Moles are sometimes dangerous - Moles can be dangerous*). (See also Palmer (1979: 152-5) on existential and dynamic modality, and Plank (1981: 67-9) for some ideas about implicit modality in general.)

(21b) Another instance of implicit modality is tied up with particular syntactic constructions which by and large represent ModE innovations: secondary subjectivizations of the two kinds illustrated in (b1) and (b2) both suggest very strongly, or even necessitate, a notion of dynamic possibility.

(b1) *Our self-contained flats sleep two to sixty-two people*; *This garage will park more than 200 cars*; *A Volkswagen seats eleven adults*

(b2) *Kant reads easily*; *Bloggs never photographed very well*; *A good tent puts up in about four hours*; *I don't kill so easy*

As long as such secondary subjectivizations were impossible in English, an explicit modality expression was necessary to convey the intended modal readings (e.g. '2 to 62 people can sleep in our flats', '11 adults can sit in a VW', 'Kant can be read easily', 'Bloggs could never be photographed very well', etc.), or different predicates had to be chosen (e.g. 'A VW holds 11 adults'), or different constructions with the referent whose disposition or potentiality is at issue not in the subject relation (e.g. 'it is (not) easy to read Kant/to kill me').

(22) The opposite of implicit modality is redundant modality, viz. the presence in a sentence of more than one modality expression where one would seem to suffice, or also of one modality expression where none seems required. Redundant modality can also be illustrated from the history of English. When comparing them with their most natural ModE translation equivalents, some kinds of complex sentences in OE not unusually give the impression of containing a (pre-)modal too many: with an expression from the spheres of dynamic or deontic modality in the main clause, an additional (pre-)modal in the subordinate clause, likewise expressing dynamic or deontic modality (rather than the original notional meaning), often appears superfluous. The following examples (mostly culled from Standop 1957) are translated less than most naturally, in order to emphasize the redundancy of the subordinate (pre-)modal: ... *se forgeaf us ða mihte ðæt we mihton gehælan adlige and untrume ealle on his naman* (Lives of Saints) 'who gave us the power that (or: so that) we could heal all the ill and diseased in his name' (the redundancy would be less obvious here if the subordinate clause really were one of purpose, rather than being a complement to the main-clause object; such ambiguities in fact seem rather wide-spread in OE); *Ða wæs he sona geornful ðæt he wolde diegellice ðone cristendom onwendan* (Ælfred, Oros.) 'then was he immediately eager that he secretly would/wanted to subvert Christianity' (cf. without redundant subordinate modal: *ðu ær sædest ðæt ðu swiðe geornfull wære hit to gehyranne* (Ælfred, Boeth.) 'you said before that you were very eager to hear that'); *Hi ðohton ðæt hie woldon hyne ofslean* (St John's Gospel) 'they planned/wished that they would kill him' (cf. without redundant modal in the finite subordinate clause: *Hi anrædlice ðohton ðæt hi hyne ofslogan* (OE Homilies) 'they persistently ...'); *Sua sua David forbær ðæt he Saul ne dorste ofslean* (Ælfred, C.P.) 'as David held himself back that he did not dare kill Saul' (cp., less redundantly: *sua sua Dauit forbær ðæt he ne slog mid his sueorde Saul* (ibid.)); *Ðu bæde me, leof, ðæt ic sceolde ðe awendan of Lydene on Englisc ða boc Genesis* (Heptateuch) 'you asked me, dear Lord, that I should translate for you the book Genesis from Latin to English'. In such subordinate-clause contexts, the modals cease to appear in ME times, with this decrease in redundant modality, thus, on the whole antedating the increases in implicit modality illustrated in (21). The joint result of the two lines of development is, then, an overall decrease in the frequency of overt modality expressions separate from full verbs.

(23) In connection with some of the changes treated previously (1, 2, 3, 9, 11, 15, 18, 19, 20), semantic developments of the (pre-)modals and certain other predicates have already been alluded to which can be characterized summarily as a loss, total or partial, of 'notional' meanings and concomitantly an acquisition or retention of meanings usually subsumed under the rubrics of dynamic, deontic, and epistemic modality (or rather mood, if we resolve to refer to these modalities, when grammaticized, as mood). Concentrating on the (pre-)modals proper and disregarding continuous later shifts in meaning once they have been more or less fully grammaticized (such as the loss of the permission and wish senses of *must* in later ME/early ModE and the acquisition of these senses by *can* and *may* respectively, or the blurring of the distinction between unverified-fact and counterfactual senses, both of which are expressible by *may (have)* in present-day American English (cf. *He may have seen them / Had the young Americans secured themselves with their climbing ropes, ... they may have survived the collapse*; observation due to Dwight Bolinger, personal communication)), as well as earlier shifts of their notional meanings antedating the OE period (see Wilde 1939: 264-7, or Standop 1957 *passim* for surveys of the traditional etymologies), it is noticeable that they lose meanings such as 'know, have carnal knowledge, have mental/intellectual power' (*cunnan*), 'be strong, have physical power/capacity' (*magan*), 'owe, be bound by an obligation' (*sculan*), or, not entirely though, 'choose, wish, be determined' (*willan*). At the same time they acquired new modal meanings, especially deontic and epistemic ones, or, more commonly, retained or slightly modified original meanings when these already were within the sphere of modality (which tended to be the case, I believe, with dynamic and deontic rather than with epistemic modality). All these individual semantic developments were gradual, and did not necessarily coincide chronologically; there nevertheless are indications that with the hard-core modals the loss of meanings with a claim to notional status ('have intellectual power to', 'have physical power/capacity to', in particular) culminated and came to a conclusion in early ModE. (See Visser (1969) for extensive documentation, and Traugott (1972: 198f.) for a tabular synopsis.) To pronounce this kind of development as such concluded, however, is slightly misleading, since it need not be confined to the limited set of the (pre-)modals, and indeed has not been in the history of English (recall 9, 18, 19, 20).

Note that strictly speaking no semantic characterization of the category of modality/mood has been provided so far, which would be necessary to determine which lexical and grammatical forms partake in its expression in

the first place. The existing general definitions of modality/mood tend to be too restrictive (cf. e.g. Jakobson's (1971:135), according to which mood "characterizes the relation between the narrated event and its participants with reference to the participants of the speech event", effectively excluding dynamic and deontic moods/modalities), or may appear over-inclusive (e.g. Plank's (1981: 66-9), in the opinion of Wunderlich (1981: 109-10)). Nevertheless, it ought to be possible to elaborate sufficiently precise notions of dynamic, deontic, and epistemic modality/mood without succumbing to either of these dangers, and this should possibly have the added advantage of shedding some light on what are the minimal degrees of differentiation in particular of grammatical mood systems. Instead of seriously pursuing this point here, let me offer, as a kind of operational definition, a rather indirect characterization of the relevant modal meanings: they ought to include at least all those meanings expressed by the verbal mood inflections, and specifically by the subjunctive, in OE. Which brings us to the last development in our story, which was amongst the first, if not **the** first, in history.

(24) As is well-known, the Germanic branch, like most others, has not preserved the full set of verbal mood inflections once available in Indo-European. Irrespective of what the 'original' inflectional mood system actually looked like and how it developed in the early daughter languages (for authoritative, though by no means uncontroversial statements see e.g. Gonda (1956) and Kuryłowicz (1964)), it is clear that of the non-indicative mood paradigms, including at least a subjunctive and an optative, and, more marginally, an injunctive (if appropriately included among the moods) and perhaps others of lesser note, only the subjunctive survived in Germanic, utilizing essentially the forms of the original optative. (Details of how the imperative, another traditional non-indicative mood, relates to these categories, formally and functionally, need not concern us.) Thus, although we find in OE regular subjunctive paradigms in the present and preterite tenses of all verb classes containing forms which are distinct from those in the indicative paradigms for most, though not for all, persons/numbers (as illustrated in Table 1), this synchronic situation must be seen in the larger context of an overall diachronic tendency towards the reduction of inflectionally expressed mood oppositions, or at least of the inflectional apparatus to express such oppositions.

| | INDICATIVE | SUBJUNCTIVE | INDICATIVE | SUBJUNCTIVE |
|---|---|---|---|---|
| P sg 1 | *lufie* | *lufie* | *drife* | *drife* |
| R 2 | *lufast* | *lufie* | *drifst* | *drife* |
| E 3 | *lufað* | *lufie* | *drifð* | *drife* |
| S pl | *lufiað* | *lufien* | *drifað* | *drifen* |
| P sg 1 | *lufode* | *lufode* | *draf* | *drife* |
| R 2 | *lufodest* | *lufode* | *drife* | *drife* |
| E 3 | *lufode* | *lufode* | *draf* | *drife* |
| T pl | *lufodon* | *lufoden* | *drifan* | *drifen* |
| | | 'love' (weak) | | 'drive' (strong) |

TABLE 1: Inflectional mood distinctions in Old English.

The employment of these subjunctive forms in OE is notoriously difficult to describe in detail, taking into account all their obligatory and optional uses in main and subordinate clauses. For our present purpose, it suffices to observe that their employment was in fact decreasing considerably even during the OE period. In particular in the preterite, where indicative and subjunctive tended to formally coincide first as a result of the attrition of inflectional morphology, but in the present tense as well, there was an increasing preference for using (pre-)modals such as *willan*, *sculan*, *magan* etc. under exactly the same circumstances which formerly had called for synthetic subjunctives. In fact, as indicated in (22) above, there even was a tendency to over-use the (pre-)modals in certain types of subordinate clauses with a modality expression in the main clause. (To avoid the impression that the (pre-) modals were strictly speaking replacing the inflectional subjunctive, it ought to be mentioned that in subordinate clauses where subjunctive-mood forms used to be employed, the (pre-)modals, when occurring, originally tended to be in subjunctive form themselves.) In ME and early ModE, phonetic attrition, analogical levelling, and morphological replacement (cf. 13c above) further obscured and eliminated most formal distinctions between the indicative and the subjunctive paradigms; and an inclination to supplant the few remaining subjunctive verb forms by the indicative, modal auxiliaries, or other modality expressions made itself felt until the present day, notwithstanding literary attempts to revive some uses of the subjunctive since the mid-19th century. (Numerous studies have been devoted to these developments; I found Wilde's (1939/40) particularly instructive, although

it is occasionally biased towards the view that the (pre-)modals from the outset merely served to replace subjunctive inflections.)

With the modals story fleshed out, we can now proceed to Lightfoot's attempt to explain the changes he has chosen to take into consideration. His 'explanation' has three main components.

(A) Changes (1)-(5) are claimed to have occurred independently of each other and of anything else in the language: they allegedly did not occur simultaneously, nor were they causally related among themselves or to other developments, so far as Lightfoot can see. But they led to a complication of the grammar of English in so far as the premodals became a group of verbs which were exceptional with respect to the grammatical regularities noted in (1)-(5). As chance developments they cannot, and need not, really be explained in any principled way — at least they do not follow from anything in Lightfoot's theory of grammar and theory of change, and are therefore provisionally assumed to be "caused by extra-grammatical factors of some kind" (p. 406).

(B) With the last of the complicating changes, viz. number (5), the grammar of English, according to Lightfoot, had arrived at a point where language learners could not continue (re-)constructing essentially the same grammars as speakers of the preceding generation: the Transparency Principle necessitated a radical re-analysis, prescribing the construction of a grammar not offending this principle while still generating a language within the limits of mutual inter-generational comprehensibility. The re-analysis, taking place shortly after 1500, consisted in the introduction of a new initial-structure category 'Modal' as distinct from '(Main)Verb', subsuming the previously exceptional subclass of verbs, the premodals. The logically independent (in Lightfoot's view) and chronologically simultaneous changes (6a-d) were the manifestations of this re-analysis. What the Transparency Principle is intended to predict actually is only that the intolerable opacity had to be eliminated somehow: it does not claim to predict how. That a new category 'Modal' (and 'Auxiliary') was introduced and that it took the form it did was entirely a matter of chance, or historical contingency, as far as Lightfoot's theories of grammar and change are concerned; other solutions, such as regularizing the modals to bring them in line again with normal verbs, could have been adopted equally well in principle.

(C) Some adaptations of the grammar were required once the category 'Modal' had been introduced: changes (7)-(9), purportedly happening immediately after the re-analysis or indeed simultaneously (thus p. 129 fn. 1), represent these adaptations. This time it is the theory of change rather than the theory of grammar, specifically the requirement that communication between generations must not break down, which is supposed to predict further changes in the wake of the modal innovation, presumably again without prescribing particulars of such grammar modifications. In spite of Lightfoot's avowedly anti-teleological stand I take it that such adaptations are to be regarded as prophylactic rather than therapeutic: presumably the idea is not that they are carried out as a consequence of an actual breakdown of communication but rather in order to prevent one.

These explanations and non-explanations strike me as a paradigm case of pretentiousness. In spite of all his rhetoric, Lightfoot is as wrong as one can possibly be in what he thinks needs an explanation, in what he thinks cannot be explained, in what he suggests constitutes an explanation, and in his belief that his theories of grammar and change predict anything. About the only thing where he is right in principle is that much, and indeed most, of the explanatory load is to be assigned to the theory of grammar, although to one invoking principles rather different from the ones he has in mind.

The chronological sequence of events suggested as an, or rather **the**, explanandum is pure fiction. No amount of idealizing and abstracting away will yield the neat pattern of scattered complications ((5) occurring last) being followed by re-analysis manifestations and simultaneous or subsequent adaptations. As to the simultaneity of the re-analysis manifestations (6a-d), I have already argued that I regard this as a single change (in the varieties where it occurs), amounting to the loss of the finiteness opposition, so that it is not surprising that what happened happened simultaneously. Practically all developments considered (ignoring the questionable case of (4)) are paradigm instances of gradual change: none is initiated and completed, affecting all and only the items on a well delimited list of modality expressions, within the time span of a few generations. Recognizing this gradualness is not necessarily tantamount to denying that the overall development of one more or less well defined group of modality expressions, the premodals, culminated, in the sense of approaching provisional completion, in the 16th and early 17th century in the standard language — which still is a time span extending over large numbers of generations. Even if things thus seem to have come to a head in these two centuries, almost all of the individual

changes overlapped considerably, some beginning, or showing signs of beginning, quite early (in fact in OE or even earlier), others gaining momentum a little or much later. If pressed to outline a rough chronological sequence myself, I would suggest the following picture, where broken lines indicate that a change is going on in the respective period, and unbroken lines that an ongoing change seems to have been particularly intensive at the time.

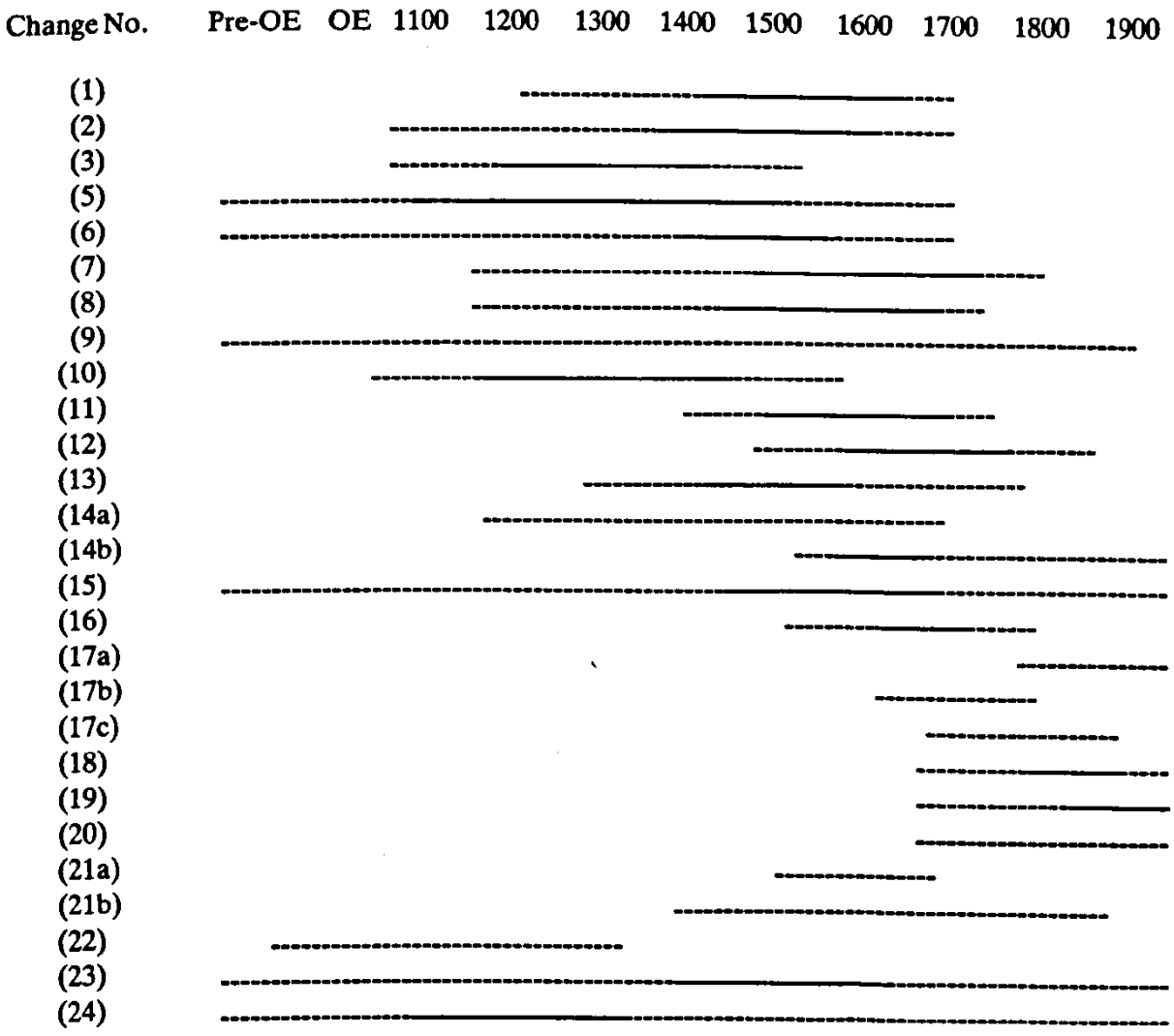


TABLE 2: Chonology of developments pertaining to the (pre-)modals.

In general, this gradualness and overlapping of individual developments does not square very well with Lightfoot's claim, dictated by his grammatical theory which recognizes only discrete lexical categories (though it should be possible to construct 'mixed' categories by employing and appropriately combining categorial features), that the premodals were re-categorized from

being verbs (albeit increasingly exceptional ones) to being 'Modals' (and, together with 'Tense', 'Auxiliaries') all at once in one catastrophic move. The picture summarized in Table 2 suggests rather that it is advisable not to dissociate, in one's analytic endeavours, categorial terms such as 'Verb', 'Auxiliary' etc. from the primary morphosyntactic and lexical phenomena on which the application of these terms is based. The usual word- or morpheme-class categories represent nothing but analytic generalizations: they are essentially only labels for sets of repertoires of morphosyntactic behaviour of words or morphemes (which may be lexically predetermined), motivated by the observation that the various behavioural repertoires tend not to differ arbitrarily in their make-up and tend not to vary without assignable limit from one word/morpheme to the other. Descriptively, the individual morphosyntactic properties of words/morphemes making up the various behavioural repertoires in particular languages, as well as (perhaps implicational) relationships between the properties characterizing the repertoires, are the primary phenomena — and it would be misleading to approach them as if they were secondary phenomena, in the sense of existing only by virtue of being properties of particular preconceived, basic and autonomous categories. Lexical and other meaningful expressions do not behave in this or that way ***because*** they are, by decree, 'Verbs', 'Auxiliaries', 'Affixes' or the like: rather, such abbreviatory category labels are applied — secondarily, as a result of one's analysis of the primary phenomena — if the behavioural repertoires of some items are sufficiently similar, and simultaneously sufficiently different from the repertoires of other items, to imply that some generalization may be in order. To avoid arbitrariness in eventually drawing, or rejecting, such generalizations, one would of course need to be more specific about the conditions under which repertoires count as sufficiently similar and different. That is, even if no two items in a language should turn out to share exactly the same behavioural repertoires, one may nevertheless wish to postulate a limited set of distinct classes, thereby ignoring behavioural properties in conflict with the classifications postulated (even if these ignored properties would in fact suggest alternative classifications), on the assumption that some features (viz. those ignored) can be shown to be less significant for, or less characteristic of, the overall sets of repertoires than others. (For example, *may* differs from the other hard-core modals in most dialects of present-day English in not allowing a following *not* to contract to *n't*, and shares this resistance to negative contraction with full verbs (recall 14b); yet, in classificational generalizations this behavioural property is usually ignored

as not important enough to impede the establishment of classes, viz. of (full) verbs and auxiliaries, which contain items whose morphosyntactic (or perhaps rather, morphonosyntactic) repertoires are neither entirely identical with those of the other items in the same class nor entirely different from those of items in other classes.)

Whatever descriptive generalizations about behavioural repertoires turn out to be feasible in particular languages, our theoretical goal must be to explain them, viz. to suggest higher-level generalizations about which descriptive generalizations are likely, and which are not so likely, to be feasible, and why. Now, it ought to be obvious that no higher-level generalizations should be possible at all if the relationships between meaningful items and the morphosyntax of their expression were entirely arbitrary, requiring fully explicit statements of the morphosyntactic repertoires associated with each meaningful item in any particular language, and if, furthermore, language-particular behavioural repertoires were able to differ in make-up from each other just as arbitrarily, again requiring fully explicit specification in individual grammars. In spite of occasional claims to the contrary, higher-level generalizations ***are*** possible, in particular if based on morphosyntactic repertoires and their composition rather than on preconceived word/morpheme-class categories. Moreover, there would seem to exist plausible functional explanations for cross-linguistically preferred associations of behavioural repertoires with particular meanings or functions. (An attempt to elucidate some of these in the case of modal 'Auxiliaries' may be found in Plank 1981, 1983.) In spite of his preoccupation with categories, 'exceptional' category features, and category-change, I can detect no trace in Lightfoot's account of an argument seeking to motivate the morphosyntactic behaviour of modality expressions in terms of their semantics and pragmatics. If it is essential for true explanations (as opposed to descriptive generalizations) in this area that form (i.e. morphosyntactic repertoire) is shown to be motivated by meaning and function, then Lightfoot's account, though claimed to be explanatory, would seem to lack essential prerequisites, no matter whether it is change or continuity in the morphosyntax of modality that is to be explained, and no matter how many or few modal auxiliary properties one would prefer to be taken into account.

It ought to be self-evident that the explanatory force of the requirement that there be no linguistic generation gap, as it stands, is nil — even if one granted (counterfactually, I believe) that generations indeed are the social units of crucial importance in this respect. Certainly in the case at issue, the

changes claimed to follow from this maxim masquerading as a 'substantive statement' in a 'theory of change' simply remain unexplained, even if they are appropriately characterized as adaptations of the grammar in the process of a categorial overhaul. This not-so-explanatory statement, at any rate, already fails on observational grounds: there are cases on record where the language of parents and their children is not mutually comprehensible, and this situation actually seems not uncommon, especially in complex societies and during processes of drastic socio-economic change. (See Giacalone Ramat 1980:552, pointing out such inter-generation linguistic gaps in urban and rural Italy.)

If one recalls the initial attempt at defining Transparency, quoted earlier from Lightfoot's Foreword, one will probably be surprised to find the Transparency Principle being invoked to explain why the English could not continue operating with the premodal verbs. The Transparency Principle was announced as requiring the distance between initial and surface structures to be minimal — but none of the changes (1)-(5) included by Lightfoot in his preparatory complicating phase in any way interfered with, i.e. increased, this distance! Thus, Lightfoot's first step in 'gradually developing' his Transparency Principle apparently is to recast the initial notion of Transparency entirely: the property of the English grammar of around 1500 which is now said to offend the Transparency Principle is that a subset of verbs, the premodals, had accumulated several exceptional features, so that it became more difficult for the language learner to figure out whether they really were true verbs. This opacity of the category membership of the premodals was eliminated as these lexical items were assigned an initial-structure category of their own. If one looks for a more precise and general specification of the degree and kind of categorial exceptionality permitted by the Transparency Principle in its revised form, one looks in vain — unless one is content to be informed that "it seems that it did not take many exception features to bring about the re-analysis" (p. 114). Claiming predictive power for vague and entirely ad-hoc impressions of this kind, pretentiously elevated to the rank of a theoretical principle, seems ludicrous. If 'not many' exception features are required to necessitate a re-structuring of grammars in the form of a diversification of their categorial framework, one wonders how generations of speakers of many languages (such as French — cf. Gross 1975) without major difficulties or catastrophies manage to get along with particular categories (such as verb) in spite of no two lexical items of this category having exactly identical repertoires of morphosyntactic behaviour. As

pointed out above, one surely would expect a theory of grammar to be a little more specific about how diverse the morphosyntactic behaviour of lexical items must be for these items to be prevented from sharing other parts of some behavioural repertoire. And one would of course also expect a theory of grammar to provide some substantive specification of categories potentially available to individual languages, or otherwise one would have to be satisfied simply to observe that a category 'Modal' appeared on the scene with an arbitrary collection of morphosyntactic properties inherited, as a matter of historical contingency, from exceptional members of the category 'Verb'.

In fact, Lightfoot sometimes gives the impression of having not really made up his mind about what Transparency, vaguely, has to do with. In telling and retelling his modals story (or rather, fairy-tale), he occasionally (e.g. on p. 114) relapses into suggesting that the re-analysis was necessary because the initial structures containing premodals were no longer "within earshot" of the surface structures. In the end he then seems to settle for a broad, but none the more specific version of the Transparency Principle: it is to characterize the limits of the permitted degree of exceptionality (the appropriate ad-hoc specification presumably being: 'Five exception features, specifically those mentioned in (1)-(5), are too much for a respectable verb') ***or*** of derivational complexity (the specification here presumably being more general: 'The distance between initial and surface structure, effected by transformations, ought to be minimal' — so far no test case). But a few pages later, a still more encompassing notion of Transparency is suggested: surface ambiguity is claimed also to contribute to the opacity of initial structures (p. 129, squeezed into a footnote; but cf. also pp. 245ff.). We are again left in the dark, though, concerning the permitted degree and kind of surface ambiguities. Only seven pages later, a further extension of the Transparency Principle is suggested in passing: quite likely it also keeps phonological derivations, rather than only syntactic ones, minimally complex. Fortunately, Lightfoot avoids further excursions into phonology, apart from admitting that phonological change may provoke syntactic change when it creates (I wonder what kind of) opacity, and from asserting, mysteriously, that this kind of interrelatedness of syntactic and phonological changes actually argues for his thesis of the autonomy of syntactic rules (p. 153 fn. 1) — not without emphasizing once more that all of his changes pertaining to the modals are instances of pure syntactic change, unprovoked by semantic or phonetic factors.

But Lightfoot's 'gradual development' of a Transparency Principle really culminates only with his suggestion (on pp. 137 and 150) that there possibly is no such principle at all: instead of an independent Transparency Principle as part of the theory of grammar there might be "a statement (perhaps in a theory of acquisition) that the child will construct the simplest possible grammar which yields an output close enough to that of his models" (p. 150). (Don't be confused when you recall having read, on p. 121, that the Transparency Principle, on the assumption that some such principle can be formulated, eventually may have to be subsumed under a more general perceptual strategy: unlike other Chomsky Grammarians, Lightfoot feels hopeful that many principles of grammar ultimately can be reduced to principles of the theory of perception; hence it doesn't really matter where you put your principles as long as they are in the theory of mind, preferably in the shaded area of Lightfoot's graphic representation of it (p. 71, if you want to know all about it).) This surprise move brings us back to square one — to the grand idea of a general evaluation metric for individual grammars coupled with the assumption that this metric in fact guides language learners (at least sometimes, when they do not happen to implement complicating changes). No noticeable effort is made, then, to further specify this evaluation metric, nor to demonstrate its significance to language acquisition. After about 160 pages we at least know that the more general notion of simplicity as defined by the evaluation metric obviously would have to subsume what so far, in Lightfoot's version of the modals story, was dealt with under the label of Transparency (which was neither very specific nor very general), and that the measuring of a grammar's simplicity would be in accordance with the so-called logic of markedness, which in effect says that individual grammars may violate conditions on rules which were formerly conceived of as absolute restrictions, but count as less highly valued, as more marked if they do so. This knowledge, however, does not in any plausible sense contribute to our understanding of the development of the English modals: neither in terms of the Transparency Principle ("simply for ease of exposition" Lightfoot continues with this term throughout the remaining 240 pages) nor of an evaluation metric does Lightfoot come anywhere near to specifying the limits of opacity/complexity which would have been approached or even infringed by the grammar of (early) ModE if there had been no re-analysis of the premodal verbs. His claim to be able to predict anything with his various theories is certainly presumptuous.

But perhaps we are asking for too much at this point. After all, Lightfoot

himself (p. 121) jocularly admits that he has introduced the Transparency Principle "more or less like a rabbit out of a conjuror's hat" towards the end of his modals story (though even this strikes me as an exaggeration: this conjuror seems to me to lack a hat as well as a rabbit), and insists that the Transparency Principle can only be elaborated in a process of inductive generalization on the basis of many historical re-analyses, each revealing something about the limits of admissible opacity/complexity. Well, Lightfoot's carelessness in spelling out precisely what his paradigm case, the alleged modal re-analysis, might reveal about these limits does not inspire much confidence in the eventual outcome of this process: in fact, nothing much has been revealed so far which one could strictly speaking begin to generalize inductively. And this exercise, furthermore, would seem to be hampered by the logical difficulty that in order to generalize from a number of re-analyses about the limits of opacity/complexity grammars tolerate, one must already know that the re-analyses at issue indeed have been undertaken for the sake of restoring transparency/simplicity rather than being complicating changes which had better been excluded from the induction. But changes do not come labelled as simplifications or complications: a priori, a grammar where some items are categorized as exceptional verbs, for instance, might be regarded as less complex than a grammar with an extra category 'Modal' covering these items — and the introduction of this category would then count as a complication. Lightfoot's rule of thumb that the simultaneity of several changes indicates a re-analysis provoked by the Transparency Principle (or the evaluation metric) might seem helpful, provided one takes more care in establishing the chronology of a chain of events than does Lightfoot in his paradigm case, but it is not absolutely reliable: first, there is always the possibility of chance coincidences; second, some of Lightfoot's further examples intended to reveal more about permitted Transparency do not really involve overwhelming quantities of roughly simultaneous changes; and third, he presents instances of simultaneous changes which in his own view are not provoked by the Transparency Principle, but are no chance coincidences either.

Space and other considerations prevent me from reviewing the other examples figuring in what Lightfoot calls inductive generalization. They include some further alleged category changes:

- the introduction of the category Quantifier in early ModE for elements which previously had been (increasingly exceptional) adjectives, with the attendant transformational re-formulations;

- the transition of *to* infinitives from nominal (NP) to verbal (VP) status in late ME/early ModE, the preposition *to* turning into a "mere grammaticized form" apparently categorized as an auxiliary (cf. p. 196, but there is no explicit discussion of this point);
- re-categorizations of serial verbs as members of minor categories such as preposition, complementizer, auxiliary, or adverb in Kwa languages, or the other way round — Lightfoot solemnly declares that the direction of these changes must remain a mystery in the absence of crucial historical records!

Some changes are envisaged as affecting the lexicon:

- the demise of the ME impersonal constructions, including subcategorization re-analyses of familiar OVS constructions such as *þam cynge licodon peran*, extravagantly claimed to contain impersonal verbs, as SVO constructions (*the king liked pears*);
- the innovation of *se*-impersonal constructions in 16th-century Portuguese, re-analyzing as active their diachronic ancestor, the *se*-passive construction;
- the loss of the middle voice in early Indo-European languages such as Greek, Sanskrit, Latin and the establishment of a passive;
- the loss of the Classical Greek subjunctive/optative verb forms and the acquisition, in the later history of Greek, of new formal means to encode essentially the same traditional mood opposition.

Also included, if rather tentatively, is the introduction of a new phrase-structure rule NP → $\bar{S}$ in early ModE to better account for certain apparent changes in the distribution of complement clauses; and, finally, the loss of a transformation Pronoun Fronting in OE, affecting the admissible (surface) constituent membership of the node COMP, which later was subject to several further changes with no apparent bearing on Transparency. Suffice it to say that the empirical analyses and theoretical interpretations offered by Lightfoot in general do not in my opinion surpass the standard set in the paradigm case of the modals. For a critical review of some of these other stories the interested reader may be referred to Bennett (1979), Fischer & van der Leek (1981), Romaine (1981), Warner (1983), and to some of Lightfoot's own subsequent works, now showing the full impact of the latest most appropriate theory of grammar.

Even with the hindsight afforded by his (1981a) later revelations, the way Lightfoot goes about inductively generalizing the Transparency Principle

still strikes me as most remarkable in view of the avowed purpose of this assemblage of case studies. There is no sustained effort at all towards elaborating such a principle, or — if it is recognized as a mirage, as Lightfoot now holds it must be — towards elaborating relevant aspects of an evaluation metric independent of that particular phantom principle. Rather, he characteristically ends each of his assorted stories with a statement to the effect that the changes just described — or more specifically the fact that changes took place at all, no matter in what form — could be seen as a consequence of the Transparency Principle, if this were formulated in the appropriate ad-hoc manner. But somehow a formulation of this principle (or of an evaluation metric) never materializes. No specific, and yet sufficiently general, conception emerges as to the permitted kind and degree of categorial exceptionality after the category-change stories, of derivational distance nor of surface ambiguity after the remaining stories. In the end, then, one is only mildly surprised to learn that it would be premature to expect the goods to be delivered in only a little more than 400 pages: "I shall make no attempt to formalize the Transparency Principle or to give a precise account of the permitted degree of derivational opacity" (p. 344). Of course those who seem to recall, from only a few hundred pages earlier, the same statement only with the negative omitted, may be slightly irritated. But there is no need to: where after all is the thematic development and the logical conclusion in the Arabian Nights, the Decameron, or the Canterbury Tales? Accept *Principles* as an equally heterogeneous, if perhaps less entertaining (unless you are one of the *cognoscenti*), collection of stories, and be glad that you are offered into the bargain a lengthy story in two parts, the myth of the introduction, around 1500, of a transformational, in addition to an already existing lexical passive in English, reflecting the innovation of NP Preposing, which admittedly is not even opaquely related to the Transparency leitmotif. (This last story has meanwhile been rewritten, in a paper conveniently published several times (Lightfoot 1979b, 1980, 1981b), like many other writings of this author including substantial parts of *Principles*.)

Among the further attractions of Lightfoot's book are numerous satirical interludes on the neogrammarians, taxonomic transformationalists, word-order typologists, syntactic reconstructionists, and others whose bad luck was of course "a failure to work with an appropriate theory of grammar" (p. 41). One can only hope that, in the case of the neogrammarians, Lightfoot's difficulties in acquiring first-hand knowledge of their theories were not comparable to those he apparently has with spelling German names and titles,

only few of which remain undistorted in the Bibliography and elsewhere. (Genuine misprints are rare; 'Bambileke' for 'Bamileke' on p. 216 (twice) presumably is to be blamed on the author rather than the printer.)

The shedding of a little light on why the (pre-)modals, and other modality expressions in English changed as they did, however, is not among this book's attractions. Whilst I agree with Lightfoot that we are faced here with a paradigm case of grammatical change (without agreeing with him, though, on the criteria of a decent description), it seems rather obvious to me what an explanation of such processes as the developments of modality expressions in the history of English, answering the question how they were possible and ideally also why they were necessary, ought to be required minimally to achieve. An illuminating explanatory account ought to be able to shed light on at least three crucial aspects of such processes of grammaticization: first, on the mere fact that grammaticizations of modality expressions are taking place at all, and that they are taking place at a particular time (typically in fact over a considerable time span: recall Table 2); second, on the selection of lexical elements which are to be grammaticized, delimiting as far as possible the range of such elements potentially available for the purposes of grammaticization in a mood system; and third, on the forms such grammaticizations may take in principle and are likely to take in particular instances. No attention at all is paid to the last two aspects in Lightfoot's account; and the attention he pays to the first is misguided because of his refusal to recognize the changes he chooses to deal with for what they are: typical manifestations of a grammaticization process, and also because of his failure to recognize many further changes as variations on the same theme. Elsewhere (Plank 1981, 1983) I have tried to outline these three components of an explanatory sketch, emphasizing in particular that reference must be made to the functions modality expressions fulfil in discourse in order to understand what forms some of them (viz. those that are labelled grammatical) are bound to take, partly sharing these forms with other grammatical(ized) categories (such as tense and aspect). Here I must content myself with simply calling to mind a relationship between changes in English, as set out above, that rather transparently bears on the question why modality expressions underwent particular changes at a particular period of history.

That obvious relationship involves, on the one hand, the development of verbal mood inflection (24) and, on the other hand, many of the developments pertaining to the (pre-)modals and other modality expressions. Originally, in Indo-European and, though with gradually decreasing efficiency,

still in Germanic and OE times, verb inflections represented the principal formal means of expressing modality distinctions, especially epistemic and deontic notions. And as this grammatical system was brought into disarray, losing much of its expressive and, vis-à-vis the indicative, distinctive power, new grammatical devices capable of subserving the same purposes in, eventually, an equally adequate manner, began, or had begun, to be elaborated, drawing on suitable lexical means of expression (viz. essentially the premodals). Note that I am not necessarily suggesting a simple unidirectional causal relationship with the reduction of the inflectional mood system, for phonetic or other reasons, precipitating the grammaticization of a set of lexical forms. I see no contradiction in admitting that causation, and at certain periods perhaps predominantly, may have been the other way round as well: without the availability of suitable alternative formal means the original inflectional mood system would hardly have been allowed to be obscured and largely abandoned. Thus, I think it would be too simplistic to conclude that the demise of the subjunctive mood in OE, or earlier or later impairments of verbal moods, actually triggered the grammaticization of the premodals. In view of the considerable length of time it took for the old mood system to wear out, a more satisfactory explanation ought to be on a larger scale and first of all invoke the familiar notion of an overall type change (which, to be sure, may be speeded up or retarded or even partly reversed depending on all sorts of circumstances): the tendency for analytic forms to replace synthetic forms in the expression of grammatical categories. Thus putting aside the question of ultimate causation of the eternal cycle of the creation and decay of morphology, we are nevertheless able to gain some insight into the rationale of the developments of various kinds of modality expressions in English: what happened to them makes sense when seen as a special case of a more general phenomenon, irrespective of the eventual explanation of that general phenomenon (viz. the exchange of synthetic for analytic coding devices). Taking this wider historical and explanatory perspective, it appears highly unlikely that any of the developments (1), (2), (3), (5) above, and presumably others as well (especially (23)) were as accidental and as unrelated to other changes as Lightfoot would have it: there can be little doubt that they are part and parcel of the same overall process.

Wilde (1939/40), Visser (1972: 789, *passim*), taxonomic transformationalist Traugott (1972: 148f.), now also Steele et al. (1981: 274ff., who otherwise subscribe to Lightfoot's version of the modals story) — in short, more or less everybody who has ever aimed at describing and explaining the

history of mood/modality in English was aware of the connection between the development of the subjunctive mood and that of the (pre-)modals. Lightfoot, boasting (on p. ix) to have actually been able to 'discover' some hitherto unsuspected simultaneity of 'superficially' unrelated changes thanks to his superior theory of grammar, is the odd man out in missing this one. And his ignorance seems intentional: in a former, otherwise roughly identical version of his modals story he still referred to the erosion of mood distinctions in a footnote (Lightfoot 1974: 247), but I can't seem to find a similar acknowledgment in *Principles*.

Noticing the pivotal role of the changes affecting verbal mood inflections and placing them in the context of an overall drift towards analyticity, however, does not quite suffice for purposes of the first component of an explanatory sketch. To account for the interrelatedness of the developments pertaining to two sets of modality expressions, verbal mood inflection and the (pre-)modals (and perhaps further lexical items of suitable meaning), a statement, or panchronic 'law', is required to the effect that each language, firstly, has to have at its disposal a (sufficiently rich) ***grammatical*** system, analytic or synthetic, for the expression of modalities, irrespective of how many ***lexical*** modality expressions are available in the same language, and, secondly, that one set of grammatical modality expressions suffices, or is at any rate utilized predominantly vis-à-vis possibly co-existing sets. (The universality of sufficiently elaborate mood systems, incidentally, need not be taken as axiomatic, but should eventually be derivable from considerations about the communicative functions of modality expressions.) If we did not hypothesize (and eventually motivate functionally) that mood, i.e. grammaticized modality, is a strongly universal category, and that it is subject to more or less severe constraints on systemic redundancy, it should have been possible for the verb-inflectional mood system to dissolve in English without the premodal verbs or other lexical modality expressions being grammaticized, or for these expressions to be grammaticized without the formerly predominant mood system passing out of use. It goes without saying that more accuracy is desirable in specifying the requirements on the expressive power which the mood system of any language must satisfy. I would speculate, for instance, that epistemic and deontic modalities are more likely to figure in the minimum mood equipment than dynamic modalities; but clearly, extensive cross-linguistic research is needed to confirm this, and to elucidate structural regularities of the kinds of modality, individually and as a system, which are obligatorily grammaticized.

Needless to emphasize that a more satisfactory understanding, or explanation, of the other aspects of grammaticization processes, of modality and other categories, also presupposes further intensive cross-linguistic research, both diachronic and synchronic. Gathering and decently describing data, and ideally of course all relevant data, certainly can be no end in itself. But for those who, as a matter of principle, are inclined to stress "depth of explanation rather than coverage of data" (Lightfoot 1979a: 76), there is also the danger of their supposed explanations getting out of their depth on account of an exaggerated neglect of relevant data. Lightfoot's version of the story of the English modals is a paradigm case of an explanation reaching new lows (instead of plumbing any depths) due to a failure to work with an appropriate set of data. If the peculiar contents and moral of the tale Lightfoot tells about the modals is to be attributed also to the theory of grammar he was working with, it can hardly be of such superior appropriateness as he untiringly proclaims it to be.*

* I have drawn on my version of the modals story in oral presentations at the universities of Hannover (June 1981), Bonn (Jan. 1982), Köln (March 1982), Konstanz (June 1982), and Stanford (June 1983). I am grateful to these audiences for questions and suggestions.

## REFERENCES

Akmajian, A., S. Steele and T. Wasow. 1979. "The category AUX in universal grammar." *Linguistic Inquiry* 10, 1-64.

Anderson, J.M. 1976. On serialization in English syntax. Ludwigsburg: *Ludwigsburg Studies in Language and Linguistics 1.*

Baker, C.L. 1981. "Learnability and the English auxiliary system." In: *The logical problem of language acquisition*, ed. C.L. Baker & J.J. McCarthy, 296-323. Cambridge, Mass.: MIT Press.

Bennett, P.A. 1979. "Observations on the transparency principle." *Linguistics* 17, 843-861.

Bolinger, D. 1980. "*Wanna* and the gradience of auxiliaries." In: *Wege zur Universalienforschung. Festschrift für H. Seiler*, ed. G. Brettschneider & C. Lehmann, 292-299. Tübingen: Narr.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Evers, A. and T. Scholten. 1980. "A Dutch answer to the Luiseño argument." *Utrecht Working Papers in Linguistics* 9, 87-101.

Fischer, O.C.M. and F.C. van der Leek. 1981. "Optional versus radical re-analysis: mechanisms of syntactic change." *Lingua* 55, 301-349.

Fodor, J.D. and M.R. Smith. 1978. "What kind of exception is *have got*?" *Linguistic Inquiry* 9, 45-66.

van der Gaaf, W. 1904. *The transition from the impersonal to the personal construction*. Heidelberg: Winter.

Giacalone Ramat, A. 1980. "Verso una teoria del mutamento sintattico." *Lingua e Stile* 15, 539-563.

Gonda, J. 1956. *The character of the Indo-European moods*. Wiesbaden: Harrassowitz.

Gross, M. 1975. *Méthodes en syntaxe*. Paris: Hermann.

Hammerich, L. 1960. "Über die Modalverben der neugermanischen Sprachen." *Zeitschrift für deutsche Wortforschung* 16 (N.F.1), 47-70.

Huddleston, R. 1980. "Criteria for auxiliaries and modals." In: *Studies in English linguistics for Randolph Quirk*, ed. S. Greenbaum, G. Leech & J. Svartvik, 65-78. London: Longman.

Hudson, R.A. 1976. "Conjunction reduction, gapping, and right-node raising." *Language* 52, 535-562.

-----. 1981. "Some issues on which linguists can agree." *Journal of Linguistics* 17, 333-343.

Jackendoff, R.S. 1971. "Gapping and related rules." *Linguistic Inquiry* 2, 21-35.

Jakobson, R. 1971. "Shifters, verbal categories, and the Russian verb." In his *Selected writings*, vol. 2, 130-147. The Hague: Mouton.

Jespersen, O. 1917. "Negation in English and other languages." *Historisk-filologiske Meddeleser* 1, 5. Reprinted in: *Selected writings of Otto Jespersen*, 3-151. London: Allen & Unwin.

-----. 1942. *A Modern English grammar on historical principles*. Vol. 6. London: Allen & Unwin.

Klöpzig, W. 1922. "Der Ursprung der *to be to*-Konstruktion." *Englische Studien* 56, 378-389.

Kuryłowicz, J. 1964. *The inflectional categories of Indo-European*. Heidelberg: Winter.

Langendoen, D.T. 1970. "The *can't seem to* construction." *Linguistic Inquiry* 1, 25-35.

Lehmann, C. 1982. *Thoughts on grammaticalization. A programmatic sketch*. Vol. 1. Köln: akup 48.

Levin, N.S. 1980. "Main verb ellipsis in spoken English." *Working Papers in Linguistics* (Ohio State University) 24, 65-165.

Lightfoot, D. 1974. "The diachronic analysis of English modals." In: *Historical linguistics* 1, ed. J.M. Anderson & C. Jones, 219-249. Amsterdam: North-Holland.

-----. 1979a. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

-----. 1979b. "The history of NP movement." *Glot* (Leiden) 2, 255-284.

-----. 1980. "The history of NP movement." In: *Lexical grammar*, ed. T. Hoekstra et al., 255-284. Dordrecht: Foris. [= *Glot* 2].

-----. 1981a. "A reply to some critics." *Lingua* 55, 351-368.

-----. 1981b. "The history of NP movement." In: *The logical problem of language acquisition*, ed. C.L. Baker & J.J. McCarthy, 86-119. Cambridge, Mass.: MIT Press.

Palmer, F.R. 1979. *Modality and the English modals*. London: Longman.

Plank, F. 1981. "Modalitätsausdruck zwischen Autonomie und Auxiliarität." In: *Sprache und Pragmatik. Lunder Symposium 1980*, ed. I. Rosengren, 57-71. Lund: Gleerup.

-----. 1983. "Function motivating form, or: Why it would be surprising if auxiliaries were, or remained, main verbs." Paper read at Stanford Iconicity Colloquium (June 1983) and at Amsterdam Workshop on Functional Causes of Language Change and Variation (Dec. 1983).

Pullum, G.K. 1982. "Syncategorematicity and English infinitival *to*." *Glossa* 16, 181-215.

Pullum, G.K. and D. Wilson. 1977. "Autonomous syntax and the analysis of auxiliaries." *Language* 53, 741-788.

Romaine, S. 1981. "The Transparency Principle. What it is and why it doesn't work." *Lingua* 55, 277-300.

Ščur, G.S. 1968. "On the non-finite forms of the verb *can* in Scottish." *Acta Linguistica Hafniensia* 11, 211-218.

Spears, A.K. 1982. "The Black English semi-auxiliary *come*." *Language* 58. 850-872.

Standop, E. 1957. *Syntax und Semantik der modalen Hilfsverben im Altenglischen. Magan, motan, sculan, willan.* (Beiträge zur Englischen Philologie, 38) Bochum-Langendreer: Pöppinghaus.

Steele, S. et al. 1981. *An encyclopedia of AUX. A study of cross-linguistic equivalence.* Cambridge, Mass.: MIT Press.

Tajima, M. 1975. "The *Gawain*-poet's use of *con* as a periphrastic auxiliary." *Neuphilologische Mitteilungen* 76, 429-438.

Thompson, L.C. 1979. "On the development of auxiliaries in Salish." *Working Papers in Linguistics* (University of Hawaii) 11:2, 67-73.

Traugott, E.C. 1972. *A history of English syntax.* New York: Holt, Rinehart & Winston.

Visser, F.T. 1970, 1972, 1969, 1973. *An historical syntax of the English language.* Vols. I, II, III:1, III:2. Leiden: Brill.

Warner, A.R. 1983. "Review article on Lightfoot (1979a)." *Journal of Linguistics* 19, 187-209.

Weinreich, U., W. Labov and M.I. Herzog. 1968. "Empirical foundations for a theory of language change." In: *Directions for historical linguistics*, ed. W.P. Lehmann & Y. Malkiel, 95-195. Austin: University of Texas Press.

Wilde, H.-O. 1939/1940. "Aufforderung, Wunsch und Möglichkeit. Die englische Sprache und die Grundlage englischer Lebenshaltung." *Anglia* 63, 209-391; 64, 10-105.

von Wright, G.H. 1951. *An essay in modal logic.* Amsterdam: North-Holland.

Wunderlich, D. 1981. "Schlussbemerkungen, Modalverben im Diskurs und im System." In: *Sprache und Pragmatik. Lunder Symposium 1980*, ed. I. Rosengren, 109-121. Lund: Gleerup.

Zwicky, A.M. and G.K. Pullum. 1983. "Cliticization versus inflection: English *n't*." *Language* 59, 502-513.