SYNCATEGOREMATICITY AND ENGLISH INFINITIVAL TO*

This paper considers the question of whether some words should be treated syncategorically — that is, introduced directly by syntactic rules, and not assigned to any lexical or grammatical category. The English infinitival marker to, which has often been treated in this way, is taken as a case in point. I argue that it should not be treated syncategorically, and thus must be assigned to some category. There are arguments that to is not an affix, particle, complementizer, preposition, ‘AUX’, or tense morpheme. The best analysis assigns it to the category V. A number of arguments are given that to is an auxiliary verb and head of a verb phrase. Several potential objections to the verbal analysis are examined and rebutted. It is pointed out that the proposal advocated is comparatively theory-neutral.

0.0 Introduction

Consider the difference in status between the elements ‘who’ and ‘?’ in written English. ‘Who’ is listed in the dictionary. It is a word. It is spelled with letters from the usual alphabet. It has a ‘part-of-speech’ classification as a interrogative pronoun. It has a meaning. In general, it occurs where the writer intends that particular meaning to be expressed. But the element ‘?’ is the exact opposite in all respects. It is not listed in the dictionary. It is not a word. It has no spelling. It has no part-of-speech status. It does not of itself have a meaning. It appears as an obligatory part of a certain syntactic construction, the (independent) interrogative clause, which in part it defines. To put all this another way, ‘?’ is syncategoractic, while ‘who’ is (I shall assume) not.
Generative grammarians have introduced an explicit way of treating some selected words as syncategorematic like ‘?’ rather than as lexical items like ‘who’. The technique is to introduce the selected words directly by phrase structure rule or transformation, not by lexical insertion. The list of words in English that have been given this special treatment is quite long. For example, among the comparatively limited number of illustrative examples of rules in Chomsky (1957) we find proposals for introducing and, be, by, do, en, have, ing, not, n't, so, to, very and the boundary symbol ‘#’ syncategorematically. The larger grammar in Burt (1971) does the same thing with most of these and adds for, of, or, that, there, and ‘poss’ (i.e. possessive ‘s’).

I believe that this analytical option has been exploited in an unprincipled and arbitrary way, and that it is an inherently undesirable option anyway. It is undesirable for two reasons. First, it introduces irreducibly parochial (language-particular) elements into the syntactic rules of the language instead of assigning them to the natural repository for such parochiality, the lexicon. This ought to really worry people like Chomsky (1981), Wexler (1981), and Pinker (in press), who want to claim that the set of distinct humanly possible grammars is actually finite, but is to some extent worrisome for anyone who wants to maximize the universality of nonlexical rule formulations in grammars. Second, it formalizes a distinction between words in a language for which there is absolutely no warrant in terms of the intuition of the native speaker: one cannot elicit from speakers any sign of an intuitive recognition of a difference between one set of words and another that corresponds to the difference between the words generative grammarians have introduced syncategorematically and those they have not. Indeed, one cannot even elicit such a recognition from grammarians. Each generative grammarian chooses a different set of words to make syncategorematic. Some will make passive be syncategorematic and some will say it is a member of the category verb; some introduce by transformationally and others assign it to the category preposition; and so on. Hence my charge of arbitrariness and lack of principle in the way syncategorematic introduction of words has been used.

There is a way to make a principled distinction between formatives introduced syncategorematically and formatives introduced via the lexicon, due to Langendoen (1979: 8). Langendoen proposes simply that “A morpheme is uncategorized if and only if it is an affix...” Putting this another way, we can adopt the following definition of the notion ‘affix’:

1. An affix is a terminal symbol that has a sister.

Under this assumption, if there is a rule of the form *V → Vstem ing*, the terminal symbol ing will be an affix by definition. If we assign John’s the structure of [N [N John ‘s], we define the possessive morpheme as an affix, but if we assign it the structure [N [N John] [Poss ‘s]], we define it as a lexical item belonging (perhaps uniquely) to the category Poss. We can envisage the principle applying both to outputs of the phrase structure rules and to outputs of the transformational component of a transformational grammar.

This proposal enables us to draw a clear distinction in theoretical terms between the notions affix and word2 as shown in (2).

2. (a) An affix is a terminal symbol that has a sister.

(b) A word is a terminal symbol immediately and exhaustively dominated by a lexical-category-labeled node.

To illustrate these definitions, I give two possible candidates for the structure of a verb-particle combination like give up.

3. (a) [vp [v [vstem give Vstem ] up [v] vp ]

(b) [vp [v [vstem give Vstem ] yl [p up p] vp ]

PHILLIM: Syncategorematicity— 183
According to my definitions, *up* is represented as an affix in (3.a), but as a word in (3.b). *Give* is not an affix in (3.a) because it belongs to the category *Vstem* — i.e., there is a rule *Vstem → give*, so *give* has no sister.

Analytical decisions as to which items in a language should be treated as affixes, and which as words, are by no means easy to make. In fact, it is very important to appreciate that answers to questions in this domain can be both difficult to arrive at and surprising, even in a well-studied language with which one is thoroughly familiar. For example, Zwicky and Pullum (1982) present evidence that *not* and *not* are synchronically unrelated, *not* being not a simple clitic form of the word *not* but rather a previously unrecognized inflectional affix indicating the finite negative form in the paradigm of an auxiliary verb.

It is also important to realize that decisions as to category membership for items that are placed in the category of words must sometimes be made on the basis of rather subtle and indirect evidence, and can also be surprising. For example, in English, *Mailing* (to appear) has shown that syntactic evidence reveals *near* to be an adjective taking NP complements (as in *near the wall*), although it has commonly been treated as a preposition, and she further shows that the items *like* and *worth*, usually treated as adjectives taking NP complements (see e.g. Huddleston (1976: 244)), are both prepositions, despite their nonlocational and thus rather unprepositionally meanings.

With the foregoing discussion in mind, I now want to turn to a specific analytical question of a sort similar to the ones just discussed, and show that an intuitively rather peculiar decision turns out to be the best supported one. The item I want to discuss is the infinitival marker *to*. I shall begin with a proposal one might easily consider making, namely that infinitival *to* is a prefix marking the infinitive forms of verbs, and argue that it is mistaken, and that *to* must be a word. I shall then consider the question of what tree-geometrical configuration it appears in (i.e. what the bracketing of infinitival phrases is), and go on to ask what grammatical category it should be assigned to. I shall settle on a rather unexpected answer, and present arguments for its correctness.

### 1.0 Why *to* is not an affix

From the policy of some traditional grammars in calling *to* the infinitive form of the verb whose present tense forms are *am*, *are*, *is*, one might assume that there was some prospect of analyzing *to* morphologically as a verbal prefix. (It is difficult to find clear quotes on the topic. For example, Jespersen's use of the phrase "a mere 'empty' grammatical appendix to the infinitive" (1940: 154) does not make it clear whether he regards *to* as a prefix or a word. Many traditional grammars are similarly inexplicit.) But a prefixal analysis of *to* can be rejected immediately on syntactic grounds. *To* can be separated from its accompanying verb, by an adverb in the familiar 'split infinitive' construction as in (4.a) or by Right Node Raising as in (4.b), or (not quite so felicitously) by VP Fronting as in (4.c), or by VP Deletion as in (4.d).

(4.a) These are the voyages of the Starship Enterprise; its five-year mission: to explore strange new worlds, to seek out new life and new civilizations to boldly go where no man has gone before.

(4.b) McCoy wouldn't like to, and he probably won't, become the sort of person that Spock is.

(4.c) Starfleet Command has instructed me to proceed without delay, and proceed I intend to, Mr. Sulu.

(4.d) I've never met a Klingon, and I wouldn't want to.

No prefix can be separated from its verb stem in the way that *to* is separated from *go*, *become*, *proceed*, and (implicitly) *meet* in these examples.

The issue of whether *to* is syntactically a subpart of a word should not be confused with the quite separate one of its phonological constituency. Zwicky (to appear) has studied the latter in great detail, establishing that infinitival *to* is required to be phonologically clitic on an adjacent item in a way that imposes interesting constraints on the ways in which it can be stranded. But his results are not in conflict with my claim here that *to* is syntactically an independent constituent.
An alternative to treating *to* as an affix would be to reject the Langendoen characterization of affixhood and to treat it as a separate word that belongs to no syntactic category but is introduced directly as a terminal symbol by a branching phrase structure rule. It is difficult to bring evidence directly to bear against this proposal. My objections to it are of a theoretical nature. Basically, I believe it is undesirable to postulate that some words are introduced by the grammar and not the dictionary. To say that the *to* of infinitival phrases was introduced by some rule such as the rule ‘VP → to VP’ postulated by Brame (1978) and Bresnan (1978) is to say it is not in the lexicon at all, and should not be expected to share any properties with any other item, since it belongs to no lexical category and thus cannot form a natural class with some set of lexical items. This is an intrinsically pessimistic position, and a theoretically weak one. It fails to derive any syntactic properties of *to* from other statements in the grammar. Its linear position is stated ad hoc by a rule that could just as well have stipulated the reverse order, and likewise for all its other properties.

The synctagematic-word proposal about infinitival *to* is assumed in the work of Chomsky. For instance, on p. 170 of Chomsky (1980) we find a tree which would need a rule ‘S → NP to VP’ to admit it. The odd thing about finding it in this context is that in the same book (121 n. 42; 268 n. 43; 273 n. 7; 276) we find Chomsky suggesting more or less explicitly that the number of distinct grammars permitted by linguistic theory may be finite. But the number of distinct permissible phrase structure rules cannot possibly be finite once we allow arbitrary terminal symbols to appear in branching rules, because the set of all possible terminal symbols in human languages is not finite. For Chomsky, therefore, it introduces an inconsistency with something else he would like to maintain if he permits rules of the sort needed to introduce *to* in a branching rule without its own category. And even if one does not believe in either the feasibility or the theoretical value of a finite set of possible grammars (as I do not; see Pullum, to appear), there is a question of whether we can obtain some reasonable kind of typology of rules of grammar. When we say that we categorize *with* as a member of the category P, and assign *with him* the structure [P P NP], we are relating the English construction to closely similar ones in other languages to which we assign an identical structure. When we introduce *to* directly under some branching node we are saying there is no similarity between *to* in English and any items in other languages; a specific formative has to be named in the rule that stipulates the form of construction in which it occurs.

Apparenty meaningless items have been treated synctagematically in transformational grammar before, of course. One example is the supportive *do of I do (not) like it*. But although Chomsky made such an analysis familiar, he never answered the obvious objections to it — that it denies the identity between supportive *do/does/did/done/doing* and the ordinary transitive verb forms *do/does/did/done/doing*, and so on (for a brief summary of some of the difficulties besetting the kind of analysis Chomsky advocates, see Gazdar, Pullum and Sag 1982, section 3.1).

From now on, I shall assume that *to* must belong to some category of words, and that both affixal and synctagematic-word treatments are wrong. Notice, however, that not everything about the Brame-Bresnan suggestion is wrong: I have argued against its assumptions regarding labeling, but it seems to assign the right constituent structure bracketing, as I shall argue in the next section.

2.0 The bracketing of infinitival complements

For a four-word phrase like *want to see you* there are eleven possible groupings into constituents. This is shown in (5.).

\[ (5.a) \]

\[ (5.b) \]
It is rather easy to determine which of these is the most plausible for phrases of this sort in English. Sentences like (4.b, c, d) show that we want what follows to (here, see you) to be a constituent, and sentences like those in (5.) suggest the same about the substring to see you.

(6.a) What I want is to see you.

(b) To see you is what I want.

There is only one structure in (5.) that represents both see you and to see you as constituents, namely (5.k). And this result is unchanged even if we adopt Chomsky's assumption (1981 et passim) that infinitival clauses really have phonologically null syntactic subject NP's, modulo the change needed to include the extra daughter in the infinitival constituent. I shall therefore assume that we want a bracketing like (5.k), without arguing further against apparently unsupported alternatives. I shall also assume from now on that an analysis that puts to in some category and thus relates it to at least some other items is inherently preferable to one that distinguishes it from all other lexical items by directly mentioning it in a phrase structure rule or a transformation.

3.0 Why to is not a particle

Occasionally one finds to referred to as a 'particle'. Bresnan (1971) is quite explicit in this, using the abbreviation 'Prt' normally used for postverbal particles (see e.g. Jackendoff (1977)), and assuming 'some such configuration as' \[ \_Vp \_Prt \_to \_V \ldots \] (see p. 7). But there are no convincing parallels between infinitival to and the kind of post-verbal particles that participate in 'Particle Movement' behavior. There is a particle in this sense with the same form as to, as in the phrases come to 'regain consciousness' and bring NP to 'cause NP to regain consciousness', but this item seems quite irrelevant to the structure of infinitival clauses. Particles have been convincingly argued by Emonds (1972) to be intransitive prepositions, and I discuss the idea that to is a preposition below.
4. Why ‘to’ is not a complementizer

Implicit in the title of Postal and Pullum (1978) is the claim that to is a complementizer, i.e. that to is to VP’s as that is to S’s (a claim that can also be read into the Brame-Bresnan proposal in (2.) above). The fact that to is never followed by a subject NP is not really a problem even for those who reject ‘orphan VP’s’ and posit an underlying S node for every surface VP node, because to would be the complementizer specific to those clauses that lose their subjects through the operation of rules like Raising and Equi or their analogs, whereas that would be a complementizer that guaranteed that its S sister would not undergo those rules.

Despite having assumed in 1978 that to was some sort of complementizer, I have become convinced after more careful reflection that this cannot be so. I take that and whether to be paradigms of what complementizers are supposed to be like in English. The behavior of to contrasts with these items sharply in at least one domain, that of stranding. To can be stranded under ellipsis, but that and whether cannot:

(7.a) It is possible to beat the computer at three-dimensional chess, but it is not easy to Q.
(b) *It is possible that the computer made an error but it is not likely that Q.
(c) *It may be true that Spock just beat the computer, but at the moment it is unclear whether Q.

To can also be stranded by fronting of its complement, but this is out of the question for that and whether:

(8.a) Proceed, I intend to. (cf. (1.c))
(b) *I can do it, I’m convinced that.
(c) *You can do it, no one knows whether.

Such facts lend no support to an analysis of to as a complementizer.

5.0 Why ‘to’ is not a preposition

Another proposal that might be defended is that to is a preposition when it has a VP complement, just as it is when it has an NP complement. (This proposal relates to the previous one in that there is a well-known connection between complementizers and prepositions in many languages. Some grammarians, e.g., Joseph Emonds and Nigel Vincent, have argued in yet-to-be-published work that there is no category distinction between complementizers and prepositions, supporting the position taken by Jespersen (1924: 88).) This idea is adopted by Starosta (1977). Attractive though it might seem, there is a lot telling against it (though it does capture the linear order facts correctly; see section 8.1). I shall review the main arguments briefly.

First, there are no other prepositions that take uninflected VP’s as complements. (All prepositions take -ing-inflected VP’s, but that is simply because gerundive VP’s have the syntactic behavior of NP’s as observed long ago, e.g., by Emonds ((1976) and earlier references cited there.)

Second, verbs that take PP complements with to do not ipso facto take to+VP. For example, the existence of the construction relate NP to NP as in (9.) does not guarantee that there will be a parallel construction relate NP to VP as in (10.):

(9.) One can often relate a scientist’s theoretical proposals to his particular economic role in society.
(10.) *One can often relate a scientist’s theoretical proposals to have a particular economic role in society.

Third, a verb that takes to+VP complements does not ipso facto take a PP with to:

(11.a) I want to go home.
(b) *I want to my home planet.
Fourth, modifiers specific to PP's, like right (see Emonds (1972)) do not occur modifying to+VP:

(12.a) He went right to the captain.
(b) *I wish right to see the captain.

Fifth, modifiers to to+VP such as adverbs do not occur modifying PP's with to:

(13.a) I agree to boldly go where no man has gone before.
(b) *I agree to boldly transportation to strange new worlds.

This argument is not defeated by claiming that in (13.a) the adverb is inside the VP so that the PP structure does not have an adverb as immediate constituent, because if that is the case then the construction that does not split the infinitive is the problem: it must involve a PP-initial adverb in cases like (14.).

(14.) Your instructions, Kirk, call for you boldly to go where no man has gone before.

But it seems incorrect to allow adverbs as left branches of PP constituents. Although adverbs can be found immediately before PP's as in (15.a), the failure of clefting in (15.b) suggests that there is no adverb+PP constituent.

(15.a) They carried the value system of American capitalism boldly to an entirely new galaxy.
(b) *It was boldly to an entirely new galaxy that they carried the value system of American capitalism.

What is true of adverbs is also true of the negative element not in this connection. Starosta (1977), who assumes that not actually is an adverb, notes this problem, and admits that it poses a problem to which he has no solution (see the discussion of his example (186)).

Sixth, predicates that allow to+VP as subject do not ipso facto allow PP's with to:

(16.a) To survive an ion storm will be tough even for the Enterprise.
(b) *To within a thousand miles of a black hole will be tough even for the Enterprise.

And this is not because PP's do not occur as subjects. They do, as Frederick Newmeyer has pointed out to me, and they even invert under subject-auxiliary inversion:

(17.a) To the Andromeda system is as far as we can go.
(b) Is to the Andromeda as far as we can go?

Seventh, the PP-with-NP construction exemplified in (18.a) does not allow a to+VP constituent in place of the PP.

(18.a) To the transporter room with him!
(b) *To get out of my sight with him!

(This argument, like the fourth, owes a debt to Emonds (1972), where the construction in question is used to argue that 'particles' are intransitive prepositions).

Eighth, if to is a preposition and to+VP is a PP, it is unclear what structure can be assigned to for-to complement clauses. It would seem necessary to postulate either S expanding as NP-PP (if for is a complementizer) or S expanding as PP-PP (if for is also a preposition). Both seem decidedly strange, since normally we expect a deep structure S to immediately dominate a VP node.

Ninth, prepositional to fails to show the contraction on to verbs that infinitival to exhibits, as pointed out by Postal and Pullum (1978: 17):

(19.a) I'm gonna explore strange new worlds.
(b) *I'm gonna Epsilon Delta IV.
Tenth, as Zwicky and Levin (1980) observe, infinitival to cannot bear high stress when it is stranded under ellipsis. The preposition to can, however. Compare the following (where capitalization indicates heavy stress):

(20.a) *It would be easy not to like McCoy, but then again, it would be easy TO.

(b) For the last time, Mudd, who did you give it TO?

These ten arguments suggest that prepositional to and infinitival to are not the same thing, I am aware of no syntactic arguments tending in the other direction.

I should say a word, however, about the proposal for a mixed analysis that can be read fairly easily into the remarks of the OED (Volume T. p. 87, columns 2-3). The claim is made there that although some occurrences of to constitute a mere ‘sign’ or prefix of the infinitive”, nevertheless it is still a preposition after an intransitive or passive verb. Although superficially he proceeds to speak has the same structure as he chooses to speak, the OED claims that “in the latter, to speak is the equivalent of speaking or speech, and in the former of to speaking or to speech”. This seems to be interpretable as a claim that where a predicate is subcategorized for a dative PP, any to+VP it allows has the structure [PP [p to] VP], but where to+VP follows a transitive verb, it has the structure [NP to VP]. No supporting argument is given for this analysis, but presumably the argument would be, in transformational terms, that strict subcategorization specifications are simplified. The analysis seems to me to have all the faults of several of the other analyses I dismiss in the present paper. It postulates two different analyses of the to that introduces infinitives: one as a preposition taking a VP and one as a syncategorematic element within NP. Thus the objections of Sections 1 and 4 would for the most part hold against it simultaneously.

The OED analysis might be further refined by positing a rule NP → to VP, and letting verbs like proceed select ‘— to NP’ while verbs like choose select ‘— NP’. But this gives deep structures of the form [VP proceed [PP [p to] [NP to VP]]. A deletion rule would be necessary to remove one or the other of the to elements (i.e. either to → 0/to, or to → 0/to). This is exactly analogous to the analysis of for-to complements and for-PP’s after the verb hope to be found in Chomsky and Lasnik (1977), and a critique along lines analogous to the critique in Brame (1980) would seem to be in order. I cannot see that there is anything to be gained by analyzing proceed to VP differently from choosing to VP. Certainly, proceed will need a dual subcategorization, but it will clearly be necessary to have some dual subcategorization in any imaginable analysis of English: for instance, even if we said that postverbal clausal constituents were always dominated by NP, a verb like believe would still need a dual subcategorization: NP, or PP with in (cf. believe this, believe in it, but *believe this in it*).

6. Why ‘to’ is not an ‘AUX’

Many transformational grammarians, I suspect, would say that they assume to is an ‘AUX’. But in fact the literature seems to contain no instances of analyses that bear this out. Where the category ‘AUX’ is assumed (as in Chomsky (1967) and Akmajjan, Steele and Wasow (1979), to cite two examples out of hundreds), it is a branching node, not a lexical one. No item is listed in the lexicon as belonging to the category ‘AUX’. Although to does appear under ‘AUX’ in surface structure in some of these analyses, it is placed there by a transformation, and in most cases (see e.g. Stockwell, Schachter and Partee (1973; 595)) it will have sisters under ‘AUX’ in at least some derivations, so that to will not always bear the ‘is a’ relation to ‘AUX’ analyses of to. There are only analyses which treat to syncategorematically but happen to have an ‘AUX’ node dominating it in derived structure.

Most linguists who postulate ‘AUX’ have failed to make it sufficiently clear where ‘AUX’ fits in their general theory of syntactic categories. Jackendoff (1977) makes a valiant
effort, postulating a category M" that has M (modal) as its ultimate lexical head, but there are numerous problems with his analysis. Others who give a fairly clear idea of what category inventory they are assuming do not posit a branching 'AUX' node, but assume that auxiliaries are members of the category [+V, -N]; see e.g. Fiengo (1980) and Lapointe (1980). But assuming, not too controversially, that a category cannot be both a branching node label and a lexical category, it would appear that there cannot be both a phrase structure rule 'AUX' — Tense do' and a lexical item (to or any other) that has 'AUX' as its lexical category. I therefore dismiss the many transformational analyses that put to in an 'AUX' node as nondistinct from the ones that treat to syncategorically (e.g. by ad hoc transformational insertion as in Rosenbaum (1967)). That is, I regard the 'VP' — to VP' of Brame and Besnan, the 'S — NP to VP' apparently assumed by Chomsky, the 'AUX' — M (to)' and 'PROG — be-ing (go to)' apparently assumed by Chomsky and Lasnik (1978), and all similar approaches, to be variations on a single theme: X — Y to Z, where either Y or Z is non-null. If there were any accounts in which to was treated as a member of a lexical category 'AUX' (and as far as I know there are none), then apparently to would be the only member of that category. It certainly cannot be claimed to have, say, the same behavior as the modals. And to say that to is the only member of its category is to say that a special category has been invented specially for it, and that it is unrelated to any element known in English or other languages. To adopt that view would be to adopt a counsel of despair.

7. Why 'to' is not 'Tense'

Related to the view that to is under an 'AUX' node is the proposal that to fills the slot for tense in an infinitival clause. This is advocated in Sag (1976) and in Bach (1982), though neither considers the claim to be particularly crucial. About this idea I shall say merely that I do not believe it has secure motivation for the notion that so-called tense affixes of English (-s, as in walks and -ed as in walked, of which the former seems better regarded as an exponent of third person singular agreement) are syntactically independent elements that fall in a natural syntactic class with to. They have the syntax of suffixes. To does not. The criterial property of the class of affixes assumed (though not in fact defined except by means of a list) in Chomsky (1957) is presumably participation in 'Affix Hopping', i.e. the 'Auxiliary Transformation' of that work. But to does not fall together with -s and -ed by this criterion. And semantically, there is clearly no argument to link to with tense; it does not in any sense relate a skeletal propositional structure to a time reference. Hence the prospects for motivating the idea that to belongs to a category Tense seems very doubtful.

8. Why 'to' is a Verb

I now turn to the position that I think is correct. I believe that to is, as strange as this may seem, a previously unrecognized member of the subclass of verbs known as auxiliary verbs. This idea was first put to me in correspondence by Paul M. Postal in early 1978. Later I found that Richard A. Hudson had independently arrived at the same conclusion (see now Hudson (1982)), and after I had written an early version of this paper it was brought to my attention that Fiengo (1980, 194, n. 23) observes: "It appears optimal to consider 'to' a member of the category [+V, -N, +AUX]." The claim I shall defend is original with Fiengo, Hudson, and Postal, not with me. What I shall be doing here is to provide some straightforward arguments, couched in basically transformational-generative terms, for a view which these three linguists have quite independently formed (from the standpoint of three different theoretical frameworks). The arguments I shall offer are not abstract syntax arguments like the ones that were once used to argue that adjectives were verbs, that and was a verb, and so on. I am defending the view that to is an auxiliary verb in surface structure, on the basis of straightforward syntactic and phonological generalizations at a superficial level of structure. Indeed, the case I am making is even stronger when it is reinterpreted in certain purely "surface" theories of grammar (see Gazdar, Pullum and Sag 1982). Differences between competing frameworks are not at issue here.
8.1 Linear position

One clearly established generalization about English constituent order is that in an X′, the X (i.e. the head) is leftmost. The positioning of to falls under this generalization if it is a V, and head of a V′, but otherwise the facts of (21.) have to be stipulated, despite the fact that they parallel those in (22.) through (25.).

(21.a) \[ _{-\text{to see it}} ]_{V} \\
(21.b) \*\[ _{-\text{see it to}} ]_{V} \\

(22.a) \[ _{-\text{see it happen}} ]_{V} \\
(22.b) \*\[ _{-\text{it happen see}} ]_{V} \\

(23.a) \[ _{-\text{sight of it}} ]_{N} \\
(23.b) \*\[ _{-\text{of it sight}} ]_{N} \\

(24.a) \[ _{-\text{cognizant of it}} ]_{A} \\
(24.b) \*\[ _{-\text{of it cognizant}} ]_{A} \\

(25.a) \[ _{-\text{of it}} ]_{P} \\
(25.b) \*\[ _{-\text{it of}} ]_{P} \\

Notice however, that this argument would also count in favor of the view that infinitival to is a preposition; it does not distinguish between them.

8.2 Subcategorization

A second argument can be made on the basis of subcategorization facts. Consider the italicized parts of the following examples.

(26.a) They could be trying to fool us.
(26.b) He wants to help get things straightened out.
(26.c) Are we going to let him tell us what to do?
(26.d) Let's make the Klingons eat their words.

It does not matter whether we assume that that they are baseform VP's (as in Gazdar, Pullum and Sag 1982) or clauses with null subject NP's and uninflected verbs. Let us call them for the moment uninflected complements. It is a fact that only one category, V, can take uninflected complements. There are no adjectives, prepositions, or nouns that can occur with complements like the ones italicized in (26.). Hence if we analyze to as a verb, it fits into a subclass alongside make, let, help, the modals, and so on, whereas if we were to put it in some other category it would introduce unprecedented subcategorizational behavior into that category.

8.3 Ellipsis

A third argument derives from the formulation of the rule for VP ellipsis. Whatever its exact character, the rule that allows VP positions to be null and interpreted by reference to the linguistic context must state that the element immediately preceding the missing VP must be either one of the elements listed in (27.a), i.e. the elements traditionally called auxiliary verbs, or the item shown in (27.b).

(27.a) be, can, could, dare, do, have, may, might, must, need, ought, shall, should, will, would
(27.b) to
This class could be specified by means of a disjunction (either an auxiliary verb or to). It could also be done by means of a system of rules that (a) guarantees that the items in (27.) are the only ones that turn up under the 'AUX' node, and (b) makes VP ellipsis sensitive to a preceding 'AUX'. (Notice that the latter idea, which is implicit in Bresnan (1976: 17), would have the disjunction embedded in it, if there were no explanation of why it is just the items in (27.) that turn up under 'AUX'.) But the proposal I am defending collapses (27.a) and (27.b) into a single lexical subcategory, the category of auxiliary verbs, without any disjunction, however disguised. The simplification this effects in the statement of VP ellipsis constitutes an argument for the verbhood of to.

8.4 Stress and stranding

Another argument emerges from the stress facts already mentioned. Zwicky and Levin note that there is a small class of items that cannot bear stress when they are stranded by VP ellipsis. They are: (a) infinitival to; (b) the infinitive form of the verb be; (c) the infinitive form of the perfect auxiliary have; and, for British English only, (d) the infinitive form of supportive do. The facts are illustrated in (28.).

(28.a)  "It would be easy not to like McCoy, but then again, it would be easy TO Ø. (=17a.1.)

(b)  "I wouldn't be upset not to be chosen as the next Science Officer, but on the other hand, I wouldn't be upset to BE Ø.

(c)  "I didn't enjoy battling against the Gorn; I'd have to be crazy to HAVE Ø.

(d)  [British English] "The United Kingdom hasn't yet agreed to support the Federation, but it could DO Ø at any time.

[Grammatical if could is stressed instead of do.]

Zwicky and Levin say:

We will suppose in what follows that the complementizer to and the infinitive auxiliary verbs have and be constitute a grammatical class, for which we coin the name infinitoids.

For Zwicky and Levin the members of this ad hoc class of 'infinitoids' cannot be simply infinitives, however, because they explicitly assert that "to is not a verb", though without providing any warrant for this assertion. But under the analysis I am advocating, the class of infinitoids is simply the entire class of base-form auxiliary verbs (the modals being absent because they do not have infinitive forms). The availability of the right natural class constitutes a clear argument for treating to as a verb.

8.5 Negative placement

A further argument involves the facts of negative (not) placement. Not, when functioning as a sentential or verb phrase negator, occurs in the three environments listed in (29.).

(29.a)  After a finite auxiliary verb, as in has not gone.

(b)  Before any nonfinite verb, as in not having gone.

(c)  Before to, as in not to have gone.

What I am proposing is that to is a nonfinite verb. This collapses (29.b) and (29.c), which yields another argument for the analysis I advocate. The remarks of Fiengo (1980: 73) about placement of not "before tenseless verbs" suggest he too regards negative positioning as an argument for the verbhood of to, for he cites the example Not to have been on time was a pity, and appends the previously quoted note saying that to belongs to [+V, -N, +AUX].

8.6 VP ellipsis and negation

A sixth argument, closely related to the third and the fifth, concerns VP ellipsis in negated sentences. Although not can follow a finite auxiliary that is stranded by VP ellipsis, it is impossible to have not after a stranded nonfinite auxiliary:

(30.a)  *By three o'clock I will have finished but you will have not Ø.

(b)  *Tomorrow at this time I will be working in the engine room, but you will be not Ø.
Exactly the same thing is true with *not* placed after stranded *to*:

(31.) *You usually pay a lot of attention to what McCoy says, but you ought to not.*

If *to* is a nonfinite auxiliary verb, this is exactly as expected, since any general device blocking the sequence "nonfinite auxiliary — *not* — missing VP" will block (31.) in the same way as it blocks (30.).

8.7 **Contraction**

My seventh argument comes from facts of *to* contraction and the contexts in which it occurs. In Postal and Pullum (1978) it was observed that this phenomenon, which gives rise to forms like *wanna, hafta, gonna, oughta, gotta,* and *sposta,* is found only where the subject of the lower clause is nondistinct from the subject of the matrix clause (in a sense that can be made precise in terms of relational grammar; see Postal and Pullum (1978: 23ff)). What it means in terms of standard transformational grammar is roughly that *to* contraction is found only where Subject to Subject Raising or Subject-Controlled Equi have applied. Frantz (1977) had independently noticed this, and had further offered a connection with universal grammar by showing that the same conditions characterized something else: the class of contexts in which Equi-Subject Clause Union is found in languages like Blackfoot. In Equi-Subject Clause Union the boundary between a matrix clause and its immediately subordinate clause is obliterated. In languages like Blackfoot the collapsing of the two clauses is particularly clear because the verbs of the two clauses amalgamate morphologically (see Frantz (1976)). But the construction has also been postulated for languages like French, Spanish, Italian, German, and Dutch (often under the name 'Verb Raising'); see Aissen and Perlmutter (1976) on Spanish 'Clause Reduction', for example. Frantz suggested that there was a connection between clause union and *to* contraction, but did not spell this out in detail. The details can now be supplied, under the present proposal about *to.* We can simply say that *to* contraction is clause union. The rule that attaches *to* to a governing matrix verb is Verb Raising. Where it does not apply, the structure of an example like *I want to live* is as shown in (32.a) Where it does apply, we get the structure shown in (32.b) (I make the assumption in these representations, and from now on, that S is the maximal — i.e. two-bar — projection of V.)
In (32.b), the sequence want to is a (derived structure) V; in (32.a) it is not, and want and to are in different VP's.

We could now give an account of to contraction constraints as follows. Some speakers—a majority, apparently—allow the pronunciation wanna only where the syntactic structure is as shown in (32.b), i.e. where want and to are dominated by the same V node. These speakers get no possibility of contraction in Who do you want to feed the chickens? because the verbs want and to have different end-of-cycle subjects (to put it in classical TG terms), so the 'equi-subject' condition for clause union (Verb Raising) is not met. Other speakers (the speakers of the "liberal dialects" mentioned in Pullum and Postal (1979) allow the pronunciation wanna wherever want and to are linearly adjacent, and thus accept Who do you wanna feed the chickens? This is a simple dialect-specific sensitivity to surface constituency in a morphological rule, with greater phonological integration corresponding to greater syntactic cohesion. As we would expect, even the dialects that have the condition ignore it in very rapid and casual speech. Nothing about the existence of the 'liberal dialects', which ignore the restriction all the time, is surprising. And nothing about the restriction itself is peculiar, either. It is a straightforward example of surface syntactic form affecting phonological interpretation.

This differentiates the present account from those of Chomsky (1980) and Jaeggli (1980), which are attempts to improve on previous trace theory descriptions by invoking the distinction between 'case-marked traces' (basically those left by wh movement from a (nonembedded) position in S) and other traces. Jaeggli is able to deal with examples like There will hafta be a hole in it, while accepting that have takes a sentential complement and that Raising from the embedded subject position has taken place, because he assumes that Raising leaves a non-case-marked trace. But he does not mention the liberal dialects, and apparently has no way of dealing with them (cf. Postal and Pullum (1982) for further discussion). He assumes a principle of universal grammar that says morphological rules cannot operate across case-marked traces. This makes the overly strong claim that the constraint on to contraction will hold for all dialects. The account offered here, which does not invoke traces at all, does not have that defect.

Clearly, the description of to contraction just sketched can only be entertained if to is a verb. To the extent that it is preferable to other accounts, we can use it as support for the analysis of to as a verb.

9.0 More on 'to' as a verb

None of the above arguments is compelling enough to sway a hardened conservative with a mind already set against the idea of there being verbs in English that had not previously been recognized as such. An empiricist who believes that syntactic analyses should spring out of the raw facts will not be convinced, for the arguments I have put forward rest on subtle simplifications in the grammar at a fairly abstract level. But it is important that while the seven points I have made could be argued to provide at least a little support for categorizing to as a verb, it seems that alternative categorizations receive no support at all. However small the degree of plausibility in my arguments might be, they are the only arguments in this domain that have any plausibility at all.

Moreover, I shall now argue that there are no valid arguments against the categorization of to as a verb. It might seem initially as if there were a great many. Almost everything that an auxiliary verb does (like bearing tense, inverting in questions, carrying contracted negation, appearing in tags, and so on) to fails to do. One might think that all the characteristic behavioral properties of auxiliary verbs furnish arguments that to does not belong in this class. But in fact things only retain this appearance as long as we overlook the distinction between finite and nonfinite (tensed and nontensed) verbs. Once that distinction is made, a simple observation essentially clears up all the peculiarity of to at a stroke: to lacks the tensed and participial parts of the verbal paradigm, and always appears in its base form, in base-form contexts.

I suggest that to is a kind of empty auxiliary verb with the function of marking infinitival verb phrases (or clauses). It is always (by definition) the head of an infinitival VP; that is, in
effect, what it is for. The sort of analogy one should keep in mind is with dative PP's in English: they are always headed by the meaningless preposition to (a homonym of the item we are considering in this paper, as it happens); that is what it is for.

Suppose we identify infinitival VP's with a feature INF, and assume an 'orphan VP' syntax of the sort advocated by Brame, Bresnan, and others, so that verbs can have VP complements (rather than only S complements, as in Chomsky's work). I will write 'V [INF]' for the label of a VP with this feature, 'V [BSE]' for VP's whose verb is in the base form (i.e. VP's like be a doctor). A verb like try will be subcategorized to take a V[INF]. The verb to will be the only one in the lexicon that can be the verb in a V [INF], and will itself select a VP with the features [+BSE, -INF]. Thus tries to be kind will be a V [-BSE, -INF]; to be kind a V [INF]; and be kind a V [BSE]. Only the verb to can follow try; and any base-form verb other than to itself can follow to.

The reliance on defective paradigms in the above proposal can hardly be said to be unjustified. Every grammar of English is going to have to allow for definiteness in verbs, and in fact defective verbs are rather commoner than is generally acknowledged. Consider the paradigms given in (34.) for the underlined verbs of the examples in (33.).

(33a) Spock is my second in command.
(b) You should beware of becoming emotional, Captain.
(c) I don't believe a Vulcan can understand what it's like.
(d) We must give first consideration to the lives of the crew.
(e) They are just raging to be set loose for their shore leave.
(f) The whole ship is riven with dissension.
(g) Spock strode into the transporter room.
(h) We can't be sure whether to risk the wrath of Khan.
(i) This planet used to be inhabited.
(j) These hull sections were wrought from an alloy unknown to your scientists.

The full paradigm exhibited by the copula is not exhibited by any of the others. A wide variety of patterns of gaps are exhibited here. (One pattern not seen is for a verb to have missing agreement forms within a given tense; but Russian has well known examples of this: see Halie (1973: 7-8)). There are verbs without participles, and verbs with only participles, and verbs with no base form, and verbs with only the base form, and so on. It seems to me that the existence of defective verbs in English, including ones that are very common as well as ones that are restricted to a limited range of expressions, is well enough established. Many more of them could be cited if one were to drop the traditional but unmotivated view that pairs like go/went or am/was are suppletive forms in the paradigm of a single verb rather than pairs of defective verbs whose partial paradigms complement each other. Thus the degenerate paradigm of to cannot itself militate against an analysis that fits it into the language as a defective, tenseless auxiliary verb.

10. An objection: the commutation principle

Rodney Huddleston has suggested to me an objection to the claim that to is a verb. It depends on a principle that he feels is implicit in previous grammatical practice, a principle that I shall call the commutation principle. The principle states that a lexical item L shall not be assigned to a grammatical category C unless there are at least some sentences containing L in which some other lexical item in C could replace L without altering grammaticality. The commutation principle tells against the categorization of to as a verb because there are no sentences containing to in which a verb could be substituted for it without destroying grammaticality. I want to claim that the commutation principle is a rule of thumb that works because of the overwhelming overlap in grammatical properties that typically exists within a lexical category, but it is not a theoretical principle, and we should expect to find exceptions in it.

Note first that the commutation principle apparently cannot be adhered to by those grammarians who espouse an analysis of English auxiliaries along the lines of Chomsky (1957). The item ought shows by its lack of third person
singular -s inflection, its invertibility in polar interrogatives, and its obligatory finiteness that it must be placed in the category M (Modal). But there is no other modal that can be substituted for it in a grammatical sentence, because it is unique among the so-called modals in requiring a following to:

(35.) I don't think she ought (*can/*dare/*may . . . ) to be so trusting.

Hence defenders of the M category, at least, cannot legitimately invoke the commutation principle in arguing against the thesis maintained here.

Consider also perfect have: this is quite generally accepted as a verb — certainly by Huddleston, to make an ad hominem point, but also by defenders of AUX-node analyses, e.g. Emonds (1978). Yet no verb in English can substitute for have in, for instance, Doc has broken his tricorder.

I do not think, in fact, that the commutation principle can be relied upon under any assumptions, in the description of the auxiliary system or anywhere else, though I shall have to cite Gedankenexperimenten in a number of cases to show why. Consider, for example, the evidence provided by Dixon (1977) concerning the occurrence of the category A (Adjective) in languages of the world. Dixon finds that some languages have an extremely small class of items belonging to the category A; Hausa (Chadic; Nigeria) has twelve, Igbo (Kwa; Nigeria) has eight, Mulluk-Mulluk (Australian; Northern Territory) has only seven. Imagine a language with seven adjectives in which just two of them appeared prenominally in NP's, the rest being restricted to predicative or postnominal attributive position. Now imagine this language losing one of its two prenominal adjectives by linguistic change. It now has an adjective such that no other adjective can be substituted for it without causing ungrammaticality. Is this such an implausible scenario? I do not think so; yet the commutation principle, interpreted as a theoretical principle, says it could never happen.

Even within English grammar, in areas other than the verbal system, there are facts that make the commutation principle suspect. Huddleston himself has pointed out to me that there are no members of the category (NP[-PRO]) that can substitute for the dummy pronouns it and there, nor can they substitute for each other, yet it is standard practice to assume that they are indeed pronouns (as the evidence from tag-question formation suggests: It's clear that he's right, isn't it?; There's a problem, isn't there?). Another example is provided by Maling's analysis of near, already cited. If she is right, there is an adjective in English that in many occurrences cannot be replaced by any other adjective, even a synonymous one:

(36.a) I left it near (*close) the tree.

(b) This is nearer (*closer) the tree than that.

(c) Pick up whichever is nearest (*closest) the tree.

Because it happens that near can be used intransitively as well as transitively, we can also find examples in which commutation is possible (It is very near. It is very close), but this, as far as I can see, could just as well have been otherwise.

And consider the list of verbs that subcategorize for a prepositional phrase with (a meaningless occurrence of) the preposition of. It seems to be quite small:

(37.) He never approved (*wanted/*expected/*understood/. . . ) of that.

Would English be impossible to learn if the class in question shrank through linguistic change down to one, and that one happened to require the of phrase obligatorily? Again, I do not think so.

I hold, therefore, that it is not correct to reject the notion that to is a verb solely on the grounds that it constitutes an exception to the generalization implicit in the commutation principle. The idea that linguistic theory provides a set of deterministic procedures of analysis of the general sort that the commutation principle suggests is an idea that was rejected many years ago after Chomsky's widely accepted critiques of discovery-procedure methodology in linguistics, and I believe that rejection was appropriate.
11. Conclusion

The conclusion of this paper is that there are good reasons for analyzing the infinitival marker to in English as a verb, and no good reasons for not doing so. This conclusion has been argued for in terms of fairly conservative transformational-generative assumptions, but seems likely to be transferrable to almost any theoretical framework for syntax. To could certainly be treated as a verb (i.e. as the head of a P-arc) in relational or arc pair grammar; it is treated as a verb in Hudson's (1982) word grammar; and it is labeled V and analyzed as the head V' in the X-bar phrase structure system of Gazdar (1982) and Gazdar, Pullum and Sag (1982). What would be involved in incorporating a verbal analysis of to into a government-binding syntax (Chomsky 1981) is not something I have investigated, but might merit attention.

Since previous treatments of the syntax of infinitival to have in large part been poorly motivated and inexplicit, I believe the suggestion argued for in this paper should be given close attention. There may be sound reasons for rejecting the idea explored in this paper in favor of some alternative, though I do not know of any. If there are, it would be worth taking the trouble to show this in detail, and make the alternative fully explicit, rather than assuming that something along the lines of earlier generative analyses (where there are any) can be accepted. It will be unfortunate if currently uncontroversial but ill thought-out analyses continue to be maintained on the basis of a default principle of pure conservatism.

NOTES

1 Many thanks to the people who talked with me about the topic of this paper — both those who liked the idea and those who hated it. Among them are Gerald Gazdar, Jorge Hankamer, Richard Hudson, Polly Jacobson, Ellen Kaiser, James McCawley, Fritz Newmeyer, Ivan Sag, and especially Paul Postal. An earlier treatment of this topic was presented at the LSA Annual Meeting in San Antonio, Texas, in December 1980, while I was a Visiting Professor in the College of Arts and Sciences at the University of Washington. It appeared in that university's Working Papers in Linguistics 6 (1981) 56-72 as 'The category status of English infinitival to'. The present paper supersedes this, and was completed at the Syntax Research Centre, University of California, Santa Cruz, in the summer of 1982. My thanks to the SRC's Karen Wallace for her able editorial and research assistance.

2 It might be that the notion clitic could also be defined along roughly these lines:

   A clitic is a terminal symbol dominated by a lexical-category-labeled node with exactly one sister, of the same category as its mother.

I shall not follow this up here.

3 Even the set of 'short' terminal symbols — say under ten segments — is infinite if Sampson (1974) is right that there is no finite inventory of possible segment types. See Pullum (to appear) for further discussion.

4 American English has supportive do only in finite contexts. British dialects seem to have it in nonfinite ones as well, in examples like He may have done.

5 The function $f(n)$ giving the number of ways in which a string of $n$ terminal symbols can be bracketed has the following recursive definition:

   \[ f(1) = 1; \]

   \[ f(n) \text{ for } n > 1 \text{ is the sum of all } f(i_1) \times f(i_2) \times \ldots \times f(i_k) \]

   such that \( k > 1, \)

   \( \text{ and } (b) \text{ for all } i_j, i_j < n, \)

   \( \text{ and } (c) \text{ for all } i_j, i_1 + i_2 + \ldots + i_k = n. \)

Thanks to David Watt for figuring this out.
*Orphan VP* is Postal's term for a VP that does not have deep structure S node that immediately dominates it (and is thus its 'mother'). VP complements in the VP are necessarily orphan VPs.

Carlson (1979: 36) appears to offer a resolution of this difficulty when he proposes that to is a "pro-AUX". Carlson's discussion of infinitive constructions is interesting, but his data are purely semantic, and I believe his semantic analysis, under which to "is assigned values by the surrounding context from the domain of the modals (which ... include the tense markers)", is fully compatible with the syntactic analysis I shall propose below.

I have other grounds for rejecting the 'AUX' analyses that I have seen so far, but I shall not review them here. For defense of my view that it is a mistake to postulate an 'AUX' category, see Pullum and Wilson (1977) and Pullum (1981).

That is, to set up a category for to alone would be to give up the search for a motivated, universal set of grammatical categories. Where a category is set up, some kind of cross-linguistic basis should be available for it in principle. Akmajian, Steele, and Wasow attempt to provide such a basis for the category 'AUX' (failing, in my view; see Pullum (1981)). But note that the attempted definition of 'AUX' that they put forth cannot by any stretch of the imagination identify to as a likely member.

Notice that it should not necessarily be assumed that the authors mentioned would endorse the arguments I give.

Actually, this ignores the placement of the negative particle not, which is discussed below. The slight oversimplification at this point does not affect the argument.

Incidentally, we can now state the restriction that would prevent clause union from attaching all subordinate verbs to their matrix verbs instead of just attaching to: clause union in English applies only where the complement is [+INF]. If the attachment is limited to the verbs want, have, going, ought, got, and supposed, as Postal and Pullum (1978) assumed, then there are also restrictions what the matrix verb may be. If instead, following Jacobson (1982), we take to contraction as a syntactic phenomenon to be much more general, the special behavior of the verbs listed above being phonological, then we might want to say that clause union is entirely general in English where the prior conditions (subject sharing, adjacency, and [+INF]) are met.

Ellen Kalisse has pointed out to me that there might be other instance of phonological contraction in English that are contingent on clause union in the syntax, particularly in the case of the contraction of have that yields you shoulda done it, you mighta been killed, they couldn't've seen us, etc. The modal verbs that allow reduction of a following have to an enclitic pronounced [və] or ə appear in the same syntactic configurations as other subject-to-subject raising verbs (cf. Pullum and Wilson (1977)). It is possible that the phonology of have contraction is to be predicted on the basis of syntactic adunction of nonfinite have to a preceding verb by the same process that affects to. The modals happen accidentally to be the only verbs (other than to) that can take as their complement a VP whose verb is perfect have in its base form.

Lexical items keep getting lost from Australian languages, as it happens, because of the prevalence of a belief that a word becomes taboo on the death of a person with a similar-sounding name.

REFERENCES


Emonds, J.E. 1972. 'Evidence that indirect object movement is a structure-preserving rule.' *Foundations of Language* 8. 546-561.


Frantz, D. 1977. 'A new view of to contraction.' *Work Papers of the Summer Institute of Linguistics.* University of North Dakota Session 21, 71-76.


Halle, M. 1973. 'Prolegomena to a theory of word formation.' *Linguistic Inquiry* 4, 3-16.


Langendoen, D.T. 1979. 'Numeralology.' Unpublished paper, Graduate Center, City University of New York.


Pullum, G.K. 1981. 'Evidence against the "AUX" node in Luiseno and English.' *Linguistic Inquiry* 12, 435-463.

Pullum, G.K. To appear. 'How many possible human languages are there?'


Sag, I. 1976. 'A logical theory of verb phrase deletion.' *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society.* 533-547.


July 5, 1982