LITMUS TESTS, THE BLOOMFIELDIAN COUNTERREVOlUTION,
AND THE CORRESPONDENCE FALLACY*

Arnold M. Zwicky
Ohio State University

1. Litmus tests.

1.1. Introductory remarks. The literature of generative grammar contains many implicit references to, and explicit discussions of, tests for various properties of syntactic elements in particular languages. Most of these tests involve the manipulation of linguistic examples—substituting, permuting, adding, removing, or combining elements and noting the grammaticality, paraphrase relationships, appropriateness in context, or some other linguistically relevant property of the result. Tests for idiomaticity can serve as typical examples of the sort of thing I am talking about.

Since an idiom is a construct, in some language, whose meaning is not predictable, on general principles of that language, from the meanings of its parts, the meaning of an idiom must be associated with this construct as a whole. It follows that an idiom should show considerable internal integrity: we would expect it to be made up of certain parts (not admitting alternatives); just these parts (not fewer or more); these in specific orders (without variants, facultative or meaningful, or interrupting
material); and with the whole relating as a unit (not part by part) to other constructs with which it combines. In very clear examples of idiomatic constructions, all of these expectations are satisfied. So, for the idiom take a powder 'leave': substituting alternative words for those in the idiom eliminates the special meaning of the idiom--

(1) grasp a powder, take the powder, take a dust adding or subtracting parts also eliminates the special sense--

(2) take powder, take a large powder

alternative orders, too, eliminate the special sense--

(3) A powder was taken.
(4) A powder we are taking.
(5) A powder is easy to take.

so does interrupting material--

(6) Quentin took, as we all watched in amazement, a powder.
(7) Robin will take, we must surely suppose, a powder.

and relating other elements to parts of the idioms will eliminate the special sense as well--

(8) Shirley took a powder before Tanya took one.
(9) Uriah took a powder, and then Victor did the same to a tablet.

Then, since so many constructs that are clearly idioms show all of these properties, we can turn matters around and use the properties as tests for idiomaticity: we examine a particular construct by seeing how fixed or 'frozen' it is—whether its parts are invariable, or whether they show the freedom of open syntactic constructions; whether transformations can move some part of the
construct away from the rest, or move other material into the construct; and whether the construct is an 'anaphoric island' (Postal 1969), in the sense that parts of the construct may not serve as antecedents for anaphoric elements, nor may parts of the construct be replaced by anaphoric elements. These tests for idiomaticity, and some others, are illustrated in Sadock 1972; the interruption test is treated by James 1972:164.

There are also, of course, all sorts of psycholinguistic tests, which use systematic observation or experiment to generate data of various kinds (recall of material, response times, accuracy of judgments, and the like) which can be treated statistically and from which inferences about linguistic structure can be made. My interest here, however, is in the more purely linguistic tests, like those for idiomaticity--tests for which I will use the term litmuses or litmus tests. Perhaps I should say that the tests used in linguistic analysis are in some ways more like medical diagnostic tests or like tests taken by students in a course than they are like chemical litmus tests; the information they give is often less than we would like, or contradictory, or uninterpretable. For tests of idiomaticity, some indication of the size and scope of these difficulties can be gotten from a reading of Fraser 1970a. I consider the general problem in section 3.2 below. In the meantime I continue to use the attractive chemical metaphor, without meaning to suggest that linguistic litmus tests are guaranteed to decide any question that is put to them.
1.2. Some examples. Linguistic litmus tests have been used by generative grammarians in several different areas. In the following subsections I survey some of these areas, with examples selected from the literature.

1.2.1. Tests for essentially semantic properties. Here I refer to such properties as idiomaticity and ambiguity, which are defined semantically but have significant syntactic reflexes. Several litmuses for idiomaticity have already been mentioned in section 1.1 above. Tests for ambiguity are described at some length in Zwicky and Sadock 1975, with bibliography; I will mention just one type here, that of transformational potential.

The idea behind this type of test is that the two interpretations of an ambiguous sentence may correspond to structures that are differently affected by the application of some transformation, while the single interpretation of an unambiguous sentence necessarily remains through transformational operations. In loose but memorable terms, it is possible for some range of meaning of an ambiguous sentence to 'disappear' when a transformation applies, but the full range of meaning of an unambiguous sentence is untouched. It follows that a disappearance indicates ambiguity, rather than vagueness, generality, or lack of specification with respect to some aspect of meaning. Suppose, for instance, that we are not convinced on other grounds that

(10) They noticed her duck.

is ambiguous, and we wonder if (10) might not merely happen to cover
a broad range of meanings including 'fowl' and 'ducking down'
understandings, just as,

(11) Paula laughed at the kangaroo.

covers a broad range of meanings including 'recent past' and 'remote
past' understandings. If our suspicion is correct, then the two
understandings of (10) should maintain themselves through all
transformational operations, just as the two understandings of (11)
do. But no. Several transformations, including passivization,
clefting, pseudoclefting, and topicalization, eliminate the 'ducking
down' understanding of (10):

(12) Her duck was noticed by them.
(13) It was her duck that they noticed.
(14) What they noticed was her duck.
(15) Her duck they noticed.

Compare (12)-(15) with the results of the corresponding operations
on (11):

(16) The kangaroo was laughed at by Paula.
(17) It was the kangaroo that Paula laughed at.
(18) What Paula laughed at was the kangaroo.
(19) The kangaroo Paula laughed at.

The sentences (16)-(19) all can be understood as referring to the
recent past or the remote past. We conclude from these transforma-
tional potential tests that (10) is indeed ambiguous, but that we
have no evidence from these tests that (11) is.

1.2.2. Tests to determine boundaries between units. Among
the boundaries to be determined are those between words and those
between constituents.
To determine whether given sequences of morphemes constitute words, linguists have referred to various properties of morphemes within words: the morphemes have a fixed ordering, alternative orders normally being associated with differences in cognitive meaning; they show a high degree of mutual selection, much higher than that between separate words; words as wholes are immune to interruption by parenthetical expressions; words as wholes are anaphoric islands; syntactic rules of movement and deletion do not apply to proper parts of words; some phonological processes (those of 'internal sandhi') apply only within words and not between words, while others (those of 'external sandhi') apply only between words; morphemes with no independent accent are affixes rather than separate words; and bound morphemes, as well as morphemes in construction with bound morphemes, are affixes rather than separate words. A number of these tests are discussed, in connection with some particularly troublesome cases, in Zwicky 1977: secs. 1-3. For a very simple example, consider how we might show that danced in

(20) They danced.

is a word. We can refer to the fact that the past tense morpheme -ed is a bound morpheme and has no independent accent, hence is an affix; to the fact that the phonological rule of progressive voicing assimilation that applies to this morpheme is a rule of internal sandhi; to the fact that parenthetical interruption is out of the question--

(21) *They dance, as I supposed you know, -ed.
and to the fact that -ed is immune to the deletion rule that forms reduced coordinations—

(22) They danced and played \[\Rightarrow\] They dance and played.

To determine whether given sequences of words form constituents, linguists have referred to another set of tests. Most often they have depended upon the hypotheses that the only sequences of elements that can be moved (in some cases, deleted) by transformations are those that form single constituents, and that some deletion rules are constrained to leave constituents. Both lines of evidence are used by Fraser 1965: sec. 2.3 to distinguish the constituency of the two constructions

(23) a. \(V \text{ Prt NP: } [V \text{ Prt}][NP]\) (look up the information)
   b. \(V \text{ Prep NP: } [V][\text{Prep NP}]\) (talk about the situation)

Fraser observes that only in (23b) does the postverbal construction prepose in questions:

(24) a. \(^*\)Up what did he look?
   b. About what did they talk?

and that only in (23b) can the postverbal construction be treated as a unit for the purposes of reduced coordinations:

(25) a. \(^*\)He looked up the information and over the files.
   b. They talked about the situation and on the issues of the day.

In the first case, what is moved must be a constituent. In the second, what remains as a reduced conjunct must be a constituent.

A related test (Levin 1976: sec. III) uses the fact that fragment answers to information questions must be syntactic
constituents. So, the question-answer sequence

\[(26)\] Q: What did you throw out?
A: The French toast.

is fine, but

\[(27)\] Q: What happened to the French toast?
A: I threw out.

is bad, because the answer consists of an NP and a V that do not
together form a constituent.

Still other tests for constituency are distributional in
nature, using generalizations about syntactic structure—using, for
instance, the hypothesis that the only categories involved in the
strict subcategorization of some element are those in coconstituency
with that element (Chomsky 1965: sec. 2.3.4 of ch. 2); or using the
observation that most constituent types occur in positions that can
be filled by single words from some syntactic category (an observation
made much of by Harris 1946).

Next, there are tests for the closeness of construction between
constituents, using various hypotheses about the consequences of
relative closeness—that relatively close constituents cannot be
separated by intonational marks, by substantial pauses, by parenthe-
tical material of various sorts, or even by modifying phrases, so that
positions where such separation is possible are locations of major
constituent breaks. Fraser 1965: sec. 2.3 uses two tests of this sort
to distinguish the constituency of the two constructions in (23)
above: he observes that the postverbal construction in (23a) can
be separated by parenthetical material, while the postverbal
construction in (23b) cannot--

(28) a. He looked up, without a reply, the information.
b. *They talked about, without a reply, the situation.

and that the verb cannot be separated from the postverbal construction by a manner adverbial in (23a), though it can in (23b)--

(29) a. *He looked quickly up the information.
b. They talked loudly about the situation.

The data in (28) indicate that the postverbal construction in (23a) is not a close one, while the corresponding construction in (23b) is. The data in (29) indicate that the verb and following word are closely constructed in (23a), but not in (23b).

Finally, there are tests for constituency that appeal to phonological phenomena, in particular to the lack of accent and phonological reductions characteristic of words that are cliticized to other words. Given the general principles governing cliticization and accent placement in a language, these phonological phenomena can be used to argue that particular words form a constituent with the words to which they are cliticized. On grounds of this sort, Postal 1974: sec. 4.4 argues that it forms a constituent with the preceding verb in

(30) Max believes it to be true.

but not in

(31) Max believes it is true.

and Fraser 1976: sec. 1.1 argues that off forms a constituent with the preceding verb in

(32) She ran off the pamphlets.
but with the following noun phrase in

(33) She ran off the stage.

1.2.3. Tests for relationships between parts of units. The units in question may be words, phrases, or sentences, and the relationships in question are various. Under this general heading I include such morphological matters as the distinction between derivational and inflectional affixes, which stand in different relationships to the roots with which they are combined (for a recent brief discussion of the distinction, see Aronoff 1976: 2-3; more extensive lists of tests differentiating the two sorts of affixes are to be found in Nida 1949: 99-100 and Bazell 1953: ch. VI), as well as a variety of syntactic relationships, a few of which I will mention briefly.

First, there are tests to distinguish between several types of coordination with and, types that differ in the way the individual conjuncts are related to their verbs (see the discussion in Stockwell, Schachter, and Partee 1972: 298-324). For example, there are tests designed to separate 'phrasal conjunction', in which coordinate NPs are understood as a unit--

(34) Amelia and Belinda are similar.

from 'sentential conjunction', in which coordinate NPs are understood as separately related to the verb--

(35) Carlo and Dennis are foolish.

Among these litmus tests are the possibility of two-sentence paraphrases--
(36) a. Amelia is similar, and Belinda is similar. 
    [\( \neq (34) \)]
b. Carlo is foolish, and Dennis is foolish. 
    [\( \equiv (35) \)]

the possibility of a paraphrase with one conjunct appearing in a
sentence-final prepositional phrase--

(37) a. Amelia is similar to Belinda. \( \equiv (34) \)
b. Carlo is foolish to Dennis. \( \neq (35) \)

and the occurrence of quantifiers like both and each--

(38) a. *Both Amelia and Belinda are similar.
    b. Both Carlo and Dennis are foolish.

(39) a. *Amelia and Belinda are each similar.
    b. Carlo and Dennis are each foolish.

Second, there are tests to distinguish between coordination and
subordination as the relationship holding between two constituents
of similar type within one sentence. Li and Thompson 1973, for
example, consider the serial verb construction in Mandarin Chinese
(sentences with a subject NP followed by two predicates, each of
the form V(NP)) and ask whether the second verb is subordinated to
the first or coordinate with it. They observe that Ross constraints
on movement (Ross 1967) prohibit individual conjuncts from being
moved by transformations of a certain type, while at least some
subordinate clauses can be moved; mobility is therefore a litmus
for subordinate status. As it happens, Li and Thompson conclude that
the serial verb construction is ambiguous, with both coordinate
and subordinate (purpose) structures.

Third, there are tests to determine which of the parts of a
construction is the head, and which the dependents or modifiers.
This distinction has played a fundamental role in many diverse theories of grammar, and is not without significance even in transformational grammar; Robinson 1970 surveys the matter with examples from writers working in a variety of frameworks, including Harris 1946, Ross 1967, and Chomsky 1970. Among the tests for headship are independence (an element like wallabies in the piquant wallabies, which can stand alone as a constituent, is the head of the construction), characterizing function (an element like raced in raced across the freeway, the properties of which essentially determine the external distribution of the constituent, is its head), and selectional determination (an element like asked in has asked Mark if pigs can fly is the head of the construction because the choice of verb determines which auxiliaries and objects can occur with it rather than the reverse); for some recent discussion, with an additional transformational test, see Akmajian and Lehrer 1976.

As a final illustration of tests for relationships between parts of syntactic units, I offer the litmuses for grammatical subjects. These have been enumerated and discussed at some length by Keenan 1976, and would require no further comment except that Keenan appears to view the properties he discusses as part of a definition of the notion 'subject of', rather than as external evidences of subjecthood (see the criticism in Johnson ms. 1976: sec. 0.3 on just this point). But we may interpret Keenan's list as a set of litmuses, each providing some evidence as to whether a particular NP
shows the typical signs of a grammatical subject—lack of case
marking, control of verb agreement, serving as the antecedent for
anaphoric constructions, position as the leftmost NP in a sentence,
and so on.

1.2.4. **Tests to identify the syntactic category of particular**
constituents. In this group fall tests to distinguish N from NP,
VP from Pred, NP from S—or simply to determine if some constituent
is an NP, a VP, an S, and so on. The literature on generative
grammar is rich in such tests, among them the Lakoff-Ross 1966 do so
test for VPs, Chomsky’s 1970 arguments that ‘derived nominals’ are
NPs rather than Ss, and Perlmutter’s 1970b demonstrations that
expletive there is an NP. I give now a few comments on these three
illustrations.

G. Lakoff and Ross 1966 maintain that do so is a proform that
stands for VPs and for VPs only. Thus, it may stand for a VP—

(40) I approached the kangaroo after Margaret did so.
(41) The kangaroo yawned, and then the dingo did so, too.

(the underlined portion is the antecedent for do so) but not for
less than a VP—

(42) *I gave them a dollar, and she did so a pound.
(43) *I fell off the bed before the baby did so out of
the crib.

while certain adverbials may act as if they are outside the VP—

(44) Ellen left because the talks were so boring, but
I did so because I had to go to the bathroom.
(45) Francis sliced a salami with a ripsaw while I was
doing so with a meat cleaver.
Chomsky 1970 confronts the category status of 'derived nominals' while comparing them to gerundive nominals; constructions like Steiner's analysis of translation are seen in contrast to Steiner's analyzing translation. Chomsky says that only derived nominals

(46) have the internal structure of noun phrases. Thus we can have such expressions as the proof of the theorem (*the proving the theorem, with a gerundive nominal), John's unmotivated criticism of the book (*John's unmotivated criticizing the book), and so on. Correspondingly, the derived nominals cannot contain aspect; there is no derived nominal analogous to John's having criticized the book. Furthermore, many derived nominals pluralize and occur with the full range of determiners (John's three proofs of the theorem, several of John's proofs of the theorem, etc.). And derived nominals, in fact, can appear freely in the full range of noun phrase structures. For example, the sentence John gave Bill advice is just like any other indirect object structure in that it has the double passive (advice was given to Bill, Bill was given advice). It is difficult to see how a transformational approach to derived nominals can account for the fact that the structures in which they appear as well as their internal structure and, often, morphological properties, are those of ordinary noun phrases.

(Chomsky 1972: 20)

Continuing the discussion later in the same article, Chomsky uses the failure of certain transformations (for instance, the movement of particles over object NPs--Chomsky 1972: 26-7) to apply to derived nominals as an argument for their being NPs rather than Ss. Finally, he notes

(47) that although gerundive nominalization applies freely to sentences with verb phrase adjuncts, this is not true of the rules for forming derived nominals. Thus we have (15) but not (16):
(15) his criticizing the book before he read it
(because of its failure to go deeply into
the matter, etc.)

(16) *his criticism of the book before he read it
(because of its failure to go deeply into
the matter, etc.)

(Chomsky 1972: 27)

In his discussion of expletive there, Perlmutter 1970b observes
that this word

(48) behaves like an NP with respect to transformational
rules in that it inverts in questions (Was there a
commotion?), shows up in tag questions (There was a
commotion, wasn't there?), shows up with so (Joe said
there would be a commotion, and so there was), undergoes
it-replacement [what is now known as raising--AMZ]
(We expected there to be a commotion), and undergoes
the passive transformation (There was expected to be a
commotion).

(Perlmutter 1970b: 116, fn. 6)

1.2.5. Tests for properties (subcategorizations and super-
categorizations) of particular constituents. Here the collection
of litmus tests that have been used by generative grammarians is
enormous, and I can do no more than give a wide sample of citations,
with a few illustrations. It is convenient to group these by the
constituent concerned; I will mention properties of S, NP, N, VP,
V, Quant, and Prt, though this list does not cover all the constituent
types, or putative constituent types, that have been treated in the
literature.5

1.2.5.1. S. Among the properties of S for which there are
tests are the distinctions between negative and positive sentences
(Klima 1964, Baker 1970a), between free relative clauses and embedded
interrogative clauses (Baker 1970b), and between interrogative and exclamatory clauses (N. McCawley 1973, Elliott 1974). Thus, Elliott notes that the complement constructions in examples like

(49)  
\[
\begin{align*}
  a. & \text{ I wonder how Joe saves money.} \\
  b. & \text{ It's marvelous how Joe saves money.}
\end{align*}
\]

behave differently with respect to a number of tests, including occurrence with the adverb ever--

(50)  
\[
\begin{align*}
  a. & \text{ I wonder how Joe ever saves money.} \\
  b. & \text{ *It's marvelous how Joe ever saves money.}
\end{align*}
\]

and introduction by the wh word whether--

(51)  
\[
\begin{align*}
  a. & \text{ I wonder whether Joe saves money.} \\
  b. & \text{ *It's marvelous whether Joe saves money.}
\end{align*}
\]

The complement type in (49a) is interrogative, the one in (49b) exclamatory.

1.2.5.2. NP. Most prominent of the tests for subcategorizations of the constituent NP have been those to distinguish various sorts of nominal constructions, in particular the gerundive (or factive) nominal versus the action nominal versus the 'derived nominal' versus the concrete nominal (see Lees 1960, Fraser 1970b, Chomsky 1970, and many others). In the previous section we saw some litmuses separating gerundive from derived nominals; among the tests to distinguish gerundive from action nominals, as in

(52)  
\[
\begin{align*}
  a. & \text{ Jacob's designing (the igloo) startled us.} \\
  b. & \text{ Jacob's designing (of the igloo) took only three days.}
\end{align*}
\]

respectively, are the absence of of following the nominal in (52a) versus its presence in (52b); the requirement that the nominal in
(52b) describe an activity, while the nominal in (52a) need not--

\begin{enumerate}
  \item the soda's containing saccharin
  \item *the soda's containing of saccharin
\end{enumerate}

and the possibility of an alternative to (52b) in which the prenominal possessive NP appears in a postnominal by-phrase, while there is no such alternative to (52a)--

\begin{enumerate}
  \item *The designing the igloo by Jacob startled us.
  \item The designing of the igloo by Jacob took only three days.
\end{enumerate}

Other subcategorizations of NP have been tested in the literature, as when Hankamer 1973: 33-4 (disputed by Channon 1974) attempts to distinguish 'titular NPs' from true NPs, as in

\begin{enumerate}
  \item We elected Karen head of the board of trustees.
  \item We introduced Lars to the head of the board of trustees.
\end{enumerate}

respectively, and when Postal 1966 (variously amended and criticized by Thorne 1972, Sommerstein 1972, Morgan 1972, Grannis 1974, Thorne 1974, and Milsark 1977) proposes to separate definite from indefinite NPs by means of such tests as occurrence in the existential there construction--

\begin{enumerate}
  \item *There was Melanie on the front porch.
  \item There was a friend of mine on the front porch.
\end{enumerate}

1.2.5.3. N. Probably the most often cited litmus tests for subcategorizations of the category N are those distinguishing mass from count nouns, as in this convenient summary from Quirk et al. 1972: 128--
<table>
<thead>
<tr>
<th></th>
<th>count</th>
<th>mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>zero (or no) article:</td>
<td>*bottle</td>
</tr>
<tr>
<td>b.</td>
<td>indefinite article:</td>
<td>a bottle</td>
</tr>
<tr>
<td>c.</td>
<td>indefinite quantifier:</td>
<td>*some bottle</td>
</tr>
<tr>
<td>d.</td>
<td>plural:</td>
<td>bottles</td>
</tr>
</tbody>
</table>

There are also tests for the subclass of N comprising pronouns, or for particular types of pronouns, illustrated in Sadock's 1975: 26-7 discussion of why the English personal pronouns do indeed constitute a class of words with similar behavior, and in Stahlke's 1976: sec. 2 arguments that the relative that does not share significant properties with the (clearly pronominal) wh words, hence is not a pronoun.

1.2.5.4. **VP.** By far the subcategorizations of the category VP that have been most pursued in the literature on generative grammar are the distinction between active and stative predicates (G. Lakoff 1966, further developed in Lee 1971, Valesio 1971, Ross 1972a, Cruse 1973, Sag 1973, and Dowty 1975, among others) and the elaborated distinction between predicates of state, activity, accomplishment, and achievement (based on Vendler 1957; see the survey in Dowty 1972: ch. 2). Some of the tests for these distinctions are clearly semantic; thus, the fact that

(58) Norbert walked for an hour.

entails

(59) Norbert walked at each time during an hour.

while

(60) Olive painted a picture for an hour.

does not entail
(61) Olive painted a picture at each time during an hour. indicates that the predicate walk is an 'activity' predicate (in Vendler's terms), while paint a picture is an 'accomplishment' predicate. Other tests are (at least apparently) syntactic; so, the fact that ask questions can occur in the progressive in English--

(62) Three people are asking questions.

while constitute a trio cannot--

(63) *Three people are constituting a trio.

indicates that ask questions is an 'active' predicate, while constitute a trio is a 'stative' predicate.

1.2.5.5. V. Among the tests for subcategories of V are those for verbs taking different types of complements (Kiparsky and Kiparsky 1970, Bresnan 1970, and Hooper 1975, and many other sources), those for verbs taking different combinations of Fillmorean case relations (for instance, the verbs treated in Fillmore 1970), and an assortment of tests for other verb classes (see Zwicky 1971 on manner-of-speaking verbs and Zwicky 1973 on connection-of-ideas verbs). As an illustration, we can take some of the Kiparskys' litmus tests for factivity, including both semantic tests, like the fact that

(64) They regret that Paula likes eels.

t entails

(65) Paula likes eels.

and syntactic tests, like the obligatory nature of that in (64)--

(66) *They regret Paula likes eels.
and by the immunity of NPs within complements of verbs like regret to extraction—

(67) *What do they regret that Paula likes?

To (64)-(67), compare examples with think instead of regret—

(68) They think that Paula likes eels.

does not entail (65); and both

(69) They think Paula likes eels.
(70) What do they think that Paula likes?

are substantially better than (66) and (67).

Another important set of tests for subcategories of V comes in separating 'true' verbs from modals and other auxiliaries, a project that has been pursued by many writers and is nicely summarized in Pullum and Wilson (to appear), who note seven criteria distinguishing the two subclasses of verbal elements (while maintaining that they belong to the same large syntactic class, despite Jackendoff's 1972: 100 claim that modals and true verbs show 'totally different syntactic behavior').

There are finally, supercategorizations of V, properties that verbs share with other syntactic categories. The relationship between 'true' verbs and auxiliaries can be seen in this light, but the supercategorization of V that has attracted the most attention in the literature is the relationship between verbs and adjectives, following G. Lakoff's 1970: app. A discussion of properties shared by the two.

112
1.2.5.6. **Quant.** I devote this subsection and the next to citing a few studies in which properties of minor, closed syntactic categories have been investigated. Among these is Jackendoff's 1968 division of English quantifier words into two classes—

(71) a. some, each, few, which, all, both
b. a few, many, one, three

primarily on the basis of whether the quantifiers permit articles in front of them—

(72) a. *the some/each/all men
b. the many/three men

1.2.5.7. **Prt.** In Emonds 1972a it is argued that the English minor syntactic categories Prt and Prep belong to a single super-category (as Emonds puts it, particles are 'intransitive prepositions'), on the basis that particles and prepositional phrases share many properties, including modification by the emphatic word right—

(73) We put the spices right on the meat.
(74) Let's look it right up.

and preposability—

(75) Into the house he ran.
(76) In he ran.

1.2.6. **Tests for markedness of categories.** One final collection of litmuses is designed to determine which of two opposed morphological or syntactic categories is marked and which unmarked, the marked category being, in an intuitive sense, the more special, less normal, more complex of the two. These tests have been treated at some length in several places (notably by Greenberg 1966: esp. ch. 3, Schane 1970, and Comrie 1976: ch. 6), so that an extended
discussion is not necessary here. We expect the marked category to be realized by morphological material, while the corresponding unmarked category need not be (thus, the noun plural in English is realized as a suffix, while there is no affix for the singular); instances of the marked category are normally less frequent in texts than instances of the unmarked category; and it is the unmarked category that appears in positions of neutralization (hence, the interrogative word who, which can be used indifferently for individuals or for groups, has singular number agreement).

1.3. The source of litmus tests. It should be noted that one might come across a litmus test in either of two ways--by prediction from general, theoretical considerations, or by happy accident, by noting that one sort of linguistic behavior correlates with another. In the preceding sections I have pointed out a number of tests that follow from other principles; the transformational potential tests for ambiguity in section 1.2.1 are of this sort, as are Li and Thompson's use of a Ross constraint to distinguish coordination from subordination (section 1.2.3) and Perlmutter's arguments for the NP status of expletive there (section 1.2.4). But others, like Emonds' use of emphatic right (section 1.2.5.7) to lump together the categories Prt and Prep, surely are not. There is something of an analogy here in the determination of numerical values for various chemical properties: there are so-called 'theoretical values', calculated from theoretical formulas, and 'empirical values', obtained by actual measurements.
It could even happen that the very property being tested for might have become of interest only through the happy accident that someone noticed a correlation between two properties. But we would expect such discoveries to be rare, since most properties are of interest because they arise from semantic categorizations (like negativity, factivity, and ambiguity) or from surface categorizations like constituency, which are relevant for perception and for phonological organization. I return to these matters in section 3.4 below.

2. The Bloomfieldian counterrevolution.

2.1. Introductory remarks. The utility of linguistic litmus tests has been questioned by Dougherty 1974, who claims that the fact that they refer to properties of strings in surface structure indicates that they belong to the methodology of a 'pregenerative view' of linguistic research. Consequently, according to Dougherty, the use of these tests by adherents of generative semantics signifies a 'Bloomfieldian counterrevolution', a return to the segmentation-and-classification linguistics that was overturned by Syntactic Structures.

Dougherty's attacks on litmus tests are directed against Postal's 1966 discussion of definiteness in NPs; against Fillmore's 1968, 1969 case proposals, which include the delineation of a number of verb classes; against the G. Lakoff and Peters 1966 distinction between phrasal and sentential conjunction; and against the treatment of coordination in J. McCawley 1968. 6 There have been
several responses to this work of Dougherty's and to Dougherty's other work directed against generative semantics: Botha 1976a and 1976b, and now the very long treatment in Botha 1977, all of which take up issues other than those I want to focus on in this article; and J. McCawley 1975, a reply specifically to Dougherty 1974, and a paper that makes several of the points I will put forth below. In addition, a metatheoretical discussion closely related to Dougherty's has now appeared in Culicover 1977, which is a critical treatment of argumentation in Postal 1966 (again) and Sommerstein 1972 on pronouns and articles, Ross 1969 on adjectives and noun phrases, and Carden 1970 on quantifiers and adjectives.

Before turning to a discussion of Dougherty and Culicover on litmus tests, I must say a few words about the conclusions drawn and analyses advocated by Postal and the other authors examined by Dougherty. Postal, as part of a larger attempt to argue that English personal pronouns are, at a more remote level, definite articles, gives a number of diagnostic environments for indefinite NPs and for definite NPs. Now, I am not defending Postal's larger claim—indeed, I would agree with Thorne 1972 that it is more likely that the is an unstressed form of deictic that (parallel to Perlmutter's 1970a derivation of a(n) as a unstressed form of the number word one\(^7\)) than that it represents more remote it-one, parallel to the one; Sommerstein 1972 suggests a related analysis, with definite articles deriving from remote pronouns, rather than the other way around. Nor is it necessary for me to uphold Fillmore's case
grammar in all its details, or to believe that the distinction G. Lakoff and Peters make applies to the examples they cite and for the reasons they give, or to accept any of the analyses of coordinate structures advanced by J. McCawley. All that I am defending is the utility of litmus tests in arriving at and arguing for analyses of particular languages.

2.2. Discovery procedures. Dougherty explains his objections to the Postal, Fillmore, Lakoff-Peters, and J. McCawley articles as follows--

(77) In a discussion of discovery procedures, Chomsky 1957 indicates that procedures of segmentation and classification, i.e. slot filler tests, diagnostic environments, etc., are useless if our goal is to construct a generative grammar. Chomsky argues, quite persuasively to my mind, that the linguist must seek to evaluate grammars which generate the sentences of a language and not attempt to construct substitution tests to discover facts about the surface structure distributional similarities of strings.

(Dougherty 1974: 263-4)

But what Chomsky says, in Syntactic Structures, Current Issues in Linguistic Theory, and elsewhere, is not really that procedures and classification are useless if our goal is to construct a generative grammar (though to be fair to Dougherty, Chomsky's early writings on this question are spirited to the point of overstatement, so that it would be easy to mistake his intent). Rather, his object is to repudiate the view that linguistic theory is identical with the specification of a discovery procedure for grammars (a view in which 'theory is equated with practice' (Teeter 1964: 204)), as in this quotation from Chomsky 1966 (which Dougherty himself cites on page 266)--
(78) [the] fundamental assumption [of structural linguistics] is that procedures of segmentation and classification, applied to data in a systematic way, can isolate and identify all types of elements that function in a particular language along with the constraints that they obey. A catalogue of these elements, their relations, and their restrictions of 'distribution' would, in most structural views, constitute a full grammar of the language.

(Chomsky in Reibel and Schane 1969: 6)

To reject the identification of theory with method is not to reject particular pieces of method. Chomsky's criticism is appropriate when directed towards views of linguistic litmus tests that take them to be definitional for properties and to be embedded within a discovery procedure, but not necessarily when it is directed towards other views. It can even be argued that the structuralists attacked by Chomsky had a much less stringent notion of the connection between theory and method than Chomsky attributes to them (just as I want to argue that the generative semanticists attacked by Dougherty have a much less stringent notion of this connection than he attributes to them). Miller 1973: 123, in fact, notes that 'in the work of [Bloomfield, Harris, Wells, and Hockett] a distinction is carefully drawn between the actual process of discovering the structure of a language and the business of describing a structure which has already been discovered'—though it must be granted that the theoretical pronouncements and practice of these linguists are sometimes divergent, and their theoretical work is not entirely consistent, so that their intent is not always clear.

At any rate, there is no reason to see in references to litmus
tests an abandonment of the Chomskyan revolution in linguistics.\textsuperscript{9} It is a misinterpretation of the articles Dougherty cites to see in them the sort of definitional discovery-oriented criteria that characterize works like Fries' *Structure of English* (Fries 1952). As J. McCawley puts it,

(79) Dougherty's objections [to Postal 1966] seem to be in the realm not of methodology but of etiquette: he is objecting not to the content of Postal's treatment but to Postal's having chosen a mode of presentation that reminds hyper-sensitive transformational grammarians of a linguistic school that they find offensive.

(J. McCawley 1975: 152)

Indeed, reading Postal the way Dougherty would have us read him should oblige us to read Chomsky's 'Remarks on Nominalizations' the same way. But we have already seen, in the quotation in (46) above, that Chomsky 1970 uses a variety of differences in internal structure, external distribution, and applicability of transformations to distinguish between gerundive and 'derived' nominals and to argue that the latter are NPs. An uncharitable reading would take these remarks by Chomsky to constitute a purely 'structuralist' definition of NP by reference to internal structure, external distribution, and transformability. Similarly, reading Dougherty 1970a in this spirit should lead us to conclude that he had reverted to pregenerative methods--

(80) In the FSR Hypothesis, the interpretation of a coordinate conjunction is determined by features assigned to the coordinated node. Two of these features, \texttt{[\texttt{individuation}] and \texttt{[+totality]}}, are defined in terms of properties of the distributive quantifiers each, all, and both. These features have (roughly) the following interpretation:
[+individual]: The adverbs alone, singly, individually, and independently can occur with a quantifier having this feature. This feature emphasizes the independent, individual action of each of the constituents of the conjunction, and suggests that a conjoined sentence paraphrase exists.

[-individual]: The adverbs alone, singly, individually, and independently cannot occur with a quantifier having this feature. No conjoined sentence paraphrase exists. The elements of the conjunction are like the elements in a series bound together by plus signs. Often and and plus are interchangeable conjunctions with [-individual] quantifiers: Two plus/and two is four.

[+totality]: The adverbs together, simultaneously, en masse, at once, in chorus, etc. can occur with a quantifier bearing this feature. The surface-structure quantifier is all. The feature emphasizes that the conjunction is to be considered as a unity; the conjuncts act mutually.

[-totality]: The adverbs simultaneously, en masse, etc. do not occur with a quantifier having this feature. The conjunction is not to be considered as a unity; the conjuncts do not act mutually.

(Dougherty 1970a: 868)

But surely this would be to mistake Dougherty's intent, and in fact he adds to the discussion in (80) the note that his 'definitions are merely suggestive, since the real content of the features comes from the role they play in the grammar'.

2.3. Constituency test arguments. Culicover 1977 considers a collection of litmus tests that have been used to argue that two superficially distinct categories (pronouns and articles, adjectives and noun phrases, quantifiers and adjectives) are 'really the same', that is, belong to a supercategory. Culicover sees these constituency test arguments (his term) as taking the form

(81) Classes A and B display the same distribution of grammaticality judgments with respect to some test T. Therefore, in the structures in which T is applicable, A and B are the same type of underlying constituent, and are distinguished by some feature if necessary.

(Culicover 1977: 66-7)
and says that

(82) Users of [(81)] appear to be of the opinion that the greater the number of tests satisfied by classes A and B, the better the argument that A and B are the same type of constituent...
If classes A and B fail to satisfy a particular test this does not necessarily lead to the conclusion that A and B cannot be the same type of constituent for someone who employs [(81)].

(Culicover 1977:67)

In addition, he maintains that (81) is a (covert) part of the evaluation metric assumed by syntacticians like G. Lakoff, Postal, and Ross. This last point I would like to dismiss fairly quickly, since here Culicover seems to be assuming that the reasons linguists give to other linguists for selecting one analysis over plausible alternatives necessarily represent pieces of an evaluation metric. I see no reason to think that evidential questions should find direct reflection in theoretical assumptions like those embodied in an evaluation metric, even for linguists who subscribe whole-heartedly to the idea of such a metric (as many of those criticized by Culicover do not).

Otherwise, the burden of Culicover's discussion is that the arguments he considers for lumping categories together are often weak (a point made very cogently by Schachter 1973), in particular that there are in most cases items that do not pass the tests, as when various adjectives fail Ross' 1969 tests--

(83) a. Peter is smart, but he doesn't look it.
   b. *Peter has become careful, but Susan will never become it.

and that the writers criticized sometimes miss generalizations in a
way that undercuts the conclusions they would like to reach. With these criticisms I have no quarrel of principle. But they in no way show that linguistic litmus tests are valueless—only that, as tools, they must be carefully chosen and carefully used. On the question of counterexamples, I shall have more to say in sections 3.2. and 3.3 below.

2.4. Support for litmus tests. In my remarks this far I have maintained a distinction between methods or techniques of arriving at linguistic descriptions, on the one hand, and the content of descriptions and theories, on the other—a distinction I have explored in a number of articles (especially Zwicky 1975). Now I wish to defend the claim that such points of methodology as litmus tests are worth serious investigation. Perhaps I should say that my own interest in the subject comes from observing that the decisions linguists make between alternative analyses are often not dictated by theoretical considerations, but rather reflect certain methodological biases on the part of the investigator. Beyond this, there are at least three reasons why litmus tests should be taken seriously.

First, a general point. The working linguist, like any working scientist, needs orderly methods for attacking his tasks and some accepted forms for presenting and defending his results. The phenomena to be described are so complex that we cannot rely entirely on unanalyzable lucky guessing as the source of hypotheses, and the amount of material to be communicated is so great that we cannot expect writers to construct a new mode of organization for each argument. This is not to deny that great advances require
brilliant strokes.

What Dougherty says on this point is that a generative linguist (84) must 'invent' new grammars and test them against existing grammars. Theory construction and hypothesis testing are the main tools to achieve rational progress. Hypothesis testing is a general method employed to choose the superior grammar from a field of alternative grammars which make conflicting claims about a given range of data. The method requires the alternative grammars to be compared with each other and with the data.

(Dougherty 1974: 272)

But the invention of new grammars is not a totally random process. Hypothesis testing is, of course, a standard procedure in generative grammar (and in structuralist approaches too, to judge from works like Harris 1960). However, unless 'theory construction' and 'hypothesis testing' are understood in a very loose and inclusive manner, they cannot serve as the only methods in grammar construction and validation.

Next, the working linguist needs ways of establishing the content of superficial syntactic representations, semantic representations, and (in some theories) deep syntactic representations. Hypothesis testing alone will not suffice here, since these are the things the hypothesized grammars are supposed to generate or relate; they are the given of analysis. Though linguists have been inclined to take for granted their own assertions that a sentence has such-and-such as its constituent structure and that it has such-and-such as its semantic representation, both these 'external' aspects of grammatical organization are in fact highly abstract, arrived at only by complex
chains of inference from actual observations, and difficult to validate in detail. Hence, there are tests for these 'givens', as in sections 1.2.1 and 1.2.2, and frequent doubts as to whether these constructs are real and, if real, whether they are determinate (as in Haas' 1973 complaints about surface structure, or many philosophers' worries about the idea of semantic representation).

Finally, we would like to make scientific hypotheses as fallible in detail, in addition to as a set, as we can—in the linguistic case, to ask that grammars be as vulnerable in parte, in addition to in toto, as we can make them. Using litmus tests can serve this end for us, by encouraging us to predict connections between properties, rather than merely generating the correct surface structures or achieving the correct pairing between surface structures and semantic representations. Sometimes such predictions will turn out to be wrong, of course; I mention a few cases in the next section.

Note that litmus tests are used not only by operationalist structuralists and generative semanticists, but also by transformational grammarians in work antedating intimations of the split between generative and interpretive semantics (for instance, Chomsky's 1965: 101 arguments for the ambiguity of He decided on the boat), by transformational grammarians in work intended to be neutral to this controversy (Baker 1970b, Elliott 1974), and by clear interpretivists (Chomsky 1970, Dougherty 1970a). The historical source for this approach to analysis is undoubtedly traditional grammar; in most cases, the properties under examination have been believed to be
grammatically relevant for centuries, and it is easy to find argumentation in the compendious grammarians (sometimes there are extensive lists of criteria, as in Buysen's 1959 study of negative elements in English, cited by Klima 1964 and Baker 1970a). Particular litmus tests might be badly chosen or badly applied, but surely the widespread occurrence of litmus tests in linguistic investigations indicates an equally widespread belief in their utility. It is difficult, in fact, to see how we could do without them, as heuristics and in public argumentation.

3. The correspondence fallacy.

3.1. Introductory remarks. A more interesting criticism of litmus tests (not made by Dougherty but certainly intended by Culicover) is that no two tests for the same property give quite the same results. Indeed, there is now a sizable body of work on the failure of perfect correspondence between different tests used by transformational grammarians--for instance, R. Lakoff's 1973: 689-93 discussion of Kiparsky and Kiparsky 1970 and Karttunen 1970, 1971a, 1971b on factivity, Ross 1972b on the distinction between nouns, verbs, and adjectives, Ross 1973a on NPs, and Ross 1973b on NPs and Ss.

However, to conclude from this that tests should not be used, or that different tests always detect different properties, would be to fall into one form of the correspondence fallacy discussed by C. E. Bazell 1952, whose article deserves quotation at length--

(85) The prototype of these fallacies is the assumption that two distinct (sets of) criteria will necessarily lead to isomorphic analyses...
But it is of course no fallacy to assume (at least as a working principle) that two units coincide more often than not. Nor is it a fallacy to assume that it is possible to devise, a posteriori, two sets of criteria which will lead to very similar results. It may be possible to gear the articulatory criteria [for phonemic analysis] to an acoustic analysis, or distributional criteria [for morphemic analysis] to a semantic analysis, providing a good deal is known of what these analyses would be. The fallacy lies in assuming a one-one relation between the results of criteria which have not been selected with this end in view.

The complementary fallacy consists in the supposition that, since different criteria lead to different results, these criteria have necessarily to be taken separately. In determining the phoneme-inventory one must use articulatory criteria alone (earlier Bloch), or acoustic criteria alone (later Bloch), but never both together. Morphemes may be determined on a distributional basis or (perhaps) on a semantic basis; but in absence of correspondence these are simply different sorts of units, maybe with a certain correlation (C. F. Hockett). The recognition of the inevitability of incomplete correspondence leads to the misrecognition of the value (which is not "merely heuristic") of probable correspondence, in determining the units.

But there are fallacies deriving more directly from the prototype. Such is the fallacy of supposing that some single necessary and sufficient criterion will lead, not to a unit fixed by some other criterion, but rather to some unit already recognized either by traditional grammar, or by the "linguistic conscience", or by linguists engaged in field-work who have set up the unit for purely practical purposes. Definitions have been sought for the "word" on this assumption (A. Martinet, K. Togeby). Sometimes (most often in America) a complex, necessary and sufficient, set of criteria replaces the single criterion. The complementary fallacy is sheer scepticism: the unit cannot be defined at all (Daniel Jones for the phoneme, A. S. C. Ross for the morpheme). (Bazell 1966: 271-2)

The version of the correspondence fallacy that I am concerned with here is, of course, the complementary fallacy mentioned by Bazell.

Now it may very well turn out that different tests do select classes that are without question distinct (but happen to overlap).
Thus, Lee 1971 and Sag 1973, among others, have argued that the progressivity test for stative predicates is distinct from the remaining ('activity') tests. As a matter of fact, I doubt that the various tests for definiteness in NPs select anything like the same class.

Aside from this possibility, there are various difficulties that can arise in applying particular tests. I now survey these briefly before returning to the difficulty referred to by Bazell, the failure of perfect correspondence.

3.2. **Difficulties in using litmuses.**

3.2.1. **Limited domain.** The domain of applicability of a test will have some limitations, and examples in which we have a special interest might not fall within this domain. For example, I have occasionally worried about whether the constituent structure of hadn't and similar contracted auxiliaries was ternary--

\[(86)\]

\[
\begin{array}{c}
\text{V} \\
\text{V} \\
\text{PAST} \\
\text{NEG} \\
\text{have}
\end{array}
\]

or binary--

\[(87)\]

\[
\begin{array}{c}
\text{V} \\
\text{V} \\
\text{NEG} \\
\text{PAST} \\
\text{have}
\end{array}
\]

but I know of no test for constituent boundaries that will decide the question. This difficulty arises frequently when we try to
determine constituent structure below the level of the word.

Another example: Zwicky and Sadock 1975: 23-5, 31-4 point out that tests for syntactic ambiguity are, by and large, inapplicable when the meaning of one putative structure is included within the meaning of another, as in the intentional and nonintentional (better: not necessarily intentional) understandings of

(88) Chuck dropped the tea tray.
(89) Doris slid down the stairs.

3.2.2. Fuzzy results. The sharpness of results varies from case to case, primarily because informants find it difficult to make clear judgments about some examples. I am myself unsure about a number of examples in which constituents have been extracted from factive complements, for instance

(90) Who does Johann know Max killed?

Sentences like (90) ought to be ungrammatical, given the Kiparskys' analysis.

In the same vein, many informants find it hard to judge whether examples like

(91) It wasn't a dog Linda saw, it was a bitch.

are acceptable or not. If dog is ambiguous between 'canine' and 'male canine', then (91) ought to be acceptable, according to a contradiction test for ambiguity (discussed by Zwicky and Sadock 1975: 7-8).

There is, of course, now a considerable literature on grammatical judgments and their attendant problems, much of it surveyed in Householder 1973.
3.2.3. **Exceptions.** Exceptional elements may give anomalous results. For instance, although singular proper nouns normally lack articles in English, there are a few that require the article the (The Hague, The Ginza, etc.). We can still use lack of an article as a test for proper nouns (in the singular), but we must note that there is a list of idioms that violate an otherwise useful generalization about English. Asyntactic idioms, like by and large and trip the light fantastic, will frustrate nearly any generalization, in fact.

3.2.4. **Interfering factors.** Interference of various other kinds will sometimes prevent tests from giving usable results. It can happen in the laboratory that the acid dissolves the litmus paper. In the linguistic case, what happens most often is that although a test predicts that some sequence should not occur, it might nevertheless occur as the output of a rule, or the exponent of some construction, other than that for which the test was devised. Thus, singular proper nouns do occur with articles in constructions like

(92) That's not the Chomsky I know.
(93) I know two Chomskys.

And, although occurrence in existential constructions is a litmus for indefinite (as opposed to definite) NPs--

(94) There had been a gorilla on the porch.
(95) *There had been the gorilla on the porch.

there is a 'list' construction in which sequences like (95) occur--

(96) What a morning! she thought. There had been the gorilla on the porch, the narcotics bust, the winning lottery ticket, and the fire in the aviary.
These are cases in which something we expect to be unacceptable turns out to be possible in another context.\textsuperscript{10} There are also cases in which the examples are indeed unacceptable, but for some reason other than the one being considered; this is one of the main lines of criticism Bolinger 1971 advances against the analysis of the verb \emph{remind} in Postal 1970. For instance, Bolinger 1972: 537 examines a putative constraint on coreference in certain complements to verbs of saying or thinking. According to Postal it is such a constraint that accounts for differences like

\begin{quote}
(97) John says Lucille impressed him as being similar to Mary in that she was very emotional.
(98) *John says Lucille impressed me as being similar to Mary in that she was very emotional.
\end{quote}

Bolinger agrees with the judgments, but maintains that the problem with (98) is the use of \emph{say} in a situation where the speaker is better qualified to give the information in the complement than the subject of the sentence is, not a matter of a coreference constraint at all. Consequently, the restriction illustrated in (97) and (98) could not be used as a test for experiencer NPs in complement constructions, as it could be if there was the sort of constraint Postal posits.

Factors not strictly part of grammar may also interfere. For instance, although occurrence in subject position is a test for being an NP, that sequences are sometimes unacceptable as embedded subjects—

\begin{quote}
(99) *For that Marcia came to surprise Martin would astonish me.
\end{quote}
We might conclude from this, as Emonds 1972b: sec. 1.5 does, that 
that S sequences simply are not NPs. But many have seen the problem 
with (99) not as a grammatical restriction, but rather as 
unacceptability resulting from a perceptual difficulty manifested 
by (99); so Ross 1967; sec. 3.1.1.3.1 describes the construction as 
one 'excluded on the basis of [a] very general output condition on 
performance', and Langendoen 1970: 99 sees it as a case of 'sentence 
unacceptability arising from internal complexity'.

3.2.5. **Variously significant results.** The interpretation of 
negative results is different for different tests. Some tests for 
the occurrence of property y are of the form if x then y, others 
of the form if x then not y, and a few of the form x if and only 
if y. Negative results are significant only in the last case. 
Otherwise, we can conclude nothing from the fact that property x 
does not occur. Thus, from the fact that factive verbs don't 
permit the raising of negatives--

(100) I didn't regret Hugh was at the party. #
(101) I regretted Hugh wasn't at the party.

and the fact that **say** doesn't permit this raising

(102) I didn't say Irene was the winner. #
(103) I said Irene wasn't the winner.

we learn nothing about whether or not **say** is factive.

Similarly, we observed in section 1.2.1 that transformational 
potential tests indicate that

(10) They noticed her duck.

is ambiguous, because one of the understandings of (10) disappears
when various transformations are applied, but that from the fact that

(11) Paula laughed at the kangaroo.

maintains both its 'recent past' and 'remote past' understandings through the application of these transformations we conclude merely that we have no evidence that (11) is ambiguous (not that (11) is definitely unambiguous with respect to the two types of past tense). Even if we could somehow know that no transformation of English affected the two understandings of (11), we could still not conclude that (11) is unambiguous (though the discovery would certainly support that conclusion); the way this litmus works is that if application of a transformation eliminates an understanding then the sentence is ambiguous.

3.3. **Center and margin.** Let us return now to Bazell's discussion of the correspondence fallacy. Bazell's response to the failure of perfect correspondence is to say that for each category or property there is a **central** class of elements, and various classes that include this central class plus some **marginal** elements. Here is his discussion of the morpheme--

(104) ...it is merely the narrow Harris-like sense of the term morpheme that gives rise to this necessity [for a variety of terminological distinctions]. The boundaries of the unit having been fixed once and for all, the word is not any longer available, although the new unit has precisely the same nuclear members. And the many other units having the same nucleus will similarly claim special terms.

The solution is easy. We maintain the term morpheme for all units having the same nucleus. "Productive morpheme" can be used to describe the
category of units defined by the criterion of productivity, "distributional morpheme" for the category defined by distributional criteria, and so on. Similarly morpheme–alternants may be described as alternating morphemes. For this the term morph may be considered a useful abbreviation. But it will not be wrong to use the term morpheme for any of these units, providing the context is clear.

(Bazell 1966: 274)

And in reference to an attempt by Martinet to define langue, Bazell maintains that 'it is surely enough to define the centre of the linguist's interests. As for the marginal cases, the linguist may occupy himself with one or another of them as he likes' (283).

Bazell's use of the center/margin distinction here is especially interesting in connection with recent proposals for the hierarchic arrangement of linguistic elements (among them Quirk 1965, Fraser 1970a, DeCamp 1971, Bailey 1973, the Labov papers collected in Labov 1972a, 1972b, and the Ross papers cited in section 3.1 above). Since Ross' squishes are perhaps the easiest of these proposals to relate to the center/margin distinction, I will discuss them in some detail.

A squish is a two-dimensional matrix which is itself the product of two linear orderings--one ordering of linguistic elements with respect to some property, and one ordering of tests for this property. I will consider Ross' 'fake NP squish' (Ross 1973a) for a concrete example. The property in question is being an NP; the elements examined are, in order,

(105) animates (Harpo), forces of nature (the heat) and concretes (the rock), events (the concert), abstracts (his evasiveness), tack in take Det tack, headway in
make headway on, 'weather it' with predicate adjectives, the dummy it with extraposed clauses, 'weather it' with verbs, expletive there, tabs in keep tabs on, heed in take heed of.

and the nineteen tests applied include, in order,

(106) occurrence in tag questions, as the head of a relative clause, with conjoined VPs, with the possessive suffix, as the antecedent for subject deletion in adverbial gerundive clauses (Before leaving, Herbert uttered many foul oaths), and in double raising-to-subject constructions (Irene is likely to be shown to have known all along).

The intersection of these two hierarchies is a matrix of the form

<table>
<thead>
<tr>
<th>item₁</th>
<th>test₁</th>
<th>test₂</th>
<th>test₃</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>item₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>item₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

in which the cells are filled with symbols indicating whether or not a particular test applies to a particular item. If the hierarchies are arranged like those in (105) and (106) above, then the northwest corner of the matrix will be the 'positive corner', the southeast corner will be the 'negative corner', and the dividing line between the positive and negative parts of the matrix will be approximately the diagonal running from southwest to northeast. And the way the items and tests are arranged in the hierarchies is so that matrix will show just this pattern—if the data will submit to such an arrangement (not all will, of course). What this means is that the items being tested are close to the positive corner—are more central

134
members of the class being tested—according as they pass more of
the tests, and the tests being applied to the items are close to
the positive corner—are more central tests—according as they
select more items. Animates are very central NPs, heed a very
marginal one. Occurrence with tag questions is a very central test,
double raising-to-subject a very marginal one.

Bazell's account, however, is consistent with there being
many distinct sorts of marginal elements with the same center, so that
the elements are not necessarily arrangeable in a hierarchy.
Indeed, Bazell's treatment of the morpheme concept in (104) above
makes it clear that this is his view, that elements are arranged
around a center not in implicational, hierarchical, or squish
fashion--

(108) 

but rather in a multiple overlap pattern--

(109) 

And, in fact, there is a good reason to think that the typical
pattern for items and tests is more like (109) than (108). Ross'
squishes have a relatively large number of cells with marks that
do not fit the expected pattern, which suggests that the data he
is examining simply are not the product of two hierarchies like (108); see the extended criticism in Hindle and Sag 1975: sec. 2.

The pattern in (109) is now a familiar one from studies of word meanings (see particularly Rosch 1973, Labov 1973, and Fillmore 1975), where the referents of a word cluster around prototypical cases and diverge in various dimensions from these prototypes. I will pursue this analogy a bit in the next section, after reconsidering the question of where litmus tests come from.

3.4. On the motivation of litmus tests. Even when a set of data does fit well into the squish paradigm, the resulting arrangement of items and tests is only an organization of the facts and not an explanation of them. If the nearness of an item or test to the positive corner of the squish is to correspond to its centrality, then we must ask why these elements are arranged as they are. Presumably the most central tests are those intrinsically connected with the property in question. But what of the others? Ross himself has articulated the problem quite clearly--

(110) ...so much of [the Fake NP Squish] is unexplained. The ordering from top to bottom is not such a problem: in Navaho, such hierarchies as that among the first four rows [animates, forces of nature and concretes, events, and abstracts] are well documented, and the fact that some idioms are more frozen than others is also well known. But why should the syntactic processes that head the columns be arranged in just the way they are? Why, for example, should get-passives be choosier than be-passives, rather than vice versa? Why should the pronouns that appear in tags be less choosy about the noun-phaseness of their antecedents than other pronouns? Why should Being DELETION be so much choosier than To Be DELETION, when they often
produce identical output sequences? And why should PROMOTION be so much choosier than RAISING, when they seem to be so similar in function? (Ross 1973a: 128)

Thus, in some cases, like the first four classes of items in (105), there seems to be a principled relationship between the degree of centrality of a class and its behavior with respect to litmus tests. For the first four classes of items in (105), their centrality is related to their semantics (with animate beings, especially human beings, as the 'best' or prototypical things for referring expressions to refer to, and other things less good according as they have fewer of the salient properties of human beings—sentience, intentionality, activity, concreteness, and so on). Systematic relationships like this are what led traditional grammarians to see the parts of speech and other syntactic categorizations as 'notional' or semantic at root, an idea revived by Lyons 1966 in reaction to structuralist and early generative assumptions that syntactic categorizations were autonomous of semantics, and an idea pursued by L. Anderson 1974 in connection with squishes; L. Anderson 1974:54 speaks of 'quasi-universal tendencies for related meanings to be expressed by words of the same "grammatical category"' and looks on a category as 'a "family" of words with "core" members, and "peripheral" members where meaning-relations produce conflicting tendencies'.

In other cases, however, differences in the centrality of items and tests look quite arbitrary. The relative frozenness of
idioms is an accident of history, and it would be silly to search for deep explanations for their order in a sequence like (105). But there may well be many intermediate or mixed cases, motivated to some degree but not entirely. In this vein there is the proposal by Newmeyer 1972, 1974 to predict some aspects of the behavior of idioms from their semantic structure, to predict, for instance, that kick the bucket 'die' fails to undergo passivization on the ground that the idiom is semantically intransitive, with no object argument in its semantic structure to correspond to the bucket; hence

(111) The bucket has been kicked by many careless mountain climbers.

lacks the fatal sense of kick the bucket.

These matters are further complicated by the fact that the properties being tested for are themselves of different sorts—from semantic categorizations to superficial ones, with perhaps many possible intermediate types. On the one hand, there are tests for properties like semantic ambiguity; on the other, tests for surface constituency. But for many syntactic categories and properties, it is not entirely clear where they fall. Certainly, it must be possible for categories and features to be introduced by rule, as Schachter notes in his comments on syntactic categories and their relevance to the Universal Base Hypothesis--

(112) ...Chomsky attributes potential universality only to categories of the base; that is, only to categories which are found in pre-transformational structures. Thus the possibility is left open that
various syntactic categories needed for the characterization of the surface structures of particular languages need not be included in a universal vocabulary of category symbols at all. For example, the category relative pronoun must evidently be distinguished from other syntactic categories in an analysis of surface structures in English and certain other languages. But if, as is generally believed, relative pronouns have no place in base structures and arise only as the result of transformations, they need not concern those seeking to establish a universal set of syntactic categories.  

(Schachter 1973: 139)

Finally, recall that it is possible to come across a good test for some property without knowing why it works. I noticed some years ago, for example, that the idiom be to 'visit, see, go' is confined in standard English to occurrence in the perfect--

(113) I have never been to Australia.
(114) We had been to Chicago many times before we realized it was not the state capital.
(115) Having been to Santa Cruz only once, I naturally would like to return.
(116) *Right now I am being to Columbus.  
(117) *We were to Vienna last year.  
(118) *She will be to Essex soon. 

and therefore can serve as a test for whether a form of have is a perfect; the grammaticality of (115) shows that the there represents the perfect, although have in adverbial gerundive constructions does not always do so--

(119) Having seen Norman yesterday, I doubt that we can trust him for the job.
(120) *I have seen Norman yesterday. 

I see no deep reason why this be to should occur only in the perfect.  
The semantically very similar visit, see, and go have no such restrictions on them; compare (117) with
We {visited saw went to} Vienna last year.

So the constraint appears to be an arbitrary syntactic condition on this idiom, and therefore of a different character from other tests for the perfectivity of have, which are semantic in nature (for instance, the use of occurrence with adverbials like since Monday as a litmus for perfectivity). The be to test was discovered by chance and has no motivation I can see.

Central tests of unknown motivation illustrate how hard it can be to give an account of the intrinsic connection between tests and the properties they are criterial for, even though we can sometimes derive tests from theoretical principles or from generalizations about the grammatical structure of a language, so that the motivation of the tests is entirely clear.

Notes

*This is a revision of a reading version presented in May 1977 at the Michigan State University Linguistic Metatheory Conference. Earlier versions of this paper were given at the University of Pittsburgh in December 1973; at the annual meeting of the Linguistic Society of America in San Diego in December 1973; at the Ohio State University in January 1974; and at the University of Sussex in July 1976. Members of all five audiences have given me sharp criticisms and useful suggestions; I am especially indebted to Lloyd Anderson, David Dowty, Anil Gupta, George Lakoff, John Lyons, James McCawley, Jerrold Sadock, Aaron Sloman, and Richmond Thomason. My work was supported in part by the John Simon Guggenheim Memorial Foundation and in part by the Royal Society.

1. I have heard this use of the term attributed to Zellig Harris in his classroom presentations, but I have not found it in his published writings, where the relatively neutral terms tests and criteria appear.
2. Though showing the direction of determination in such cases is not itself always an easy matter.

3. Significant amendments to this statement have been made by several writers, especially Bouton 1970 and S. Anderson 1971.

4. There is a similar list of tests in Ross 1973a: 96.

5. For lexical categories like N, V, and Prt, the properties in question have ordinarily been described by the use of features, or by decomposition into primes of some sort, whereas for construction types like NP, VP, and S, the properties have been described structurally, by some characteristic configuration of the constituents or by the presence of a special marker. Despite this difference in the mode of description, the two cases are similar (they are treated in exactly parallel fashion by Hudson 1976: ch. 2, building on the sort of systemic grammar outlined in Hudson 1971), and the same sorts of tests have been suggested for both.


7. This, too, has been disputed; see Yasui 1975.

8. In fact, I think several of their arguments are faulty. See Zwicky and Sadock 1975: 15 for a discussion of an invalid transformational potential argument in the Lakoff-Peters article. The Lakoff-Peters arguments that depend on disambiguation with quantifiers like each and both probably fail as well, in which case they would be cases of invalid special distribution arguments (Zwicky and Sadock 1975: 12-4).

9. And I share with J. McCawley a grave uneasiness about this use of the word 'revolution' as applied to the history of linguistics, at least insofar as revolutions in linguistics are supposed to be the sorts of things one is faithful to, abandons, leads counterrevolutionary movements against, and so on.

10. Bolinger 1968 lists a variety of examples in which context beyond the sentence is relevant to judgments of grammaticality.

11. As it happens, Ross arranges his hierarchies so that the northeast corner is the positive one, but this difference in mode of presentation affects no matter of substance.
References

Akmajian, Adrian and Adrienne Lehrer. 1976. NP-like quantifiers and the problem of determining the head of an NP. Linguistic Analysis 2.4.395-413.


Botha, Rudolph P. 1976a. On the Rosetta Stone strategy: or how to camouflage nonvalid arguments. Distributed by IULC.

Botha, Rudolph P. 1976b. 'The theory comparison method' vs. 'the theory exposition method' in linguistic theory. Distributed by IULC.


Channon, Robert. 1974. The NAC and how not to get it. NELS 5. 276-98.


Corum, Claudia, T. Cedric Stark-Smith, and Ann Weiser (eds.). 1973. You take the high node and I'll take the low node. Chicago Linguistic Society, Chicago.


Fraser, J. Bruce. 1965. An examination of the verb-particle construction in English. Ph.D. dissertation, MIT.


Johnson, David E. ms. 1976. On Keenan's definition of 'subject of'.

146


Newmeyer, Frederick J. 1974. The regularity of idiom behavior. Lingua 34.4.327-42.


Sommerstein, Alan. 1972. On the so-called definite article in English. LingI 3.2.197-209.

Stahlke, Herbert F. W. 1976. Which that. Lg. 52.3.584-610.


Zwicky, Arnold M. 1977. On clitics. Distributed by JULC.