ON CLITICS

Arnold Zwicky

OHIO STATE UNIVERSITY

Reproduced by the Indiana University Linguistics Club.
February, 1977

Write for our latest publications list
Indiana University Linguistics Club
310 Lindley Hall
Bloomington, In. 47401
1. Introduction. Most languages -- very possibly, all except those of the most rigidly isolating type -- have morphemes that present analytic difficulties because they are neither clearly independent words nor clearly affixes. The problem is recognized, at least as a difficulty in terminology, in traditional language descriptions, where certain elements are set apart from the ordinary words and affixes of the language by being labeled clitics (pro- or pre-, en- or post-), particles, or pre/post-positions (for some recent general discussion, with French, Latin, English, and Italian examples, see Matthews 1974:166-9). The problem was also recognized by structuralists, who saw such elements as challenging the principle that the levels of linguistic structure (in particular, the levels of morphology and syntax) are sharply separated; Pike 1945, for instance, discusses quite clearly the analytic difficulties presented by English fast speech forms end, especially, by a set of Mixtec morphemes.

For generative grammarians these difficulties were obscured for some years, since the traditional domain of morphology was assumed to be apportioned between syntax and phonology. With a return to the traditional position that morphological structure and syntactic structure obey different principles by and large, as do morphophonemics and phonology proper, has come the realization that there are borderline cases. So in Aronoff (1976:3-4), where it is proposed that inflection is a 'syntactic' matter while derivation is a matter of word formation rules belonging to the lexicon, we find a brief discussion of clitics (with examples from Classical Hebrew, English, Syriac, and Navaho), which are distinguished from both inflectional and derivational affixes but assumed to be closely related to the former. The generative grammarian's striving for both precision and generalization in linguistic descriptions has, in fact, led to the uncovering of a host of analytic problems, which I will try to survey in this paper.

I begin by looking at a particularly complex case from the Austronesian language Madurese, as treated by Stevens (1971). On the one hand, a Madurese reduplicative morpheme, symbolized by Stevens with an R, shows alternative orders with respect to certain other morphemes, in particular the causative prefix pari: from the base kumpul 'gather', both
I - pa - pul-kumpul  
passive causative R base

and i-pul-pa-kumpul

(for some speakers, also pul-i-pa-kumpul)

can be formed, and both mean 'kept on being gathered'. In this respect, R shows the syntax of an independent word, for, though alternative orderings without difference in cognitive meaning are very common in constructions formed from independent words (as in so-called 'free word order' languages), within words the ordering of elements is rigidly fixed (Perlmutter 1970:234, citing Postal);

Ordering: Alternative orders of morphemes within a word are associated with differences in cognitive meaning.7 This bit of evidence from word-internal syntax is supported by the phonology of R; R acts...very much like a separate word phonologically; various rules which are effective across morpheme boundaries within a word (for example, a type of vowel harmony and syllabification, are not effective across the boundaries separating R from adjacent morphemes) (Stevens 1971:421). Here we appeal to a distinction between rules with different domains:

Internal Sanshi: In many languages there is a set of phonological rules of internal sanshi: these apply only within words, never across boundaries between two words.8

On the other hand, the morpheme R shows at least two of the syntactic properties of an affix, as opposed to an independent word: it is a bound morpheme, in the sense that it cannot occur in isolation (even, presumably, as an answer to a question) and in the sense that it cannot be set off from the morphemes with which it is in construction by parenthetical material; and it occurs in words between two clearly affixed morphemes, the passive i and the causative pa. The first of these properties is not specifically stated by Stevens, though it can be inferred from his description of R as a prefix; here we appeal to the principle:

Binding: Bound morphemes are affixes.

The second property depends on an assumption that syntax and morphology are not intermingled:

Construction with Affixes: A morpheme in construction with an affix is either a base or an affix.
Finally, there are two matters that would bear on the affixes. Independent word decision for R, but for which Stevens¹ brief note does not contain the relevant information. The first of these is syntactic, and relies on the following general hypothesis:

Rule Immunity: Proper parts of words do not undergo rules of deletion under identity.²

In particular, affixes are immune to such rules, so that we do not find derivations like

yellowish or greysih → yellow or greyish

dancing and singing → dance and singing

Though I know nothing whatsoever about conjunction reduction in

Madrugas, I think it is safe to suppose that R behaves like an

affix in this respect, and does not delete under identity to an

R in a conjunct.

The second consideration is phonological. Though Stevens
does not mention accent in his squib, I will suppose that R does

not bear accent, in which case there is phonological evidence for

its affixal status, using the general principle:

Accent: Morphemes that do not bear an independent accent are

affixes.

2. Types of clitics. Now it is clear that the six prin-

ciples above, whatever their general utility, cannot all be held

simultaneously. Of them, Binding, Construction with Affixes, and

Accent are the most susceptible to attack, and indeed there are

many cases in which these three principles as a group have

traditionally been assumed to need refinement or amendment. In

the following subsections I describe three classes of exceptional

cases.

2.1. First class: cases where an unaccented bound form acts

as a variant of a stressed free form with the same cognitive

meaning and with similar phonological makeup. Traditionally the

unaccented bound unit is called a conjunct, conjoint, or conjunc-
tive form, the stressed free unit a disjunct, disjunctive, or dis-
junctive form or the unaccented bound unit is said to be

critic (literally, 'leaning on'), weak, or dependent, in context-
ual alternation with free, strong, or independent forms. The

weak or critic pronoun of many Romance and Slavic languages are

standard examples; thus French conjunct pronoun me [ma] 'me'

and je [le] 'him', versus the corresponding disjunct pronouns

moi [mwa] and lui [ly] and Serbo-Croatian enclitic pronouns

me
In 'to them' and ei 'to you (sg.)', versus the full pronouns
njima and tabi. Similarly, Egyptian colloquial Arabic enclitic
pronouns [n]a '2nd sg. masc.' and [na] '3rd sg. fem.', versus
the independent pronouns [e]mi and [hya] Mitchell 1962:
52-3 (the enclitic pronouns are objects or possessives, while
the independent pronouns are normally subjects, though they can
be used with possessive enclitics for emphasis); and Modern
Greek conjunct pronouns [me] 'me' and [se] 'you (acc. sg.)'
[se] 'you (gen. sg.).', versus disjunct pronouns [esóna] and
[eosóna], respectively.

The disjunct form in such languages is typically used
whenever an accented pronoun is called for on syntactic or seman-
tic grounds, in particular when a pronoun is to be emphasized or
when it must stand in isolation (as in answers to questions); so,
In Modern Greek:

[esóna] [esóna] 
'see you to you I give an apple'

[(sók] [sók] 
and, in answer to [plandó toló] 'To whom are you saying it?':

[esóna] 
'To you (sg.)'

[(sók] [(sók] 
(examples adapted from Householder at al. 1964:82).

Notice that the phonological relationship between the weak
and strong forms in these examples is not straightforward; it is
unlikely that the weak forms are related to the strong ones by
phonological rules of any generality. Notice also that emphatic
uses of strong forms may, in some languages and under certain
circumstances, cooccur with clitic forms, as in the Modern Greek
example above and in French je ve vois jus [l'ouve 1uff] 'I see
him'.

This French example also illustrates the fact that conjunct
citics often show special syntax: in French declarative sentences,
conjunct object pronouns are obligatorily placed before the verb
despite the fact that French declarative word order is SVO, objects
ordinarily coming after the verb:

Je vois Jean [l'ouve 35]
'I see John'

*Je Jean vois
Je je vois [ʒi wœ]  
'I see him'

*Je vois [ʒœ]  
'I see*

Notice, finally, that the French conjunct pronouns behave like affixes with respect to rule immunity:

Je connais Jean et je crains Jean  
'I know John and I fear John'

* Je connais et crains Jean  
'I know and fear John'

but

Je le connais et je le crains  
'I know him and I fear him'

* Je le connais et crains  
'I know him and fear him'

2.2 Second class: Cases where a free morpheme, when unaccented, may be phonologically reduced, the resultant form being phonologically subordinated to a neighboring word. Criticization of this sort is usually associated with stylistic conditions, as in the casual speech criticization of object pronouns in English; there are both formal full pronouns and casual reduced pronouns:

<table>
<thead>
<tr>
<th>Full</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>He sees her</td>
<td>[hɛ sɛz hɛ]</td>
</tr>
<tr>
<td>She met him</td>
<td>[ʃi met hɪm]</td>
</tr>
</tbody>
</table>

As in the case of the clitics discussed in the previous section, the full form is the one that appears under emphasis or in isolation:

She met him   | [ʃi met [hɪm]]

(Who is it?) Him | [hɪm]  

* [hɪm]  

[ʃi]  

[ɛ]  

[ɛ]  

[ɛ]  

However, those clitics showed special syntax and opaque phonology, while these have ordinary syntax (the reduced forms occur in the same positions as the full forms) and ordinary phonology (by and large, the rules relating the two forms are of general applicability in the language; but see section 7.1 below). Although the line between the two types of clitics is not always clear, it is useful to have separate terms for the two cases. I suggest the (nonstandard) terms special clitics for the first type and simple clitics for the second.

From the historical point of view, it should not be surprising that there is difficulty in drawing the line between simple and special clitics, since special clitics are often the remnants of an earlier system of simple clitics; this point is made by Givón (1971: 356-7) with respect to the clitic pronouns of modern French and Spanish, the ordering of which can be taken to reflect the object-before-verb order of earlier Romance, and by Hale (1978:339-44) with respect to the pronominal special clitics in the Australian language Valkiri, which he suggests have evolved from simple clitics, with the neighboring language Warramunga showing an intermediate stage (in Warramunga the clitic pronouns are merely unstressed variants of independent pronouns, but have moved into 'second position', after the first (nonpronominal) constituent of the sentence). After the development from independent word to clitic, the next step is, of course, the incorporation of clitics into morphology proper: what is a clitic at one stage is reinterpreted as a derivational or inflectional affix at the next.

2.3 Third Class: Cases where a morpheme that is always bound and always unaccented show considerable syntactic freedom, in the sense that they can be associated with words of a variety of morphosyntactic categories. Frequently, such a bound word is semantically associated with an entire constituent while being phonologically attached to one word of this constituent, and ordinarily the bound word is located at the very margins of the word, standing outside even inflectional affixes. Examples of bound words are the Latin particle -que 'and', the Tagalog particles, and the English possessive morpheme.

The Latin -que is a conjunction and is associated semantically with a word, phrase, or clause, though it is attached to the first word in such a unit:

duasque hui legionum donata
\[\text{two and there legions (he) enrolls}\]

'and (he) enrolls two legions there'

(De Bello Gallico, cited by Hale and Bick (1966:sec.307))
The Tagalog clitic particles all have adverbial meaning and are associated semantically with an entire sentence; the interrogative particle ba, which corresponds to no free or accented morpheme in the language, is typical. The particles, ba included, are attached after the first (nonclitic) word in a sentence, and since the word order in Tagalog is relatively free, this could be any of a variety of words:

- Nakita ba ni Juan si Maria ngayon?
  - Has seen agent topic today
  - Hawai ba nakita ni Juan si Maria?
  - Has Juan seen Maria today?
  - Hindi ba nakita ni Juan si Maria ngayon?
  - Hasn't Juan seen Maria today?

(examples adapted from Schachter (1974)).

The English possessive morpheme is associated semantically with a noun phrase, though it is attached phonologically to the last word of the noun phrase (which is not necessarily the head of the construction, or even a noun):

- Germany's defenses
- The Queen of England's hat
- The woman I talked to's arguments

The possessive morpheme can follow inflectional suffixes, as in

- the woman I interviewed's arguments
- which parallels in this respect the behavior of Latin -um, as in
  - arma viriique
  - 'arms and the man (acc.)'

3. Endoclitics. With this rough delineation of three types of clitics (special clitics, simple clitics, and bound words), all of them bound unaccented morphemes that sometimes are in construction with affixes, we return to the problem posed by the morpheme R in Madurese. Of the three types of traditional clitics, R is most like a bound word, since it seems to correspond to no free or accented morpheme in the language. However, R occurs not at the margins of words, where all the other
clitics illustrated so far have occurred, but rather within words, between affixal morphemes. We appear to have a case here of a clitic that is neither a proclitic -- preceding the word to which it is attached -- nor an enclitic -- following the word to which it is attached -- but rather is an endoclitic (the term is nonstandard), or 'infixed clitic', interrupting the morphemes within the word to which it is attached.

Numerous further cases of endoclitics have been brought to my attention, all of some interest from the historico-grammatical point of view. The Estonian emphatic morpheme -ki (—ki), for instance, has the syntactic freedom of the typical bound word, and in addition (like Malagasy a) fails to condition at least one rule of internal sandhi ([a] fails to assimilate to [a] before -ki, though [a] regularly assimilates to velar words internally; see Lehnste 1960:39). The morpheme is normally enclitic. However, when added to interrogative words (making them indefinite), -ki may either follow or precede a number of case suffixes: the allative of misä 'something' (misä 'what' plus -ki) is either millalip, with -ki following the allative suffix -ta, or millep, with this order reversed. According to Lehnste, to whom I owe all of these observations about Estonian, school grammars prescribe the order with the enclitic -ki, but many speakers, Lehnste included, prefer the other order in certain of the cases. And in some cases only the internal placement of -ki is possible, as in the allative kusälo (or kusägoogle) of kusat 'somewhere' (from kus 'where' plus -ki).

Here we have, transparently, a morphological change in progress, with -ki coming to be treated more and more as a suffix attached to the base.

Turkish presents us with another illustration. According to Lewis (1967:108), in the third person plural of verbs, 'alternative forms can occur, with the plural suffix preceding or following the personal suffix': thus, ağvoxar and ağvolzar 'if they are taking (plural -lar, third person conditional -sa). Here we have an apparent violation of the ordering principle. But, in fact, the personal suffixes are clitic forms of the verb 'to be'; under certain conditions the personal suffixes alternate with fuller forms (for -ta, -ta), and, in addition, these suffixes are exceptional in not bearing stress, which normally falls on the last syllable of the word (hence, ağvoxolar but ağvolzar).

Still another case involves the pronoun clitics of the New Guinea language Huo. These are normally prefixed to nouns and verbs;

There exists, however, a small number of extremely common nominal and verbal roots in which these pre-fixes, depending on their phonological shape, are either (a) obligatorily infixed (b) optionally infixed or (c) absorbed into the root. (Malan ms. 1976)
Thus, from ma "hand" with the clitic g 'your (sg.), g-ma', but from hamu "person who shares our name" with za' "we (du.), ra'-za", from za' and za' "we (du.), ra'-za", but from hamu' and za', either za'-hamu' or ha-ra'-amu' (examples from Haiman).

The alternative orders permitted to R in Madurese, to -ki in Estonian, to the third person suffixes in Turkish, and to some prefixes in Hua would be extraordinary indeed for ordinary affixes, but alternative orders are by no means uncommon for special clitics and bound words. Following Perlmuter's claim that the analogue of the ordering principle above holds for the combination of a word with all its clitics, a wide variety of counterexamples have been adduced; some of these will be surveyed below. They include cases in which clitics have alternative orders with respect to one another and also cases in which a clitic may occur in both orders with respect to its host (the word to which the clitic is attached). In any event, it is clear that the ordering of elements within a group (a host plus all of its clitics) is not so rigid as the ordering of elements within a word. And, apparently, this relative freedom of occurrence may carry over to clitics that move inside the word.

I have so far argued that Madurese R is a clitic, in that by the first two principles -- alternative ordering and lack of internal sandhi -- it behaves like an independent word, while by the remainder -- binding, construction with affixes, rule immunity, and lack of accent -- it behaves like an affix. With this discussion as background, I turn now to a survey of the empirical and analytic problems surrounding clitics, illustrated with examples from a variety of languages. These problems fall naturally into four large areas: the (synchronic) source of clitics, their external syntax (with respect to their hosts), their internal syntax (with respect to one another), and their phonology. I take up these areas in turn.

4. The Source of Clitics. Here there are at least two questions to be asked about the clitics in any particular language, or about the source of clitics universally:

- Which elements in a sentence are realized as clitics?
- When is this cliticization blocked, when is it optional, and when is it obligatory?

4.1 Simple Clitics. The answer to the first of these questions is particularly straightforward for simple clitics: apparently, any word that can appear unaccented has the potential to cliticize to a neighboring word. As a result, the list of syntactic categories that can appear as clitics is the familiar
list of categories that may, or must, appear without accent in various languages:

(a) auxiliaries, in particular modals and the verbs of periphrastic constructions (verbs of being, becoming, possessing, doing, wanting, going and coming, causing, etc.);

(b) personal pronouns, or redundant expressions of these bearing marks of person, number, and other agreement categories;

(c) determiners;

(d) 'dummy' nouns, like the English one in this one;

(e) prepositions and postpositions;

(f) conjunctions and complementizers;

(g) adverbial words, among them negatives, place and time adverbs, adverbs marking sentence type (interrogative, quotative, imperative, etc.), emphatic adverbs (including items meaning 'even' and 'only'), epistemic adverbs (indicating degree of speaker's belief in a proposition), and narrative adverbs (indicating temporal or logical sequences).

Examples of simple clitics of all seven types can be found in English: (a) would in 2nd object; (b) he in What is he [(i)] to do? (c) a(n) in an answer; (d) one in dialectal big'un [big'] and this'un [dis']n; (e) for in for hours [(f)warz]; (f) and in and you [(yoo)] and to in I wanna go; (g) not in haven't. Sometimes the clitic words already form units with their hosts (are already 'dependents of a head', in the sense of Selkirk (1972: sec. 2.1)); this is true of determiners (a(n) above), prepositions (for above), and conjunctions (and above) in English. But in the other cases the clitic words are not in close syntactic relationship with their hosts, and there is a disparity between syntactic constituency and the accentual units I have termed 'groups'. I do not know of any interesting hypotheses about universal conditions on the direction of cliticization in the remaining cases or on the contexts which would permit or require cliticization in these cases -- though these are certainly important areas for research -- and I conclude, with other investigators of simple clitics, that language-particular rules of clitic attachment must be stated, even for simple clitics.

The theoretically most satisfying assumption about such rules would be that they apply, as a group, to the output of all syntactic transformations and yield the structures to which
phonological rules of reduction and deletion apply (in other words, that they are 'readjustment rules', in the scheme of Chomsky and Halle (1968-69-II et alii)). In at least one case, English co-
emcliticization, it seems clear that the clitic attachment is
globally constrained (Lakoff 1970); the ability of so to
criticize onto certain preceding verbs depends on whether or not
a constituent intervened between the verb and so at an earlier
point in the derivation:

\[
\text{He \text{is the man} \text{I want to waste}}
\]

In addition to whatever global constraints there are on
rules attaching simple clitics, a variety of syntactic condi-
tions determine their applicability. English not, for instance,
attaches optionally to a preceding modal or auxiliary verb and
is blocked from attaching to any other word:

\[
\begin{aligned}
\text{You left but} & \text{ought not to laugh. They tried not to laugh.} \\
\text{Re has not} & \text{is not clever.} \\
\text{Those not} & \text{*Thosen't served should complain.}
\end{aligned}
\]

\[
\begin{aligned}
\text{If not satisfied, you should return} \\
\text{the book.}
\end{aligned}
\]

Beyond the syntactic conditions on the attachment of simple
citics, there is one phonological condition I have been assuming,
namely, that to be eligible for attachment a word must lack stress;
this 'cliticize unstressed words' position is to be distinguished
from a 'destress cliticized words' analysis, in which clitic
attachment results in automatic destressing of the clitic. In
favor of the general 'cliticize unstressed words' approach is the
fact that several global constraints on cliticization have turned
out to be global constraints on destressing. Among these are
some constraints on the cliticization of auxiliary verbs in
English (discussed well in Baker (1971)), including the much-
discussed condition that the auxiliary not be followed by a
'deletion site':

\[
\begin{aligned}
\text{The \{party is\} in room 16.} \\
\text{I wonder when the \{party is\}}
\end{aligned}
\]

Also among these constraints is the condition that not fails to
cliticize to a preceding modal when the modal is understood with
wider scope than the negation (Horn (1972:ch. 4)).
We could not do it, you know.

[ambiguous: (we could (not x))]

or (not (we could x))]

We couldn't do it, you know.

[only: (not (we could x))]

The generalization that covers these cases is that whenever some condition prohibits a word from losing stress, then that word cannot cliticize to a neighboring word, consequently is not subject to phonological reduction.

There is, finally, some evidence from English negation that rules of clitic attachment should not be thought of merely as readjustment rules, but must be seen as belonging to a larger set of postcyclic rules, including as well free deletions and minor movements (Pullum 1976: ch. 4, in fact, suggests that all rules forming phonological words are postcyclic): this is the familiar fact that the inversion of subject and verb in questions carries along a not that is cliticized to the verb but not one that is independent:

*I have not helped. *

*I haven't helped.*

This fact is explained naturally if clitic attachment is permitted to apply before inversion.9

Turning away from English, we see that most of the sources of simple clitics can be illustrated from Old Irish:

Words which are not themselves fully stressed are attached either (a) to the following words as pro-

clitics, or (b) to the preceding word as enclitics.

(a) includes the article, possessive pronouns, and prepositions before words governed by them: pre-

positions...and infixed personal pronouns before verbs: the forms of the copula...; often also conjunc-

tions before verbs.

(b) includes certain demonstrative particles...and the emphasizing particles...Certain conjuncts used

in principal clauses...are not stressed either.

(Thurneysen 1968:sec. 41)

Not all of the Old Irish clitics are simple clitics, however:

the personal pronouns, classed by Thurneysen with proclitics,
are in fact 'infixed', that is, endoclitics, thus showing special cliticization. In addition, cliticization of the personal pronouns is virtually obligatory, so that the stressed forms of these pronouns are usable only as predicate nominals, in clauses without verbs, after the interrogative pronoun, and after an 'and' (Thurneysen 1976). With this apple we can move to the source of special clitics and bound words.

4.2. Other Clitics. As we should expect on historical grounds, the list of categories of words that can attach to neighboring words as simple clitics serves also as a list of possible sources for special clitics and bound words. Auxiliaries, negatives, and personal pronouns are the most common types of special clitics and bound words, but very rich systems occur in some languages -- for instance, in Tagalog (Schachter 1974), which has in addition to a full set of agent and topic clitic pronouns a group of eighteen clitic particles, including an interrogative marker, a reported speech marker, politeness particles, a conjunctive particle meaning 'because', and various adverbials (translated as 'for a while, yet', 'only, just', and 'too, either', for example). A new difficulty arises when we consider systems like this: the notion of syntactic source of clitics can no longer be taken for granted, as it was for simple clitics; rather, for each individual clitic or class of clitics some analysis is needed to determine the appropriate remote structure. Arguments that clitic pronouns originate in the same positions as full NPs can undoubtedly be constructed for any language that has clitic pronouns. The pronouns appear in their surface positions by means of a movement rule or rules (or if the clitics are in complementary distribution with full forms of pronouns, as in Tagalog and Old Irish) or by means of an agreement rule (if the clitics are redundant expressions of categories of NPs appearing elsewhere in sentences, as in Wolof (10)). For the remaining special clitics and bound words, there is frequently no class of nonclitic constituents to which a clitic can be assigned, so that its syntactic source is not at all clear. In most of these cases these problematic clitics are, semantically, sentence modifiers and are naturally assignable to a constituent (like COMP) in construction with an entire S (if not to an underlying higher S). This is a natural analysis for the Tagalog ka, for instance (which is an optional indication of interrogative sentence type, rising final intonation being the only obligatory indication), for the interrogative clitic ya of Russian, and for the three sentence clitics of the Australian language Djirrbal (Dixon 1972: 122-3): Interrogative -mo-, -yga (which indicates 'that the statement should be correct but that there does remain an element of doubt'), and -yia (which indicates 'the statement is definitely true').

Next, consider the conditions on placement rules for special clitics. We have already seen that languages differ with regard
to when items may or must appear as special clitics; personal pronouns filling the subject or direct object slot in a full Old Irish sentence must appear as clitics, while the corresponding personal pronouns in most of the modern Romance and Slavic languages may appear in the full form or may cliticize; the following may both occur in Serbo-Croatian (Brown 1974:19-40):

\[
\begin{align*}
\text{Da ti dam knjigu?} \\
\text{conj. to you I give book (clitic)} \\
\text{Should I give you the book?}
\end{align*}
\]

\[
\begin{align*}
\text{Da dam knjigu tebi?} \\
\text{to you (full)} \\
\text{Should I give you the book?}
\end{align*}
\]

There are even cases where elements appear in the same form they have as clitics, but do not move into clitic position; thus, Frelsh and Halvorsen (1972:37) say that in their dialect of Norwegian, "[S[bject] clitics must undergo the rule, but [O[bject] clitics may in certain cases remain in the position of non-clitic Object NPs even when the phrase marker satisfies the SO of Clitic Movement," and Hale (1973:312-3) says of Yalbiri that the auxiliary optionally remains in sentence-initial position, rather than moving into second position, if it is more than one syllable long, and obligatorily remains there in certain negative constructions.

Although lack of accent ought to be a precondition on the appearance of special clitics and bound words, as on simple clitics, anomalously accented clitics have been reported. In Modern Greek, an enclitic pronoun receives stress when it follows a verb with antepenultimate stress and precedes another clitic (hence [Oóse] 'give(me)', [Oósēnu] 'give me', but [Oósēmu] 'give it to me'). Examples adopted from Warburton (1970a:38), and in the Philippine language Bikol, according to Stevens (1971:421), the disyllabic clitics do not lose their stress to a preceding nonenclitic element.

The Modern Greek case properly belongs to a later section of this paper, since it involves the assignment of stress within the group, rather than within the word; we may assume that syntactically unstressed [mu] is cliticized, and that it later receives stress by rules of stress placement applying within the group. The Bikol facts, on the other hand, look genuinely problematical. Apparently the disyllabic clitics have not been fully absorbed into the clitic system, meanwhile exhibiting many of the properties of independent words (they bear stress, are freely ordered with respect to one another, may move from the usual position of clitics to attach to a verb elsewhere in the sentence, and do not condition a rule lengthening certain preceding vowels, although monosyllabic clitics do). The Bikol case is
reminiscent of the placement of some accented particles in
certain of the older Indo-European languages: the Latin post-
positive conjunctions autem 'however', vērō 'in fact', and
digitus 'therefore', for instance, which bear accent but are
placed like clitics, after the first word of a clause (male
and
Duck (1966:sec. 310-11); and similar accented particles in
Sanskrit, among them dāna 'just', ādetā 'indeed', nā 'now', and
vādi 'indeed' (Nadkarni (1961:sec. 1911).

Finally, as with simple clitics, I turn to the issue of
where rules of clitic movement and clitic agreement are located
in a grammar. Pullum's (1976: ch 4) position, cited earlier, is
that all rules of cliticization are postcyclical. The arguments
he cites for postcyclicity (Kayne (1975) for French, Quirk
(1972) for Portuguese, Perlmuter (ms. 1973) for Spanish, Poyne
and Carden (1974) for Persian) all, in fact, involve special
clitics. An interesting question then arises as to what sorts
of rules can apply after rules of clitic movement and agree-
ment. At least two large classes of such rules would be of
interest: rules affecting (that is, moving, deleting, or
attaching elements to) whole groups, and rules affecting indi-
vidual clitics or sequences of them.

Though I do not have enough examples to make generalizations
here, a few observations might be of value. I know of no rules
affecting whole groups, although there are some affecting entire
strings of clitics (without the host): when a language has a
primary location for its clitics, and another location under
certain conditions, then analyses usually assume that the clitic
string is assembled in its primary location and moved elsewhere
under those conditions. For instance, in Modern Greek clitic
pronouns precede all verb forms except the imperative, which they follow:

[αὑτὸς] 'I(11) bring it to you'
[ἀποκριτο] 'Bring it to me!

It is natural to assume that the clitics are assembled praverbally
and then moved in the imperative. Similar analyses are standard
for Spanish and Italian, where (with various exceptions and
complications) the clitics are praverbal with finite verb forms,
postverbal with nonfinite forms:

Spanish: Mi vecina me las da

'Her neighbor gives them to me'

Le gusta darlas

'She likes to give them to me'
Italian: *To porto a Maria*

'I'm taking it to Mary'

*Porto a Maria*

'Take it to Mary!'

See the next section for further discussion.

Rules affecting individual clitics are, for the most part, rules that would traditionally be described as 'morphological' rather than 'syntactic'. Typically, they mention the phonological makeup of specific clitics. Of such a morphological character are the haplography rules in many languages, which alter sequences of phonologically identical or near-identical clitics:

**Spanish:** 1a(′a) 1a(′a) 1a(′a)

1 2 3 4 2

**Italian:** si si → es 2

1 2

**Ewe:** m[ ] m[ ]

1 2 1 2 (Kubo 1973: 118)

**Japanese:** to to → to (Kuno 1973: 118)

and rules creating portmanteau versions of clitic sequences:

**Tagalog:** [Au / Au] [Au / Au] [Au / Au] [Au / Au]

[Schachter 1974: 100]

**Pashto:** Ce ye → Ce

Tephey (1975: 57)

(e.g., to ye → te)

**Albanian:** ṡe e = ta (Meyer 1966)

and rules metathesizing sequences of clitics, like the Ewe rule discussed by Clements (1973), which moves certain clitic pronouns (obligatorily in most cases, optionally for the third plural mō) over a following negative mē or imperative mō. I do not know of any 'syntactic' rule affecting an individual clitic: the Malhari imperative formation rule, which deletes a second person clitic, leaving a full NP elsewhere in the sentence (Malari 1973: 326-7). See the next section for another possible class of such rules.
Certain syntactic and morphological rules can thus be seen to apply after clitic placement. Another logical possibility for rule interaction — though not a possibility consistent with the separation of syntactic rules from phonological rules, with all of the former applying before all of the latter — would be for some indubitably phonological rules to apply before clitic placement. The two putative cases of this type that have been brought to my attention both involve endocllitics, and in each case it is not clitic placement per se that is at issue, but rather the infixing of clitics within their hosts. That is, we may maintain that the order of application is

clitic placement (a syntactic rule)
phonological rules
Inflection (a morphological rule)

(thus keeping syntax and phonology separate, but permitting some morphological rules to apply after some phonological rules). The cases are the Hua endocllitics, which Haiman (es, 1975) argues are not moved into roots until after reduplication (reduplication itself not applying until after some fairly low-level phonological rule); and the Pashko clitics, which Teggy (1975: sec. 2) maintains do not move within verb forms until after the application of a contraction rule.

5. External Syntax of Clitics. The previous section was concerned with the conditions under which cliticization takes place. Now I consider the relationship between clitics and their hosts, about which there are at least three significant questions:

a. To which elements in the sentence are clitics attached?

b. Are they attached as proclitics, enclitics, or endoclltics?

c. Are there alternative possibilities for the location of clitics?

I have already touched on these questions as they concern simple clitics. Perhaps all that need be added here is that some simple clitics may appear as enclitics in one context and proclitics in another. This is the way the English complementizer to behaves:

I've got to [gärə] run.

For us to run [torɪn] now would be foolish.
Special clitics and bound words are another matter. To begin with, they accumulate at certain points within the sentence. The best general hypothesis I can suggest about where these clitics appear is that clitics whose source is within a particular constituent (an NP or an S) move either to one of the margins of that constituent or to the head of that constituent (the N or the V). Thus we find NP clitics associated on the surface with the entire NP, as in the English possessive construction in the daughter of the regiment's happiness, where the clitic comes at the end of the NP, or the Modern Greek possessive construction in

'τα πατριάρχια
the father my

'καλή φίλη
the good your sister

'your (sg.) good sister'

where the clitic is attached after the first word of the NP (examples adapted from Thumb (1964: sec. 142); or associated with the head N of the NP, as in the earlier English construction in my lord of Norfolk's counsel) (Paston Letters, cited by Breeze (1975: 15)).

And we find S clitics, both pronominal and adverbial, located either at one of the ends of the S or attached to a V. Of the options, probably the least common is sentence-final position, though this is where the adverbial and sentence-type particles of the Non-Rhmer language Cherau are located. These particles are generally monosyllabic and their form is also restricted, in that there is never more than one consonant in [the onset], the vowel quality tends toward a central position, and the final consonant tends to be weakly articulated" (Thomas (1971: 33); they indicate yes/no questions and imperatives, as well as adverbials translated as 'so, thus', 'already', 'only', 'else' and 'so on.'

The most common location for sentential clitics is probably at the beginning of the sentence, although in this case the clitics seem rarely to appear literally in sentence-initial position but instead are moved into 'second position', following some accented unit. Halo (1973: 340), having noted that Wabiri clitics appear in second position, adds in a footnote that this ordering principle is
operative in many of the Uto-Aztecan languages of North America, in at least one of the Algonquian languages (namely Abnakį...), and to a limited extent in the Athabaskan language Halkoñą; its operation in Serbo-Croatian is described by Browne (1974). In Papago, a Uto-Aztecan language of the American Southwest, the surface positioning of the auxiliary appears to be identical to that of Walbiri... The principle is, in fact, known as Wackernagel’s law and its operation in Indo-European is the subject of a long paper by Wackernagel (1992).

'Second position', however, means different things in different languages. In Walbiri, it means following the first (nonclitic) constituent, so long as that constituent is immediately dominated by $\$.$

\[
\text{wawiri kapi-na pura-mi kangaroo future I cook nonpast}
\]

'I will cook the kangaroo'

\[
\text{wawiri njampu kapi-na pura-mi this}
\]

'I will cook this kangaroo'

(dute from Hale (1973:312.4); clitics are underlined.)

In Serbo-Croatian, 'second position' means, in sentences beginning with a predicate, after the first accented word of the predicate --

\[
\text{Doputojao je bio ranije arrived was.had earlier}
\]

\[
\text{Bio je doputojao ranije}
\]

'He had arrived earlier'

but, in other sentences, after the first word or the first constituent, where the constituent may be of any sort:

\[
\text{Predsjednik ja Tainu danas doputojao president aux. Tainu today arrived}
\]

\[
\text{Predsjednik Tainu ja danas doputojao}
\]

'President Tainu arrived today'

(Serbo-Croatian examples supplied by Wayles Browne.)
In Tagalog, 'second position' means strictly after the first (nonclitic) word of the sentence:

\[ \text{hindi ko aina nakita ngayon} \]
\[ \text{not i him/her has seen today} \]
\[ \text{agent topic} \]

'i haven't seen her today'

Finally, in Pashto 'second position' means, in sentences beginning with a verb, after the stressed vowel of that verb, so that the sequence of clitics will all be treated as endoclitics:

\[ \text{axistô ba va} \]
\[ \text{buy it will he} \]

\[ \text{a-ba-ye-xistô} \]

'He would be buying it'

(Tegey 1975:578)

but, in other sentences, after the first stressed constituent of the sentence. Doubtless there are further interpretations of 'second position' in other languages. What is important here is that it is some accented unit (a constituent of S, any constituent, a word, a syllable) to which the clitics are attached; a language may permit some freedom as to the point of attachment; and verb-initial sentences may have a special treatment, as in Serbo-Croatian and Pashto.

Cliticization to V rather than S may involve precisis, enchisis, or endoclitis. Very commonly, clitics attach in front of the verb under one set of conditions, and after it under another, as we have already seen. Free alternative orders of host and clitic (for at least certain combinations of clitics) are also possible, as in Romanian

\[ \text{bîlatul 1-oa da-o} \]
\[ \text{boy 10 will give to him} \]

\[ \text{bîlatul 1-o-oa da} \]

'The boy will give it to him'

(Netzron 1976 ms., citing Olsen 1928:60). I have offered similar examples in Old Provençal (by Distor Wanner) and in Cretan dialects of Greek (by Angeliki Melkouti-Dracman).
Endoclsis usually occurs in alternation with proclisis or enclisis, as in the cases of Estonian and Turkish mentioned above (where an enclitic moves into the host noun), or in the following example from Albanian (where normally proclitic pronouns move into the host, as in Huc): Albanian clitic pronouns are proclitic to verbs, except in the affirmative imperative, where they appear after the verb stem (and before the second person plural suffix -ni):

\[
\begin{align*}
\text{aa dërgon} & = u + e\ dërgo-nu \\
& \text{10 DO send you (sg.)} \\
& \text{1.pl. 3.sg.}
\end{align*}
\]

'You send it to them'

\[
\begin{align*}
\text{mos ju dërgoni} & = mos i + e\ dërgo-ni \\
& \text{neg 10 DO send you (pl.)} \\
& \text{3.sg.3.sg.}
\end{align*}
\]

'Don't you (pl.) send it to him!'

dërgojani = dërgo + i + e + ni

'Send it to him!'

(Data supplied to me by Jerry Morgan).

Even given that a clitic must attach to a verb, there may be more than one candidate verb, and in such cases a language may choose one or another of the alternatives, or it may permit variant forms. The issue arises in combinations of modal or auxiliary plus main verb, and in verb-plus-complement-verb constructions. A few illustrations: the Albanian pronominal clitics are placed between most modes and the verb, but in the perfect (formed with the auxiliary 'to have') the clitics precede the auxiliary:

\[
\begin{align*}
do\text{ ta dërgoj} & = do\text{ tō + e\ dërgo}\text{j} \\
& \text{future subjunctive 10 I send} \\
& \text{3.sg.}
\end{align*}
\]

*do\text{ ta dërgoj}

'I will send it'

e\text{ kam dërguar} = e\text{ kām dërguar} \\
10 \text{ I have sent} \\
3.sg.

*\text{kām e dërguar}

'I have sent it'

(Data again supplied by Jerry Morgan)
in Norwegian, 'in sentences where there is both a special finite auxiliary verb and an infinitive lexical verbal form in a single S of derived structure, Su[bject] clitics...move to the finite form, while O[bject] clitics move to the infinite form' (Fretheim and Halvorsen (1972:23)):

Här har han ikke gitt henne dat?

Has he not given \textit{to} \textit{her} it

'Hasn't he given it to her?'

In Old French, object clitics attached to a modal verb, while in modern French, they attach to the main verb:

OFR: Elle a \textit{voulu} aller

HFr: Elle \textit{voulait} \textit{aller} aller

'She wanted to leave'

(from Morin (1975:392)).

And in Spanish, clitics can move up or 'climb' certain sequences of complement verbs:

Quisiera poder estar haciéndomelo ahora

I wish to be able to be making me it now

'1 wish I could be making it for myself right now'

Quisiera poder estándomelo haciendo ahora

Quisiera poder haciéndome hacerlo ahora

Me lo quisiera poder hacer haciéndolo ahora.

(examples from Soldán (1974:13)).

Standard assumptions about these alternatives and variants for locating clitics are that they involve either distinct rules of clitic movement (move clitics to such-and-such a position under certain conditions, otherwise to another position) or else movement rules applying in sequence (first move clitics to such-and-such a position, then move some of them to another position). I have little to add to this discussion, except to point out that if the second type of analysis is opted for -- clitic attachment, followed by clitic displacement -- then it must be possible for syntactic rules to pick out specific clitics or classes of them,
which is to say that clitics must be distinguished in some way from genuine suffixes and independent words at the point in derivations at which clitic movement transformations apply (and the rules moving these clitics would be 'syntactic', in the sense of the previous section).

6. Internal Syntax of Clitics. In this section I examine relationships among clitics. Two issues concern us here:

What conditions are there on combinations of clitics -- on how many may occur in a group or on which particular ones may co-occur?

What conditions are there on the sequencing of clitics with respect to one another?

These are questions on which an enormous amount of scholarly effort has been expended (the locus classicus is Perlmuter (1970)), and I have little of substance to add to this discussion, but merely survey the main points in the literature.

On the one hand, it seems that no language permits the clitics in a group to occur in all possible orders. On the other, some languages have rigid requirements on the ordering of clitics. The analytic question is whether the observed orders are to be described by clitic placement rules (which successively move clitics into their surface location, as in Fretheim and Hatvani's analysis of Norwegian, or Emonds' of French) or by syntactic rules which permit any order whatsoever, plus a positive surface structure constraint (which filters out all orders not satisfying its conditions, as in Perlmuter's treatment of Spanish and French, and Tegey's of Pashto).

The major motivation for surface structure constraints comes from instances in which a morpheme with two or more distinct syntactic sources, or in which two or more homophonous morphemes, have identical conditions on occurrence -- in which case distinct clitic placement rules would have the same effect and obey the same constraints, and generalizations would be lost. This is the line of argument Perlmuter gives for an SSC in Spanish, with what he maintains are several sources for the morpheme no as the hinge of the argument. The Pashto case is particularly clear in this regard, since two morphemes which are obviously fortuitous homophones, the modal do 'should' and the second person singular personal pronoun do, obey exactly the same (rigid) conditions on occurrence with respect to other clitic morphemes of the language, including a constraint against sequences of the 'same' clitic (here, *do do*).
Whatever the appropriate descriptive device — and this might well be different for different languages — there is a rich collection of phenomena that deserve careful study and comparison, with an attempt to draw some generalizations from these data. A survey of only a few systems of clitics is enough to demonstrate that conditions on clitic sequencing involve both morphosyntactic categories and at least some aspects of the phonological makeup of individual clitics. These are frequently mixed in the same system; thus, the Spanish ordering formula

ao-second person-first person-third person

identifies one of the elements involved (ao) by reference to its phonological identity, the others by reference to their morphosyntactic categories (in this case, person). Similarly, the constraints on clitics in Tagalog are in part phonological, referring to the number of syllables in a clitic, and in part morphosyntactic, referring to the difference between pronouns and (essentially adverbal) particles; simplifying somewhat:

monosyllabic pronouns—particles—disyllabic pronouns

(thus, for the monosyllabic pronoun ko 'i (agent)', the particle na 'already' and the disyllabic pronoun aya 'he/she (topic)', only one order is possible: ko na aya; (Schachter (1974:98))). The conditions may, of course, be entirely morphosyntactic, as in Norwegian (Fretheim and Halvorsen (1972:24)):

subject-indirect-object-direct object—sentence adverb—verb particle

and in Czech (George and Tomon (1976:235)):

auxiliary-reflexive-from dative-indirect object—direct object

And they may refer solely to morphemes identified by their phonological form, as in one version of the Pashto constraint:

\[ \text{no hā \{-(a)m\} no de ye} \]

(adapted from Tegey (1975:572), where no is listed as 1 sg. and ye as 3).

The ordering of clitics within a group may involve reference to 'second position', as in Old Irish, which calls upon a special place-holding particle when no other suitable initial element is available:
Where the verb is preceded by conjunct particles..., the pronoun is attached to the last of these and the accented falls on the element immediately preceding...Where there is no conjunct particle the pronoun is attached to the first proposition or verbal particle of a compound verb...Where neither a conjunct particle nor a preverb...precedes the verb, the verbal particle no...is inserted before it for the purpose of infixing the pronoun.

(Thurneysen (1946:sec. 410))

What needs to be examined in these systems is which morphosyntactic categories figure in ordering constraints (person and case are prominent, as is the pronoun/particle distinction) and what ordering relations obtain amongst these categories (pronouns before particles, Su-10-DC, and 2nd-1st-3rd are all frequent orders). And, of course, which phonological properties of morphemes figure in ordering constraints (to my knowledge, only the phonological identity of morphemes and the number of syllables ever need mention in ordering constraints, though many other phonological properties -- vowel height, presence of a nasal, and the like -- are logically possible candidates for ordering principles) and what ordering relations obtain amongst morphemes with these properties ("shorter before longer", or "Pāñelī's principle" (Cooper and Ross (1975:78-9)), is almost surely a universal, in the sense that if a language constrains clitic order by reference to number of syllables, the requirement will be shorter-before-longer and not the reverse).

The conditions on clitic order in a language may be absolute, as in Spanish and Pashto, or they may permit alternative orders with no concomitant differences in meaning, as in Tagalog, where within certain subclasses of clitic particles there is free order:

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Clitic</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>lana</td>
<td>po</td>
<td>sala</td>
</tr>
<tr>
<td>three only</td>
<td>politeness</td>
<td>they</td>
</tr>
<tr>
<td>ranker (topic)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'There are only three of them, sir'

Ratio po lana sala

(Schachter (1974:110),

or in earlier Italian (and in some modern dialects), where various pairs of clitic pronouns showed free alternative orders face, Wanner (1974:162-5), and Hatzron (ms. 1976), which contains a variety of Romance examples, plus Czech and Somali as well).

Although it is common for the conditions on clitic order to be essentially the same whether the clitics occur before a verb or after it -- this is the case in Spanish, Italian and Modern Greek,
for instance -- there are languages in which proclitic order and
enclitic order are substantially different. In modern French,
in particular, the ordering for proclitics has third person pro-
nouns preceding first and second person pronouns, while the
ordering for enclitics (which appear in positive imperatives) has
first and second before third:

Tu me je donne
you 10 DO give
me it

‘You give it to me’

Donne-le-moi
give DO 10
me it

‘Give it to me!’

In such cases, it is assumed (as in Snoedsen (1975)) that a single
order underlies both of the observed orders and that one or more
movement rules yield the different surface orders. These rules
are then ‘syntactic’ rules applying to certain individual clitics.

7. Phonology. This extremely complex topic revolves around
essentially two questions --

(a) How are the phonological forms of clitics related
to the forms of nonclitic elements?

(b) How are clitics phonologically integrated with their
hosts?

7.1 Simple Clitics. Since clitics are, among other things, mor-
phemes with no independent accent, they should show the regular
phonological concomitants of a lack of accent. We therefore
expect the clitics in a language to undergo the same reductions,
deletions, and assimilations, and under the same conditions, as
other unaccented syllables in that language. And so it is with
the simplest of simple clitics [kən], [h], [h], and [ʔə] as
successive clitic versions of the English modal can, [km] in
its full form; [tə] or [tə] for the complementizer or preposi-
tion to, [t] in its full form; or [t] for the pronoun has
[h] in its full form. In each example the clitic forms are
derived from the full forms by processes (in the sense of Stampe
(1973)) of considerable generality in English: the sequence for
can, for instance, shows the ordinary reduction of unaccented
vowels ([a] to [a], as in mechanical—mechanistic) and then the
same development seen in has/on [(bəkən), [(bəkən), [(bəkən), [(bəkən)].
Other simple clitics show special phonology, as in the reduction of German preposition-plus-article combinations (zu dem zu am, zu der zu der, etc.). The English clitic not, for instance, loses its vowel in can't, what's, won't, and don't, though genuine suffices that are phonologically similar to not, like nees, do not lose their vowels. Moreover, not lacks an ordinary reduced form [nɔt], so that this elision is not only special but also obligatory when not is enclitic. The indefinite article a(n) also shows special phonology, whichever of its two unstressed forms, [a] or [ən] is taken to be mere bias: an insertion of [n] between vowels is not paralleled elsewhere in English phonology, nor is a deletion of [n] before consonants (without a transfer of nasalization to a preceding vowel). In such cases we may say either that there is a minor phonological rule, one that applies only to certain specific morphemes, or that the different forms are in virtual suppletion, with both of them (like the two forms [wəm] and [wem] for the verb come) listed in the lexicon for English, each with its own context of occurrence. When the application of a rule becomes obligatory in certain clitics (as has the deletion of [a] in clitic not, or, for many speakers, the deletion of [ə] in hortatory let us), a third course may be open: the combination of clitic and host may be treated as a lexical unit, one that realizes certain combinations of abstract elements.

Which of these treatments is the correct one in a particular case is not always obvious. For let's (as well as the Welsh emphatic negative mere) I have maintained (Zwicky 1972) that the third analysis, as a frozen lexical unit, is the correct one, on the grounds that many speakers now permit expression of the pronoun that is (historically) embedded in this hortatory formula:

Let's (us (you and me)) yo and get some pizza.

The historical development of simple clitics in these cases is fairly clear: what is at one stage of the language a regular casual speech form is perceived at the next as forming a lexical unit. This reinterpretation may then preserve remnants of earlier casual speech phonology -- the reduction of us in let's, for example, though clitic us (as in they saw us) no longer reduces. Reinterpretation is facilitated when the use of casual speech forms extends into other styles (so that the failure to use the forms marks speech as formal). The reduction of not in English has extended in this way, so that it is not surprising to find peculiarities in its phonology. These peculiarities, in fact, extend to the forms taken by modals serving as hosts to clitic not: shan't and won't lacking the [i] of shall and will.
and want and don't showing vowels unpredictably related to the vowels of will and do.

So we see that hosts as well as clitics may show exceptional phonology. The fact that in the particular examples we have been looking at both host and clitic are phonologically exceptional supports an analysis of English cliticization in which can’t, shan’t, won’t, and don’t are treated as lexical units. Special phonology for hosts is not invariably associated with special phonology for clitics, however; modal going, as in

We’re going to

goose on the goose house.

shows an unparalleled shift from [ə] to [a] to [o] in the form with enclitic to, though the reduction of to is perfectly normal for English.

When we turn from the phonology of clitics and hosts individually to the phonological properties of entire groups, matters become still more complex, even for simple clitics. The most straightforward situation would be if clitic attachment always created genuine ‘phonological words’, units to which all the relevant (segmental and prosodic) word-internal vowels of the language applied, and to which no other rules applied. Indeed, very often we see word-internal rules applying within groups: the reduction of intervocalic [n] to [n] or [n] in wanna from wanna is the same as that in parties, for example. We have already seen, however, that in some groups special rules apply: further examples show special interactions between host and clitic, as when we and give assimilate to a following clitic me ([bɛmœ] and [gɛmɪ]) while phonologically similar verbs do not (bec me is not pronounceable as [bɛmɪ], or have me as [hɛmɪ]).

The simple clitics in English seem, on the other hand, to undergo all applicable word-internal rules. To see this, we must clarify what is to count as a relevant rule of word phonology. The matter is made complex by the fact that many languages, English among them, have affixes falling into two classes (sometimes more than two classes have been posited, as in Stanley (1973)) with respect to their phonological behavior. In English, all the regular inflectional suffixes and some derivational suffixes, for instance -ness, fall in one class of affixes (which Chomsky and Halle (1968) assume are separated from their stems by a single word boundary, #), the remainder into the other class of affixes (which Chomsky and Halle assume are separated from their stems by the ordinary morpheme boundary, *); similarly for Dutch derivational affixes discussed by Booij (1976). In English the first suffixes have no effect on stress, while some of the second
(like -le and -lety) do; /reg/ is realized as [ŋ] before suffixes of the first sort, as it is in word-final position, but as [œŋ] before the other suffixes; and so on. What is important for us here is that simple clitics show the phonological behavior of affixes of the first, or 1, type. One indication of this fact in English is that enclitics never condition a stress shift; embarrassment preserves antepenult stress even with clitic objects, as in

It never embarrasses

[than [m]]
[her [f]]

7.2 Other Clitics. Special clitics and bound words are much more peculiar phonologically than simple clitics. To begin with, as I pointed out in section 2.1 above, the phonological relationship between a disjunct form and a conjunct form can be very remote; consider two of the Norwegian oblique pronouns (Froholm and Halvorsen (1972:3)):

<table>
<thead>
<tr>
<th>written</th>
<th>disjunct</th>
<th>conjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td>'him'</td>
<td>ham</td>
<td>[n]</td>
</tr>
<tr>
<td>'her'</td>
<td>hene</td>
<td>[nønø]</td>
</tr>
</tbody>
</table>

or the first person singular subject pronoun clitic in Walbiri (Hale (1973:315)) — disjunct notJo, conjunct -yga. Even where there is some visible connection between the forms, as with French moi/me and jaiz/le, the forms usually cannot be regularly related within the language. As a result, most investigators of special clitics seen to have assumed that there is no rule-governed phonological relationship and that the distribution of forms is essentially a lexical matter. In some cases, words that may occur both pronomically and enclitically have different forms in the two positions: thus Modern Greek third person plural feminine accusative [tis] proclitic, [tes] enclitic.

Hosts, too, may adopt special forms, as in Old Irish, where verbs take a special 'conjunct flexion' (Thurneysen (1946: sec. 38)) when they occur with the proclitics (negative and interrogative markers, certain conjunctions, and the combinations of prepositions with the relative particle) known as 'conjunct particles'.

The most striking, and puzzling, aspect of the phonology of special clitics and bound words, however, is the extent to which groups show word-internal phonology. The subject is very intricate: some special clitics and bound words act as if they were associated with #like simple clitics#, others as if they were associated with #: some groups show the application of phonological rules peculiar to sequences involving clitics.
We have already seen examples of special clitics and bound words evidencing what I shall call # behavior: Madurese -k and Estonian -ki both fail to condition rules of internal sandhi. Similarly, the Sanskrit enclitic on 'and' shows # behavior, since its hosts have the same forms before ca as they do before independent words beginning with c. The Spanish clitic pronouns show # behavior in that they have no effect whatsoever on the position of stress in their hosts: dándonos 'giving us then' mantains its initial stress even with two clitics (Harris 1959:115). And the Turkish clitic personal endings do not bear stress (aşkışınız), even though the regular word-internal stressing is on the final syllable.

Though most special clitics and bound words show # behavior, some are phonologically integrated to the extent that they show + behavior. Thus, in Latin the accent placement rules apply to the group rather than the word; the bound words -que 'and' and interrogative -ne, and the special clitic cum 'with', may therefore cause an accent shift:

virum 'the man (acc.)' virāque 'of the man'
vidēs 'you see' vidēsas 'do you see?'
cum vōrib 'with you (pl.)' vōribasum 'with you (pl.)'

In Modern Greek, as we have already seen, stress placement works (in a rather complex way) on groups as a whole; in addition to the stressing in [fēmēto] 'bring it to me!', there is a special stressing when an enclitic is attached to a word with antepenultimal stress: [tōmērōpas] 'the man, person', but [tōmērōpas] 'our man, person'. Note also the assimilation of [s] to a following voiced [n], which is typical of word-internal phonology in Modern Greek; Barbourton (1970:112) cites this assimilation in particular when he declares that 'when we say that proclitics and enclitics attach themselves phonetically to the following and preceding word respectively, we mean that the resulting clusters are subject to the phonological rules of morpheme medial clusters.' Wabiri enclitics also behave like ordinary suffixes with respect to the rule that 'a suffixal high vowel assimilates to a preceding final high vowel' (Hale 1973:313). And in Catalan (Johnson Uts. 1974), the [-] of the infinitive normally deletes in word-final position, but remains when a clitic follows:

[kontè b6] 'to sing well'
[kontè rə] 'to sing it (fem.)'

and [nt] normally simplifies to [n] word-finally or before a consonant, but (in some styles) remains before a clitic:
[fén oddf] 'doing this'
[partant u] 'bringing it'

But there are innumerable complications in this picture. Even in Modern Greek, special variants of enclitics, not with proclitics. And in Turkish, the enclitics in general show a behavior with respect to stress but *behavior with respect to vowel harmony (the interrogative particle mi, for instance, is stressless and takes its vowel harmony from the word to which it is attached: dogrú 'true', dogrú mi? 'true?'; bpyšün 'today', bpyšün mi? 'today?'; varšün 'tomorrow', varšün mi? 'tomorrow?' - examples adapted from Lewis (1967:105)).

In still other cases special rules (other than the 'morphological rules' described earlier) apply across a boundary associated with a clitic. Two examples of special facultative variants from Modern Greek should suffice: first, the second person singular genitive proclitic [su] often reduces to [a] in speech when it precedes a direct object pronoun beginning with [t] [zutafóra] 'I bring it to you' → [zutafóra]; second, the first consonant of a proclitic pronoun often voices after the clitic particles [ta] and [na] - [zututop] 'I'll tell it to him' → [zututop] (examples adapted from Householder et al. (1964:82-4)).

The complexity of the phonology of groups is on the same order as the complexity of the phonology of compounds. In both cases, some combinations show a behavior, some show a behavior, some show a behavior in one respect and a behavior in another, and some show the application of special rules. There are languages in which the compounds divide sharply into two classes with respect to their phonological properties, and in which these two classes can be associated with the boundaries # and +. Allen (1975) argues persuasively for such a division in Welsh, for example. But in Welsh the division into # compounds and + compounds, though motivated by several independent phonological and syntactic properties, has no syntactic basis: some compounds act one way, some the other, and there seems to be no general principle governing the assignment of forms to one category or the other (compare the # compounds cyn-a-faces 'ex-mayor', from cyn 'former', preceding and maxi 'mayor', with the + compound cyn-ideulu ('prototype, from cyn and ideulu 'image'), although Allen does point out that the + compounds are, on the whole, semantically less transparent than the # compounds. There are similar problems in predicting the stress patterns of English N-N combinations, where the differences between Forestress (Plitch Street) and afterstress (Plitch Avenue) has sometimes been taken to involve the distinction between genuine compounds (with internal #) and phrases (with internal #); see the discussion in Zwicky (1973). Apparently, in Welsh and English, some compounds have one boundary lexically, while others
have another. There is no question here of writing rules that
demote ¥ to + (or ¥ ¥ to ¥) under certain stable conditions
(and even if it were possible to devise such rules, they could
be objected to on general theoretical grounds, as in Pyle (1972)).

If clitic groups, which exhibit the same sort of phonolog-
ical peculiarities as compound words, are to be analyzed parallel
to compound words, then in at least some cases we would need to
list groups in the lexicon, complete with the appropriate bounda-
ries or equivalent marks. In any event, it cannot be assumed
that the phonology of clitics in a language follows in any simple
or direct way from syntactic rules of cliticization plus general
rules denoting ¥ ¥ to ¥, or ¥ to + (which is not to say that
the groups in some languages, or some subclasses of the groups
in a particular language, might not submit to such an analysis).
FOOTNOTES

"Paper read at the Third International Phonologie-Tagung at the University of Vienna, 2 September 1976; an abbreviated version is to appear in Phonologie 1978, the proceedings of that conference. A tiny precursor of this article was distributed in dittos in January 1975. I have benefited enormously from the criticisms, questions, and factual contributions of those who read this earlier version and of those who have discussed clitics with me since then; among these intellectual supporters are Arlene Berman, Wolfgang Dressler, Robert Hetzron, Robert Jeffers, Lise Lehtisa, Patrizia Donegan, Jerry Horns, Geoff Pullum, Jerrold Sadow, Jochem Schindler, David Stemple, Dieter Wanner, and Ann Zwicky. I am, of course, much indebted to David Perlmutter, whose work on clitics has largely defined the field and stated the central issues.

1 Various potential counterexamples to this principle have been drawn to my attention. Robert Hetzron has pointed out to me, for instance, that the alternative plurals pomodoro and pomodorò for Italian pomodoro 'tomato' have what might be taken as the same morphemes in different orders. Similarly, for some English speakers, brooders-in-law and brother-in-laws. Clearly, these examples are of compounds, with alternative principles for the placement of an inflectional morpheme belonging to the compound as a whole. As a class, compounds often show partly word-like, partly phrase-like, syntax and phonology (see footnotes 2 and 3 below, and sec. 7.2), and in this respect they resemble clitic constructions, though I will assume the two types of 'semi-words' can be distinguished.

2 Here, too, compounds -- and compound-like offshoot constructions (for instance, the English stressed prefixes mis-, dis-, re-, over-, and so on) -- may show external rather than internal behavior, as when a stop at the end of the first element of a Sanskrit compound assimilates in voicing to the segment beginning the second element ।। (period of six days), from ।। and ।। (six and aha 'day'), paralleling word-final (but not word-internal) stops, or when English has aspirated voiceless stops after the of stressed mis- and dis- (in mistranslate and disconnect), paralleling word-initial (but not word-internal) voiceless stops.

3 Again, compounds and certain affixed forms resembling compounds may show exceptional behavior, as in German Nütz- und Abendessen 'lunch and dinner' and Ein- und Ausgang 'entrance and exit', or English book and magazine 'deltas and pro- and anti- Marxists.'

4 The bulk of Sivón (1971), concerned with this and development, gives examples where morpheme ordering within words reflects an earlier syntactic order.
This term was suggested to me by Robert Hetzer as a substitute for my earlier support.

Another nonstandard term. My use of it here is not to be confused with other uses this word has had in linguistics. Nor is the group to be confused with the phonological word or phonological phrase, in Chomsky and Halle (1968) or other works.

An unsuccessful attempt at framing such a condition is Postal's hypothesis that cliticization is, in general, blocked across major constituent boundaries, particularly across clause boundaries. Processes like cliticization would be permitted only for those constituents which are, at late levels of derivation, sisters of certain limited kinds of constituents. (Postal (1974:106))

Postal is maintaining that there is a syntactic difference between I believe it is true and I believe it to be true (it is true is an S in the first example, while it is a surface direct object in the second), that the difference corresponds to a difference in clitic behavior (it is a (simple) clitic in the second example but not in the first), and that a universal constraint accounts for the difference. But the putative constraint is violated by other simple clitics in English (as in the one you absolutely probably going to win, where it cliticizes to the verb of an embedded clause), and there is not much reason to think it holds for special clitics or bound words either (Postal himself cites counterexamples with the English possessive morpheme).

Lakoff's treatment has been widely disputed (I will not review this literature here), but the alternatives that have been suggested seem to me to be notational variants, each coding globality in its own way.

If the attachment of not is optional, then the ordering attachment — inversion is paradoxical, violating as it does the principle of Obligatory-Optional Precedence (Ringen (1972)). However, if the attachment of not is obligatory (that is, it not must cliticize whenever it lacks stress), then attachment would be expected to apply before inversion, by Proper Inclusion Precedence (Sanders (1974)). Since I know of no compelling evidence against the assumption that not-attachment is obligatory, I will assume that it is; perhaps other attachments of simple clitics are as well.
In the latter case, further analysis is needed, for the agreement rules might be one that straightforwardly copies some morphemes into the clitic position, or it might be one calling for a clitic agreement in certain abstract features with the source NP. Halle (1973) makes this point very clearly for Welsh, which he argues has an abstract agreement rule.

12Cherna is an SVO language, so that there is no difficulty in seeing that these particles are genuinely sentence-final. In an SVO language like the Slouan language Hidatsa (Matthews 1965: ch. 3) it is hard to tell whether certain morphemes are S-final or V-final. The Hidatsa mood morphemes appear to be V-final, however, since in certain moods the subject may follow the particle (this is so for the 'period' particle c, the report particle nano, and the quotative particle wakoc) while in two other moods the subject must follow the particle (for the quotative (a)n and the imperative (a)ka).

13In YSO languages it may be hard to distinguish S-initial cliticization from V-initial cliticization. For example, in modern Welsh the affirmative particle y (e), the question particle a, and the negative particle ni (d) all come at the beginning of sentences, and hence also come before verb forms. There are a few cases in which other elements can precede one of these particles (the question words hir 'where', pa bryd 'when', and sut 'how' can precede y(e), for instance), but on the whole, when a constituent is moved in front of the verb the particle y(e) is not used at all and the question particle (in the form a) and the negative particle remain in sentence-initial position (Bowen and Rhys Jones 1960:ch. 23). I conclude that these three particles are S-initial clitics.

14Dinnsen's 1972 reanalysis of the Spanish case involves two SCC's, one referring to so and to the category of person, the other referring to the case categories: reflexive-benefactive-dative-accusative.
REFERENCES


Wackernagel, Jacob. 1892. *Über die Gesetze der indogermanischen Wortstellung.* Indogermanische Forschungen 1:335-436.


