LINGUISTICS OF THE TIBETO-BURMAN AREA
VOLUME 14:1 — SPRING 1991

On the word in Angami Naga*

P. P. Giridhar
Central Institute of Indian Languages
Mysore, India-570006

ABSTRACT

This paper seeks to expound and juxtapose two alternative approaches to linguistic wordhood as applied to Angami Naga, a Tibeto-Burman language of North East India. In the first of these approaches the criteria, which are laid out with ample exemplification, apply across the board, which results in the definition of only one kind of wordhood (viz. lexical wordhood). In the other approach different criteria apply in response to different levels (viz. the phonological, intraphrasal and interphrasal), resulting in a six-way typology of the Angami word. There is furthermore an attempt to define compound lexical units. Compounds, however, do not lend themselves to being analysed in terms of the second approach. Although this paper is basically empirically inclined, theoretical exegesis perfuses data-elucidation all along. We conclude that the disjunct approach explicates the phenomenon of the word in Angami in a more fruitful perspective than the conjunct approach.

****

---

* (a) Angami Naga, phonemically /âŋgāmī nâgā/ belongs to the Western subgroup of the Naga group of the Tibeto-Burman family of languages. It is spoken in the hilly state of Nagaland in North-East India. Angami is a tone language, but tone is not marked here because it is irrelevant to the central thrust of this article. Intonation does bear on the problem of the word, but we lack sufficient data to take it into consideration here.

(b) Drs. E. Annamalai, D. P. Pattanayak (my linguist-colleagues) and Mr. Vishwanath Mirle (my redoubtable graduate teacher) have given ungrudgingly of their time and acumen in commenting on a first version. My genial Angami friends Khose Sale, Atha Vizol, and Zakielie Iralu, the willing celerity of whose response has been a great goad in my attempt at plumbing the depths of the language, never hemmed or hawed when accosted with a request for data. My daughter Usha helped tie up some loose ends in presentation, and converted my forbiddingly squiggly calligraphy into an incredibly neat manuscript. My devout thanks to and ready absolutions on all of them.

(c) This is a recension of a paper which was presented to the Third International Conference on South Asian Languages and Linguistics, 13-16 January, 1982, Mysore, India. I'm grateful for some piquant, if skimpy, feedback I got there.

0.0 To sift lexical units (words, compounds included) from the more numerous syntactic units (phrases) on the one hand and the less numerous sublexical units (bound forms) on the other, has seldom been an easy task in any language.[1] As it happens, some of the criteria which make for the strainer are necessary without being sufficient, some sufficient without being necessary, none both necessary and sufficient, some neither. Two criteria may prove dissonant to each other in deciding the morphological, lexical or syntactic status of a linguistic structure, in which case one criterion may override the other, or the criterion that holds the stage in some specific cases may itself be overridden by others in certain others, or a single criterion may prove inconclusive in which case criteria apply in concert... In a word, there is about the defining criteria of the word a seemingly messy criss-crossing asymmetry.[2]

0.1 I shall, in this exposition, sketch out and juxtapose two alternative approaches to wordhood, neither of which is admittedly entirely novel. The comparative juxtaposition will be more implicit than explicit as the differences between the two approaches will be quite obvious. I will prepare the ground for such an exercise by first laying out some well-defined and well-known criteria. These will establish under what conditions (a) phonic material is to be considered sublexical or bound, and (b) adjacent

---

[1] In languages where the rung of the word can be motivated in the linguistic-structural ladder, word segmentation is not an easy matter (especially in nonagglutinative languages). Any definitions of the boundaries of the word which are restricted to a gnomic sentence or two rarely, if ever, turn out to be unimpugnable. There is undoubtedly more to the linguistic unit of word than can be encapsulated in a couple of laconic sentences. Further, it is difficult to square the notion of a language with no well-motivated level of the word with the formidable position that the word is a primordial category that "mediates between linguistic entities (phrase, sentence, text) and (nonlinguistic) ontologic entities," so that prototypically — ontogenetically and phylogenetically — the speaking animal (*homo loquens*) is a lexical rather than a syntactic animal. Such languages also go against the position that "all regular word-formation rules are word-based." (Aronoff 1976:21)

[2] Although most of the examples that are discussed in this article as compounds were listed under the rubric "word compounding" in my grammar of Angami, there were no heuristics advanced for doing so except the native speaker's intuition — which I dubbed "a rather dubious" criterion (Giridhar 1980:90). Although possibly a necessary criterion, it is never a sufficient one. Intuitional, introspective evidence is typically shored up by some kind of formal-structural evidence in language. Surely, there must be something in what the intuitions are about, which brings them into being in the first place. If the intuitions cannot be buttressed by any formal-structural evidence, one should doubt (the authenticity of) the intuitions rather than give them the benefit of the doubt. Which heuristics should override which, however, is determined both by metatheoretical and intuitional considerations. Even if the criteria behave erratically, there must presumably be some organizing principle which makes a (compound) lexical unit a (cultural) cognitive reality.

morphemes are to be construed conjunctively, as sub-constructions of a single distributional framework of morphemes (i.e. as a single word), or disjunctively, as constituents of different distributional frameworks (i.e. as different words). These conditions will set "autonomous" units apart from "syntagmatic" units, to use Trnka's terminology (Trnka 1960, quoted in Marchand 1969). In the first approach, these criteria apply indiscriminately across the board, without regard to levels (see Section 1). The underlying principle behind this way of looking at lexical distribution and categorisation is "*once a word always a word*," or "*once an affix, always an affix.*" After noting their interactional behaviour, I shall suggest (1.7) a hierarchisation of the heuristics to circumvent the conclusion of the squishiness of the Angami word, a conclusion one is willy-nilly led to by the dissonant character of the criteria broached in our opening paragraph. Angami compoundhood (section 2), however, turns out to be a discrete rather than a squishy notion, because of the mutually agreeable nature of the structural attributes. The further question of deriving the compounds thus arrived at in terms either of transformational or lexical (derivational) rules is outside the scope of this study, as are the questions of morphological word-formation processes and lexemes. Nor will the paper address the theoretically pregnant topics of "word as articulated thought" rather than as a formal unit, and of the relative psychological reality of the two approaches. Changing tack, the analyst can explore an alternative theoretical possibility, viz. that the word is a variable notion and that word divisions could vary in response to linguistic levels, with distinct criteria applying at different levels. In this approach, there is a *de novo* consideration and definition of the boundaries of the word at each linguistic level. Section 3 expounds this theoretical posture.

1.0 The following characteristics governing the delimitation of the concept of "word" that defines the "upper and lower limits of morphological structure" (Nida 1949:102) in Angami keep simplex words and compounds separate from affixes and syntagms. Criteria 2.1 through 2.4 concern only compounds, i.e. complex words which are derived by combining two or more roots.

1.1. Noninterruptibility

That a linguistic unit which cannot be interrupted is a lexical unit seems a reasonable assumption.

1. umhi - u
   eye indsg[3]
   "the eye"

2. tshũ - lie
   do imp
   "do"

Apparent counterexamples are 1a and 2a:

1a. umhi keza-u
   big
   "the big eye"

2a. tshũ pevi-lie
   well
   "do well"

It might be argued that *keza* "big" and *pevi* "well" interrupt the structures in 1 and 2 respectively, and therefore 1a and 2a are counterexamples to 1 and 2. However, the fact that in Angami inflectional affixes close the phrase — NP or VP (branching leftward or rightward) — buttresses a possible initial hypothesis that 1a and 2a are not the interrupted versions of 1 and 2, and hence are not counterexamples to our statement.[4] The question, however, still remains whether 1 and 2 are morphological structures and *-u* and *-lie* bound morphemes. Even the

---

[3] Abbreviations with their expansions and symbols are: ? = gloss uncertain; 3 = third person (coded in the verb or lexically marked in the subject); - = morpheme boundary (word boundary is represented by space); Δ = the preanalytic, presystematic indeterminacy about (juncture at) linguistic boundaries; ũ = lower-high retracted unrounded central vowel, the retraction giving it a cock-eyed quality (inadequately described as a central vowel in Giridhar (*op. cit.*:7)). This cock-eyed quality (for lay ears, that is) is even more conspicuous in Mao Naga, a sister language); ɲ = palatal nasal; ŋ= velar nasal; ==> = "can possibly become"; =/=> = "cannot possibly become"; Ø = zero; acc = accusative case marker; cmn = common gender; cndl = conditional mood marker; cnt = continuous participialiser; cntrsv = contrastive particle; cnj = conjunctive participialiser; dat = dative case marker; def = definite suffix; dimnv = dimunitive; emph = emphatic particle; exrt = exertive mood marker; fem = feminine gender signaller; fut = future tense marker; fng = future negative indicative; gnr = generic proform; hab = habitual; imp = imperative mood suffix; indsg = individuated singularity marker; loc = locative; masc = masculine gender indicator; nom = nominative case marker; pl = plural marker; prf = perfective aspect marker; prg = progressive aspect marker; prpsv = purposive marker; prs = present tense marker; prtpl = participialiser; pst = past tense marker; Q = interrogative marker; qut = quotative; sbj = subject; vr = valency role marker which indicates the number and nature of "players". See Giridhar (*op. cit.*:63-65) for a brief sketch.

[4] This of course could be true only of subordinate or attributive phrases, and not of co-ordinate phrases. In *lesũda mu likhuo-u* "book(s) and the bag", *-u* closes only the morphological structure that it forms along with *likhuo* "bag" and not the whole phrase. In fact, as the gloss shows, *lesũda* in the above phrase could mean "books".

knowledge that *keza* "big" in 1a is a word (*za* "be big") does not help us as to the wordhood of *u* "the" which may very well close a phrase but still be a word. In different languages, elements opening or closing a phrase need not be affixes but may be words, as indeed the English articles are. The article in English opens rather than closes the NP, English being premodifying rather than postmodifying (except for right-recursive relative clauses and highly marked postmodifications like *Alexander the Great*). This, indeed, turns out to be the case with some units in Angami.

3. umhi puo "one eye"

3a. umhi keza puo "one big eye"

4. a tshũ tyo
   I-nom do fut "I will do"
   (=Ø)
4a. a tshũ pevi tyo "I will do well"

Puo "one" in 3 and *tyo* "will" in 4, both of which close the phrase, are independent words by other criteria. It is the phonetic criterion and its syntactic correlates which decide the wordhood of 1 and 2. It seems necessary before accepting it as a valid heuristic principle to say what constitutes interruptibility, since the notion is not so self-evident as to be assumed without discussion. If **a** Δ **b** is a linguistic sequence, and if **x** which is a linguistic element cannot intervene to make it **a** Δ **x** Δ **b**, then **a** Δ **b** is a word (**a**-**b**), not a phrase (**a b**). *Vo-krũ* [pig-fem] "female pig", for instance, is a word because no linguistic material can come between *vo* and *krũ*. By this token, participial nouns (e.g. *ke-za-u* [participializer-be big-the one (masc)] "the big one", *phu-ketyo-pfũ* [marry (sbj:fem)-participializer-the one(fem)] "the one (human female) who is marrying"); the speaker-proximate deictic form (*ha-nunu* [here-from] "from here"), *uko* "we (incl pl)"; verbs coded for the person of the subject NP, (*puo-ni* [he-be] "happy (sbj: 3 sg)"); cardinals containing morpholexically obtained forms of decades (*hie-pengu* [ten-five] "fifty", *hie* being the suppletive allomorph of *kerũ* "ten"); ordinals (*kenie-u* [two-ordinal marker] "the second"); lexical morphological units (e.g. *thepu* "to lend", *lesũ* "paper", rũmho "be beige, grey", *puozerũ* "seam in *lohe*, the most common Angami shawl"); derived causatives (e.g. *pe-mhe* [causativiser-go out (as fire)] "extinguish"); derived adverbs (e.g. *pe-vi* [adverbializer-be good] "well"); derived verbal participles functioning as nominal postmodifiers (e.g. *ke-ti* [participializer-be black] "black"); verbs reduplicated to express the iterative aspect (*pu-pu* [say-say] "say again and again" *vo-vo* [go-go] "keep going", adjectives (partially) reduplicated for a

distributive function (*kekri-kri* [different-different] "different (spoken of multitudinous things)") are all words. Now consider

5. puo dukhri-lie
   he-acc kill-imp
   (=Ø)
   "kill him"

and the formally related, semantically identical

5a. puo dukhri-wa-lie
                vr
   "kill him"

or

6. sũu pro lho
   it-nom fly ftng
   (=Ø)
   "It will not fly"

and the formally related, semantically identical

6a. sũu pro-lie lho
            vr
   "it will not fly"

-wa and *-lie* are valency-role markers in the verb and are optional, as 5 and 6 bear out. Now are we to consider *wa* and *lie* as material which interrupts the verb complexes in 5 and 6? Clearly not. Paradoxical as it may seem, even though there is grammatically determined potential room for affixes, e.g. for *-wa-* in *dukhrilie* "kill", there is no potential pause between *dukhri* and *-lie*. Suffixes have immutably determined places in the linearized verb morphology. In 5a and 6a, *-wa* and *-lie* are only taking their appointed places in morphotactic space rather than interrupting such space. This is true also of infixes, which "interrupt" root morphemes. Further, these affixes are not PARENTHETICALLY added, and such "expansion" is not potentially unlimited as it is with phrases. In the case of "interruption" by compound elements, the "interruption" is uniquely limited, as in the "expansion" of *umhiu* "the eye" to *uphi-mhiu* "leg-eye (=the ankle)" where the "interrupting" compound element is uniquely *-phi* "leg". This unique limitedness is both paradigmatic and syntagmatic, i.e. no other member of the form class to which the "interrupting" material *-phi* belongs can replace it, and no other linguistic material can either precede or follow it. A parallel in English is the "expansion" of *blacks* to the compound *blackbirds*. In the case of "interrupting" phrasal elements on the other hand, the "expansion" is potentially unlimited. In the "expansion" of *umhiu* "the eye" to *umhi ketiu* "the black eye", for instance, not only does the phrasal element *keti* "black" belong to an open ended lexical class, but admits of linguistic

elements preceding or following it. An English parallel would be the "expansion" of blacks to the phrasal black birds. These then are cases which do not illustrate interruption, though they apparently seem to do so. We have now delimited the range of the word "interruption" in three ways by saying that

(1) affixes can close syntactic phrases,
(2) they can take their appointed places right in the middle of morphotactic expanses, and
(3) "interruption" by nonaffixes or free forms which are compound elements is uniquely limited both vertically and horizontally.

Incidentally, while both kinds of interruption can add new dimensions of meaning to the interrupted chunk (cf.8-8a for examples of interruption in the delimited sense and 7-7a for examples of interruption in the non-delimited sense), none can change the semantic head of the construction. Thus, the analyst can *a priori* rule out the possibility that the English *blackbirds* (and *black birds*) are cases of interrupted *blacks*. However, such delimitation, correct as it seemingly is, does not take one very far, because irrespective of how the term is defined, to rule that certain interruptions are not interruptions proper, one has to invoke the wordhood status of the interrupted and interrupting elements. If a sequence cannot be interrupted at all in any sense, it is a word structure rather than a phrase structure. If a sequence is interrupted, one could not, *ipso facto*, conclude that it is a phrase structure rather than a word structure. The surprising conclusion that follows is that while Noninterruptibility is necessary, sufficient and therefore definitive, Interruptibility is not. Interruptibility and Noninterruptibility are not parts of the same statement. If **a** Δ **b** is not interruptible, it is a word all right, but if **x** is a suffix and **b** is one too, then **a** Δ **x** Δ **b** would be a word (**a-x-b**), not a phrase (**a x b**); if, on the other hand, **x** is a word, irrespective of whether **b** is a suffix or a word, **a** Δ **b** would be a phrase (**a b**), not a word (**a-b**). In empirical terms

7. puo vor-te
   he-nom come prf "he has come"
   (=Ø)

can be "expanded", resulting in

7a. puo vor lle-te
   exrt "he has managed to come; he has made it"

7b. puo vor ba-te "he has come and sat"

On the other hand,

8. puo vor tyo
will "he will come"
can be "expanded" to give
8a. puo vor ba tyo "he will come and sit"

Both *vor-te* (of 7) and *vor-lie-te* (of 7a) are words, but neither *vor ba-te* nor *vor ba tyo* is. These conclusions are reached not on the basis of the criterion of interruptibility, but on the basis of the status of the interrupted and interrupting chunks as to wordhood. If *-te* (7) and *-lie* (7a) are affixes, then *vor-lie-te* (7a) would be a word. If, on the other hand, the "interrupting" *ba* is a word, *vor ba-te* (7b) would be a phrase irrespective of whether *te* is a word or not. *Te* in *vor ba te* (7b) is a suffix while *tyo* in *vor ba tyo* (8a) is a word. But both *vor ba te* and *vor ba tyo* are phrases. Further, once the working grammarian is sure about the monomorphemic status of a linguistic element, this heuristic is of no avail. *-lie*, the exertive mood marker, *-ro*, the conditional protasis-clause marker, and *tyo*, the future tense marker are all monomorphemic, and tautologically, noninterruptible. Since monomorphemic linguistic utterances may be either words or nonwords, the conclusion that this criterion leads one to is misleading. By other criteria, only the last one turns out to be a word while the former two turn out to be nonwords. The criterion does not, of course, come into play in case of monophonemic morphemes, e.g. a syllabic consonant with morphemic content like *n* "your".

Scholars (e.g. Kramsky 1963:37 ff, Matthews 1972:98 fn 2) have insisted on the extraneous character of the interrupting material. That a word structure "cannot be interrupted by rearranged material from elsewhere within a given sentence" (Mathews:Ibid),[5] but only by a fresh piece as in 5a and 6a, seems to make good sense at first. (Notice that the word "interruption" is used in the ordinary sense, not in any restricted or technical sense.) However, Tibeto-Burman languages in general, and Angami in particular, falsify Matthews' statement endorsing Kramsky and others. Consider:

9. no haki vor-lie "you come here"
you here come imp

9a. no vor haki-lie "you come here"

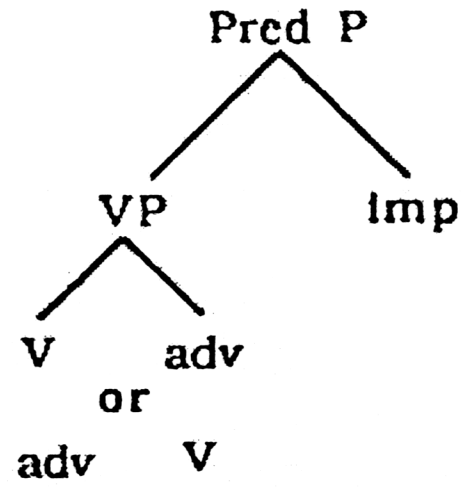
---

[5] It is surprising that Matthews should stress that "in the sentence it is possible, even normal, for a syntagm to be interrupted by certain extraneous elements", while in a footnote on the same page, the word "extraneous" is explained to be extraneous to the sentence and it is averred that morphological structures can be "interrupted" only by extraneous elements.

10. mhai thu -sũ -lie "write quickly"
quickly write vr imp

10a. thu-sũ mhai-lie "write quickly"

Vor-lie in 9 and *thu-sũ-lie* in 10 are words, morphological structures, and the interrupting material, viz. *hakí* "here" in 9a and *mhai* "quickly" in 10a, are not extraneous to the sentence, but are from elsewhere in the sentence. *Vor-lie* and *thu-sũ-lie* are, in fact, "interrupted by rearranged material from elsewhere within a given sentence". That the "interruption" or rearrangement results in a redrawing of word boundaries is beside the point. The following phrase marker may be said to express what is happening.



This is a syntactic phrase marker, which does not represent information about morphology. For instance, the fact that the imperative marker and the verb root together make up a word in 9 is not representable in a syntactic phrase marker.

Angami has what could be suspected to be "classifiers" — consistently recurrent morphemes which mark a semantic class. For instance, *u-* and *a-* mark the semantic classes of body parts and kinship terms respectively. These "classifiers" are dropped in specifiable contexts, e.g. they are deleted when a pronoun or proper noun precedes:

11. ubuco "wrist"

11a. puo buco "his wrist"

11b. atha buco "Atha's wrist"

12. apuo "father"

12a. n puo "your father"

12b. zakilie puo "Zakielie's father"

Now the problem is to decide whether the preceding elements in the a and b examples above are *paradigmatic* replacements of the semantic class markers by the possessive (pro)nouns, or *syntagmatic* additions with a morphophonological deletion of *u-* and *a-*. The heuristic criterion of Noninterruptibility does away with all possibilities except that of syntagmatic addition. *Ubuco* "wrist" is not interruptible while *puo buco* and *atha buco* are:

11a.	puo buco	"his wrist"
11c.	puo puo buco	"his father's wrist"
11b.	atha buco	"Atha's wrist"
11d.	atha puo zeu buco	"Atha's father's friend's wrist"

As pointed out earlier, one should, however, know *a priori* that the interrupting materials are words belonging to an open-ended lexis. Incidentally the above is a classic example of morphological structures or words being internally not expandable, as against phrase structures which are. This also lends analytical support to the orthographic decision taken by the community (Sekhose 1973:16, 17) that a word of the type *ubuco* be written as one "complete" word while an utterance like *atha buco* be written as two words. As we will see, the criterion of Substitutability also lends incontrovertible evidence in favour of our conclusion. In sharp contrast, the grammatically conditioned variability of the pronominal copy in the verb exemplifies paradigmatic replacement. Angami verbs descriptive of mental, emotional states are coded, in optimal grammatical pleonasms, for the person and number of the subject NP.

13a.	a	a-ni	ba	"I am buoyant"
	I-nom	I-be happy	prg	
	(=Ø)			
b.	no	n-ni	ba	"you are buoyant"
	you-nom	you-be happy		
	(=Ø)			
c.	puo	puo-ni	ba	"he is buoyant"
	he-nom	he-be happy		
	(=Ø)			

d. uko u-ni ba "we are buoyant"
we(incl.pl) we-be happy
-nom(=Ø)

Clearly, the variability of the grammatically coded person-number prefix in agreement with the subject NP is a matter of syntactically triggered paradigmatic replacement. *Puo* "his" of 11a and *n* "your" of 12a are "additive roots", while those in the verb complexes of 13b and c are "replacive prefixes". The latter are in fact bound alternants of free forms. Notice also that the grammatically conditioned choice is between members of a highly restricted, closed paradigm. Supporting our conclusion is the structural attribute of Noninterruptibility. The verb roots in 13 a-d are not interruptible in any sense, which suggests their lexical-morphological status.

COMPOUNDS

Compounds are invariably noninterruptible internally cohesive, even if complex, units.

14. lesū - kepro "kite"
paper flying

15. thevo - ki "pigsty"
pig house

16. nyo - kra "chalk"
mud be-white

17. dzū - vū "swim"
water beat

*14a. lesū keza kepro "big kite"
big

*15a. thevo keza ki "big pigsty"[6]

*16a. nyo keza kra "big chalk"

---

6 The Angami form corresponding to the allegedly ambiguous English "big pig house" is not ambiguous: the modified compound can refer only to the big enclosure. The language makes use of a phrase to express the portliness of pigs:

thevo keza $\left\{ \begin{matrix} \text{-ko} \\ \text{-u} \end{matrix} \right\}$ ki "the big $\left\{ \begin{matrix} \text{pigs'} \\ \text{pig's} \end{matrix} \right\}$ house"

*17a. dzü kemekou vü "swim in cold waters"[7]

Contrast examples 14-17 with

18.	lesü paper	pro-ketyo fly-cnt	"flying paper"
19.	nyoba	kekra	"white clod"
20.	tsali song	tshü do	"sing (a song)"
21.	dzü water	rülu pour-onto	"to have a bath"
22.	vor	ba-lie	"come and sit"

which can all be interrupted:

18a. lesü keza proketyo

19a. nyoba keza kekra

20a.	tsali song	puo one	tshü-sü-lie sing-vr-imp
21a.	dzü	kele hot	(se) rülu-lie with
22a.	vor come	haki here	ba-lie sit-imp

---

[7] For the way the concept is expressed, see example 73. The participial form of *dzü·vü* "swim" viz. *dzü·ke·vü* is not a counterexample for reasons set forth in the following pages. In a nutshell, they are:

(a) the positional relationship of *ke·* with its linguistic context is grammatically conditioned. It can only be prefixed to verbs, and *vü* is the verbal component of the compound;
(b) the "interposition" by ke- is nonparenthetical;
(c) it is uniquely limited; and
(d) *ke·* cannot be from elsewhere in the sentence.

The last of the reasons holds here although it is not a universally valid postulate, and is, in fact, not generally true in Angami, as we shall see.

The latter (exs. 18-22) are thus syntactic units, rather than lexical units. If a phrase is not expandable, it is because of an adventitious conceptual constraint rather than a linguistic-structural one.

1.2 Potential Pause

There is no potential pause between components of a lexical unit. They cannot be temporally stretched except in hesitations which are triggered by speaker-indecision, in pedagogically oriented enunciations of utterances and in pathological language behaviour. If the sequence **a Δ b** =/=> **a b**, then **a Δ b** is **a-b**, not **a b**. In 1 and 1b below

1. umhl-u "the eye"
   eye-indsg

1b. dzū-vū "to swim"

as opposed to

23. umhl hau "this eye"
    eye this

20. tsall tshū "sing"
    song do

there is no pause between the two components. In contrast, if the speaker so wishes, he could parenthetically pause between the two elements in 23 and 20. This criterion is buttressed by

(a) 1.1. viz. that parenthetical material can be interposed in a phrase and not a lexical unit. As Sadock (1980:303) puts it, "error correction can take place only at the boundaries of words."

(b) the Elicitability test which says that only word-sized units can be elicited through the interrogative *kedlpua* "what", in case the interlocutor did not hear the speaker well in the first instance. Thus, one can have a repetition

23. umhl hau
    eye this

but not

1. umhi-u
   eye-indsg

in response to

23a. umhi kedipua...?
   eye what

One elicits the (whole) word viz. *umhi-u* "the eye" by such interpolated interrogation, not the nonword *-u* "the". This also works the opposite way round. The possibility of 23a suggests that *umhi* is a word too.[8] If a sequence like

24. lesũ -ki "school"
   paper house

is a compound, which indeed it is, a question like

24a. shupuo ki (ga) "whose house?"
   whose house Q

is not likely to yield 24 as an answer, whereas such a question is perfectly possible with a phrase like

25. vizoi ki "Vizoi's house"
   Vizoi house

24a. shupuo ki (ga) "whose house?"

This criterion as defined above would not be sufficient. Disyllabic morphemes which happen to be bound (eg. *-ketsũ*, the gerundializer), one could argue, are phonologically cohesive in that there can be no pause

---

8 a. In some forms of spoken English at least, this is a compelling reason to consider the determiner *the* a word. An interpolated "the what?" is possible in response to a declarative like "I am basking in the balmy bracing opulence of my consciousness" because *the* is an integral linguistic unit, a construction rather than a subconstruction.

b. Pike and Pike (1977/80:113) consider *the* and *a* to be grammatical words but phonological nonwords (in other words, clitics) because "(they) are not stressed and normally are not isolatable." I am not sure stress group should be considered criterial in all cases. The Elicitability test (and the Substitutability test) point(s) rigorously to the fact that *the* and *a* are NOT phonologically bound. The same is true of the unstressed, weak forms viz. [tu] and [t ə] of the English preposition/infinitiviser *to*, the stressed strong form being [tu:]. Phonologically bound phonic material does not pass these tests. A neat solution is to call the combination of the unstressed, "weak" material plus the stressed form it leans on "a phonological phrase".

between the two syllables, and hence are words. But the question to ask is not whether *-ketsū* is in itself a word, but whether it is a word in a natural utterance like

26.	themie	dukhri-ketsū	vi		mo
	man-nom(=Ø)	kill-grn	be good	-prs(=Ø)	neg
	"killing human beings is not good."				

In the context of a constitute, Potential Pause is sufficient. The Chomsky and Hallean definition of the word reflects this: "a string of formatives (one or more) contained in the context # # # # and containing no occurrences of # #." (Chomsky & Halle 1968:12-14)[9]

Bound morphemes would be integral words in metalinguistic statements. This is predicted by the operation of this heuristic in sentences. The criterion, putatively the most overarching of the criteria of wordhood in language, is not necessary in Angami, since there is at least one case where it is overridden. *Tyo*, the future tense marker is isolable. But there is no potential pause between it and the phonic material preceding it, a case of isolability dominating Potential Pause.

A final point that is germane here is that morphophonemic alternation does not give one the lead as to the lexical wordhood of the interacting linguistic chunks, for sandhi could be either word-internal or word-external.

27.	vor + ro	===>	voro	"if (someone) comes"
16b.	nyoba + kra	===>	nyo-kra	"mud-be white (=chalk)"
11a.	puo + ubuco	===>	puo buco	"his wrist"

27 exemplifies intraword or internal sandhi (specifically, loss of inter-morphemic gemination), while 16b and 11a illustrate crossword or external sandhi (loss of the final syllable and of the initial vowel). Absence of sandhi changes, on the other hand, is not diagnostic of (compound) wordhood vis-à-vis phrasehood. While compounds may be either parathetical or nonparathetical, phrases may also consist of either phonologically noninteracting or interacting free morphemes.

---

9 Bh. Krishnamurthy, discussing this definition in the National Seminar on Morphology (Jan. 6-7, 1984), Hyderabad, India, dubbed it "circular": that Chomsky and Halle *a priori* decided that determiners in English were words and then went on to define the word the way they did. I, for one, do not find any circularity in the definition *per se* inasmuch as it is an empirical issue whether a phonemic sequence contains an occurrence of # #.

1.3 *Isolability*

Linguistic material which is isolable either from an immediate linguistic context (i.e. from part of a sentence) or from a larger linguistic context (i.e. from a sentence), is a word. If **a** Δ **b** ==> **b**, **a** is a word. Linguistic material which is isolable from part of a sentence can perform the *potential phrase function*, i.e. it can stand for the whole phrase, while linguistic material which is isolable from the sentence can perform the *potential sentence function*, i.e. it can stand for the whole sentence. The former is exemplified by the future tense marker *tyo*, and the future negative marker *lho*.

28. puo sodu n kɪ nu vor tyo
he-nom(=Ø) tomorrow your house to come fut
"he will come to your house tomorrow"

28a. puo tyo
he-nom(=Ø) fut

29. puo sodu n kɪ vor lho me
ŋg ŋ
"will he not come to your house tomorrow?"

29a. puo lho
"he will not"

Tyo and *lho* stand for the predicate phrases in 28 and 29 respectively. Isolability of the second kind is illustrated by all lexical-morphological units, which can stand alone as answers to questions about extralinguistic reality[10]. This is what is meant by Matthews (*op.cit.*96) when he calls the word "minimal sentence construction" and Pike and Pike (1980:11) when they speak about "isolability". Neither of them, however, brings up the question of isolability at the phrase level.

D1 A: no kira vo ga
you(sg)-nom(=Ø) where go-pst(=Ø) Q
"where did you go?"

B: kewhira
Kohima
"Kohima"

---

[10] This is a nonrestrictive relative clause.

D2 A: n mlz gl kebau kedlpuo ga
your table on that-which-is what Q
"what is that which is there on your table?"

B: lesūda
book
"a book"

Linguistic material in the response slots of the above verbal exchanges stand for whole sentences. A subword could not perform either the potential phrase or the potential sentence function, so that the criterion of isolability is a sufficient criterion to distinguish wordhood from affixhood. Sentential isolability, however, is not sufficient, as phrases can also stand for sentences. Thus in dialogue D2, a phrase could also fill the response slot:

D2: A: n mlz gl kebau kedlpuo ga
"what is that which is there on your table?"

B: (1) lesūda "a book"
(2) phizo lesūda "Phizo's book"
(3) phiso zeu lesūda "Phizo's friend's book"

Isolability of either kind is not necessary, since notionally dependent forms (e.g. noun attributes, the disjunctive and conjunctive coordinators, etc.) are not isolable but are clearly words by other criteria. For example, *kevi* "good" as in

30. lesūda kevi "good books"

is not isolable. In fact, it has no existence except in the company of the noun it modifies. But it is a word by the criteria of Noninterruptibility and Potential Pause. Further, phrasal isolability conflicts with the criterion Potential Pause. There is no potential pause between *vor* "come" and *tyo* "will" in 28, but *tyo* is isolable as we saw, and is clearly an integral element. In the case of *tyo*, isolability overrides Potential Pause. In the case of postpositions, on the other hand, Potential Pause overrides Isolability. The sequence of a noun and the following postposition can be separated in time and hence the postposition is a word. But the postposition is not isolable. Isolability conflicts with Elliptibility, as we shall see.

1.4. Elliptibility

A syntactic test of lexical wordhood, this principle states that phonic material which can be deleted on identity with a trigger either in a linguistic or pragmatic context is a word. Ellipsis is a communicative option[11] available for the speaker to delete linguistic elements which are contextually retrievable. If **a Δ b Δ c ===> a b Ø , c** is a word. In the following interaction, the ellipted material symbolised by Ø constitutes a word each:

D3 A: kiezotuo kira vo -zhie ga
Kiezotuo-nom(=Ø) where go prg Q
"where is Kiezotuo going?"

B: Ø (=kiezotuo) n ki nu Ø (=vo-zhie)
your house to
"(=Kiezotuo) Ø (=is going) to your house"

Subwords are not elliptible. *-zhie* is the bound morpheme in D4:

D4 A: kiezotuo kira vo_1 $-zhie_2$ ga
"where is Kiezotuo going?"

B: a Ø (=kiezotuo) n ki nu Ø (=vo-zhie)
"(=Kiezotuo) Ø (=is going) to your house"
*b. Ø (=kiezotuo) n ki nu $vo-Ø_2$ (=zhie)
"Ø (=Kiezotuo) Ø (= -is -ing) go to your house"

A third type of response by B, viz.

*c. Ø (=kiezotuo) n ki nu $Ø_1$ (=vo)-zhie
"Ø (=Kiezotuo) is Ø (=going) to your house"

is not possible either, leading one to the conclusion that *vo* "to go" is not elliptible and hence is a subconstruction rather than a construction. vo in D4 B: c cannot be ellipted, not because it is intrinsically non-elliptible, but because it is a part of a word in the sentence under consideration. It forms a word together with *-zhie*, the bound morpheme. Supporting evidence comes from constructions where the linguistic material in the immediate

---

[11] There is an important (Chomskian) hedge to all statements which relate language structure to the function of communication viz. that language is an emergent, stimulus-free, creative instrument of self-expression. Its role as a social semiotic, a mediator of social purpose is an incidental, overlaid function.

linguistic context of *vo* is not phonologically bound, but is free and isolable (*tyo* in the following example), and where *vo* is a word, not a part of a word.

D5. A: n nyo vo tyo me
your son go ft Q
"will your son go?'

B: a nyo Ø (=vo) tyo
"my son will Ø (=go)"

Notice that the isolability of an element entails the ellipsis of certain other elements in the sentence. The response in Dialogue 5 (D5) illustrates at the same time the elliptibility of *vo* and the isolability of *tyo*. The converse viz. that ellipsis entails isolability is not true.

Elliptibility is not necessary, since it is often sensitive to the syntactico-semantic nature of the linguistic material in question, not solely to its status as to wordhood. Phonologically independent but linguistic-structurally and notionally dependent linguistic material cannot be ellipted. For instance, the individual constituents of modifier-head-modifier constructions are not subject to ellipsis:

D6 A: tiso ki kesa kira ba ga
Teiso house new where be-prs(=Ø) Q
"where is Teiso's new house?"

B: a Ø (=Teiso's new house) lunu ba
there
"Ø (=Teiso's new house) is there"

*b. Ø (=tiso) ki kesa lunu ba
"Ø (=Teiso's) new house is there"

*c. tiso Ø (=ki) kesa lunu ba
"Teiso's new Ø (=house) is there"

*d. Ø (=tiso ki) kesa lunu ba
"Ø (=Teiso's house) new is there"

*c. tiso ki Ø (=kesa) lunu ba
"Teiso's Ø (=new) house is there"

In the above dialogue, *tiso*, *ki* and *kesa* are constituents of an NP. None of them could be deleted to the exclusion of the others. But all of them are words by other criteria.

Ellipticibility is not sufficient because it is diagnostic of wordhood vis-à-vis affixhood, but not of wordhood vis-à-vis phrasehood. Not only words but phrases (D7 below) and sentences within sentences (D8 below) may be ellipted as well:

D7. A: tefü pengu-ko kira ba ga
dog five-pl-nom(=Ø) where be-prs(=Ø) Q
"where are the five dogs?"

B: Ø (=tefü pengu-ko) hanu ba
"Ø (=the five dogs) are here"

D8. A: saville dimapur nu vo ke idi
Savilie-nom(=Ø) Dimapur to go down-pst(=Ø) qut

n nu shupuo pu-sü
you to who-nom(=Ø) say-pst(=Ø)-vr
"who told you that Savilie went down to Dimapur?"

B: Ø (=saville dimapur nu vo ke idi) (a ki) a puo pu-sü
"my father said (to me) Ø (=that Saville went down to Dimapur)"

Ellipticibility conflicts with the criteria of Potential Pause and Isolability. Postpositions illustrate its conflict with Potential Pause:

D9. A: savilie shupuo kinu priesa se vor ga -pst
Savilie-nom(=Ø) who from money with come Q (=Ø)
"who did Savilie bring the money from?"

B: *savilie selie Ø (=kinu) priesa se vorü
"Saville brought money Ø (=from) Selie"

kinu "from" is not elliptible but is a word by the criterion of Potential Pause. The conflict between Ellipticibility and Isolability is illustrated by *iho* "will not" and manner adverbs:

31 a puo themuo sa lho
he-nom(=Ø) meat have-more ftng
"he will not have more meat"

b puo lho
"he will not"

32. puo themuo sa
have more-pst(=Ø)
"he had more meat"

Sentence 31b illustrates the isolability of *lho* which stands for the whole predicate phrase. Sentence 32 does not illustrate its elliptibility because 31 and 32 are not derivationally related. On the other hand, manner adverbs are elliptible (B:a), but are not isolable (B:b):

D10. A: puo $tshū_1$ $pevi_2$ -lie me
he-nom(=Ø) do well vr Q
"did he do well?"

B: a puo tshū $Ø_2$ -lie
"he did Ø (=well)"[12]

*b. puo $Ø_1$ pevi-lie
"he $Ø_1$ (=did) well"

---

[12] a. This is not possible in English:
A: Has he done well?
B: (a) Yes, he has done well.
(b) Yes, he has.
*(c) Yes, he has done.

b. By this test, *the* and *a* in English could be nonwords, contrary to the conclusion averred in footnote 8:
A: Is there a book/Is the book there?
B: (a) Yes, there is a book/the book is there.
(b) Yes, there is/it's there.
*(c) Yes, book is there/ there is book.
Quite possibly, the Elicitability test overrides both the (Piklan) phonological test of stress group and the syntactic test of Ellipsis. One could of course take the tack that we took as regards the nonisolability of modifiers and postpositions viz. their notional dependence on heads of construction. The formal (not notional in this case) and unilateral dependency relationship between determiners and head nouns rules out both elliptibility and isolability for the former.

1.5 *Substitutability*

A first approximation of this heuristic principle is that if **a** can substitute for **b**, and if **b** is a word, **a** is a word too. By this criterion, *buco* of *ubuco* "wrist" is established as a word, although *buco* can never be established as an utterance in its own right.

33. ubuco "wrist"

33 a. atha buco "Atha's wrist"
b. atha zeu "Atha's friend"

Since *buco* occupies the same linguistic role (viz. the ultimate element in an NP string) and more importantly, plays the same role as *zeu* (viz. the head of an NP), *buco* may be said to be in a relation of substitution with *zeu*. Since *zeu* is a word preanalytically, *buco* is a word too. Similarly, possessive pronouns are established as words:

34. apuo "father"

34 a. saville puo "Savilie's father"
b. a puo "my father"
c. n puo "your father"
d. puo puo "her father"

a "my", *n* "your", and *puo* "her" occupy the same prehead position and have the same function (viz. nominal attribute) as *savilie*, a proper noun. They are therefore in a relation of substitution with *savilie*. Given that *savilie* is a word, *a*, *n*, and *puo* are words too, and 34b-d are to be construed disjunctively, i.e. as phrases.

If on the other hand **a** can substitute for **b**, and if **b** is an affix, **a** is an affix too.

35 a. ki -u "the house"
house indsg

b. ki -nie "the two houses"
inddu

c. ki -ko "the houses"
indpl

-u immediately follows the noun root, which is its location in linguistic space. It marks individuation (sg), which is its linguistic role. *-nie* also immediately follows the noun root and marks individuation (du). Given that *-u* is an affix, *-nie* is an affix too. Similarly, *-ko* of 35c is an affix since it immediately follows the noun root and marks individuation (pl). Consider now the case of *puo* "one" as in

36. ki puo "one house"

puo follows the noun root all right, but does not play the same linguistic role as *-u* of 35a. does, i.e. it does not mark individuation. It expresses a cardinal number, not grammaticalised number. *puo* is therefore not in a relation of substitution with *-u*. So no judgement could be pronounced regarding its status as to wordhood using the criterion of Substitutability.

The substitutes in 35 belong to a morphological substitution class (MSC) while those in 33 and 34 belong to a syntactic substitution class (SSC). A crucial difference between MSC's and SSC's is that MSC's do not allow expansion while SSC's do. *ko-yo* as in

35 d. ki-ko-yo "the small houses"

where *-yo* marks diminution does not substitute for *-ko* of 35c. *-ko* and *-yo* belong to different MSC's as they play different linguistic roles. *-ko* belongs to a three-member MSC (*-u* and *-nie* being the other members), and *-yo* to a single-member MSC. But SSC's allow unlimited expansion. In 34, *savilie* could be replaced not just by single-word utterances like *a* "my", but also by phrases as in

34 b.	1	a puo	"my father"
	2	a zeu puo	"my friend's father"
	3	a zeu kimie puo	"my friend's wife's father"

All of these phrasal substitutes occupy the prehead position and function as genitival attributes to nominal heads. Substitutability, therefore, is not sufficient. (Syntactic) substitutes could either be words or phrases. A first rephrasing of the principle of Substitutability would state that if **a** can substitute for **b** and if **b** is a word, **a** could be either a word or a phrase.

The discrepancy in expandibility between MSC's and SSC's calls into question Nida's seventh structural contrast between morphological and syntactic structural levels (Nida 1949:102) which states that "morphological structures have a more limited substitutability of expanded expressions". As

pointed out in the foregoing paragraph, members of an MSC cannot be expanded. "Goings on" is an expansion of "go on", but not an expanded substitute. In contrast, the phrase "a dumpy girl" is both an expansion and a substitute, or an expanded substitute for the word "girl". Clearly, it should be a discrete either-or statement, not a gradient more-less statement. Impugnably, Nida's is a gradient more-less statement. In Mao Naga, a Tibeto-Burman language closely related to Angami, case markers may be bound morphemes or postpositions.

36 a. ai doktar-yi nie
I-nom(=Ø) doctor-DO see-pst(=Ø)
"I saw the doctor"

b. ai doktar he tae
I-nom(=Ø) doctor near go-pst(=Ø)
"I went to (lit. near) the doctor"

-yi, the direct object marker as in 36a, and *he*, literally "near" marking the allative in 36b, obviously belong to a substitution class: the former to a morphological substitution class and the latter to a syntactic substitution class. *doktar-yi* and *doktar he* have "equally" limited substitutability of expanded expressions.

Whatever expanded expression can substitute for the former can substitute for the latter as well, as 36c and d. show. That is, both the verb phrases in 36a and b have equally limited substitutability of expanded expressions. Whatever can expand 36a can also expand 36b, but the former is a morphological structure and the latter is a syntactic structure:

36 c. ai doktar mamül kazhü thopfü-yi nie
good-looking all-DO
"I saw all the handsome doctors"

d. ai doktar mamül kazhü he tae
"I went to all the handsome doctors"

It appears that when elements have grammatical meaning, expanded substitutability has more to do with the linguistic role the element plays than with its status as to wordhood. "Substitutability of expanded expressions" in the case of *doktar-yi* (36a) and *doktar he* (36b) is determined crucially by the case-marking function of *-yi* and *he*. It has little to do with the fact that *-yi* is an affix and *he* is a postposition, a word.

Substitutability, however, is necessary. A major difference between this heuristic principle and the others is that in the case of Substitutability, one needs to know *a priori* the wordhood status of another piece of phonic material, whereas the other heuristics do not need such knowledge. Note that Isolability and Substitutability are in conflict in cases like *buco* "wrist" and *puo* "father". *buco* and *puo* are not isolable but are words on the criterion of Substitutability. Substitutability is used in a slightly different sense in the second approach.

1.6 Potential Mobility

Phonic material which is potentially mobile across linguistic space is a word. If **a** Δ **b** ==> **b** **a**, then **a** and **b** are words. The Angami *sodu* "tomorrow" and *a* "I" in 37 and *mhatsü* "to eat food" and *rükri* "fast" in 38 are words because they can move across an utterance.

37 a. a sodu vor tyo
I-nom(=Ø) tomorrow come fut
"I will come tomorrow"

b. sodu a vor tyo

38 a. a mhatsü rükri si -ya
I-nom(=Ø) eat food fast always hab
"I always eat my meal fast"

b. a rükri mhatsü si-ya

In contrast, *-u* the definite article of *lesüda-u* "the book" as in 39; *-lie*, the imperativizer of *vor-lie* "come" as in 40; and *-krü*, the feminine gender marker of *ʃükrü* "female dog" as in 41, are not words because they are not mobile, but are subject to a rigid linear order:

39 a. lesüda -u kira ba ga
book-indsg-nom(=Ø) where be-prs-(=Ø) Q
"where is the book?"

*b. u-lesüda kira ba ga

40 a. haki vor -lie
here come imp
"come here"

*b. haki lie-vor

41 a. puo fü-krü kenie ba
he-nom(=Ø) dog-female two be-prs(=Ø)
"he has two dogs"

*b. puo krü-fü kenie ba

The sequential order of elements within a lexical unit is invariant.[13] By our operational definition of this heuristic principle, not only *-u*, but *lesüda* in 39 would not be a word either. This is met by the same argument that was adduced in the case of Elliptibility (1.4). *lesüda* in 39 is part of a word, not a word in itself. Whole words, not parts of words, are scrambled across. This is confirmed by sentences where *lesüda* constitutes a word:

42 a. hako lesüda
these-nom(=Ø) book-pl-(=Ø)
"these are books"

b. lesüda hako

Mobility is not necessary, since phonic material which obviously constitutes word-sized units by other criteria need not be potentially mobile. The constituent elements in the subject NP in 43 are not positionally free:

43 a. a ki kezhivi kewhira ba
my house beautiful Kohima be-prs(=Ø)
"my beautiful house is in Kohima"

*b. kezhivi ki a kewhira ba

a "my" and *kezhivi* "beautiful" cannot move across the phrase, but are still words. The Angami Naga verb-complex is constrained by a rigid morpheme order:

44 a. a rü pevi tyo
I-nom(=Ø) write well fut
"I will write well"

---

[13] Cases apparently refuting this universally valid metatheoretical postulate (or empirical regularity) of fixed morpheme order within the word have been reported (Stevens 1971) from Madurese (Austronesian, Indonesia) and Bikol (Southeast Luzon, Philippines). See Stevens for an exposition of the reasons why they are not completely cogent countercases.

*b. a pevi rü tyo

*c. a tyo pevi rü

*d. a tyo rü pevi

But each of the three constituents of the verb-complex in 44 is a word by various other criteria. Potential Mobility is not sufficient either, as phonic material subject to mobility may constitute a phrase as well. That is, Potential Mobility may keep the word and the bound morpheme apart, but could not keep the word and the phrase apart. Thus, in the variant pair of sentences

45 a a^{1} puo^{2} puo^{3} ze^{4} kese^{5} tyo^{6}
"I^{1} will^{6} meet^{5} with^{4} his^{2} father^{3}"

b. puo^{2} puo^{3} ze^{4} a^{1} kese^{5} tyo^{6}

the moveable linguistic chunk viz. *puo puo ze* which changes its position with *a* "I" is not a word but a three-word phrase by other, more dominant criteria.

Potential Mobility conflicts with both Potential Pause and Isolability. Postpositions are marked by potential pause at their boundaries but are not mobile. *tyo*, the future tense auxiliary is isolable but not mobile. What is more, mobility may be sensitive to the nature of the lexical unit, not exclusively to its woodhood. *si*, the aspectual adverb meaning "always", for instance, does not show the same degree of mobility as *mhatsü* "to eat a meal" and *rükri* "fast" do. The sentence

38 c. a si rükri mhatsü-ya
I-nom(=Ø) always fast eat

is not as felicitous as

b. a rükri mhatsü si-ya

The positive and the negative sides of a heuristic principle must be polar opposites. Gradience invalidates its heuristicity. The word "mobility" has perforce a slightly different sense in the second approach to wordhood.

1.7 Of the six heuristics laid out in the foregoing section, only one viz. Potential Pause is phonetic, the rest are all syntactic except that

Substitutability can be morphological as well. While Elliptibility and Mobility are neither necessary nor sufficient, there is no heuristic which is both necessary and sufficient. Potential Pause and Isolability are sufficient without being necessary. Substitutability is necessary without being sufficient. Potential Pause agrees only with Noninterruptibility in both directions. That is, whenever a morpheme boundary is marked by potential pause, parenthetical material can interrupt the sequence of morphemes (the delimited definition); wherever parenthetical material can intervene, the site of intervention is characterised by potential pause. Potential Pause need not agree with Isolability in either direction, i.e. whenever a sequence is marked internally by potential pause, none of the elements need be isolated, and whenever an element is isolable, there need be no pause between it and an adjacent element. Isolability works at cross purposes with Potential Pause (1.2), Elliptibility (1.4) and Substitutability (1.5). Potential Mobility does not agree with Isolability and Potential Pause, and what is more, it may vary in degree in response to the nature of the lexical unit, having nothing to do with its status as to wordhood (1.6). The Angami word is a squishy phenomenon in that there is a clash among the criteria that define it. Let us fashion a hierarchy to resolve the clash:

Isolability/Potential Pause
Isolability/Substitutability
Elliptibility
Noninterruptibility
Potential Mobility

A hierarchy may be implicational either in the sense of entailment or domination. The two are of course typically related, as they seem to be here. Entailment implies that a criterion lower in the order of precedence entails the one(s) above in a relation of inclusion. Thus, Potential Mobility entails Interruptibility. Interruptibility entails Elliptibility, Elliptibility entails Isolability Substitutability, and so forth. Domination implies that in case of dissonance, the criterion above holds the stage overriding the one(s) below it. Thus, if an element is substitutable for another but there is no possible pause in the speech continuum that it constitutes along with the preceding or following element, it is an infra-word unit rather than a word; if a sequence of linguistic elements is rigidly ordered but is interruptible or is characterised by potential pause, it exemplifies a phrase structure rather than a word structure; if an element is isolable but not elliptible, it is a word, Isolability overriding Elliptibility.

Hierarchised, some of the criteria, then, dovetail into an agreeable whole, but only some. It is difficult to rank Isolability vis-à-vis

Substitutability and Isolability vis-à-vis Potential Pause, because Isolability and Substitutability override each other in different instances, as do Isolability and Potential Pause. This is indicated in the proposed hierarchy by placing them on a par. And entailment is, only typically, not uniformly, unidirectional. Ellipsibility, for example, only typically entails Isolability, i.e. whatever is elliptible need not always be isolable. The adverb *pevi* "well" is elliptible (dialogue D6), but is not isolable. All this notwithstanding, it is possible to formulate a complicated, inelegant rule which could lead one to the discrete Angami word. In fact, except when it is overridden by Isolability, Potential Pause will determine word boundaries. Whether the unit we are trying to define is discrete or nondiscrete, this approach is narrowly conceived, blinkered and to that extent, attenuated in value.

2.1 Supervening Meaning

Compounds, unlike phrases, may have a meaning which is not a straight sum of the meanings of its constituent elements, a meaning which is at some level absolute rather than composite. The word "may" is used advisedly: this is a sufficient, not a necessary criterion. Compounds with such supervening meaning and their phrasal parallels are ideationally different. *nyoba kekra* "white clod of earth" and *lesū proketyo* "flying paper" are the phrasal parallels of *nyo-kra* "mud-be white (=chalk)" and *lesū kepro* "paper-flying (=kite)" in 16 and 14.

It is possible to say

16 b. no nyo-kra keti ngu-wa-te me
you-nom(=Ø) mud-be white black see vr prf Q
"have you seen a black mud-be white(=chalk)?"

but not

*16 c. no nyoba kekra keti nguwate me
"have you seen a black white clod?"

It is possible to say

14 b. lesū -kepro hau pro -lie lho
paper flying this-nom(=Ø) fly vr ftng
"this kite will not fly"

but not

*14 c. lesü proketyo hau prolie lho
flying
"this flying paper will not fly"[14]

That is to say, the compounds *nyo-kra* "chalk" and *lesü-kepro* "kite" differ in morphemic content from the members of the corresponding syntactic sequence classes. They have truth values and ontologies which are different from their phrasal congeners. As part of the process of acquiring a new semantic identity in conjunction with the combining component, nonheads of noncompositional compounds shed their "sense" and retain only "reference". "Sense", to quote Lyons (1968:424), "is the place in the system of relationships a word contracts with other words in the vocabulary" whereas "reference" is "the relationship that holds between words and the things, events, actions and qualities, they stand for" (ibid:427). *nyo-kra ketl* "black mud-be white (=black chalk)" is possible because *kra* "be white" no longer has its meaning defined by a network of relationships of contrast in which *ketl* "black" is a member: *kra* "be white" as part of the compound

---

[14] There are some borderline cases. There is a question mark about the compound status of *dzü keza* "water-big (=ocean)" going by this test. The first sentence is not possible. If anything, says my informant, the second is possible:

* *hau* *dzü* *keza* *ketsü* *puo*
this-nom(=Ø) water big small one
"this is a small big-water (=ocean)"

? dzü keza hau tsü
water big this-nom(=Ø) be small-prs(=Ø)
"this big-water (=ocean) is small"

It is nevertheless a compound because of the following evidence:

kijü nu (dzü keza) penggu ba
world loc (water big) five-nom(=Ø) be-prs(=Ø)
(*dzü)
(water)
"there are five (big-water (=ocean))s in the world"
(*water)

dzü keza "water big (=ocean)", if our conclusion about its compoundhood is right, runs counter to the argument towards the end of 2.1, where the compoundhood of *dzü-le* "water-be hot (=hot water) will be brought into sharp focus in terms of a comparative juxtaposition of *dzü kele* "water hot (=hot water)" and *dzü-le*. *dzü-za* "water-be big" exemplified in a sentence like

dzü-za u la vi se
water be big-nom(=Ø) us for be good very-prs(=Ø)
"water-be big (=kind of salty water) is very good for our health"

refers to a kind of nutritious brine that is found deep down in the earth. It appears that *dzü keza* and *dzü za* do not pattern the way *dzü kele* and *dzü-le* do, since both of the former pair are compounds, unlike the latter pair. This leaves one with an additional perspective on (though not with an additional structural type of) the Angami compound, which will not be further examined in this paper.

has no "sense" vis-à-vis *keti* "black". Even if a hypothetical compound *nyo-ti* "mud-be black" were possible, and in the event that such a compound were noncompositional, *ti* "be black" would still have no sense vis-à-vis *kra* "be white" or vis-à-vis any member of its semantic class in particular or any member of the vocabulary in general.

The test of (negative) equational or definitional sentences where part of the compound is repeated as part of the predicate points to this kind of exocentric meaning that a compound acquires. It is possible, for instance, to say

46. uphi-mhi umhi mo
leg eye eye not
"leg-eyes (=ankles) are not eyes"

but not

*46 a. a umhi umhi mo
"my eye is not an eye"

A problem about this test is that at some level of meaning, the sentences strained out by it seem perspicuous.

47 a. a mhi umhi mo
"my eye is no eye"

47a, for instance, is possible as an utterance by a blind man. Further, the (im)possibility of a sentence like

47 b. "tears (eye-water) are not solid"

does not say anything about the compound status of "tear (eye-water)", because unlike "leg-eye (=ankle)", which is not an eye, both tears and water are liquid. Notice, however, that the compound in 47b (viz. eye-water) is compositional. The test is generally valid for noncompositional compounds. 47a seems a special case. Contrast 48 with 48a:

48. lesũ -nyũ nhanyũ mo
paper leaf-nom(=Ø) leaf neg
"paper-leaf (=page) is not a leaf"

*48 a. shūwhuo nyū nhanyū mo
a citrus fruit
"shūwhuo leaf is not a leaf"

Noncompositional compounds are nearly always structurally paralleled by phrases, unless semantic considerations rule them out (cf. *?lesū nhanyū* "paper's leaf"). *mhi-rho* "eye-open (=open eyes)", for instance, has no phrasal counterpart, whereas *nyo-kra* "mud-be white (=chalk)," *fū-ti* "dog-be black (black dog)," *lesū-kepro* "paper-flying (=kite)" do. This parrallellism of lexis and syntax is evidence of, or at least constitutes a *prima facie* case for hypothesizing noncompositionality.

When the noncompositionality of the meaning of a compound is not obvious, these tests can be particularly decisive. Angami has a class of compounds of the structural type N+V. *dzū-le* "water-be hot (=hot-water)", *fū-ti* "dog-be black (=black-dog)", and the already discussed *nyo-kra* "mud-be white (=chalk)" exemplify this class. The lexical *dzū-le* "water-be hot", a canonical example of this class of compounds, has a structural parallel in the phrasal *dzū kele* "hot water". *dzū-le* does not quite have a whole-different-from-the-sum-of-its-parts meaning. But consider

49. dzū - le le mo luo
water be hot-nom(=Ø) be hot neg yet
"water-be hot (=hot-water) is not hot as yet"

where something is predicated of water which is hot to a degree. A particular kind of object (viz. *dzū-le*) someone has asked for is not as hot as one would like it to be. Note that *le* "be hot" of *dzū-le*, the subject in 49, evinces no reaction to the sentential constituent which is its own token and which heads the VP (viz. *le* "be hot"). But 49a below with the phrasal *dzū kele* "hot water" is a blatant contradiction in terms:

*49 a dzū kele le mo luo
water hot-nom(=Ø) be hot neg yet
"hot water is not hot as yet"

kele is a true adjective. It has the function of making a more inclusive class a less inclusive one. It modifies, from the outside as it were, the nonlinguistic entity symbolized by *dzū* "water". The modification is word-external, making *dzū kele* "hot water" a kind of water, not one of various kinds of objects in the universe; whereas the predicative modification in *dzū-le* is gestalt-internal, making the word complex a notionally unitized atom, an ideational primitive, one of various kinds of objects in the universe,

rather than one of various kinds of water. In a situation where the speaker is identifying (50a) or choosing between hot and cold water (50b-c), *dzū-le* is incongruous. The attributive, identificatory *dzū kele* is the structure to use:

50 a hau dzū (kele)
this-nom(=Ø) water (hot)
(*-le)
(be hot)
"this is hot water"

b. a dzū (kele) se tyo
I-nom(=Ø) water (hot) use fut
(*-le)
(be hot)
"I will use hot water (rather than cold water)"

c. puo $dzū_1$ (kele) se $Ø_1$ rūlu -wa tyo
she-nom(=Ø) water (hot) using; wash; vr fut
with pour
(*-le) onto
(be hot)
"she will bathe with hot water (not with cold water)"

On the other hand, when talking about habitual dispositions (51a) or when choosing not between hot and cold water, but between hot water and rice-beer (51b), for instance, *dzū kele* is odd:

51 a. puo $dzū_1$ (-le) se dzū rūlu-wa -ya
she-nom(=Ø) water (be hot) with water hab
(*kele)
(hot)
"she habitually bathes with (hot-water)" (*hot water)

b. a zhu krie tyo no dzū
I-nom(=Ø) ricebeer-acc(=Ø) drink fut you-nom(=Ø) water

(-le) krie -lie
(be hot)-acc(=Ø) drink imp
(*kele)
(hot)
"I will drink ricebeer, you drink (hot-water!)"
(*hot water)

Finally, in contrast with the phrasal *dzũ kele*, *dzũ-le* passes all tests suggestive of a lexical entity. It cannot be stretched in time and cannot be interposed. Neither of its constituents can be ellipted or isolated. It thus satisfies the overarching structural requirement of compound lexical units viz. syntactic insularity (see 2.6). Unlike *dzũ kele*, it cannot be modified. *se* "very" is the postmodifier in 52:

52. pekrio dzũ {kele} se se puo-ni mo
Pekrio-nom(=Ø) water {hot} very with he-like neg
{*-le}
{be hot}
"Pekrio does not like very {hot water}"
{*hot-water}

Unlike *dzũ kele*, *dzũ-le* is closed for identity erasure. Compare 50c, where *dzũ* of *dzũ kele* acts as an antecedent triggering the deletion of *dzũ* of *dzũ rũlu*, with 51a where *dzũ* of *dzũ-le* does not.

2.2 *Unique Combinability*

The combination of linguistic elements in a compound tends to be unique, in that one of the constituents need not occur freely with other elements that seem quite similar semantically and formally to its fellow-constituent. Put more technically, no subset of the set of compounds in a language constitutes an internal distribution class. The (compound) word for "well" is

53. dzũ -khou
water storage place

But it is not possible to have

*53 a. lhako khou
rice

to mean "place where rice is stowed away". *khou* does not combine liberally with (the semantically relevant) members of the form class to which *dzũ* "water" belongs. I say "semantically relevant" because a sequence like *tl khou* "place where skies are stowed away" is not possible, one could argue, not because *khou* cannot combine well with *tl* "sky", but because the semantic result is balderdash.

Related to the above is another diagnostic test, viz. that synonyms[15] can not be interchangeably used in compounds while they can very well be in phrases. In syntactic groups, the slots are invariably open-ended, which is why, unlike compounds, they are substitution frames. For instance, the gerundialiser in Angami is the prefix *ke-*, as in *kedukhri* "killing" derived from *dukhri* "kill", which freely varies with the suffix *-ketsū* as in *dukhriketsū*:

26 a themie dukhriketsū vi mo
man killing be-good neg

b themie kedukhri vi mo
killing
"killing (humans) is not good"

Despite the free variation between *-ketsū* and *ke-*, only *lesū-kepro* is a compound. *lesū pro-ketsū* is not:

54. lesū- kepro / *proketsū vi mo
paper flying flying be-good not
"the kite is not good"

55. lesū proketsū vi mo
"to fly paper/the kite is not good"

This structural attribute of compounds is not necessary. Although *mha*, literally "material thing", combines with a number of verbs to yield compounds, as 56a-d shows,

56 a mha-tsū "to eat food"
mha-tshū "to do"
mha-cha "to cook food(?)"
mha-dzū "to be poor"

the lexical class of the combining element is not open-ended. *mha* does not combine with *rūnyū* "to listen(to)", for instance:

57. *mha-rūnyū "to listen (to)"

The heuristic is, however, sufficient, if the analyst keeps clear of phrasal unique combinability. The following illustrates unique combinability

15 As of now, I have no examples of synonyms.

which is phrasal. In Angami, verbs may be lexically marked, for instance, for the direct object. *metī* "wash hands", *chu* "wash face", *kenu* "wash mouth", *rūlu* "wash (the whole) human body", *khrū* "wash other individual body parts (eg. head, leg), utensils etc.", and *menyī* "wash clothes" illustrate such lexical marking. None of these verbs, except *menyī* which is marked for a class of direct objects and *khrū* which is marked for a class of (sub)classes, can take the interrogative direct object:

58. no kedipuo {*metī} ba ga
you-nom(=Ø) what-acc(=Ø) {*chu} prg Q
{*kenu}
{*rülu}
{menyī}
{khrū}
"what are you {*hand-} washing ?"
{*face-}
{*mouth-}
{*human body-}
{cloth-
body part-
utensil-}
etc.

Because *metī* "wash hands" combines uniquely with *ujie* "hand", *ujie metī* "wash hands" is not to be construed as a compound. The simple reason is that *ujie meti* does not constitute a lexical unit, the primary prerequisite for compoundhood: it can be interrupted (eg. *n jie₁ kerhu meti-lie* "wash your dirty hands"), ellipted (Ø ₁ *meti-wa-te* "(already) washed") etc. *dzū-khou* "well", on the other hand, is a lexical unit by these criteria.

2.3 *Syntactic Insularity*

Compacted structures that they are, compounds are autonomous islands in a syntactic sea. They are self-contained, seamless wholes which lead syntactically insular existences. This is evidenced by the syntactic fact of modification and the syntactic processes of anaphora, identity erasure etc.

2.3.1 *Modification*

Attributive phrases in Angami are either premodifying or post-modifying or both at the same time:

59.	umhi	"eye"
59 a.	n umhi	"your eye(s)"
b.	umhi kezhivi	"beautiful eyes"
c.	n umhi kezhivi	"your beautiful eyes"

Nouns within compounds cannot be modified (60a-b below), but those in phrases can be (61a-b):

60.	mhi -rho -lle eye open imp	"open the eyes"
*60 a.	mhi kezhivi -rho-lle beautiful	"open the beautiful eyes"
*b.	n mhi (kezhivi) -rho-lle	"open your (beautiful) eyes"

But

61.	lesūda hle-lle	"open the books"
61 a.	lesūda kezhivi hle-lle	"open the beautiful books"
b.	n lesūda kezhivi hle-lle	"open your beautiful books"

Postmodification is not reliable evidence because one could argue that the impossibility of 60a is not because the part *mhi* "eye" is closed to modification, but because *mhi-rho* cannot be interrupted. Genitival premodification, however, constitutes ironclad evidence by means of which one may get around this analytical red herring.

2.3.2 *Anaphora*

Compounds are anaphoric islands (Postal 1970). Components of compounds cannot be referred to across syntactic space by anaphors. They can neither control (62a) nor be victims (62b) of pronominalization:

62 a.	a	vo	$dzũ_1$	vũ
	I-nom(=Ø)	go-pst(=Ø)	water	beat-prpsv(=Ø)
	deri	(*sūu$_1$)	mekou	te
	but	(it)	be cold	vr-pst(=Ø)
		(dzūu)		
		(water-the)		

"I went to water$_1$-beat (=swim) but (*it$_1$) was cold"
(the water)

b. dzü mekou te sūla a {dzü-}
water-nom(=Ø) be cold vr therefore I-nom(=Ø) {water}
{*süu$_1$}
{ it}

vù ŋho
beat fting
"the water$_1$ is cold, so I will not {water-} beat(=swim)"
{*it}

Phrases, in contrast, are accessible to pronominalization. (Angami does not backward-pronominalize):

63. a vo dzū$_1$ rülu deri {sūu$_1$} mekou-te
{it}
{dzūu}
{the water}
"I went to water$_1$-pour onto (=bathe) but {it$_1$} was cold"
{the water}

2.3.3 Identity Erasure

Nominal components of compounds cannot be erased under identity with their tokens elsewhere in the sentence. They can neither control nor be victims of equi-NP deletion, both of which are illustrated by each of the starred examples below.

64. a dzü -khou nu vo di dzü -vū
I-nom(=Ø) water place to go-pst(=Ø) cnj water beat-pst(=Ø)
"I went to the water-place (=well) and water-beat (=swam)"

64 *a. a dzū$_1$ khou nu vo-di Ø$_1$-vū
"I went to the water$_1$-place and Ø$_1$-beat"

*b. a Ø$_1$-khou nu vo-di dzū$_1$-vū
"I went to the Ø$_1$ -place and water$_1$-beat"

In contrast, syntagmas are of course accessible to identity erasure. 64 exemplifies forward equi-NP deletion while 66 exemplifies backward equi-NP deletion:

65. a dzū$_1$ huo krie -dI Ø$_1$ rūlu tyo
I-nom(=Ø) water some-acc(=Ø) drink cnj pour onto fut
"I will drink some water$_1$ and Ø$_1$-pour onto (=bathe)"

66. a Ø$_1$ rūlu dI dzū krie tyo
"I will Ø$_1$ pour-onto (=bathe) and drink water$_1$"[16]

2.3.4 *Absence of Contrast in Syntactic Space*

Compounds being closed constructions, their constituents are not available for contrast with other sentential constituents. Sentence 67 choosing *dzū* "water" (instead of *tIso* "Telso (a human proper noun)") from among the parenthesised alternatives is horrendous:

67. a we vo {Telso} vū tyo
I-nom(=Ø) cntrsv go-cnj(=Ø) {tIso}-acc(=Ø) beat fut
(*dzū-)
{water}

derI no vo Khose vū- lie
but you-nom(=Ø) go cnj(=Ø) Khose-acc(=Ø)
beat imp

"As for me, I will go and thrash {Telso} and you go
and thrash Khose" {*-water(=swim)}

A more telling example is the one with the compound verb *mha-dzū* "thing-not to have (be poor)":

68. puo pen dzū derI a
he-nom(=Ø) pen-acc(=Ø) not to have but I-nom(=Ø)

{pen-dzū} dzū
{pen water} not to have-prs(=Ø)
{(=Ink)}
{*mha-}
{thing}
"he does not have a pen, but I don't have {ink}"
{*-thing (=am poor)}

---

16 See however Giridhar (in the works) where the theoretically piquant phenomenon of the "syntactic compound" is elucidated in detail. A paradigm case of the "syntactic compound" is briefly sketched as type 6 on the second approach. Cf 3.2.6.

Note that *mha-dzũ* "be poor" accords well with the conjunctive *mu* "and" rather than with the disjunctive *dert* "but". The compound sentence 68a signifies assonant, complementary conjunction unlike the complex sentence 63 which signifies dissonant, contrastive disjunction:

68 a puo pen dzũ mu a mha-dzũ
"he does not have a pen and I am poor"

The import is that the (referent of the) subject does not have a pen and the speaker does not have the means to lend him one or help him in any way. The conclusion is that *dzũ* of *dzũ-vũ* "water-thrash (=swim)" and *khose* "Khose" of *khose vũ-lie* "beat Khose" in 67 on the one hand, and *pen* of *pen dzũ* "not to have (a) pen", and *mha* of *mha-dzũ* "thing-not to have (=be poor)" in 68a on the other, are not on a par. I am not sure if sentence 69 is a counterexample wherein *mha* of the lexical unit *mha-tsũ* seems to behave on par with the syntactic *n biskits* "your biscuits". Or does it?:

69. a mha -tsũ tyo (dert) no n
I-nom(=Ø) thing eat fut (but) you-nom(=Ø) your
(mu)
(and)
biskits tsũ -lie
biskits eat imp
"I will eat food/a meal (but) you eat your biscuits"
(and)

2.4 *Nonredundancy of Repeated Cooccurrence*

This is a consequence of the previous generalized heuristic principle (2.3), but could in itself be used as a valid test. It states that if the concept expressed by the elements of a sequence has been institutionalized by a language and if the institutionalized sequence and one of the elements cooccur in the sentence, then the native speaker will tend to feel that the iterated cooccurrence is nonredundant. For instance, *mhi-rho* "eye-open (=open eyes)", *mhi-dzũ* "eye-water (=tear)" are compounds because, *inter alia*, the cooccurrence of *umhi* "eye" with the former and *dzũ* "water" with the latter is not felt to be superfluous. Such repeated occurrence does not lead to conceptual redundancy:

70. puo puo mhi mhi -rho
he-nom(=Ø) his eye eye open-pst(=Ø)
"he eye-opened his eyes"

71. puo dzũ kemekou nunu dzũ -vũ
he-nom(=Ø) water cold in water beat-pst(=Ø)
"he beat-water (=swam) in cold waters"

This goes to show that compounds lead an independent insular existence in a syntactic expanse. They do not evince any response to their free-standing cognates in a sentence. In contrast, a sequence like *dzũ rũlu* "water pour-onto; wash (=to have a bath)" fails this test, as in 72:

72. puo dzũ kele se (?dzũ) rũlu -wa -ya
he-nom(=Ø) water hot with (water) wash; vr hab
(Ø 1) pour-onto

"he habitually takes hot baths".

where to retain the second instance of *dzũ* leads to odd redundancy. Sentence 51a, where *dzũ* of *dzũ rũlu* is felicitously retained is not a counter-example because, since *dzũ-le* "hot-water" is a compound, there is no antecedent to trigger the deletion of *dzũ*: *dzũ rũlu* "to bathe" is, therefore, not a compound.[17] This decision is bolstered by the principles of Potential Pause, Noninterruptibility and Ellipsibility, but contradicted by evidence from second language learning (see fn 18). In 73, *dzũ rũlu* is interrupted by *kele (se)*:

73. a dzũ kele (se) rũlu ba
I-nom(=Ø) water hot with pour-onto; wash be(=prg)-prs(=Ø)
"I am taking a hot water bath"

2.5 As is well known, suprasegmentals play a role in defining compoundhood.[18] Stress in English, intonation in Tamil, Kannada play such

---

[17] Kannada, a South Dravidian language, has a (partial) parallel in *nĩ:ru erko/huyko* "water pour onto-rflx (=to bathe)". As in Angami, the nominal *mai-ge* (body-dat) "onto the body" is conventionally, not discourse-contextually, elided. However, with the verb being marked by the reflexive suffix *-ko* in Kannada, the person of the self (the subject) is doubly clear. *nĩ:ru erĩ/huyyĩ* without the reflexive suffix would unambiguously mean "pour water onto someone/something other the person of the (referent) of the subject". Like its equivalent in Angami, *nĩ:ru erko/huyko* is not a compound: the nominal syntagma *nĩ:ru* of this syntagm is modifiable: *bisi* (=hot) *nĩ:ru huyko/erko* "take a hot water bath". Nominal components of compounds are not. The linguistic sequence can be interrupted. There are of course differences. *nĩ:ru* "water" in the Kannada syntagm is a direct object whereas dzũ "water" in the Angami syntagm could be interpreted either as a direct object or as a kind of instrument. Cf 21a.

[18] Originally, I thought evidence from second language learning would constitute a heuristic too. That is, a compound in L1 may possibly make its counterpart in L2 a conceptually unitary

a role. Stress is not distinctive in Angami and intonation has not been investigated here.

2.6 Two of the criteria that define only compounds, Syntactic Insularity and Conceptual Nonredundancy, are necessary and sufficient. Unique Combinability and Supervening Meaning are not necessary but sufficient. Intonation, not looked into here, would probably be a necessary constraint on compounds. Ranking these along a scale, as was done in the case of criteria of general wordhood (the first approach) is pointless because they operate fairly independently. None of them dominates the other(s), because, as a set, they are not locked in an internal clash. Syntactic Insularity, however, would be the most crucial heuristic principle, as it is diagnostic of the lexical wordhood of the compound. In other words, the notion of the compound word in Angami is neatly discrete, despite the marginal kind of indeterminacy indicated in fn 17. Generally, none of the criteria entails the other(s). Conceptual Nonredundancy, for instance, does not entail Supervening Meaning. The *mhi* "eye" of *mhi-rho* "eye-open", as in 70, would not be redundant in the presence of another of its tokens in its syntactic context, but there is no noncompositional meaning that it acquires as a compound. As noted, however, Syntactic Insularity entails Conceptual Nonredundancy. As for the structural attributes that characterize both kinds of lexical units (compound and noncompound), it does not make sense to ask which entails which, i.e. whether the heuristic, for instance, of Interruptibility entails that of Unique Combinability is not *ad rem*. Further, the definition of the compound does not vary in response to linguistic levels, as does general wordhood, as will be seen in Section 3. The criteria of compoundhood in any language essentially concern the compacting of two lexical units into a single structure, and this unitization has nothing to do with levels of linguistic structure. The above heuristics unearth the following structural typology. The arrow pointing rightwards indicates progressive dominance relation and the arrow pointing leftwards regressive

---

whole too. An example was *dzū rūlu* |water pour onto; wash| "to bathe". An Angami was heard saying in the Naga Pidgin

ami	pa:ni	saʃa:	karo	ashe
I-nom	water	wash	do	am

"I am having a bath"

rūlu translating as *saʃa:* "to wash" and *dzū* translating as *pa:ni*. In L2, i.e the Naga Pidgin, it means the unacceptable, facetious "I am washing water". This heuristic test is neither necessary nor sufficient. It does not seem to be a valid heuristic procedure because interference from the mother tongue is not infallibly predictable. It is at best collateral evidence corroborating what has already been established. Further, it is in conflict with the heuristics of Potential Pause, Noninterruptibility and Conceptual Non-redundancy because by all these tests, *dzū rūlu* "to bathe" fails to qualify as a lexical unit. As we shall see, the heuristics of compound wordhood, in general, agree with one another, so the fact that this characteristic is the odd man out in its recalcitrance may not be accidental.

dominance relation. That is, $\overleftarrow{NV}$ for instance, symbolizes an NV compound whose (resultant) external distribution class is V and not N. Similarly with $\overleftarrow{NN}$, $\overrightarrow{VV}$, $\overrightarrow{NV}$, $\overleftarrow{AdvN}$, and $\overleftarrow{VN}$. A possible structural gap is $\overrightarrow{VN}$.

3.0 An alternative view of the linguistic word is one which emerges from the application of different criteria to different layers of the language system, resulting in an asymmetrical notion of wordhood. That is, what is phonologically one word, for instance, may be grammatically two or more words or vice versa. In contrast, the once-an-affix-always-an-affix approach elucidated in Section 1 marks out the boundaries of *only* one kind of wordhood, viz. lexical wordhood, which is normally coextensive with phonological wordhood. The notion of a linguistic unit could vary in response to levels, the defining criteria being different at each level. "Word" seems to be unique in so far as it is the only linguistic category which admits of a multi-level definition. There is, for instance, no phonological morpheme, syntactic morpheme, etc. The word in Angami is a unit which is amenable to such a disjunct definition. As against the one kind of wordhood that the first conjunct approach ends up defining, in the disjunct approach, the six criteria operating at the three levels at which "word" can be defined, suggest a six-way typology of Angami wordhood. As we shall see, the multi-level approach to wordhood is significantly different.

3.1 Angami seems to recognise three levels at which wordhood may be defined. "Word" is discrete rather than squishy at all levels except the phonological level, where whatever is true by the first approach holds good for this approach too. The levels are: (a) the phonological level where the phonological criterion (viz. Potential Pause) and all syntactic criteria except Substitutability operate to define what could be called the "phonological word", which is coterminous with the "lexical word", Matthews' "word-form" (Matthews 1974:26); (b) the intraphrasal level (so called so for want of a better locution) where the syntactic criterion of Mobility applies to define what could be called the "grammatical word"; and (c) the interphrasal (intra-sentential) level where the test of Substitutability operates to define what it is difficult to name.[19] Essentially, if **a(x)** is a structure in a language, and if the relevant criteria operate to establish **a(x)** as an integral linguistic entity, one is speaking at the phonological level, of phonological wordhood; if either **a** or **x** of the phonological word **a(x)** is syntactically mobile, its scope ranging beyond morphology, over the syntactic expanse called the phrase, one is speaking at the intraphrasal level, of the grammatical word (the morphosyntactic word). Typically, these

[19] The expressions "intraphrasal" and "interphrasal" are woefully inelegant, if not entirely inadequate. We will make do with these in the absence of alternative holophrases.

are bound morphemes which are units of syntactic distribution. This is entirely different from Matthews' "grammatical word" (Matthews *op. cit.*:32,33). The "Past Participle of TRY", as in the word-form /traid/, is an example of his "(grammatical) word". Put in terms of a criterion of linguisticity viz. "double articulation", all units of grammatical articulation are "(grammatical) words" in Matthews' sense. They need not be in our sense. Derivational morphological units, for instance, are not. If either **a** or **x** can be isolated to expound either of the two major categorial syntactic constituents, the VP and the NP, one is speaking at the interphrasal level, of the "interphrasal word". Phonological wordhood and grammatical wordhood, however, may converge in a single lexical realisation, intraphrasal mobility then being characterised slightly differently (as seen in 1.0). Phonological words are linear units, while nonphonological words are hierarchical units. Finally English is said to distinguish between phonological, morphological and syntactic words. Thus the form compart-mentalise-ation is said to be one morphological word which contains three phonological words, and the famous Chomsky-Hallean "The-BOOK was-in-an-UNLIKELY PLACE" is phonologically three words (centered around the capitalised forms) but syntactically more. Angami has no phonological word vs. morphological word opposition, so that phonological, morphological and lexical words have identical boundaries in Angami.

3.2.1 What follows is an elucidation of the six-way typology of Angami wordhood.

The first type is a convergence of phonological and interphrasal non-wordhood on the one hand, and intraphrasal wordhood on the other, e.g. *-u*, the singularity marker; *-ko*, the plurality marker; and *-yo*, the diminutive marker, which are all illustrated (at the risk of repetition) in the following utterances:

74 a	lesūda	"book"		lesūda ketɨ	"black book"
b.	lesūda-u	"the book"		lesūda ketɨu	"the black book"
c.	lesūda-ko	"the books"		lesūda ketɨko	"the black books"

75 a	lesūda	"book"	
b.	lesūda-yo	"the little book"	
c.	lesūda ketɨ-yo	"the little black book"	

76 a	hicko	mha	-tsū	-ya	"we eat our meals"
	we(excl.pl)	gnr	eat	hab	

b. hieko mhatsū rükri-ya "we eat our food fast"
fast

c. hieko mhatsū rükri si-ya "we always eat our food fast"

77 a. tsali tshū -lie "sing"
song do imp

b. tsali tshū pevi -lie "sing well"
well

c. tsali tshū pevi se -lie "sing very well"
very

It is clear that *-u*, as in 74, can be suffixed either to the (head) noun *lesūda* "book" or to the verbal participle functioning as an adjective ("black"), its scope ranging over both the noun and the modifying verbal participle; the habitual aspect marker *-ya* could be added, as in 76, either to the verb root or to whatever adverb follows the verb root, its scope ranging both over the verb root and the adverb. They are thus not bound to particular word classes, but are grammatically free enclitics rather than suffixes: their spread is syntactic, not morphological. In contrast, the prefix *pe-* as in the manner adverb *pevi* "well" (*vi* "be good"), the verbal participialiser *ke-* as in *keti* (*ti* "be black") on the one hand, and *a-* as in *apuo* "father", *the-* as in *thevo* "pig" on the other are all phonologically as well as grammatically bound. "Phonologically bound" because of criterial attributes like Potential Pause, Potential Mobility, Isolability and Elliptibility. "Grammatically bound" prefixes rather than grammatically free, mobile proclitics, because they are "immobile" in the sense of being attached to a single morphological distribution class. They are found bound to an invariant form class, the former to the verb (root), and the latter to the noun (root). Whatever the preceding and the following linguistic context might be, *pe-* of *pevi* "well" or *-krū* of *ʃū-krū* "female dog" are related immediately and solely to *vi* and *ʃū* respectively. This difference in constituent structure status and distributional behaviour must be reflected in any description of the Angami word.

3.2.2 The second type of the Angami word, the opposite of the first, is a phonological and interphrasal word but an intraphrasal nonword. This is numerically the largest class. It may be subclassified into simplex and complex subclasses, the parameter being the characteristic of asymmetry between phonological and grammatical wordhood. The simplex type which may be either morphologically simple (subtype a) or complex (i.e. bimorphemic, subtype b) is exemplified by (a) forms which are potential

syntactic constitutes, which, in Bloomfield's classic formulation, are called "minimum free forms", the kind that bulks large in traditional dictionaries; and (b) lexical items which are morphologically complex but one of whose constituent elements is a free form and the other a uniquely dependent form which is neither an inflectional nor a derivational formative: "uniquely dependent" in that it does not occur in combination with any other (free) form in the language. Words of the simplex type are at the same time phonological and grammatical words. That is, there is no asymmetry between phonological (=lexical) wordhood and grammatical wordhood. *galho* "a dish made of rice and greens", for instance, is phonologically as well as grammatically a word, whereas *-u* as in *lesüda-u* "the book" is phonologically a nonword but grammatically a word; *pe-* as in the adverb *pevi* "well" is phonologically and grammatically a nonword. The fifth of the six types of the Angami word exemplifies the last possibility: a phonological word which is a grammatical nonword. Further, it appears that only a form which is phonologically one word and grammatically at least one word (it could be more) can function as a minimum sentence unit.

Examples of the simplex type:

Subtype (a) all lexical morphological units

78. dzü "water"
rüde "Angami month around December"
cie "now"
pfü "measure of length, the distance between the tip of the little finger and the thumb stretched outwards; handspan"
na "mate with; make love to"
puo "one"
sa "have more"

Subtype (b)

79. kewhira "Kohima, the biggest settlement of Angamis in the Angami country"

Explicatory remarks on the single member subtype (b) are in order: *ra* meaning "place, village" is an independent form, as exemplified in phrases like *kelhupilie ra* "Kelhoupielie's village", but *kewhi* is not. It occurs nowhere else in the language either as a bound form or as a free form. The word *kewhira*, of course, shares with examples of subtype (a) the type property of being at the same time a phonological and a grammatical word. The complex subtype exemplar comprises more than one morpheme, and

the number of words it may be said to contain is a function of the level of the linguistic hierarchy (of the phonetic, phonemic, morphological, syntactic or semantic rungs) that it is viewed from. Typically, it is a two-morpheme form. It is a single word at the level of phonology and two words at the level of grammar. This asymmetry between levels of linguistic structure vis-à-vis the the same linguistic unit is what separates the complex subdivision from the simplex one. *puo-vie* "his; hers", for instance, is phonologically one word but grammatically two, because *-vie* is not inseparably attached to a single filler class. Its distribution is syntactic. The following example shows that it may be added not only to the grammatical filler class of nouns but also to that of verbal participles functioning as nominal postmodifiers:

80.	puo kimie	"his wife"
	puo kimie-vie	"his wife's"
	puo kimie kezhivi-vie	"his beautiful wife's"

kezhivi "beautiful" is composed of *ke-*, the participializer, and the verb *zhivi* "be beautiful". Now, this way of treating *-vie* as a grammatically independent and syntactically pertinent form, and hence a (grammatical) word *per se*, its phonological dependence notwithstanding, seems to afford better insight into the NP than saying that *vie* closes the NP, as would be done on the first approach. *-vie* is an immediate constituent not of *kezhivi* which it "leans" on (the well-known etymological meaning of "clitic"), but of the whole phrase. So would be the case with a form like *lesūda-u* "the book", where *-u* is better treated as a grammatically integral word on the basis of its combinatorial or positional freedom, as in phrases like *lesūda ketī-u* "the black book". Specially noteworthy and a distinctively different example is the modal auxiliary *moroshuo* "must". As noted, *mo-ro-shuo* literally means "if something does not happen, it will be bad, dangerous": *mo* "not", *-ro* "if" and *shuo* "be bad, dangerous". This literal meaning is no longer present in the native speaker's consciousness. That the individual components of *moroshuo* have merged, fossilized in a larger, new identity of meaning is evidenced by its capacity to be isolated from the immediate linguistic environment, i.e. the verb root. Contrast 81b with 81d, whose ungrammaticality indicates that *mo-ro vi* is not isolable:

81	a.	a vorū moroshuo	"I must come"
	b.	a moroshuo	"I must"

c. a[1] vor[2] mo[3] -ro[4] vi[5]
l-nom(=Ø) come not cnd be good
"If[4] I[1] don't[3] come[2], it's o.k[5]"

*d. a moro vi "if I don't, it's o.k."

moroshuo is a phonological word. There is no potential pause between its constituent morphemes and it is not interruptible. Contrast 81a and b with 81e where *shuo* is preceded by silence which marks phrasal juncture:

81 e. (1) a mo-ro shuo (-te) "If I am not there, it
(2) a mo-liro shuo (-te) is bad (=I am necessary)"

That *moroshuo* is a gestalt is only half the story. The other half is that *moroshuo* is grammatically two words with *shuo* "be bad" continuing to retain the identity of its (general) external distribution class (as a verb),[20] even as it is frozen as part of another linguistic identity. Consider

81 f. a (vor) moroshuo-di vor "I should have come and I came"

g. a (vor) moroshuo tyo "I should come in the future"

where *shuo* is conjugated for the conjunctive participle by *-di* and is followed by *tyo*, the future tense auxiliary, which are external distribution facts about the class of verbs. All linguistic sequences with inflectional affixes exemplify this subclass of the second type.

The third type is statistically the smallest class. Examples of this type are words at all the three levels at which the Angami word may be defined: the phonological, the intraphrasal, and the interphrasal. These come close to being lexical-morphological units. They are isolable, but not elliptible. Despite being phonological and grammatical words, they cannot be

---

[20] A terminological aside. Nida (*op. cit.*:98 [n 25] equates general external distribution classes with the "so-called parts of speech, e.g. nouns, verbs, adjectives, adverbs and articles". But notably the equation is not complete. For instance, as Nida notes (*ibid.*:99), "a word such as *boys* does not belong to the same [external] distribution class as the simplest member of the noun class". But both *boy* and *boys* are nouns, as is indeed implicit in Nida's statement. This skewness of range is also clear in statements like "(inflectional formations) do not belong to substantially the simplest member of the class in question" (*ibid.*). Inflectional formations do, however, belong to substantially the same part-of-speech category as the simplest member of the class. Interestingly, perhaps stultifyingly, the expressions "external distribution classes" and "simplest member of the class" in the above quotation are referentially distinct. More explicitly, depending on their morphological complexity, members of the class of nouns, for instance, could belong to different external distribution classes.

sentential constitutes. This is because they have no lexical-semantic content. This falsifies the assertion made earlier, i.e. that units which are phonological and grammatical words can function as sentences. *tyo* "will", *lho* "will not", *lira*, the contingency mood marker meaning "in case; in the event of" are examples of this limited membership set.

3.2.4 The fourth type is the diametrical opposite of the third type. Exponents of this type are nonwords at all levels. All derivational affixes fall under this class: *pe-* of *pe-shuo* "badly" (*shuo* "be bad"), *ke-* of *keza* "big" (*za* "be big") illustrate the type.

3.2.5 The fifth type is a phonological word but intraphrasal and interphrasal nonword. Predictably, they cannot perform the "potential sentence" function. Particles *se* "very", *zo* "very" interrogative markers *si* and *ga*, postmodifying verbal participles like *ke-ti* "black" etc., derived adverbs like *pe-vi* "well" etc., coordinators like *mu* "and" and *mori* "or", the disjunctive subordinator *deri* "but", and postpositions like *nunu* "from (nonhuman)", *kinu* "from (human)" illustrate this type.

3.2.6 In a manner of speaking, there is no sixth type, for the exponents of this type are phonological and interphrasal words but intraphrasal nonwords, i.e. Type 2. It is, however, a class apart, since in an extremely interesting paradox, it exemplifies at once word structure and phrase structure. Exponents of this class are "morphosyntactic" words insofar as they are morphological structures which display significant syntactic behaviour. Members of Type 1 are morphosyntactic words too, as we have seen. But exemplars of Type 1 are syntactic in terms of constituent structure and distribution, but syntactic in terms of syntactic (transformational) behaviour. Let us explicate an example of this paradox: *mha-tsü* means "eat a meal". *mha*, which literally means "(material) thing" means "meal" in combination with *tsü* "eat". It is dropped when an object is specified (82b-c below). (Note that the "my" in the English translation cannot be realised in Angami with the verb *mhatsü*):

82.	(a)	a	mha-tsü	tyo	"I will eat my meal"
		I-nom(=Ø)		fut	
	(b)	a thevo-tsü	tsü	tyo	"I will eat pork"
		pig	flesh		
	*(c)	a thevo-tsü	mhatsü	tyo	"I will eat a meal of pork"

There is abundant evidence that *mha-tsü* is a phonological word. There is no potential pause between the two components. It is not interruptible.

D 11. A: no mha -tsū mhai -ya
you-nom(=Ø) gnr eat quickly hab
"you eat meals fast"

B: *u a mha kevi tsū mhai -ya
"yes, I eat good food fast"

Further, the facts (a) that *mha*, the generic pro-form cannot be elicited by a wh- question (D12); and (b) that it cannot be pseudo-clefted (84) constitute unmistakeable evidence that *mha* is not a syntactic object, and that it is indeed a componential part (like *mhi* "eye" of *mhi-rho* "open eyes") of the verb *mha-tsū* "eat a meal":

D 12. A: no kedipuo tsū ga
you-nom(=Ø) what-acc(=Ø) eat-pst(=Ø) Q
"what did you eat?"

B: *(a) a mha -tsū "I ate my meal"
*(b) a mha tsū

83. *a. n mha - tsū - lie "eat your meal"
your gnr eat imp
*b. n mhi - rho - lie "open your eyes"
eye open imp
c. n mhi mhi - rho - lie "open your eyes"
eye eye open imp
d. n galho tsū -lie "eat your galho"

*84. a ketsūlle-u mha "what I ate was a meal"

On the other hand, there is equally persuasive behavioural evidence that *mha* is part of a substitutional expanse of words and not of a distributional framework of morphemes. No nominal part of a compound can undergo syntactic processes like ellipsis, identity erasure or pronominalization (see 2.3.2.). *mha* lends itself to these with felicitous alacrity:

D 13. A: no mha_1 tsū -lie -te me
you-nom(=Ø) gnr eat vr prt Q
"have you eaten food?"

B: Ø$_1$ - tsū - lie - te
"eaten (= yes, I have)".

where *mha* is ellipted as a function of the interaction between intra-*sentential* structure and intersentential context.

85. hieko mha$_1$ - cha
we(excl.pl)-nom(=Ø) gnr cook-pst(=Ø)-cnj(=Ø)

Ø$_1$ -tsū
eat-pst(=Ø)
"we(excl.pl) cooked food$_1$ and ate Ø$_1$".

where *mha* of *mha-tsū* "eat a meal" has been erased under identity with the *mha* of *mha-cha* "cook". One could question this example of identity-erasure, saying that the sentence-final Ø-*tsū* of 85, is in fact *tsū*, and not Ø-*tsū* because *tsū* can stand alone as evidenced in D13B. The answer to this objection is that *tsū* is a transitive verb which takes either a lexically defined and morphologically realised object, e.g. *mha-* (as in D13A) or a syntactic object (as in 83d). Either kind of object can be deleted under identity with a trigger in the discourse (cf. D13B). In other words, a sentence with only *tsū* "eat" in it, without a *mha-* or a syntactic object going with it, can never open a dialogue in Angami. Sentence 85a (which is sentence 85 minus *mha-cha* "cook food"),

85 a a tsū "I ate a meal"

cannot open a dialogue, but since 85 can, one is forced to conclude that 85 is *in situ* a *mha-cha mha-tsū* which is transduced as 85 through an obligatory rule. Consider also:

86. a mha$_1$ - tsū - lie derl puo-e sūu$_1$ cha
I-nom(=Ø) gnr eat vr but he-nom it cook

-sū
vr-pst(=Ø)-prf(=Ø)
"I ate my meal but he had cooked it".

where *mha* of *mha-tsū* "eat a meal" functions as the antecedent for the pronoun *sūu* "it". Verbs like *mhatsū* are exemplars *par excellence* of the possibility in language of units belonging to the morphology-syntax interface.

3.3 It need scarcely be belaboured that the second, disjunct approach describes the phenomenon of the Angami word with more interesting results than the first approach. It has demonstrated that lexical wordhood is not the only kind of linguistic wordhood. Other kinds of word, equally valid, become realities if different criteria apply at the levels where the different kinds of word may be said to have their being. There should, however, be a unifying property that entitles a linguistic entity to be called "word", whatever the level. Mobility is such an overarching property: the Angami word — phonological or intraphrasal or interphrasal — is mobile in some sense. A nonword is not mobile in any sense.[21]

4.0 To perorate, criterial attributes of general wordhood (1.1-1.6) and compounds (2) have been discussed. Two possible theoretical positions with respect to the notion of word, i.e. (a) that of a single uniform set of criteria for all levels of language structure (the "conjunct" difinition (1.7)), and (b) that word divisions could vary in response to linguistic levels, different (sets of) criteria applying at different levels (the "disjunct" definition (3)) were explored. With no heuristic principle proving to be both necessary and sufficient, and the criteria being found at odds with one another, they were hierarchised (1.7) so that the unit they sought to define could be discrete. The attempt met with limited success and in any case could not overcome the serious limitation inherent in the first approach, i.e. the definition of only one kind of wordhood. The second approach, on the contrary, conceived of and defined various kinds of wordhood at three different levels of language structure. The dynamics of wordhood at the phonological level, however, was treated in essentially the same way by the two approaches. We did not require explicit comparison to see that the perspectives opened up by the second approach to the Angami word were more fruitful than what the first approach had to offer.

---

21 This paper for the most part works out a taxonomy rather than explicating the dynamics that lead to the taxonomy. For instance, one could see the distinction between syntactically pertinent inflectional affixes and affixes which are grammatically bound as a function of derivational history. An NP rewrite rule would formulate the grammatical independence or the phrasal pertinence of the affix in the former case, while the category would figure not in an NP rewrite rule but in a relatively later rule (e.g. a segment transformation or spelling rule) in the latter case. Conceivably, syntactically mobile affixes may represent a stage in the linguistic evolution of syntax becoming morphologised. A plausible conjecture is that a language like Angami could develop separate absolute forms for definite and indefinite adjectives, which would be a result of a syntactic category being morphologised. Now the definite article seems to belong both to syntax and morphology. In terms of physical realisation, it is morphological, but in terms of immediate constituency status and combinability it is syntactic.

REFERENCES

Aronoff, Mark. 1976. *Word-Formation in Generative Grammar.* Linguistic Inquiry Monograph Series 1. Cambridge: MIT Press.

Bloomfield, Leonard. 1933. *Language.* [1963 edition] Delhi: Motilal Banarsidass.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Giridhar, P. P. 1980. *Angami Grammar.* Mysore: Central Institute of Indian Languages.

_____ . (in the works) "Noun incorporation in Angami: notes on the syntactic compound". Paper presented to the National Seminar on Morphology, Jan. 6-7, 1984. Hyderabad, India.

Kramsky, Jiri. 1969. *The Word as a Linguistic Unit.* The Hague: Mouton and Co.

Lyons, John. 1968. *Introduction to Theoretical Linguistics.* Cambridge: Cambridge University Press.

Marchand, Hans. 1969. *The Categories and Types of Present-day English Word-formation.* München: C. H. Beck'sche Verlagsbuchhandlung.

Matthews, P. H. 1972. *Inflectional Morphology: a Theoretical Study Based on Aspects of Latin Verb Conjugation.* Cambridge: Cambridge University Press.

_____ . 1974. *Morphology: An Introduction to the Theory of Word-Structure.* Cambridge: Cambridge University Press.

Nida, E. A. 1949. *Morphology.* Ann Arbor: The University of Michigan Press.

Pike, K. L. and E. G. Pike. 1977/1980. *Grammatical Analysis.* Dallas: Summer Institute of Linguistics.

Postal, P. 1969. "Anaphoric islands." *Proceedings of the Fifth Regional Meeting of the Chicago Linguistics Society*. April 1963. Chicago: University of Chicago Press.

Sadock, M.J. 1980. "Noun incorporation in Greenlandic." *Language* 56.2:300-19.

Sekhose, R. 1973. *Spelling and Word Divisions in Angami Naga*. Kohima: Kohima Printing Press.

Stevens, A. M. 1971. "Fixed morpheme order." *Linguistic Inquiry* 2.3, pp. 420-21.