

November, 2001

The Word in Luganda

Larry M. Hyman
University of California, Berkeley
hyman@socrates.berkeley.edu

&

Francis X. Katamba
University of Lancaster
F.Katamba@lancaster.ac.uk

1. Introduction

This paper investigates the word in Luganda, a member of the Narrow Bantu subgroup of the Bantu sub-branch of the Benue-Congo branch of the Niger-Kordofanian language family which is spoken in southern Uganda.¹ The major questions posed by the paper are the following:

- a. Can the “word” be defined?
- b. If not, why not?
- c. If yes, is the “word” a universal?

As a Narrow Bantu language, Luganda is typical with its rich, largely agglutinative morphology. In addition, its extraordinarily complex segmental and tonal phonology make frequent reference to the notion of the word. Luganda provides fertile ground for addressing the above questions.

The need to disentangle different aspects or kinds of words as shown in (1) is generally recognised (Dixon and Aikhenvald, 2000).

- (1) a. the semantic word
- b. the syntactic word
- c. the morphological word
- d. the phonological word
- e. other (orthographic word, cognitive word etc.)

Semantic and morphological criteria for the word are problematic for a number of reasons. First, the amount of semantic material that can be incorporated into a word varies enormously, cross-linguistically. What is said using one word in one language may require a sentence containing many words in another. Compare the English and Luganda utterances in (2).

- (2) a. te- tú- lí- kí- bá- gùl- ir- a
 not we future it them buy applicative infl suffix
- b. ‘we will not buy it for them’

The Luganda utterance in (2a) consists of a single word containing eight morphemes.² Its English translation in (2b) by most accounts would be analyzed as seven (monomorphemic) words.

Second, a morphological definition of the word as a stem plus affixes runs into the problem of words that consist of a proclitic + enclitic (i.e. no stem).

- (3) a. byaa=mû ‘of in there’
- b. waa=ki ‘of what?’

¹ In Guthrie’s widely used classification system of Bantu languages, it is classified as Zone E.15, modified by Tervuren to J.15.

²In these and other examples, an acute accent (á) marks high (H) tone, a grave accent (à) marks low (L) tone, and a circumflex (â) a HL falling tone. Vowels lacking a tone mark are toneless and receive a H or L tone according to the phrasal tonology. Long vowels are transcribed as double throughout this study. The symbol (=) separates proclitics and enclitics from their host. Hyphens, when present, mark morpheme boundaries, though not all internal morphology is marked in the examples.

We return to these words of this type in (11) and (12) below.

Third, syntactic phrases may enter into a paradigm with words (morphological objects), e.g. comparatives of some monosyllabic adjectives appear to block the periphrastic comparative: *bigger*, *smaller* vs. (*)*more big*, (*)*more small* (see Poser 1992, who distinguishes morphological vs. word formation rules).

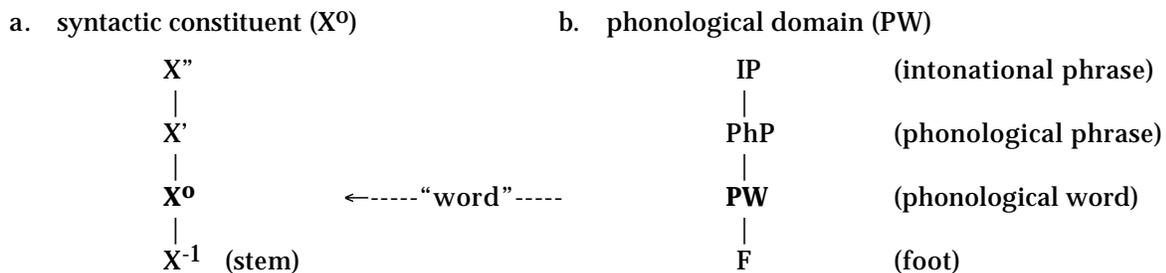
Finally, there is the notorious problem of “phrase words” (Bloomfield 1933) such as *Johnny-come-lately* in English, which constitute a single word (e.g. noun), but which have complex multi-word syntactic structure. As seen in (4),

- (4) “Phrase words” (Bloomfield 1933) such as *Johnny-come-lately*
- “orig. U.S.: (a) a newcomer; (b) = Johnny Raw; (c) fig. and attrib.” (OED Online, 2nd Edition, 1989)
 - ‘But it’s Johnny Comelately, aint it, you?’ said a young mizzen topman. (1839 C. F. BRIGGS Adv. H. Franco I. 249)
 - He may be an old barbarian, but he’s entitled to more consideration than these Johnny-come-lately’s who cruise along the coast after trade. (1924. R. DALY’ Outpost xiv. 139)
 - The Midlands are..all Johnny-come-latelys who coined money out of the war. (1952 E. COXHEAD Play Toward iii. 88)
 - Postwar planning in these United States was no Johnny-come-lately. (1946 M. SHULMAN Zebra Derby iii. 22)
 - The excessive power and renown of many Johnny-come-lately anti-Communists. (1953 Amer. Scholar XXIII. 17)

scholars are not always sure how to treat phrase words, e.g. whether to write them as one word or hyphenate them, whether to indicate their plural with *vs.* without an apostrophe etc. As we shall see below, Luganda is full of such entities, which pose even greater problems than in English.

Although a consistent definition of the word remains elusive, its place in the linguistic hierarchy is, at first blush, reasonably clear. It occurs in an intermediate position within a hierarchy of syntagmemes. Many modern linguists would place it at the position in the hierarchy shown in (5).

- (5) The word occurs in an intermediate position within a hierarchy of syntagmemes



According to Prosodic Domain Theory (Selkirk 1984, Nespor & Vogel 1986), which subjects the phonological hierarchy in (5b) to the Strict Layer Hypothesis (SLH), feet should be grouped into phonological words, which are grouped into phonological phrases, which are then grouped into intonational phrases (and finally into an utterance). This approach encounters problems, however, since, for a variety of reasons, elements do not always occupy the same position in the hierarchy. First, nesting, schematized in (6), is not uncommon, e.g. a PW can potentially consist of more than one PW, with one of the two structures indicated.

- (6) Nesting



Second, although also outlawed by the SLH, recursion, schematized in (7), does occur, due to the phenomenon of rank-shifting, which was illustrated with the Johnny-come-lately example in (4) and to which we return below with Luganda examples such as *mugenzi tázzê* ‘delinquent debtor’ (lit. ‘traveller has not returned’) in (17).

- (7) Recursion/rank-shifting
- | | |
|-----|---|
| PW | N |
| | |
| PhP | S |

Lastly, intersection of domains relevant for hierarchies may be found such that for the purposes of certain phonological processes the relevant domains may overlap or intersect. Thus, in Luganda, the clitic group (CG) and tone group (TG) intersect (Hyman, Katamba & Walusimbi 1987; Hyman 1988), as shown in (8), and cannot therefore be neatly placed in a single phonological hierarchy:

- (8) Intersection/overlapping
- | | |
|----------|----------|
| CG | TG |
| / \ | / \ |
| TG TG | CG CG |

We return to this below in §4.

For the above—and perhaps other reasons—the word in linguistics cannot be characterised in a simple manner. But that is not to say that any attempt to elucidate its position is bound to be fruitless. A possible useful strategy is to seek generalizations on how the various word-like entities function and interact in a total grammar (cf. Zwicky 1990, Dixon & Aikhenvald 2000). This involves, first, determining the criteria that define X^0 and PW in a wide array of languages, and second, comparing word-like entities of a given language to the canonical (or prototypical) word defined by the morphology.

Luganda provides a rich laboratory for studying such questions. In the following sections, we first provide a brief background to Luganda morphology (§2), followed by a discussion of the two phonological properties that have been criterial for establishing word status in the language: quantity (§3) and tone (§4). We return to the questions raised in this section in a brief conclusion in §5.

2. Background: Luganda morphology

We begin by sketching the structure of the canonical morphological word in Luganda, which is very clearly agglutinating. The Luganda canonical morphological word (CMW) is summarized in (9a).

- (9) a. CMW prefix(es) + stem
- b. Noun (augment) (noun class prefix) + stem e.g. o-mu-limi ‘farmer’
 Adj V- CV-/N- mu-limi ‘he’s a farmer’
- c. Verb (prestem) + stem
 Stem root + extensions + FV (FV = inflectional final vowel, usually -a)
- e.g. [a- bá- tà- lì- [lim- ir- agan- a]_{stem}] ‘they who will not cultivate
aug subj neg fut ROOT appl recip FV for each other’

As seen, the CMW consists of one or more prefixes followed by a stem. The typical noun and verb structures are indicated and exemplified in (9b,c).

While these structures are typical, they are not the only possibilities. Thus, as seen in (10), words can appear without prefixes:

- (10) a. noun class 1a Ø- (plural via noun class 2a proclitic ba=)
- | | | | | | |
|-------------|--------|-----|---------|-----------|-----------|
| katonda | ‘God’ | cf. | mu-limi | ‘farmer’ | (class 1) |
| ba= katonda | ‘gods’ | | ba-limi | ‘farmers’ | (class 2) |

lujúùju	‘drunkard’	cf. lu-sózi	‘hill’	(class 11)
ba= lujuuju	‘drunkards’	n-sózi	‘hills’	(class 10)
doodô	‘spinach’ (no. pl.)			
Walúsimbi	(proper name)	ba= Walúsimbi	‘the Walusimbis’	

b. imperative verbs (2nd person singular affirmative without object)

siba	‘tie!’ (2sg.)	cf. n-sibâ	‘tie me!’	
/sib-a/		tó-sibâ	‘don’t tie!’	/te-ó-/ (NEG+2sg.)
FV		mu-sibê	‘tie!’ (2pl.)	

The singular class 1a nouns in (10a) lack a prefix and show that words can be monomorphemic. Second person singular affirmative imperative verbs such as siba ‘tie!’ in (10b) consist of a prefixless stem.

In addition, (11) shows that words exist in Luganda which do not have an obvious root or stem structure:

(11) a. prefix(es) + suffix (?), e.g. demonstrative /-o/ ‘that/those’ (near hearer)

cl. 1	oyo	/o-i-o/	cl. 6	ago	/a-ga-o/	cl. 11	olwo	/o-lu-o/
cl. 2	abo	/a-ba-o/	cl. 7	ekyo	/e-ki-o/	cl. 12	ako	/a-ka-o/
cl. 3	ogwo	/o-gu-o/	cl. 8	ebyo	/e-bi-o/	cl. 13	otwo	/o-tu-o/
cl. 4	egyo	/e-gi-o/	cl. 9	eyo	/e-i-o/	cl. 14	obwo	/o-bu-o/
cl. 5	elyo	/e-li-o/	cl. 10	ezo	/e-zi-o/	cl. 15	okwo	/o-ku-o/

b. proclitic + enclitic

mu=kí	‘in what?’	kyaa=kí	‘it’s for what?’ (‘what’s it for?’)
ku=kí	‘on what?’	byaa=mû	‘those (belonging) inside’
na=kí	‘with what?’	byaa=kô	‘those (belonging) on there’

The demonstratives in (11a) appear to consist of two prefixes, e.g. class 2 a-ba-, followed by a suffix -o ‘this’. The combinations in (11b) consist of a combination of proclitic + enclitic (see §3). Assuming that the forms in (11) have word status, not all morphological words as “canonical,” thus making it difficult to provide a clear, single morphological definition of the word in Luganda.³

Any grammatical definition of the word is further complicated by at least three factors. First, the existence of cliticised forms that, on the one hand, are not autonomous words, but, on the other, enjoy greater freedom than affixes. These clitics show a considerable degree of grammatical diversity, e.g. attaching to hosts from all morphological categories. Examples of nominal proclitics enclitics are shown in (13a) and (13b), respectively (cf. §3 for verbal clitics):

(13) a.	ba= Walusimbi	‘the Walusimbis’	na= mu-limi	‘with a farmer’
	ku= ki-tabo	‘on the book’	byaa= mu-limi	‘those (cl. 8) of the farmer’
	mu= n-nyumba	‘in the house’		
b.	ki-dee =kyo	‘your (sg.) bell’	ki-dee =ki	‘which bell?’
	ki-dee =kye	‘his/her bell’		

Words containing the so-called ‘-a of relationship’ (a.k.a. connective, associative, genitive), which we analyze as a proclitic, are especially problematic and cause problems of word division in the standard orthography which are due to their unclear word status. Ashton et al (1954:104) recommend treating the forms in (14) as single orthographic words.

(14) a.	Owessaza	‘a Ssaza chief’ (a county chief)	<	(o-mwámi) owa= e-ssaza
				chief of county
b.	Oweggombolola	‘a Ggomolola chief’ (a sub-county)	<	(o-mwámi) owa= e-ggombolola

³It is likely that the clitic combinations in (11b) are formed postlexically, i.e. at the syntactic level. Since they escape the lexical morphology, it could be said that they are not morphological words. We would find it difficult to draw the same conclusion concerning the demonstratives in (11a), however.

	chief)	chief of sub-county
c. Oweekitiibwa	'the Honourable' (title)	< (omuntu) owa= e-kitiibwa person of honor

The justification for this is that we have lexicalised here nouns built up on “the possessive noun form but without an expressed antecedent”. Elsewhere, as in (15), Ashton recommends writing the proclitic as a separate word, although phonologically and grammatically it is no different:

- (15) a. wa mukazi 'of the woman' (waa= mukázi)⁴
 b. ba Mulondo 'of Mulondo' (waa= Mulondo)

It seems the distinction is meant to show that lexicalisation has taken place in (14), but not in (15).

Similar problems potentially arise in lexicalized and productively created compounds. Compounds are a special case of words that are built up using phrase structure rules to combine pre-existing words into syntactically complex words. In this respect they resemble syntactic phrases. Examples of compounds are given in (16).

- (16) a. mulwa - kujjula 'woman slow to serve food' (lit. delayer + to serve food)
 katwé - kàsa 'stupid person' (lit. small head + empty)
 nkyá - mùzi 'type of bark-cloth tree' (lit. morning + small root)
 b. mulyá - màtooke 'plantain-eater' (lit. eater + plantains)
 munywá - mwènge 'beer-drinker' (lit. drinker + beer)
 mumwá - mítwê 'head-shaver' (lit. shaver + head)

When compounding is encountered, a question that always arises is how does one distinguish between compounds and syntactic phrases? We will turn to this task shortly. But first we will consider a third category of problematic complex words, namely phrase words, illustrated in (17).

- (17) a. mugenzi - tázzê 'delinquent debtor' (lit. traveler hasn't returned)
 mwáámi - akóoyè 'easy chair' (lit. chief has tired)
 kyáálà - kímípadde 'thief' (lit. fingernail has given me)
 b. Nsí - yàléèta (proper name) (lit. country brought)
 Túlíná - ómùbéèzi (proper name) (lit. we have a helper)
 Sílivá - kùno (proper name) (lit. I will not leave here)
 c. ndábírwáá =kô 'mirror' (lit. I am seen from it)
 kyaa= kulábira =kô 'example' (lit. that to see from)
 d. Wáálábyèè =kí (proper name) (lit. what have you seen yet?)
 gwe - bátákígámbyê (proper name) (lit. one that they haven't said it to)

As seen, these forms appear to be words based on full sentences. As seen, they can have a wide range of internal syntactic structure, e.g. two word combinations of subject-verb or verb-object in (17a,b). Other comparable forms involve clitics, as in (17c,d). Finally, as seen in the glosses, many of these are proper names, some of them derived from proverbs.

Previous literature on Luganda has assumed that phonology provides a number of litmus tests for wordhood:

- (18) a. long vowels are allegedly shortened at the end of a word
 “...the final syllable of a word spoken in isolation is always short.... Within the sentence too, final syllables of words are usually short, and this fact has been of great value in assessing word division.”
 (Tucker 1962:155)

⁴ The fact that the vowel of the proclitic is always long unless it is followed by a word whose first consonant is a geminate is not reflected in the standard orthography. Cf. §3.

- b. a word allegedly may not contain more than one HL tonal sequence
 “...(at most) a single High-Low sequence occurs in every major lexical item in the language.” (Heny 1974:1)

Both vowel length (18a) and tone (18b) have been said to be indicators of word status. In the next section we evaluate the extent to which the claim that the word in Luganda can be defined phonologically is justified.

3. Quantity as a criterion for word status

Over the last 40 years, the literature on Luganda phonology has remarked on a process by which long vowels are shortened in final position. See the quote from Tucker in (18a). In this context the assumption is that “final position” means “at the end of a word”. But this conceals a crucial ambiguity. All researchers are aware that such “words”, postulated to account for final vowel length, may consist of a “full” word (or “host”) and one or more proclitics or enclitics. For instance, commenting on asomyê ‘he has read’, Stevick (1969:4) states: “...the last syllable of the isolated word is short. Before an enclitic, however, it receives the expected two moras....” We will explore this with the data in (19).

(19) Final Vowel Shortening (FVS)

a.	kulábwà	‘to be seen’	/ku-láb-u-a/	kulábwàà	kulábwà
b.	kulábwà	Walúsìmbi	‘to be seen by Walusimbi’	gliding+CL	FVS
c.	kulábwàà	=kô	‘to be seen a little’		

The word-final vowel [a] is phonetically short in (19a) where it is in absolute final position, and in (19b) where it is followed by an object NP, which is not an enclitic. By contrast, its phonological length is preserved in (19c) where it is followed by the locative enclitic =kô.

As (20) shows, proclitics also fail to undergo FVS the length of their final vowel being protected by the host:⁵

(20) a.	-aa= ⁶	‘genitive linker’	:	kikópó kyáá= Wálúsìmbi	‘cup of Walusimbi’	/ki-aa=/
				luggí lwáá= Wálúsìmbi	‘door of Walusimbi’	/lu-aa=/
b.	-ee=	‘subject cleft marker’	:	kikópò kyèè= kyáágwà	‘it’s a cup that fell’	/ki-ee=/
				luggí lwèè= lwáágwà	‘it’s a door that fell’	/lu-ee=/

It is also possible to have a daisy chain of proclitics preceding the host or enclitics following it. In either event, each clitic protects the vowel length of the clitic to its left which, as a consequence, escapes FVS:

(21) a.	two proclitics	:	kikópó kyáá= wáá= Wálúsìmbi	‘the cup of the one of Walusimbi’
b.	two enclitics	:	kikópó =kyèè =kí	‘which cup of his/hers?’
c.	three enclitics	:	yákitééséé =múù =kòò =kí	‘what did he put a little of in?’
d.	one of each	:	kyaa= musíkáá =wange	‘the one of my heir’

It is tempting to hypothesise that FVS is a rule that applies at the end of a clitic group (CG) as depicted in (22).

(22) VV V / __ lCG

Such temptation should be resisted. Rule (22) makes the right prediction in many but not all cases because it overlooks some of the subtleties of the process.

⁵ Note that the cliticisation does not induce lengthening; it only helps preserves it where is already has some reason for being present present. Observe the short vowels of ba= and ku= in (13a), for example.

⁶ Although not relevant for our present purpose, for evidence that some vocalic morphemes should be analyzed as VV and others as V, see Hyman & Katamba 1990.

An adequate account must take on board the distinction between syntactic clitic and phonological clitic (cf. Klavans 1985). The two types of clitics are not identical.

(23) Syntactic clitic phonological clitic (cf. Klavans 1985)

- a. syntactic procliticization of object cleft/relative clause marker (phonological proclitic; vs. (13b))
 (i) kikópò Walúsìmbi kye yalábà 'it's a cup that Walusimbi saw' /ki-e/
 (ii) embwâ Walúsìmbi gye yalábà 'the dog that Walusimbi saw' /gi-e/
- b. syntactic encliticization of emphatic pronouns (phonological enclitic)
 (i) yalyá kô 'he ate IT' (class 12) cf. yalyáá =kô 'he ate a little'
 (ii) yalábwá yê 'he was seen by HIM' yalábwàà =kí 'what was he seen by?'
- c. other particles, e.g. topic marker -o: mbwâ yô 'as for the dog'

In (23a), we observe that the object cleft and relative clause marker /-e/ is syntactically procliticized to the verb, hence translatable as "it's a cup Walusimbi that he saw". Despite this, /ki-e/ is realized as [kye], not as [kyee], which it would be, if it were a phonological proclitic. In the case of the syntactic enclitics in (23b), which must immediately follow the verb, the final length of the latter is not preserved, as it would be if kô, yê etc. were phonological enclitics. The same is true of other short particles such as the topic marker in (23c).

The shortening rule in (22) is inadequate for another reason. As we shall see presently, FVS sometimes applies WITHIN the CG. It is not restricted to CG-final position. A satisfactory characterization of FVS needs to account for this. To do so, Hyman & Katamba (1990) found it essential to treat nominal clitic groups (NCG's) separately from verbal clitic groups (VCG's). Furthermore, it is necessary to recognize that the mode of application of FVS is influenced by the specific SOURCE of the final vowel length. Specifically, the following three sources of final vowel length must be distinguished: a) underlying vowel length; b) monosyllabic vowel length; c) contour tone vowel length. We shall now consider these dimensions of the phenomenon in turn.

In verbs, the length of a final bimoraic (heteromorphemic) syllable which arises when the first of two vowels is glided or deleted, the second being lengthened in compensation, is preserved before an enclitic:

- (24) a. ku-lábwà 'to be seen' ku-lábwàà =kô 'to be seen a little'
 /láb-u-a/ (root-passive-FV) ku-lábwàa =kí 'to be seen by what?'
- b. ku-limya 'to make cultivate' ku-límyàà =kô 'to make cultivate a little'
 /lim-i-a/ (root-causative-FV) ku-límyàa =kí 'to make cultivate what?'

As another source of length, there is a minimality requirement that words belonging to lexical categories must satisfy. Such words must contain at least two moras. So, if a verb has a monosyllabic stem, the one stem syllable must be bimoraic. As seen in (25a), this length is preserved before an enclitic:

- (25) a. ku-mwa 'to shave' ku-mwáá =kô 'to shave a little'
 ku-mwáá =kí 'to shave what?'
- b. -li 'be, have' a-li =mû 'he has in him'
 (a stem?) bá-lí =mû 'they have in them'

The sole exception to this, seen in (25b), is the copula -li, which can be analyzed as either non-lexical or as not being a stem.

The third source of final vowel length comes from the realization of contour tones (Ashton et al 1954:424,452; Tucker 1962:157; Cole 1967:67-68,88; Stevick 1969:6; Hyman 1982:13). Setting aside an utterance-level downstepping phenomenon, Luganda has two surface tones, H(igh) and L(ow), which, conditions being met, can combine to form a HL falling contour tone. (The language does not allow LH rising tones.) Within a word this HL contour (i.e. falling tone) must be realized on two separate moras. In word-final position, if a vowel bears a H tone that H tone is always realised as a HL (falling) contour tone.) Length from final HL contour length

be the pressure exerted by verbal paradigms which are more pervasive and “tight” than nominal paradigms. The isolation forms of verbs are less complete in themselves than I nominal forms occurring on their own.

Nominalisation provides an excellent context in which to observe the asymmetry between nouns and verbs with respect to FVS. FVS will apply as it does to nouns if a verb is nominalized (including in a phrase word) as (31) shows:

- | | | | | | |
|---------|-------------------------------------|---------------------|-----|--------------------------|------------------------------|
| (31) a. | kí-kòlwa | ‘deed’ | | kí-kòlwa =kí | ‘which deed?’ |
| | < /kí-kól-u-a/ | ‘it is done’ | cf. | kí-kòlwaa =kí | ‘it is done by what?’ |
| b. | muzáddè - t-á-lyà | (proper name) | | muzáddè - t-á-lyà = kí | ‘which M-T.?’ |
| | (lit. a parent doesn’t eat) | | | muzáddè - t-á-lyà = wè | ‘his M-T.’ (*...lyàà =wè) |
| c. | mugenzi - tázzè | ‘delinquent debtor’ | | mugenzi - tázzè =kí | ‘which delinquent debtor?’ |
| | (lit. traveler hasn’t returned) | | | mugenzi - tázzé(é) =wè | ‘his delinquent debtor’ |
| d. | nnámpá - wè - n-gwá | ‘a neutral person’ | | nnámpá - wè - n-gwá = wè | ‘his N.’ (? ... n-gwáá = wè) |
| | (lit. there is nowhere that I fall) | | ~ | nnámpá - wé - n-gwá = wè | (= with H tone plateauing) |

Finally, as the final element of a compound or phrase word, monosyllabic stems undergo FVS optionally if the phrase word is a common noun and obligatorily if it is a proper name, as seen in (32).

- | | | | | | | |
|---------|--------------------------|---------------|-----------------------|---------------|-----------------|---------------|
| (32) a. | mutunda - bide | ‘bell-seller’ | mutunda - bide(e) =ki | ‘which...’ | cf. bi-dee =ki | ‘which bells’ |
| | | | mutunda - bidé(é) =wè | ‘his...’ | cf. bi-déé =byè | ‘his bells’ |
| b. | akisá - ènte | (proper name) | akisá - ènte =ki | ‘which A-E.?’ | cf. n-tee =ki | ‘which cow’ |
| | (lit. he who hides cows) | | akisá - ènté =wè | ‘his A-E.’ | cf. n-téé =wè | ‘his cow’ |

A possible interpretation of the variation in the realisation of vowel length is that speakers are unclear in (32a) whether the enclitic cliticises to preceding PW (N₃), as in (33a), or to the whole syntactic word X⁰ (N₁), as in (33b).

- | | | | |
|---------|---|----|--|
| (33) a. | $\begin{array}{c} N_1 \\ \swarrow \quad \searrow \\ N_2 \quad N_3 = ki \\ \text{mutunda - bidee} \end{array}$ | b. | $\begin{array}{c} N_1 = ki \\ \swarrow \quad \searrow \\ N_2 \quad N_3 \\ \text{mutunda - bide} \end{array}$ |
|---------|---|----|--|

We have shown in this section the final vowel length *per se* is not a reliable indicator of word status. In the next section we turn to another indicator, that has been proposed, namely tone, and see whether it fares any better.

4. Tone as a criterion for the word

We have already referred to Heny (1974) who proposes the tonal criterion of one pitch drop per lexical item to characterise the tone word (TW). In (34) a fuller set of tonal criteria for word determining word status is provided:

- (34) a. at most one HL pitch drop per TW
 b. mapping of phrasal %LH% boundary tones to toneless TWs
 c. function of TW in H tone plateauing (HTP)
 d. one overall “tonal configuration” (e.g. in verb reduplications—see (41))

In the following paragraphs we elaborate on these criteria.

Morphological words (MWs) can have one drop from H to L, as in (35a), or none (in which case they are lexically toneless), as in (35b).

(35) a.	ki-bê	'jackal'	ki-sásilo	'rubbish'	ki-jũiko	'spoon'
	ki-kópò	'cup'	ki-yulífù	'torn' (cl.7)	ki-bíinã	'society'
	ki-sikî	'log'	ki-wójjólò	'butterfly'	ki-wúúgúlù	'owl'
	ki-lókwâ	'weed'	ki-bónèlezo	'punishment'	ki-sàànikizo	'cover, lid'
	kí-kòlwa	'deed'	ki-begábèga	'shoulder'	ki-sáákáátè	'reed fence'
b.	ki-de	'bell'	ki-tabo	'book'	ki-tooke	'plantain tree'
	ki-bya	'bowl'	ki-muli	'flower'	ki-seenge	'room'
	ki-lagiro	'command'	ki-biiliti	'match(-box)'	ki-sanilizo	'comb'
	ki-papajjo	'branch'	ki-janjaalo	'bean'	ki-sumuluzo	'key'

MWs are marked for the tones with which they exit the lexical (word-level) phonology: (á) = H, (à) = L, (a) = toneless. Toneless lexical items may acquire tone postlexically.

Multiword forms are marked with the tones they carry after the application of H tone plateauing (HTP) e.g. in (41) below. Toneless moras acquire a H or L according by operation of rules at the phrase level. A major rule applying at the phrasal ensures that at the left edge of a phrase, a toneless word is realized L-Hⁿ (%L on the first mora and H% on remaining moras), as seen in (36).

(36) a.	kì-dé	'bell'	kì-tábó	'book'	kì-tóóké	'plantain tree'
	kì-byá	'bowl'	kì-múlí	'flower'	kì-sééngé	'room'
b.	kì-lágíró	'command'	kì-bíílítí	'match(-box)'	kì-sánílízó	'comb'
	kì-pápájjó	'branch'	kì-jánjááló	'bean'	kì-súmúlúzó	'key'
c.	%L	H%	%L	H%	%L	H%
		△		△		△
	kì-	pápájjó	kì-	jánjááló	kì-	súmúlúzó

By this criterion, the augment (a.k.a. initial vowel), here below in (37) e-, is a prefix for when it appears in a lexically toneless word, a low tone falls on the initial syllable and any remaining syllables receive high tone.

(37) a.	è-kí-dé	'bell'	è-kí-tábó	'book'	è-kí-tóóké	'plantain tree'
	è-kí-byá	'bowl'	è-kí-múlí	'flower'	è-kí-sééngé	'room'
b.	è-kí-lágíró	'command'	è-kí-bíílítí	'match(-box)'	è-kí-sánílízó	'comb'
	è-kí-pápájjó	'branch'	è-kí-jánjááló	'bean'	è-kí-súmúlúzó	'key'

By the same criterion, the preposed elements in (10a.i.) are proclitics since we do not get the pattern of L tone on the first syllable and H on the rest where the word belonging to a lexical category is lexically toneless.⁷

(38) a.	class 2a ba=			
	bà= kàtòndá	'gods'	bà= Mùkásá	'the Mukasas'
b.	locatives ku= (class 17) and mu= (class 18)			
	kù= kì-dé	'on the bell'	mù= kì-dé	'in the bell'
	kù= kì-tábó	'on the book'	mù= kì-tábó	'in the book'
	kù= kì-sánízízó	'on the comb'	mù= kì-sánízízó	'in the comb'

⁷We have two other tonal criteria which yield the same result: (i) tone retraction (ki-sikî vs. e-ki-sikî 'log', with retraction, vs. na= kisikî 'with a log', without retraction); (ii) Meeussen's Rule (H-H H-L), e.g. a-láb-à 'he sees' vs. tú-làb-à 'we see'; cf. né= Kígùndu 'with Kigundu', where né= is the [+augment] form of na=.

- c. na= ‘with, and’ (comitative, instrumental, associative)

nà= kì-dé	‘with a bell’	nà= Kàtón dá	‘with God’
nà= kì-tá bó	‘with a book’	nà= Mùkásá	‘with Mukasa’
nà= kì-sánízízó	‘with a comb’		
- d. genitive linker -aa; here: byaa= ‘those of’ (class 8)

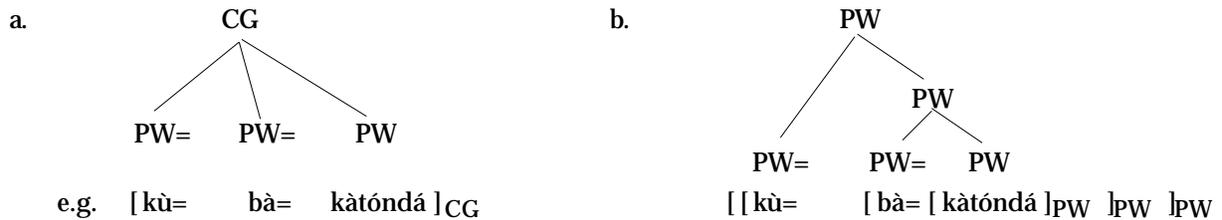
byàà= mù-ntú	‘those of a person’	byàà= Kàtón dá	‘those of God’
byàà= mù-límí	‘those of a farmer’	byàà= Mùkásá	‘those of Mukasa’
byàà= mù-lámúzí	‘those of a judge’		(i.e. ‘Mukasa’s’)

An initial string of proclitics will all remain L (plus one more L on the first s of the noun), as seen in (39), indicating again that we are not dealing with a single lexically toneless MW.

- (39) a. kù= bà= kàtón dá ‘on the gods’ nà= bà= kàtón dá ‘with the gods’
 mù= bà= kàtón dá ‘in the gods’ byàà= bà= kàtón dá ‘those of the gods’
- b. nà= kù= bà= kàtón dá ‘and on the gods’
 nà= byàà bà= kàtón dá ‘with those of the gods’
 byàà= kù= bà= kàtón dá ‘those on the gods’
 byàà= wàà= bà= kàtón dá ‘those of the one of the gods’
- c. nà= byàà= kù= bà= kàtón dá ‘with those on the gods’
 nà= byàà= wàà= bà= kàtón dá ‘with those of the one of the gods’

Two reasonable representations can be proposed to account for one or more proclitics combining in a CG with toneless MWs:

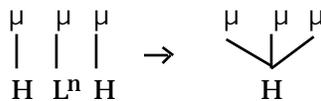
(40) Two reasonable proposed structures to account for proclitic(s) + toneless MWs



We opt for (40b) which reflects the cyclic nature of the attachment of proclitics in (39).

We shall now address the question of whether “at most one HL pitch drop” is an adequate criterion for identifying TWs. The answer will be in the negative. This is because of the extensive use of the rule of High Tone Plateauing (HTP) depicted in (41), whose effect is exemplified in (42).

(41) High Tone Plateauing (HTP)



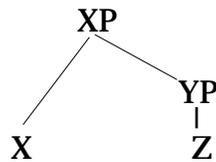
(42) a. Noun + possessive

ki-bê	+	kyaa= Walúsìmbi	ki-bé	kyáá= Wálúsìmbi	‘Walusimbi’s jackal’
ki-kópò	+	kyaa= Walúsìmbi	ki-kópó	kyáá= Wálúsìmbi	‘Walusimbi’s cup’
ki-sikí	+	kyaa= Walúsìmbi	ki-sikí	kyáá= Wálúsìmbi	‘Walusimbi’s log’
ki-lókwâ	+	kyaa= Walúsìmbi	ki-lókwá	kyáá= Wálúsìmbi	‘Walusimbi’s weed’
kí-kòlwa	+	kyaa= Walúsìmbi	kí-kólwá	kyáá= Wálúsìmbi	‘Walusimbi’s deed’
H <u>L</u>		HL	H-----H	L	

b. (Affirmative) verb + following “word”

twáá-làbà	+	Walúsìmbi	twáá-lábá Wálúsìmbi	‘we saw Walusimbi’
twáá-làbwà	+	Walúsìmbi	twáá-lábwá Wálúsìmbi	‘we were seen by Walsimbi’
twáá-gèndà	+	tútùtu	twáá-géndá tútùtu	‘we went slowly’
twáá-gèndà	+	lulí	wáá-géndá lulí	‘we went day before yesterday’
twáá-làbà	+	ki-kópò ki-nénè	twáá-lábá kí-kópò kinénè	‘we saw a big cup’
H L L		HL HL	H-----HL HL	

What the data in (42) show is that there is “at most one HL pitch drop” per tone group (TG). As seen in (42a), a N + gen=N constitutes a single TG. Similarly, in (42b), in most affirmative tenses, a verb will form a TG with the PW that follows it. We propose to define the TG as shown in (43).

(43) Definition of the Tone Group (TG): $X+Z$ (see Hyman & Katamba 1990/91, 1993)

- where: (i) X negative, imperative or infinitive of verb
(ii) Z = PW
(iii) Z does not begin with an augment

The role of syntactic information in defining TGs is crucial. As well as requiring the presence of appropriate tonal properties for a string to count as a tone group, it must also meet the syntactic characteristics specified in (43). The postlexical application of HTP is restricted to a TG. Some constraints on its application will now be considered as a way of illustrating TGs.

As seen in the following examples, HTP may also apply to compounds, whether lexicalized, as in (44a), or productively created, as in (44b).

(44) a.	mwásà + jjutè	mwásá - jjutè	‘hard, uncomfortable chair’ (lit. boil-breaker)
	HL HL	H----HL	
b.	mu-témà + bi-sikí	mu-témá - bisikí	‘log-chopper’ (chopper + logs)
	HL HL	H-----HL	

We now show that HTP is a domain-juncture rule (Selkirk 1984, Nespor & Vogel 1986). Plateauing is not permitted where a toneless word is intercalated between words with the requisite H tones as seen in (45).

(45) a.	twáá-làbà	ki-tabo	kyaa= Walúsìmbi	‘we saw Walusimbi’s book’
	[H L L]	[]	[HL]	
b.	twáá-làbà	mugenzi - tázzè		‘we saw the delinquent debtor’
	[H L L]	[]	[H HL]	
c.	mukúbà - balimi	waa= Walúsìmbi		‘Walusimbi’s farmer-beater’
	[HL]	[]	[HL]	

Our interpretation is that the intervening toneless forms in (45) are PWs, which are visible to HTP. Now compare comparable situations that arise when the intervening form is a clitic. As seen in (46), HTP is obligatory when the toneless words are proclitics which are not visible to HTP.

(46) HTP is, however, obligatory when the toneless “words” are proclitics (which are not visible to HTP— which seems to argue for the structure in (32b))

a.	twáá-làbà + byaa= Walúsìmbi	twáá-lábá byáá= Wálúsìmbi	‘we saw Walusimbi’s’
	H L L HL	H-----HL	
b.	twáá-gèndà + na= Walúsìmbi	twáá-géndá ná= Wálúsìmbi	‘we went with Walusimbi’
	H L L HL	H-----HL	
c.	twáá-géndá ná= byáá= kú= bá=Wálúsìmbi		‘we went with those on the Walusimbis’
	H-----HL		

Our conclusion is that such clitics are not PWs, rather join their host to define a complex PW. This fact, plus the mapping of left-boundary %L onto each proclitic in (39), seems to argue for the branching structure in (40b), though with the proclitics not identified as PWs.

Returning to the syntactic conditions, we find that HTP will not apply where X does not c-command Z. Thus, for example, because a subject and the verb that follows it are not in a c-command relationship, a subject-verb sequence (including phrase-words as in (47b)) will never constitute a TG. Hence they are not a viable domain for HTP:

- (47) a. Walúsímbi ya-géndà ‘Walusimbi went’ *Walúsímbi yá-géndà
 HL H L
- b. mwámi - akóoyè ‘easy-chair’ (lit. chief has tired) *mwámi - ákóoyè
 H L H L

Most of what has been said thus far concerning the identification of TGs and the application of HTP is consistently adhered to by all speakers all the time. However, we can induce situations where there will be uncertainty in how to resolve conflicts—and consequent variation. While the subject-verb sequences in (47) clearly cannot, for syntactic reasons, undergo HTP in and of themselves, when such forms are modified by a possessive enclitic, we find variations such as in (48).

- (48) A possessive enclitic (or noun) can optionally induce HTP on Subj-Verb phrase words (some speakers)
- a. mwámi - akóoyè mwámi - akóoyè = yàngè ‘my easy chair’
 H L H L ~ mwámí - ákóoyè = yàngè
- b. kyáálà - kímpáddè kyáálà - kímpáddè = wàngè ‘my thief’ (< ‘fingernail has given me’)
 H L H L ~ kyáálà - kímpáddè = wàngè
- c. muzáddè - tágúlwá muzáddè - tágúlwá = wàngè ‘my M.-T.’ (< ‘parent cannot be
 H L H H HL ~ muzáddé - tágúlwá = wàngè (proper name) bought’)
- d. lúmbè - musolô lúmbè - musoló(ó) = wàngè ‘my L.-M.’ (< ‘death is a tax’)
 H L HL ~ lúmbé - músoló(ó) = wàngè (proper name)

As seen, a possessive enclitic (or gen=noun) can optionally induce HTP on Subj-Verb phrase words for some speakers. This is because the possessive highlights the rank-shifted nature of such phrase words, allow these speakers to ignore their internal structure, at least optionally. Thus the c-command constraint on the formation of TGs is relaxed in just such cases.

Let us now compare the TG with the CG established on the basis of FVS in §3. The Strict layer Hypothesis is premised on the assumption that each layer of the prosodic hierarchy in (5) neatly and uniquely consists of elements at the next lower layer. This is not always so. In Luganda there is sometimes a conflict between the TG and the CG, which intersect or overlap (Hyman, Katamba & Walusimbi 1987; Hyman 1988), making it extremely problematic to use them in those situations as clear indicators of wordhood as depicted in the hierarchy in (5).

Let us begin by disposing of straightforward cases like (49) where there is isomorphism between TGs and CGs:

- (49) a. one TG, one CG: tú-lyáá =kô ‘we eat a little’ (cf. tú-lyà ‘we eat’)
 H-----HL (“Z” = enclitic) H L
- b. two TGs, two CGs: te-tú-ly-à mu-púunga ‘we don’t eat rice’ (two TGs because verb = negative)
 H L HL

(40)

Intersecting “Words” and “Groups” in Luganda

	TW	PW	TG	CG
proclitic + noun	1	1	1	1
affirm. verb + kô (~ mû)	2	1	1	1
(i) noun =poss. pronoun (ii) stem reduplications	1	2	1	1
neg. verb =kô (~ mû)	2	1	2	1
irreg. noun compounds (51b)	1	2	1	2
adjective =kô (~ mû)	2	2	1 ~ 2	1
noun =kí ‘which’	2	2	2	1
(i) affirm. verb + noun (ii) noun + poss. noun (iii) noun compounds (44)	2	2	1	2
other “word” + “word”	2	2	2	2

Reduplication is the final arena of the intersection of phonological domains that we will consider. It is also an area where often the application of morphological processes is circumscribed by prosodic parameters. Reduplication is highly productive in Luganda. Here we restrict ourselves to verbal reduplication, which is total.⁹ For noun-and adjective reduplication, see Hyman & Katamba 1990).

Verbal reduplication has a variety of uses including signalling an action done frequently, or ‘here and there’ or without real commitment. Here it is the phonological properties of reduplication which have a bearing on word recognition criteria that will be treated. We will begin by considering the realisation of word-final underlying length. As seen in (53), untypically of length from this source, in a verb the underlying length of the base (first part) of a reduplicated form is not preserved. By contrast, the expected length survives in the reduplicant if it is followed by an enclitic.

- (53) a. ku-sasulwa-sasulwaa =kô ‘to be paid a bit’ < ku-sasulwa ‘to be paid’
 b. ku-wúlirwa-wulirwaa =kô ‘to be heard a bit’ < ku-wúlirwa ‘to be heard’

Where the final length is induced by a contour tone that is realised on the surfaces as in the reduplicant in (54a), length is saved before an enclitic. But where the final HL contour fails to surface, the length associated with it also perishes with it as in (54b).

- (54) a. mu-wulile-wulilèè =kô ‘(you pl.) hear a bit!’ < mu-wulilè ‘(you pl.) hear!’
 mu-labilile-labililèè =kô ‘(you pl.) look after a bit!’ < mu-labililè ‘(you pl.) look after!’
 b. a-wúlílá-wúlílá =kô ‘he who hears a bit’ < a-wúlílâ ‘he who hears’
 a-lábilílá-lábilílá =kô ‘he who looks after a bit’ < a-lábilílâ ‘he who looks after’

As seen in (55), monosyllabic length is preserved in both parts of the reduplicated verb. This seems to be motivated by the requirement for words belonging lexical category to have stems that are at least bimoraic:

- (55) a. ku-mwaa-mwaa =kô ‘to shave a bit’ < ku-mwa ‘to shave’
 b. ku-lyáá-lyàà =kô ‘to eat a bit’ < ku-lyâ ‘to eat’

Bimoraic CVCV stems require an iambic base and hence become CVCVV in reduplication. Thus, the FV of the base is lengthened in (56a) to assure the iambic structure. Where there is underlying length as in (56b), it is

⁹ It is hence difficult to determine which is the base and which is the reduplicant. We will assume a base-reduplicant structure.

also preserved, but not because of the input, here /bal-u-a/, but because of the iambic condition on bisyllabic bases.

- (56) a. ku-balaa-bala =kô 'to count a bit' < ku-bala 'to count'
 b. ku-balwaa-balwa =kô 'to be counted a bit' < ku-balwa 'to be counted'

Finally, let us consider the forms in (57), which consist of non-tonic compounds followed by a possessive enclitic:

- (57) a. mu-gemera + wala 'gun' mu-gemera - walá =gwàngè 'my gun'
 (lit. preventer (from) far) ~ mu-gemérà - wàlà =gwàngè
 b. mu-tunda + bitabo 'book-seller' mutunda - bitabó =wàngè 'my book-seller'
 (lit. seller + book) ~ mutúndà - bitàbò =wàngè
 c. n-tabaaaza + bakadde 'beer' ntabaaaza - bakáddè =yàngè 'my beer'
 (lit. I make old people go to war) ~ ntabáázà - bàkàddè =yàngè

As in the forms in (51), the possessive 'my' shifts its H to the second mora of the stem in the preceding compound. But which stem? There are two in each case: the stem of the immediately preceding word, or the stem of the first word of the compound. While the first form is preferred in each case, i.e. where the H goes on the second mora of the preceding stem, speakers also frequently accept to place the H on the second mora of the first stem, as indicated.

Our interpretation of this variation is shown in (58).

- (58) a.
$$\begin{array}{c} N_1 \\ \swarrow \quad \searrow \\ N_2 \quad N_3 = gwàngè \\ \text{mugemera} - \text{walá} \end{array}$$
 b.
$$\begin{array}{c} N_1 = gwàngè \\ \swarrow \quad \searrow \\ N_2 \quad N_3 \\ \text{mugemérà} - \text{wàlà} \end{array}$$

In our view, speakers are unclear in (57a) whether the enclitic cliticizes to the preceding PW (N_3), as in (58a), or to the X^0 (N_1) as a whole, as in (58b). The difference is expressed via the landing sight of the H of the possessive enclitic.

5. Conclusion

Let us recall the workshop questions that were posed at the beginning of §1.

- a. Can the word be defined?
- b. If not, why not?
- c. If yes, is the word a universal?

When the Luganda facts are considered in the context of these questions, the answers that emerge are not simple. It is obvious that the word cannot be uniquely defined in any of the senses commonly recognised by linguists using a consistent set of criteria. Why should that be so? In our view, this is because the different criteria conflict in at least two ways:

First, there are conflicts between the different components of the word (morphology, syntax, phonology).

Second, there are conflicts even within the same component, e.g. with regard to phonological criteria vowel length conflicts with tone.; tone also conflicts with itself.

Although intuitively it makes sense to recognise words as key building blocks of language, finding consistent and reliable ways of characterising words in one language, let alone cross-linguistically is Sisyphian task. The linguist's desire to categorise and compartmentalize and to label entities neatly is frustrated because words are amorphous entities in the sense that probably there is no one point where all relevant information about a word is packaged together. While the result at any one time is a partial view of a changing situation, one needn't

fret. The excitement of such a study as the one in which we have been engaged in Luganda for a number of years has taught us much about how the different concerns of a language interact and ultimately conflict.

In producing this study, we realized how central the word has been in our previous work on the morphology-phonology and syntax-phonology interfaces in Luganda. While we cannot define, we can delimit, and we can also advance hypotheses as to why these complexities and contradictions exist. We suspect that rather than being different from other aspects of language, the problems set out for us by the workshop organizers could be applied to most any aspect of language: the sentence, the syllable, etc. Even the morpheme and the phoneme, which are presented as if easily defined, are not exempt from definitional and analytical problems. *Raison de plus to keep at it.*

References

(a) Publications of Hyman & Katamba relevant to the Luganda word

- Hyman, Larry M. 1982. Globality and the accentual analysis of Luganda tone. *Journal of Linguistic Research* 1.1-40.
- Hyman, Larry M. 1988. Direct vs. indirect syntactic conditioning of phonological rules. *Proceedings of E.S.C.O.L. 1987. Ohio State University.* 147-163.
- Hyman, Larry M. & Francis X. Katamba. 1990. Final vowel shortening in Luganda. *Studies in African Linguistics* 21.1-59.
- Hyman, Larry M. & Francis X. Katamba. 1990-91. The augment in Luganda tonology. *Journal of African Languages and Linguistics* . 12.1-45.
- Hyman, Larry M. & Francis X. Katamba. 1993a. A new approach to tone in Luganda. *Language* 69.34-67.
- Hyman, Larry M. & Francis X. Katamba. 1993b. The augment in Luganda: syntax or pragmatics?. In Sam Mchombo (ed.), *Theoretical aspects of Bantu grammar*, 209-256. Stanford: C.S.L.I.
- Hyman, Larry M. & Francis X. Katamba. 1999. "The syllable in Luganda phonology and morphology". In Harry van der Hulst & Nancy Ritter (eds), *The syllable: views and facts*. Berlin: Mouton de Gruyter.
- Hyman, Larry M., Francis X. Katamba & Livingstone Walusimbi. 1987. Luganda and the strict layer hypothesis. *Phonology (Yearbook)* 4.87-108.
- Katamba, Francis X. 1978. How agglutinating is Bantu morphology? *Linguistics* 210.77-84.
- Katamba, Francis X. & Larry M. Hyman. 1991. Nasality and morpheme structure constraints in Luganda. *Africanistische Arbeitspapiere* 25.175-211.

(b) Other

- Ashton, E.O. et al. 1954. *A Luganda grammar*. London: Longmans, Green and Co.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Cole, Desmond T. 1967. *Some features of Ganda linguistic structure*. Johannesburg: Witwatersrand University Press.
- Dixon, R.M.W. & Alexandra Y. Aikhenvald. 2000. Position paper for an International Workshop to be held at La Trobe University, 8-13, August 2000, on 'The status of "word": its phonological, grammatical, cultural and cognitive basis.'
- Heny, Frank. 1974. Ganda tone as "pitch accent". Ms. University of Massachusetts, Amherst.
- Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61.95-120.
- Poser, William J. 1992. Blocking of phrasal constructions by lexical items. In Ivan A. Sag & Anna Szabolsci (eds), *Lexical matters*, 111-130. Stanford: CSLI.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Selkirk, Elisaberth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.
- Tucker, A.N. 1962. The syllable in Luganda: A prosodic account. *Journal of African Languages* 1.122-166.
- Stevick, E. W. 1969. Pitch and duration in Ganda. *Journal of African Languages* 8.1-28.
- Zwicky, Arnold M. 1990. Syntactic words and morphological words, simple and composite. *Yearbook of Morphology* 3.201-216.

in kyaa= ku-lábira =kóó =kyaange, can =kóó be short?

Wálábyée =kí(í) =wange ‘my W.’

(25) Need for nested PW’s, e.g. in reduplications

a. noun reduplication (marks diminutive, often pejorative)

- | | | | | |
|-------|----------|----------|--------------------|--------|
| (i) | kí-kòlwa | ‘deed’ | kí-kólwá-kòlwà =kí | ‘which |
| (ii) | ki-de | ‘bell’ | ki-déé-dè-dè =kí | |
| (iii) | ki-sikî | ‘log’ | ki-sikî-sikî =kí | |
| (iv) | ki-bê | ‘jackal’ | ki-béé-bé-béè =kí | |

b. verb reduplication

c. no evidence that pro=MW=encl is a phonological constituent (see next section)