The phonology and grammar of Galo "words"

Mark Post
Research Centre for Linguistic Typology, La Trobe University

Abstract: "Words" may be independently defined and identified in Galo (Tibeto-Burman > Tani) in terms of relatively consistent and functionally well-motivated sets of phonological and grammatical criteria. However, these criteria very often fail to converge upon identification of the same formal unit; instead, we frequently find grammatical "words" which consist of two phonological "words", and phonological "words" which consist of two grammatical "words", etc. The resulting "mismatch" between "phonological words" and "grammatical words" in Galo is argued to be theoretically non-trivial, in that its existence is capable of explaining a variety of otherwise seemingly disparate facts in the synchronic and diachronic organization of Galo grammar. The facts from Galo thus support a view of language in which "word" is independently defined in phonological and grammatical terms, and in which neither type of "word" necessarily corresponds to (or is projected by) the other. Although there might be said to exist a very generalized functional pressure towards "unification" of "phonological words" and "grammatical words", such a pressure would not be expressible as a formal constraint on language grammar.

Keywords: word, syntax/prosody interface, Tibeto-Burman languages, Tani languages

1. Introduction

It is well-known that the concept "word" – taken as a generalized concept, not restricted to any particular linguistic subdomain – is far from unproblematic. Research into word prosody and word morphology has uncovered numerous ways in which particular types of unit may be more or less "wordlike" than others; similarly, it may be difficult in a variety of cases to consistently determine the boundary between one "word" and another.[1] A prominent descriptive strategy of recent decades has been to describe units which appear relatively "wordlike" in some respects, but relatively less "wordlike" in other respects, as "clitics".[2]

Despite these and related qualifications, some commonly-held assumptions remain which are seemingly not limited to any particular theoretical or methodological orientation or tradition – first, that a unit "word" may be identified, at least in some "prototypical" sense, in most if not all languages of the world, and, second (and more important for our purposes here), that such a unit will display a *fundamental unity* in both phonological and grammatical senses; that is to say, while one may speak of the

[1] The number of works in the literature which directly address questions of the nature, identification and properties of "words", whether in particular languages or more generally, is very large and would certainly include all major studies of linguistic morphology, most major studies of phonology, and most if not all large-scale descriptive grammars. Thus it would be pointless to provide a handful of citations here when the list could just as easily include a hundred or more works of equal relevance. That said, a survey of characteristic approaches to the definition of "word" is provided in Dixon and Aikhenvald (2002).
[2] Here too, the literature on clitics is vast (Nevis, Joseph et al. 1994), though a recent general survey may be found in Aikhenvald (2002).

phonological and grammatical properties of a "word" in independent terms, the overriding assumption is that these various properties will ultimately converge upon the same type of unit or set of forms (again, potentially in some "prototypical" sense).[3] Cases of "mismatch", in which the application of phonological and grammatical criteria for word-identification leads to conflicting judgments concerning the number of "words" represented in a particular string of morphemes, should be exceptional, and should be in some structural and/or functional sense "marked" in nature.[4]

The main purpose of the present paper will be to provide an illustration of a language in which such assumptions turn out to be radically untenable. In Galo, a Tibeto-Burman language of the Tani branch spoken in the North East Indian Himalaya which has been recently been extensively described for the first time (Post 2007), "words" are independently-definable according to relatively large, consistent and functionally well-motivated sets of phonological and grammatical criteria. However, such criteria very frequently fail to converge upon identification of the same unit; often, a "mismatch" between what will be called (following Dixon and Aikhenvald (2002)) *phonological words* and *grammatical words* is observed in Galo. This "mismatch" would appear to be language-general, and to not in fact be reducible to the marked properties of a particular set of forms (i.e., it is not limited to clitics). For an example of the type of data with which we will mainly be concerned – mainly, "words" which form part of the Galo predicate complex – first compare (1)-(2).[5]

(1) *`zabdù*
záp-dùu
talk-IPFV
'talking'

(2) *`zabrə́p `duukù*
záp-rə́p-dùu-kù
talk-ICEP-IPFV-CMPL
'finally starting to talk'

In (1), an Imperfective aspect marker *-dùu* 'IPFV' is suffixed to a bound verb root *záp-* 'talk'. Neither form may occur independently or in isolation, nor can they occur in any other order or be interrupted by an independent syntactic word (such as an adverbial).

[3] In the generative tradition, Prince and Smolensky observe the "universal prosody-morphology interface constraint", which states that "every lexical word must correspond to a prosodic word" (Prince and Smolensky 2002:111). They also claim that "any member of a certain morphological category (root, stem, word) must be, or must correspond to, a phonological category" (Prince and Smolensky 2002:45). I am not aware of the existence of a precise specification of the nature of any relationship between "words" in phonological and grammatical senses in the functionalist literature.

[4] I.e., such cases might be assumed, in general, to be describable in terms of some marked properties of a particular form or set of forms, which would often then be characterized (for better or for worse) as "clitics" (Zwicky 1994).

[5] Transcription follows IPA except where *c* = [t͡ɕ] and *z* = [d͡ʑ]. Tones are High/Plain ˊ, Low/Tense ˋ, or Rising-Falling ˆ, with the marker conventionally written over the penultimate vowel of a TBU.

Native speakers can readily pronounce and assign meaning to the whole when uttered in isolation, but cannot do the same for either part. A single prosodic contour unites both forms, bearing a single primary stress accent and a single tonal specification and contour. Internal sandhi (regressive voicing assimilation) obligatorily operate at the root-suffix boundary, and underlying vowel length is neutralized at the right-edge word boundary. In short, a diverse set of grammatical and phonological criteria converge upon identification of a single "word".

In (2), we find the same verb root *záp-* 'talk' followed by a longer string of dependents, including Imperfective suffix *-dùu* 'IPFV'. Here again, none of these four morphemes may occur independently or in any other order, nor may they be interrupted by any independent syntactic word. Native speakers readily pronounce and assign meaning to the whole, but are less comfortable independently assigning meaning to any subpart. However, *two* prosodic units may now be identified, bearing *two* primary stress accents and *two* independent tonal specifications and contours. Regressive voicing assimilation is observed at the right edge of the verb root as in (1), but not at the right edge of the following form (despite that segmental conditions are identical). In short, example (2), like example (1), contains a *single grammatical word*. However, in (2), the number of *phonological words* is *two*. The primary purpose of this paper, then, will be to exemplify and explain such data.

A second contention of this paper will be that the phonological word-grammatical word "mismatch" in Galo is non-trivial in both descriptive and theoretical senses, in that its acknowledgement enables explanation of a number of otherwise seemingly obscure and disparate facts in both the synchronic and the diachronic dimensions of Galo grammar. In general, data from Galo would thus support a view of language in which "phonological word" and "grammatical word" were defined in independent terms, and in which neither type of unit was viewed as a simple projection of or correlate of the other (nor would they be simultaneous projections of a third, higher-order or more generalized type of unit). Although a general functional pressure toward the ultimate, eventual, or (in some sense) "prototypical" unification of a unit "word" might be said to exist, such a unifying principle could not be attributable to any underlying feature of or constraint upon the language grammar.

The remainder of the paper has the following organization: in §2, we will review the various senses of "word" identified by Dixon and Aikhenvald in their well-known (2002) study, and clarify the senses in which terms referring to "words" will be used in the present paper. In §3, we provide a basic and highly schematic background sketch of Galo phonology and grammar, with particular attention to syllable types and structure. §4 looks more closely at the composition and phrasal functions of Galo "words" from a grammatical perspective, while §5 adopts a phonological perspective on the Galo "word". Sections §6 and §7 present the paper's primary arguments, illustrating the main evidence for a phonological word-grammatical word "mismatch" in Galo, and illustrating its consequences for the synchronic and diachronic organization of Galo grammar, respectively. Finally, in §8 a speculative account of the diachronic origin of these aspects of Galo grammatical organization is provided, and §9 includes some comments related to current efforts by the Galo community to represent "word" boundaries using a Roman-based script. §10 concludes with a summary of the presentation.

2. What is a "word"?

In their well-known survey of approaches to its definition, Dixon and Aikhenvald (2002) outline at least three senses in which the concept of a "word" has been applied in linguistic theory:

The first, and most difficult to apply in practice (whether or not it is in fact a valid concept) might be called the "general" or "psychological word".[6] This is the sense in which a native speaker of a language may be aware of a given form as representing a *minimal form-meaning unit* in that language, as in the classic description of Sapir (1921:33-34).

The second, "phonological word", is more technically describable as *a phonological unit between syllable and phonological phrase*, which is recognizable in terms of a coalescence of phonological properties, among which may be *segmental features* (internal and external phonotactics), *prosodic features* (accent and/or tone assignment, cross-segment harmonies), and *rules* (internal and external sandhi). Thus, for example, English *started* [ˈstaɹɹəd] is a single phonological word in that it exhibits a single stress accent, exhibits word-internal assimilation behaviour at the stem-suffix boundary, and so on.

Finally, a "grammatical word" is describable as *a grammatical unit between morpheme and syntactic phrase*, whose constituents are *a head plus immediate (local) dependents* (prefixes, suffixes, or compound elements) in a *fixed, continuous order* in terms of a *given semantic value*. Again, English *started* is a single grammatical word in that it consists of a head (*start*) plus a suffixal dependent (*-ed*) which can occur in no other order, and exhibits its suffix as an inflectional reflex of its function as predicate head of a tensed clause such as *I had started* (or some such), etc.

While in this paper we will be primarily concerned with "words" in the senses outlined above, it will be useful before proceeding to first define "affix" and "clitic" in opposition to "word". Without going into too much detail (and fully conceding that there may be other or better definitions, whether with respect to particular languages or in general), I will here simply stipulate a working definition which I feel to represent the consensus practice among descriptive linguists in the Greater Mainland South-East Asian tradition at least. An "affix" is a morpheme which depends grammatically on a word to which it is bound. A "clitic" is a morpheme which depends grammatically on some unit *other than* its host (usually, a phrase of which its host is also a constituent). Thus, English plural *-s* as in *dogs* is a suffix since it is a grammatical dependent of the word in which it appears, while English definite article *the* as in *the dog* [θəˈdɔg] is a clitic since it is a constituent of the noun phrase of which *dog* is head (it could just as easily be proclitic to another type of noun phrase constituent, as in *the big dog* [θəˈbɪg ˈdɔg]). For a fuller account of the description and analysis of clitics, again see Zwicky (1994) and/or Aikhenvald (2002).

Much of the remainder of the paper will be devoted to an illustration of these concepts in terms of Galo phonology and grammar. First, however, we turn to a brief overview of some relevant linguistic features of Galo. The description is based on the

[6] Though not directly identified as such by Dixon and Aikhenvald, this sense is implicit in their (2002) account of several earlier approaches.

Lare dialect of Galo as described in Post (2007), where a detailed summary of cultural-linguistic context may also be found. In the interest of brevity, only features bearing directly on the arguments of this paper are mentioned here.

3. A brief outline of Galo

Galo is a basically synthetic and agglutinating language, with statistically verb-final constituent order and three major lexical classes noun, adjective and verb (adverbs are primarily derivative). Noun phrase and predicate structures are quite distinct in Galo; most nominal operators are functional words, postpositions and/or phrasal enclitics, while most predicate operators are suffixes (3).

(3) *ŋûn nám arâa lò indù.*

[ŋunù]$_{NP.S}$ [namə́ aràa = lo]$_{NP.OBL}$ [ín-dùu]$_{PRED}$
1.PL house interior=LOC go-IPFV
'We're going inside the house.'

Galo exhibits a robust structural distinction between independent and marked dependent predicates/clauses (a.k.a. "finite/non-finite") (4). Clearcut instances of verb-serialization have not been found; however, what may once have been post-head serialized verbs now occur as a very large and productive set of bound predicate derivational formatives in Galo, as *-ŋám* 'EXH' in (4) (note that *-ŋám* 'EXH' has no homophonous verb root counterpart in modern Galo).

(4) *ŋûn bâal doŋám tó.*

[ŋunù]$_{NP.A}$ [báa-là(a)]$_{PRED.NF}$ [dó-ŋám-tó]$_{PRED.FIN}$
1.PL roast-NF eat-EXH-PFV
'We roasted (it) and (then) ate it up.'

Galo exhibits a basic unity *syllable = morpheme* among lexical roots and suffixes, as *dó-* 'eat' and *-dùu* 'IPFV', although a relatively smaller number of synchronically unanalyzable polysyllabic morphemes also occur. Basic syllable structure is **(C$_i$)V(X)**, in which C$_i$ is one of the possible initial consonants p/t/c/k, b/d/z/g, m/n/ɲ/ŋ, l/r/j, or s/h, V is an obligatory vowel a/i/u/e/o/ɨ/ə and X is either a nucleus-identical (lengthened) vowel or one of the possible final consonants C$_f$, p/t/(c/)k, (b/d/z/g,) m/n(/ɲ/ŋ), (l/)r(/j), (s,) K.[7,8] Of the preceding list of C$_f$, only the non-parenthesized forms occur in the underlying forms of morphemes; parenthesized forms represent forms which are introduced in word-medial position following application of internal sandhi processes, to be discussed below. Galo is a quantity-sensitive language, meaning that syllables are either light/monomoraic or

[7] *K* is an underlyingly underspecified consonant; its surface realization is discussed in §5.2.2 below.
[8] Discontinuous syllable-internal vowel sequences (diphthongs) *may* occur in a very small number of lexemes as a result of historical segment losses and monosyllabification (as in *aí* 'tooth' < Proto-Tani **afi*). These do not impact the present discussion and may be safely disregarded.

heavy/bimoraic, according to the presence or absence of the coda X; various morphophonological processes and prosodic features are sensitive to syllable weight, some of which will be discussed in §5.

4. Grammatical word in Galo

In Galo morphosyntax, the grammatical word constitutes an important and clearly recognizable unit whose constituents are morphemes and which may stand as a constituent of a higher-level grammatical phrase. Among lexemes, the great majority of basic (non-derived) nouns and adjectives are etymologically complex and dimorphemic, and consist either of two compounded roots or a single prefixed root. Examples are in Table 1.

Type	Class	Initial	Gloss	Final	Gloss	Term	Gloss
[PFX-ROOT]	N	*a-*	'PFX'	*kíi-*	'guts'	*akíi*	'belly/guts'
[PFX-ROOT]	ADJ	*a-*	'PFX'	*hòo-*	'long/tall'	*ahòo*	'long/tall'
[ROOT-ROOT]	N	*lə̀-*	'leg/foot'	*cə́ə-*	'digit'	*ləcə̀ə*	'toe'
[ROOT-ROOT]	ADJ	*lə̀-*	'leg/foot'	*zìn-*	'stretch'	*ləzìn*	'outstretched, of legs'

Table 1 – Formation of nouns and adjectives

In modern Galo, productivity at this basic level of word-formation is limited, although certain more productive constructions exist which also take nominal or adjectival roots as basic formatives. For example, Numerical and Adjectival classifier expressions consist of an initial classifier root formative followed by a numeral or adjective root, respectively; an example is *cɨ́r-úm* 'CLF:PELLET-three' 'three grains of (rice, e.g.)'. A very small number of basic nouns and adjectives appear to be fully (synchronically and etymologically) simplex, as *ɲíi* 'person' and *zèe* 'green/blue'; however, such forms constitute a tiny set of exceptions to the overall picture. The point to note here is that the formative roots of a complex noun or adjective in Galo are almost always morphologically bound, and under no circumstances are able to stand independently as a grammatical word (including when uttered in isolation). Exceptions are limited to an idiosyncratic set of forms which do not constitute a natural class over which any positive generalizations can be made. In short, roots and words constitute clearly distinct levels of morphological structure in Galo.[9]

Turning to predicate formation, we find that the morphological distinction between root/morpheme and word levels is even more salient. Simplifying matters somewhat, we can say that the prototypical form of a final predicate in Galo is as shown in Figure 1. As a general rule, predicate formation according to the pattern shown in Figure 1 is fully productive.

[9] The clarity of this distinction in Galo sets it apart from several more easterly Sino-Tibetan languages, such as Chinese. In Chinese, although it is also true that some roots are bound as a result of having lexicalized as formatives of particular compound words, as a general principle lexical roots are readily employed as simplex words (Chao 1968).

[[ROOT – (DER$_{1...n}$)]$_{STEM}$ – INFL]$_{WORD}$

Figure 1 – Structure of a Galo final predicate (simplified; DER = derivation, INFL = inflection)

A predicate stem minimally consists of a single, morphologically bound verb root, as *ín-* 'go', *hí-* 'die' and *mèn-* 'say'). A final predicate in Galo minimally consists of a verb root plus predicate inflection (5), while predicates in non-final, subordinated or derived functions minimally consist of a verb root plus a functional marker of some kind (6). Note that unlike some other Tani languages such as Apatani (Post 2006), it is never possible in Galo for a verb root to stand as an independent syntactic word, such as head of a simple predication.

(5) *ŋó indùu.* **ŋó ín*

ŋó ín-**dùu**	ŋó ín-
1.SG go-**IPFV**	1.SG go
'I'm going.'	

(6) *ŋó bâal dorə́.* **ŋó báa(-)dorə́*

ŋó báa-**là(a)** dó-rə́	ŋó báa- dó-rə́
1.SG roast-**NF** eat-IRR	1.SG roast- eat-IRR
'I'll roast (it) and eat (it).'	

Predicate stems may also be expanded by one or more of a very large set (> 300 members) of predicate derivations, which provide an interesting and provocative dimension to the analysis of Galo predicate grammar. While many are clearly relatable to (other) lexical roots – usually, verb roots – in modern Galo they occur as bound predicate stem-expanding formatives (Post under review). Due to their large class-size and ability to co-occur on the same predicate stem, extensive use of predicate derivations can lead to the formation of predicate words of considerable length and internal complexity (7).

(7) *tíi-ŋám-còo-mò-làa=kée!*

[imbibe-]$_{ROOT}$[**EXH-FIRST-CAUS**]$_{\mathbf{DER}}$[-IPTV.SDIR]$_{INFL}$[=HORT.POL]$_{PCL}$

'**Let** them **finish** drinking **first**, will you?'

With a handful of marked potential exceptions to be discussed in §7.4, predicate derivations cannot themselves stand as head of a grammatical word.

Finally, we can note that grammatical words are the minimal constituents of a grammatical phrase. For example, in a Galo noun phrase (schematized in Figure 2), all labelled constituents are realized by either grammatical words or other phrase types; no sub-word units are referred-to by rules or ordering constraints at the level of the phrase.

GENP – PRHD – RELC – MNOM – NOM – (RELC) – ENUM – RN – QN – PSHD – POST – (PCL)

Figure 2 – Order of common nominally-headed noun phrase elements (head is underlined)

In sum, grammatical words in Galo are clearly distinguishable from both morphemes/roots (i.e., sub-word units) and phrases (i.e., super-word units). The internal constituents of a grammatical word are fixed in terms of a given semantic value, whether they are based on a synchronically productive pattern of formation or not. Grammatical phrases refer to grammatical words as minimal constituents, but do not refer to any sub-word constituents, and native speakers are comfortable uttering and assigning meaning to grammatical words in isolation, but do not generally feel similarly toward sub-word roots (except in those few cases in which a given lexical word consists of one root only).

5. Phonological word in Galo

Phonological word properties in Galo can be usefully divided into *prosodic* (§5.1) and *segmental* features (§5.2).

5.1. Prosodic features

5.1.1. Stress and meter

While possibly not absolute, there exists a strong tendency in Galo for every phonological word to consist of a single metrical foot, and for every metrical foot to instantiate a single phonological word. Strongly trochaic (falling), a Galo foot minimally consists of a single stressed syllable, which we can provisionally take to be obligatorily heavy/bimoraic. In (8), we find seven phonological words, each realizing a single metrical foot bearing a single stress accent. Note discontinuity between the number of phonological words expressed on the surface and the number of grammatical words expressed in the interlinearization.

(8) *`abó `taníi...`mɨ̀ɨ́káa `nammə́...`gumbòk `zaalâa `jù.*

abó-taníi	mɨ́ɨ-káa-nam＝əə	gùm-bók≡záa≡là(a)	juu
father-mankind	char-TENT-NZR:RLS=TOP	lean-DOWN/SOUTH≡REAL≡NF	REP

'Abo Tani...having tried to burn them...found (the flames) actually leaned southward, so they say.'

A number of phonological processes apply at the level of the phonological word which seem to either be motivated by or to otherwise interact with stress-assignment. For example, "Triggered foot-strengthening" is a process applying at certain clitic boundaries; in it, onset-copying creates a heavy-syllable-initial foot/word, on which primary stress can easily rest on the initial syllable. In (8), the effects of Triggered foot-strengthening are observed in *-nam＝əə* 'NZR:RLS=TOP', realized [`nammə́]; another example is *tabə́＝əə* 'snake=TOP', realized [`tabbə́].

A Syncope process also appears to be motivated by stress-assignment to phonological words at the level of the phonological phrase. In it, low-sonority vowels which are nuclei of an underlyingly weak ((C)V) syllable are reduced and sometimes deleted in metrically weak (unstressed) positions. This process enables consolidation of an underlyingly trisyllabic string of morphemes into a disyllabic word. The new surface word carries an initial heavy, stress-bearing syllable following resyllabification of the erstwhile second syllable onset consonant as initial syllable coda, as in the first word of (9). For a fuller description of these and other similarly stress-related morphophonological processes, see Post (2007:§4.1.4).

(9) *`lə̂b(ə)rə̀m `geegáp `nammə́...*

ləbə̀r=əəm gée-gáp-nam=əə

foot.surface=ACC seal-STUCK-NZR:RLS=TOP

'(The stone) having (expanded and) sealed in the soles of her feet...'

5.1.2. Tone

In Galo, all lexical roots and many (though not all) functional morphemes are underlyingly specified for one of two tones, High/Plain and Low/Tense. However, it is usually not possible to determine the underlying tone of a morpheme simply by uttering it in isolation, primarily because most simple morphemes do not have the capacity to stand as an independently meaningful utterance/grammatical word (see §4). Rather, the surface Tone-Bearing Unit (TBU) is the phonological word, and underlying tones must in general be derived inductively in Galo through comparative analysis of the surface tonal contours of phonological words, according to the following set of principles:

When a phonological word has only one constituent TBU (whether because it is morphologically simple or because only one of its constituents is underlyingly specified for tone), the surface contour is a direct reflex of the single underlyingly specified tone (Table 2).

Form. 1	Gloss	Form. 2	Gloss	Word	Gloss
ɲíi	'person'	--	--	*ɲíi*	'person'
ta-	'MDIM'	*bə́-*	'snake'	*tabə́*	'snake'
ta-	'MDIM'	*bə̀-*	'sugar cane'	*tabə̀*	'sugar cane'
hɨ́ɨ-	'urinate'	*-nam*	'NZR:RLS'	*hɨ́ɨnám*	'to urinate'
hɨ̀ɨ-	'plug/clamp'	*-nam*	'NZR:RLS'	*hɨ́ɨnàm*	'to plug/clamp'

Table 2 – Direct projection of underlying root tones in phonological words

When a phonological word consists of multiple underlying TBUs, its surface tone is derived by rule according to the patterns exemplified in Table 3.

Form. 1	Gloss	Form. 2	Gloss	Word	Gloss	Pattern
lák-	'arm/hand'	*cə́ə-*	'finger'	*lakcə́ə*	'finger'	**H + H → H**
lák-	'arm/hand'	*cì-*	'left'	*lakcì*	'left hand/arm'	**H + L → L**
lə̀-	'leg/foot'	*cə́ə-*	'finger'	*ləcə̀ə*	'toe'	**L + H → L**
lə̀-	'leg/foot'	*cì-*	'left'	*ləcì*	'left leg/foot'	**L + L → L**

Table 3 – Derivation of surface tones from multiple underlying tones

This pattern appears to hold whether or not the phonological word in question is itself a grammatical word. Thus, to refer back to (2) above, although a native speaker of Galo cannot generally attribute a meaning to the phonological word *duukù*, s/he is usually able to identify its tone as Low/Tense just as readily as that of *ləcə̀ə* 'toe' in Table 3.

5.1.3. Glottal stop onset prosody

In the Lare dialect of Galo which forms the basis of this description, underlyingly vowel-initial lexemes tend strongly to exhibit a glottal stop onset, blocking resyllabification of across a phonological word boundary. Underlyingly vowel-intial functional morphemes (including postpositions/enclitics) tend not to exhibit a glottal stop onset; resyllabification across such boundaries is therefore possible (10).[10]

(10) *porók-luggóm ʔuudʲlá...ʔaɨ̵gə́ ʔalə́kə dollòm*

porók-lugó=əəm úu-dó(o)-là(a)=́ aɨ̵́=gə alák=əə dolò=əəm
fowl-crowing=ACC awake-STAT-NF=NFI1 self=GEN hand/arm=TOP paddy=ACC
ʔɨ̵dú...
ɨ̵́-dùu=́
pound-IPFV=NFI1
'After waking up at the cock's crow...they pound the paddy with their own hands...'

A glottal stop cannot occur word-internally. For example, *kók-úu* 'crow-AWAKE' 'crow someone awake; awake by crowing' – in which the result derivation *-úu* 'AWAKE' is cognate with the verb root *úu-* 'awake' in (10) – is realized [kogúu], not *[kogʔúu].

5.1.4. Intonation contour

It is common to find pauses across phonological word boundaries in Galo – usually, though not always, when such boundaries also constitute the boundary of a phonological phrase – however, it is never possible to find pauses inside a phonological word in absence of repair. In (11), the speaker hesitates at the word onset, phonetically

[10] In other Galo dialects, the glottal stop onset appears to be phonemic, and to be underlyingly assigned to some vowel-initial lexemes but not to others. Obligatoriness of the glottal stop onset in such dialects has not yet been extensively researched.

realizing an initial verb root but not realizing any further constituents of the predicate. After he settles upon an appropriate predicate form, he repairs the hesitation by repeating the predicate head; no examples of this form occur in my data in which the predicate head is not repeated (i.e., in which the speaker directly proceeds by mentioning the predicate derivation).

(11) *nó...gogbooló...**tá/...ta**pâa lammò.*

nó gók-boolo **tá/** **tá-**pàa-là(a)-mòo
2.SG call-COND **listen/** **listen-**ATTN-ABIL-NEG
'If you call, she...she surely won't hear.'

Now compare (12), in which the speaker corrects himself "mid-word", by deciding to employ a Change-of-State aspectual suffix rather than a Stative suffix. Notice that this time the initial constituents of the predicate word are *not* repeated or "repaired". Rather, the speaker seemingly privileges *phonological* (not grammatical) word-boundaries.

(12) *allò-rôə nè*

allò-roò nè
tomorrow-day.after.tomorrow IRR.TMP.PUNC
*caatə̂r **dóo/...dag**êe bə́…*
càa-tə́r-**dó(o)/-dàk**-ée = bə́
ascend-TO.END-**STAT/-COS-**IPFV.DISJ=AVZR
'After they come up tomorrow or the next day…'

5.2. Segmental features

5.2.1. Assimilation sandhi

Word-internally, regressive assimilation sandhi of two types are obligatorily observed in Lare Galo: *voicing* and *place*.[11] Voicing assimilation causes all oral stop codas to the initial syllable of a phonological word to be voiced when followed by a voiced second syllable onset. An example is the second word of (11). Place assimilation causes nasal codas *-m* and *-n* to assimilate in place to certain following consonants: *-n* to labial and velar consonants, *-m* to velars only. For example: *ɲóm-káa* 'swallow-PF' 'swallowed' (realized [ɲoŋkáa], not *[ɲomkáa]) and *ín-pɨ̀ɨ́* 'go-REACH' 'reach' (realized [impɨ̀ɨ́], not *[inpɨ́ɨ́]). While assimilation processes may be sporadically observed across phonological words occurring within the same phonological phrase, as when speaking rapidly, such processes are never obligatory and are generally "undone" in clear speech.

[11] In other Galo dialects (such as Pugo), regressive nasal assimilation is also observed.

5.2.2. Realization of underspecified consonants

In Lare Galo, two incomplete or "underspecified" consonants occur in the underlying forms of native Galo morphemes. The first, a fricative *h*, is realized [h] word-initially (as *hɨ̀ɨ́dàa* 'stick') and word-medially when following a vowel *and* when standing as onset of a heavy (CVX) syllable (as *ɨ́hɨ̀ɨ́* 'wood'). Word-medially when following a consonant and/or when standing as onset of a light (CV) syllable, *-h* surfaces [s] (as *namsùu* 'stinky' and *ɨ́sì* 'water'). Also consider the behaviour of the Reflexive suffix *-hí* 'REFL' in (13)-(14). In (13), *-hí* 'REFL' follows a consonant-final morpheme as the final syllable within a phonological word; its initial fricative is therefore realized [s]. In (14), *-hí* 'REFL' also follows a consonant-final morpheme within the overall grammatical word of which it is a constituent; however, it occurs as *initial syllable* of the phonological word of which it is a constituent, and accordingly surfaces with initial [h].

(13) *zapsí toké!*
záp-**h**í-tó=kée
talk-REFL-IPTV.ODIR=HORT.POL
'Talk to yourself!'

(14) *zabmín hitᵊké!*
záp-mí**n-h**í-tó=kée
talk-RECP-REFL-IPTV.ODIR=HORT.POL
'Talk amongst yourselves!'

Underspecified consonant *-K* reflects a Proto-Tani syllable-final consonant of uncertain form (but which may have been **-ɕ*). In modern Lare Galo, *-K* fully assimilates to any following consonant when occurring word-medially. Word-finally, it surfaces [k]. When followed by a vowel word-medially, it surfaces [g], seemingly reflecting [k] followed by Regressive voicing assimilation (cf. §5.2.1). Examples are *cíK-nam* 'throw.spear-NZR:RLS' 'to throw a spear', realized [ci**n**nám] and *cíK-ùp* 'throw.spear-SHATTER' 'throw a spear such that something shatters', realized [ci**g**úp]. In (15), note that the final consonant of verb root *zíK-* 'melt' is realized [d] following full assimilation to the following, phonological-word-internal Imperfective suffix initial. In (16), note that although the Imperfective suffix *-dùu* 'IPFV' still occurs in the *same grammatical function*, the final consonant of cognate Result derivation *-zíK* 'melt' is realized [k]; this is because *-K* and *-dùu* 'IPFV' now occur *across a phonological word boundary*.

(15) *plastikə́ zi**dd**ûu kú.*

plastík = əə zí**K-d**ùu-kú
plastic(<Eng)=TOP melt-IPFV-CMPL
'The plastic is now melting.'

(16) *dooɲə́ plastikə́m amzí**k d**uukù*

dooɲí = əə plastík = əəm ám-zí**K-d**ùu-kú
sun=TOP plastic=ACC roast-MELT-IPFV-CMPL
'The sun is melting the plastic.'

6. Good fences make good neighbours; or, how I learned to stop worrying and love the disconnect between "grammatical word" and "phonological word" in Galo

In the preceding two sections, we reviewed some of the grammatical and phonological properties of words in Galo. In the process, we also identified a number of cases in which the number of morphemes identifiable in a particular string resolved into different numbers of words according to grammatical and phonological criteria. As long as it was thought that the general principle of a unified category "word" had to be maintained, this would take on the appearance of a *problem*: which set of criteria is to be privileged as the primary basis for a language-general definition of "word"? If the other set of criteria is discounted as a means of identifying "words", then what sort(s) of unit do these criteria identify?

With reference to (16) above, let us briefly entertain two possible alternatives to the analysis of a grammatical word/phonological word "mismatch" that we have identified above: first, one in which phonological criteria were discounted, while grammatical criteria were treated as the sole measure of word boundaries, and second, one in which phonological criteria were treated as primary.

If grammatical criteria were treated as the sole measure of word boundaries, then (16) should be re-transcribed as (17) below.

(17) *dooɲə́ plastikə́m **amzíkduukù***

dooɲí = əə plastík = əəm ám-zíK-dùu-kú
sun=TOP plastic=ACC roast-MELT-IPFV-CMPL
'The sun is melting the plastic.'

Assuming that tone- and stress-assignment, as well as boundary phonotactics and rule-applications could at least provisionally be handled through reference to another type of unit (say, a prosodic foot), we would still be left with a situation which conflicts wildly with the intuitions of Galo native speakers in many cases. Compare (18), which represents an analogous case drawn from the nominal lexicon, a type of classificatory compound with a fixed formative structure AB-BC (in which "B" represents a root held in common).

(18) *tapêk-perrò*

ta**pèk-per**rò
leech-jungle.leech
'jungle leech'

In (18), the initial and final terms share a common **bolded** root *pèK-* 'leech', seemingly reflecting PT **paɕ*. As can be seen, in the initial term *tapèk* 'leech', *pèK-* surfaces [pek], while in the second term *perrò* 'jungle leech' it surfaces [per], reflecting what we have described as the word-final and word-medial behaviour of syllable-final *-K* respectively (§5.2.2). Phonologically, then, (18) exhibits the same seemingly word-oriented characteristics as does the predicate in (17), and Galo speakers are quite clear on potentially independent "word" status of both *tapèk* and *perrò* in (18). Writing *"tapêkperrò"* as a single "word" would render the string practically unparsable to a Galo native reader.

To consider the second possible "unifying" solution, in which phonological criteria were privileged in determining "word" status, we might re-transcribe (17) as in (19).

(19) *dooɲə́ plastikə́m amzík duukù*

dooɲí=əə	plastík=əəm	**ám-zíK**	**dùu-kú**
sun=TOP	plastic=ACC	**roast-MELT**	**AUX.IPFV?-CMPL**

'The sun is melting the plastic.'

The challenge with respect to (19) would be to determine the grammatical status of Imperfective aspect marker *-dùu* under the assumption that it was functioning as head of a grammatical "word". One possibility might be to analyze it as some sort of auxiliary verb root – which, from a historical perspective, is perhaps not a silly sort of proposal at all.[12]

There are two problems with this possible solution. The first is that it would be left to explain why it is that a form with a single function 'Imperfective' should surface as a suffix in a sentence like (15), but as an "auxiliary" in a sentence like (19), when the only apparent difference between the two sentences is that in (19), the predicate stem has been derivationally expanded.[13] Furthermore, analysis of *-dùu* as an auxiliary head in

[12] Post (2007:§2.1.4) has suggested that the full set of Tani non-perfective aspect markers (as well as many if not all other predicate inflections) may derive historically from a series of uninflecting post-head auxiliary verbs – in this case, *dùu-* 'sit; stay; exist (animate); be at (for an item construed as 'sitting')'. Auxiliary-like behaviour of a seemingly cognate form *duŋ* is still represented in Pagro Mising, a Tani language of the Eastern branch (Post forthcoming 2008).

[13] This assumes, of course, that it is necessary to analyze *-dùu* as a "suffix" in (15), inasmuch as it enables a bound verbal root to stand as a grammatical word. Continuing to analyze *-dùu* as an "auxiliary" in such conditions would require an assumption that an "auxiliary" were able to compound directly to a lexical verb root, forming a single grammatical word. This would seem to require a very different definition of

(19) would lead to the proliferation of literally hundreds of potential auxiliary heads in the language – all of which were also capable of being suffixed or compounded to a bound verbal root. This is because of the large number of predicate derivations which are available in Galo (§4). Table 4, note that the predicate derivations in **bold** occur word-finally in the first column, but word-initially in the second column, despite that their semantic values remain unchanged. This is due simply to the fact that the predicate stems in the second column have all been previously expanded by a separate Result derivation *-kák* 'CLEAN'.

*rɨglɨ**ɨ** dù*	'wanting to wash it'	*rɨkkák l**ɨɨ**dù*	'wanting to wash it clean'
*rɨk**kên** dù*	'easy to wash'	*rɨkkák k**en**dù*	'easy to wash clean'
*rɨk**sí** dù*	'washing oneself'	*rɨkkák h**i**dù*	'washing oneself clean'
*rɨg**ŋám** dù*	'washing everything'	*rɨkkák **ŋam**dù*	'washing everything clean'
*rɨg**bəə** dù*	'still washing it'	*rɨkkák b**əə**dù*	'still washing it clean'
*rɨg**ŋóo** dù*	'habitually wash it'	*rɨkkák **ŋoo**dù*	'habitually wash it clean'
(...)		(...)	

Table 4 – "Suffix/Auxiliary" alternations in the predicate complex

The position adopted in this paper is that there would be little point in developing a complex morphosyntactic analysis to explain why, for example, Desiderative derivation *-lɨɨ* 'DESD' should have a different word-level grammatical status in the first column of Table 4 than it has in the second column. A far simpler and, to my mind, more insightful account would acknowledge that its grammatical status is unchanged; it is a bound predicate formative in both cases. The only difference is that it occupies the third syllabic position in a grammatical predicate in the second column; it therefore occurs as the second syllable of a disyllabic phonological word in the first column, and as the initial syllable of a separate phonological word in the second column. In other words, when phonological and grammatical criteria for the identification of "words" in Galo are considered separately – and when their "unification" is neither required nor expected – these and other tricky analytical problems simply go away.

7. Implications and effects of the grammatical word-phonological word "mismatch" in Galo

In §6 it was argued that separate identification and analysis of "phonological words" and "grammatical words" leads to a simpler and, arguably, more insightful analysis of Galo grammar than one in which a more general, unified concept "word" were thought primary. The contention of this section will be that the grammatical word-phonological word "mismatch" in Galo is not simply a matter of analytical preference or descriptive consistency, however. Rather, it will be argued that the mismatch has deep implications for the overall organization of Galo grammar.

"auxiliary" than is generally assumed by most syntactic theories, and would also require an assumption that syntactic rules could access word-internal morphology. I do not wish to pursue this possibility here.

7.1. "Functor fusion"

As in a number of other Tibeto-Burman languages, it is very common in Tani languages to encounter usually disyllabic forms (or once disyllabic forms) which appear to represent historical fusions of two previously simplex functional morphemes. Dual and plural pronouns very often incorporate formatives which, while synchronically unanalyzable as such, would once have occurred as postposed functional nominals (often, the numeral 'two' and a noun meaning something like 'group'; see Post (2007:§7.1.3)). Sometimes, two previously co-occurring postpositions – a phenomenon not unlike "double case" (Plank 1995) – become fused into a single form with complex functionality; for example, consider the general Galo Ablative postposition *lokə̀* 'ABL', which seems to derive historically from the sequence *lo* 'LOC' plus **kə̀* 'GEN/ABL'.[14]

In Galo, we find literally dozens of such forms. Most noticeable – due to their large quantity and high frequency of use – are the so-called "demonstrative postpositions" described in Post (2007:§7.4). Deriving historically from sequences of the form DEMONSTRATIVE + POSTPOSITION and retaining the deictic and relational-marking functionality of both types of term, modern Galo demonstrative postpositions are no longer analyzable as productive formations. For example, *tolò* 'DST.LOC.UP' derives from the collocation *tə̀* 'DST.UP' + *lo* 'LOC'; it exhibits irregular progressive vowel-harmonization, and can be used both pronominally and in pre-head positions (like all demonstratives but unlike all simplex postpositions). While potentially explainable any number of ways, it is very likely that such fusions were at least partially motivated by the likelihood that such collocations would have been frequently uttered as independent phonological words; this hypothetical process is schematized in (20).

(20) **dolúu təlò* → *doolúu tolò*

doolúu	tə̀=lo	doolúu	tolò
village	DST.UP=LOC	village	DST.LOC.UP

'up in the village'

Such fusions are perhaps even more salient when they take place across word boundaries, or even across constituent boundaries; such instances can lead to complex structural and functional reanalyses. For example, consider the two Lare Galo Concessive subordinators *(-)dakkòm* 'CONC' and *(-)la(a)cìn* 'CONC'. To a great extent semantically equivalent, the first form derives from a collocation of Change of State aspectual

[14] **kə̀* does not occur as a postposition in modern Galo, although it is reconstructed as such by Post (2007). Modern Galo reflexes include the pronominal Genitive suffix *-kə̀*, as well as the Genitive postposition *gə* (seemingly reflecting an earlier process of lenition). A seemingly cognate Genitive postposition *kə* is observed in Pagro Mising (Post field notes).

suffix *-dàk* 'COS' plus a Minyong-derived Additive particle *kòm* 'ADD'.[15] The second quite similarly derives from Non-final suffix *-là(a)* 'NF' plus native Galo Additive particle *cìn* 'ADD'. Both may occur as clause-subordinating suffixes to an uninflected predicate stem, with the basic overall sense 'Although/despite that CLAUSE, CLAUSE' (21). When suffixed to a disyllabic stem, the Concessive suffix is realized as an independent phonological word (22); this follows the basic behaviour of any relatively lengthy predicate word, such as those exemplified in Table 4 above.

(21) *ə̂mbə rɨdâkkom, nokkə̀m zərjâa rə́.*

əmbə	rɨ-**dakkòm**	nó-kə̀=əəm	zə́r-jàa-rə́
ANAP.PADV	happen-**CONC**	2.SG-GEN=ACC	spin-MORE-IRR

'That (being the case) **notwithstanding**, my (top) will spin longer than yours (will).'

(22) *ŋûn kaamáa **dakkòm**, doolúu hɨgɨ́ dookáa hidù!*

ŋunù	káa-máa-**dakkòm**	doolúu	hɨgɨ́	dóo-kaahí-dùu
1.PL	have/exist-NEG-**CONC**	village	SPRX.IND	LOC.EXIS.INAN-APLENTY-IPFV

'**Although** we don't have any (*koobu-curgen* ornaments), there's plenty available in this village!'

Interestingly however, both *(-)dakkòm* and *(-)la(a)cìn* are also able to occur as adclausal *noun*-subordinating particles, with the basic sense 'despite (n.); (n.) notwithstanding' (23). Naturally, from this function they can then be extended to marking an argument NP with a concessive sense 'even' (24).

(23) *ə̂ᵍ **dakkòm**, bulù...tukâa bulù, apúk-anág bə́...caamâa rə́.*

əgə	**dakkòm**	bulù	tukkáa	bulù	apúk-anák=bə́	càa-máa-rə́
ANAP.IND	**CONC**	3.PL	blackie	3.PL	hasty=AVZR	ascend-NEG-IRR

'**Despite** that, Tuka and all them won't move in in such a rush.'

(24) *hɨɲɨ̀...ərə́k go **dakkòm** alərə́ əmbóolo...*

hɨɲɨ̀ɨ	ərə́k=go	**dakkòm**	alə́-rə́	ə́m-boolo
this.year	pig=IND	**CONC**	good-IRR	say-COND

'If (the shamans) say that this year **even** a pig will do (then that's what we'll go ahead and sacrifice).'

There are two points to note here. The first is that there is no evidence (nor would there seem to be any likelihood) that Change of State aspectual suffix *-dàk* 'COS' has ever been capable of functioning independently as a noun marker; therefore, it seems unlikely

[15] Not accepted by most modern Lare Galo speakers as a simplex Additive particle, the form seems to have entered Galo via the Minyong-bordering Pugo dialect, together with a large number of other Minyong forms. However, Lare speakers accept and use the complex fused form.

that noun-marking uses such as those in (23)-(24) could have arisen compositionally. The second is that, although it is conceivable that (22)-(24) could all represent instances of a single subordinating particle/word *dakkòm*, capable of both predicate and nominal scope, this analysis is untenable in (21) (where *-dakkòm* licenses a grammatical predicate by suffixing directly to a bound verbal root) – and the functional values of *-dakkòm* in (21) and (22) are quite clearly identical.

Ultimately, then, the suggestion made here is that frequent mention of *-dakkòm* in contexts in which it occurred as an independent phonological word such as in (22) would in fact have *encouraged* reanalysis of *dakkòm* as an independent *grammatical* word, ultimately leading to extensions in its functionality. This hypothetical development is sketched in Table 5; in Table 5, note that while the grammatical status of *(-)dakkòm* is hypothesized to have undergone several developments, its phonological value remains unchanged throughout.

subordinated inflected predicate	káa-máa**-dàk=kòm**	[kaamáa **dakkòm**]	→
subordinated predicate stem	káa-máa**-dakkòm**	[kaamáa **dakkòm**]	→
subordinated nominal	əgə **dakkòm**	[əgə **dakkòm**]	→
particle-marked nominal	ərək go **dakkòm**	[ərək gó **dakkòm**]	

Table 5 – Development of noun particles from predicate subordinators via reanalysis of independent phonological words as grammatical words (for glosses and translations, cf. (22)-(24))

To summarize, there exist large numbers of disyllabic functional morphemes in Galo which seem to derive historically from fusions of previously independent, monosyllabic morphemes (suffixes or simplex functional words). While diverse in grammatical origin, the suggestion here is that the common thread accounting for their development is that all would have occurred frequently as independent phonological words; to the extent that speakers were able to assign a single functional value to the collocation, relative phonological independence would have encouraged reanalysis of such forms as independent grammatical words.

7.2. "Versatile particles"

Particles in Galo constitute a large and diverse set of forms and functions. Most occur constituent-finally (or as enclitics to a major constituent), and can be effectively analyzed as pertaining more or less to the predicate or noun phrase areas of the grammar, according to subtype. However, a relatively small number of particles have a more heterogeneous distribution; they are described as "versatile" particles, and are divided by Post (2007:§13.5) into Emphatic and Adverbial subtypes. Both subtypes are to a great extent able to follow any major syntactic constituent; Emphatic particle *(ə)ɪ́* (◌́ represents an extra-high tone) occurs three times in (25), first following a postpositionally-subordinated clause and the second two times following each of two copula complements in an appositive coordination. In each case, the function of Emphatic *(ə)ɪ́* is basically to

draw attention to the marked constituent, as though it were this and no other that was intended, or as though to cast an especially high degree of focal importance.

(25) *aŋŋí lokkə̀ hikaî maarʔmə́ (…) ânə bəədâk lokkəí́*

aŋŋíi lokkə̀ hikai-máa-nam＝əə anə̀ bə́ə-dàk lokkə̀＝**(ə)í́**
bit ABL.SRC teach(<Ind)-NEG=TOP mother bear-COS ABL.SRC=**EMPH**
annəí́ abbʷə́í hobʔgə̀
anə̀＝əə＝**(ə)í́** abó＝əə＝**(ə)í́** hobə́＝əgə̀
mother=COP.IPFV=**EMPH** father=COP.IPFV=**EMPH** mithun=ANAP.IND
moodîi lo...rəŋóo nà zaatə̀.
moodìi＝lo rə́-ŋóo-nà zaatə̀＝əə
mountain=LOC exist-HAB-NZR:SUB nature(<Ind)=COP.IPFV
'(Due to our) not teaching them from when they're young...***right*** from when they're born, be they ***female***, be they ***male***, these mithuns are mountain-dwellers by nature.'

Another signal attribute of versatile particles is that they are all capable of occurring *inside* a grammatical predicate word of which they are *not* themselves a structural constituent, "interrupting" the predicate with basically emphatic or attention-drawing functionality (26) (cf. also (8) above). In Post (2007), the non-standard symbol ≡ is used in this context, to denote a boundary which is neither precisely a word/clitic boundary nor precisely a suffix boundary, but which rather represents the word-internal *imposition* of a clitic at a suffix boundary.

(26) *hodûm hórʔcìn rəkú **əí́** maané.*

hodùm-horə́＝cìn rə́-kú**≡(ə)í́≡**máa＝né
barking.deer-boar=ADD live/exist-CMPL**≡EMPH≡**NEG=DECL.ADM
'Even wild game was ***nowhere*** to be found, see (because a tiger had scared them all away).'

In (26), the full form of the grammatical predicate word is *rə́-kú-máa* 'live/exist-CMPL-NEG'; Admonitive declarative particle *né* is treated as an enclitic which is structurally outside the predicate word (it can also mark noun phrases, for example). That *(ə)í́* 'EMPH' occurs within, and not at the boundary of, a grammatical word, is clear from the fact that *rəkú_* is not in fact a grammatical word boundary (27).[16]

[16] Completive aspectual suffix *-kú* is a "Secondary predicate inflection" which is not capable of licensing a finite predicate word, unlike "Primary predicate inflection" *-máa* 'NEG', which is.

(27) **hodûm hórᵊcìn **rəkú**.*

hodùm-horə́ = cìn	**rə́-kú**
barking.deer-boar=ADD	**live/exist-CMPL**

Furthermore, and crucial for our purposes here, Emphatic particle *(ə)ɪ́* 'EMPH' can *only* interrupt the predicate at a *phonological word boundary*. Since *rəkú* (as in (26)) constitutes one phonological word, not two (i.e., *rə́_kú* is not a phonological word boundary), the sentence in (28) is unacceptable. (29) simply shows that this is not a problem related to the relative ordering of *(ə)ɪ́* and *-kú* (*rə́-máa* 'live/exist-NEG' would constitute an acceptable final predicate on its own).

(28) **hodûm hórᵊcìn **rəɪ́** kumá.*

hodùm-horə́ = cìn	**rə́≡(ə)ɪ́**≡kú-máa
barking.deer-boar=ADD	**live/exist≡EMPH**≡CMPL-NEG

(29) **hodûm hórᵊcìn **rəɪ́** má.*

hodùm-horə́ = cìn	**rə́≡(ə)ɪ́**≡máa
barking.deer-boar=ADD	**live/exist≡EMPH**≡NEG

In sum, Versatile particles are capable of occurring within a grammatical predicate word, "interrupting" it despite not being a grammatical predicate constituent. However, predicate interruption respects *phonological* word-boundaries which occur *within* the grammatical predicate word; failure to respect phonological word-boundaries results in an unacceptable utterance.

7.3. Multiword predicates

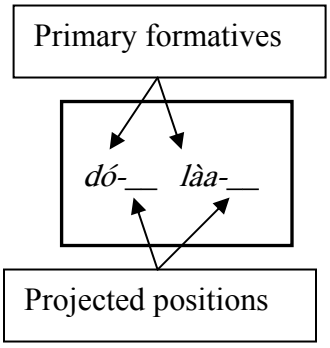
A number of productive constructions exist in Galo which are based on the full or partial repetition of certain formatives in a grammatical predicate whose surface output consists of at least two "words". Neither precisely a serial verb construction nor precisely a complex predicate (in the ordinary senses of these terms) the resulting Galo "Multiword predicate" harbours a considerable amount of underlying structural complexity which belies their often playful, expressive rhetorical quality and presents challenges to morphological analysis at the word level.

The most common type of multiword predicate is built around at least one of three possible types of "primary formative": a Discontinuous compound verb (DCV), Discontinuous predicate derivation (DPD) or an Expressive semi-reduplication (ESR). Examples of each are first given in Table 6.

Type	Example	Gloss	F1	Gloss	F2	Gloss
DCV	*dó-…làa-*	'subsist'	*dó-*	'eat'	*làa-*	'take'
	pɨ́-…pàa-	'make a living'	*pɨ́-*	'craft'	*pàa-*	'get'
DPD	*-pàa…-là(a)*	'WHATEVER'S AVAILABLE'	*-pàa*	'ATTN'	*-là(a)*	'ABIL'
	-kúp…-lék	'HELTER SKELTER'	*-kúp*	'UPSIDE DOWN'	*-lék*	'RIGHTSIDE UP'
ESR	*-bə́ə…-jə́ə*	'DURATIVE'	*-bə́ə*	'DUR'	*-jə́ə*	'RDUP'
	-pèn…-jèn	'SEPARATE'	*-jén*	'SEP'	*-jén*	'RDUP'

Table 6 – Primary formatives of Multiword predicate constructions

Given two primary formatives, a Multiword predicate is then formed in which two formative positions are projected according to the following constructional template (Figure 3; for morpheme glosses, refer to Table 6).



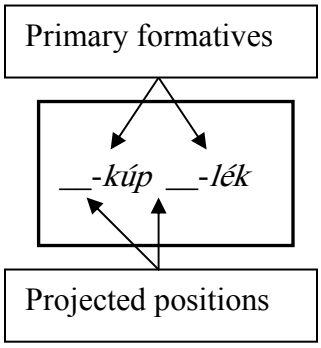
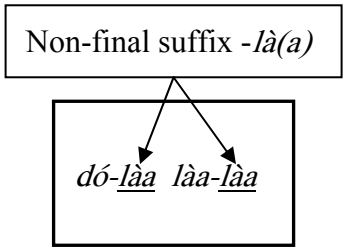


Figure 3 – DCV Template **DPD/ESR Template**

The projected formative positions are then filled by *separate iterations* of the corresponding predicate constituent – in the case of a DPD/ESR, a preceding verb root; in the case of a DCV, a following predicate derivation, predicate inflection, or other predicate dependent (Figure 4).



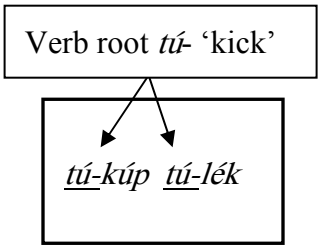


Figure 4 – DCV Projected positions filled **DPD/ESR Projected positions filled**

Once both the lexically-specified and projected formative positions of a Multiword predicate are filled, any remaining predicate formatives simply occur in turn (30)-(31).

(30) *dolâa laalâa kú*

dó-là(a) làa-là(a)-**kú**
eat-NF take-NF-**CMPL**
'came to make a living'

(31) *tukúp tulék ká*

tú-kúp tú-lék-**káa**
kick-OVERTURN.1 kick-OVERTURN.2-**PF**
'kicked it over'

Phonologically, there is no question about the number of "words" represented in Multiword predicate; minimally, there are two, and in examples such as (30)-(31), there are three. Grammatically, however, the number of "words" is difficult to assess. Note that in each case there are *two* grammatical predicate heads represented, whether these are lexically-specified (as in a DCV) or projected (as in a DPD/ESR). However, there is only *one* set of grammatical predicate dependents, whether these receive separate or individual iterations, as the Non-final and Completive suffixes of (30), respectively. In short, there is only *one* underlying grammatical predicate, which exhibits a complex mapping onto the surface phonological form. This mapping relation might be represented as in Figure 5.

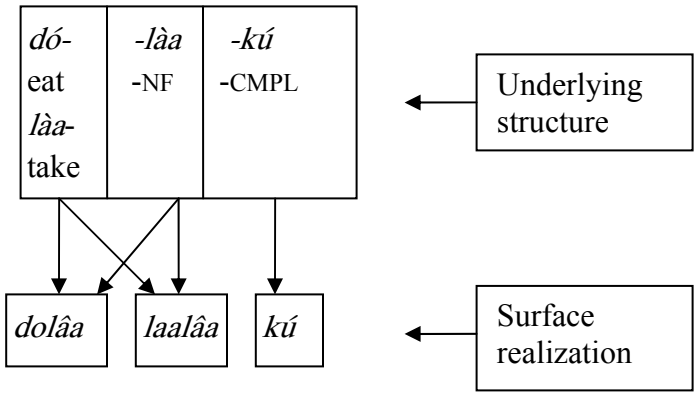


Figure 5 – Mapping relation between underlying and surface structures of Multiword predicate (cf. (30))

7.4. "Auxiliation" in polar question responses

The languages of Greater Mainland South-East Asia (GMSEA) commonly lack terms signifying general agreement or disagreement with the presupposition of a polar question (equivalent to English "yes" and "no"); instead, responses to polar questions often involve full or partial repetition of the predicate in positive or negative polarities. Responses to polar questions are in turn regularly applied as tests for grammatical predicate (or predicate head) status in GMSEA languages (Enfield 2004, among many others). Galo is no exception (32)-(33).

(32) *A: tacên duurè?*

tá-cèn-dùu=ree
listen-KNOW-IPFV=PQ
A: 'Do you understand?'

(33) *B: tacên dù.*

tá-cèn-dùu
listen-KNOW-IPFV
B: 'Yes, I do.'

Importantly, the response in (33) requires repetition of the predicate head; it is not possible, for example, to simply reply "*dǜ*" (treating the Imperfective suffix as though it were an auxiliary-like predicate head). However, a small number of predicate derivations – seemingly, only three out of the hundreds available – appear to license a different kind of response. A question containing one of the Desiderative derivation *-lɨ̀ɨ́* 'DESD', Ability derivation *-là(a)* or Reflexive derivation *-hí* may be answered by treating the predicate derivation as though it were an auxiliary verb-like predicate head, omitting the predicate root (34)-(35).

(34) *nó əpâk **la**rəì?*
nó əpàk-**là(a)**-rə́=(ə)ì
2.SG discard-**ABIL**-IRR=PQ
'Will you be **able** to quit (smoking)?'

(35) ***la**rə̀.*
Ø-**là(a)**-rə́
Ø-**ABIL**-IRR
'Sure I will.'

(36) *nó əpâk zirəì?*
nó əpàk-**zí**-rə́=(ə)ì
2.SG discard-**BEN**-IRR=PQ
'Will you throw it away for him?'

(37) **zirə́.*
Ø-zí-rə́
Ø-BEN-IRR

Other predicate derivations do not license this type of response (1)-(37).[17]

It is currently unknown whether this irregular behaviour in Lare Galo polar question responses is more likely to represent a morphosyntactic conservation from a proto-language or an innovation in Galo or one of its ancestor languages, since we currently lack adequate comparative data from other Tani languages to enable corroboration of an internal reconstruction. For present purposes it will not be necessary to resolve this point; the important thing to note here is that the construction itself is sensitive to the manner in which a Galo grammatical predicate of three or more syllables is divided into phonological words. This fact is made plain by way of the phonetic realization of the Ability derivation *-là(a)*. Like Non-final suffix *-là(a)* (30) and a handful of other predicate formatives, Ability *-là(a)* is subject to the irregular but pervasive Lare Galo process of Third syllable truncation (TST). In TST, a qualifying morpheme with an underlyingly long rhyme surfaces with a short rhyme when occupying the third syllabic position in a predicate string; in other positions, the rhyme surfaces with the conservative long form (again, compare (30)).[18] Note, then, that (35) exhibits the *truncated* form.[19] In

[17] (37) is unacceptable as a response to (1). If interpreted as a sentence headed by the verb root *zí-* 'give' – the certain historical source form of the Benefactive suffix – it would be grammatically acceptable; however, the semantic value would then be quite different, meaning 'I will give (it to someone).' As such, it would represent a pragmatically marked non-sequitur to (1), since it would not address the question concerning 'discarding'.

[18] Ability *-là(a)* is reconstructed as PTs **laŋ*, the regular Lare Galo reflex following Final velar nasal deletion with compensatory lengthening is *-làa*.

other words, the form of an irregular polar question response taking a predicate derivation as "head" is not built-up compositionally from the morphemes in question; rather, it takes the (irregular) phonological form of the *question* as the basis for its own structure.

The important point for present purposes is that this grammatical outcome seems to be conditioned not exclusively by the grammatical facts of the predicate structure, but also (and perhaps especially) by the regular division of the predicate structure into *phonological words*. It is possible, and indeed likely, that frequent utterance of Ability, Desiderative and Reflexive derivations as phonological word initials in sentences like (34) has in part led (or may be leading) to their reanalysis as auxiliary verbal heads – whose use is, however, at present limited to certain constructions.[20]

8. A diachronic perspective: Rhythm and the synthetic drift of Tani

The preceding subsections have illustrated some seemingly disparate facts about Galo grammar which are all argued to relate in some way to the presence of a "mismatch" between phonological words and grammatical words in Galo. Since this phenomenon is not particularly commonly-identified across languages, we might wonder how it came about in Galo.

In two important and far-reaching papers, Donegan and Stampe (1983; 2004) put forth a theory of morphosyntactic change in which typological shifts often assumed to result from "language contact" (via some unspecified mechanism) are suggested to be more directly caused by a shift in prosodic organization (which, however, may itself derive ultimately from language contact). Primarily with reference to Mundan, a branch of the Austro-Asiatic language family, Donegan and Stampe argue that a basic shift from iambic (rising) to trochaic (falling) rhythmic organization can account for a variety of typological differences between Mundan and modern-day Mon-Khmer languages (which are argued to more closely reflect the basic typology of Proto-Austro-Asiatic). Among the observations they make, which may be generalized and cast as predictions, are that a language which undergoes a shift to trochaic (falling) rhythm should develop suffixes/postpositions, synthetic/agglutinating structures, a (C)V(X) syllable canon, stable, geminate clusters, stable, monophthongal vocalism, harmonic prosodies, and register rather than contour tones. The basic form of the principle is first illustrated in Figure 6.

---

[19] This is independent of the position of the Ability derivation in the questioning predicate, thus is apparently attributable to conventionalization of the response form rather than productive "mirroring" of the question form on a given occasion of use.

[20] In principle, such data could be viewed as a counterexample to the unidirectionality principle of grammaticalization theory, inasmuch as an "auxiliary verb" would be generally viewed as "less grammatical(ized)" than a derivational predicate formative (Traugott 2001). Whether this point were conceded or not, it would seem to have no bearing whatsoever on the validity of the unidirectionality hypothesis as such, inasmuch as a principled functional explanation for the counterexample would seem to be available. Counterexamples which can be explained on functional grounds in fact strengthen, rather than undermine, a generalized functional principle such as unidirectionality in grammaticalization – which has *never* been framed (contra Campbell (2001)) by its proponents as an inviolable formal constraint on language grammars.

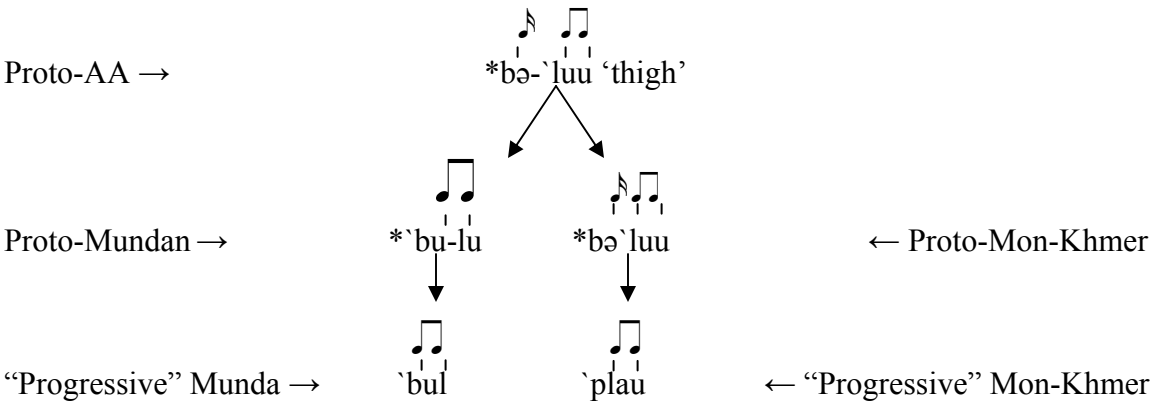


Figure 6 – Rhythm and the opposite typological drifts of Munda and Mon-Khmer (adapted from Donegan and Stampe (1983:346))

Although research into historical-comparative Tibeto-Burman prosody is not yet sufficiently advanced to enable sweeping characterizations of the type offered by Donegan and Stampe for Austro-Asiatic, available data from languages of the Tani branch and some not-too-distant neighbours offer support for Donegan and Stampe's claims in almost every respect. Like Mundan, many modern Tani languages (including Galo) exhibit a basic (C)V(X) syllable structure, are suffixing/postpositional, exhibit geminate clusters and monophthongal vocalism, progressive harmonization, regressive consonant lenition and register/word rather than contour tones, and have undergone coda-reductions at both syllable and (increasingly, over time) word levels. Several more South-Easterly Tibeto-Burman languages exhibit a typology more closely comparable in these respects to Donegan and Stampe's Mon-Khmer (Figure 7).

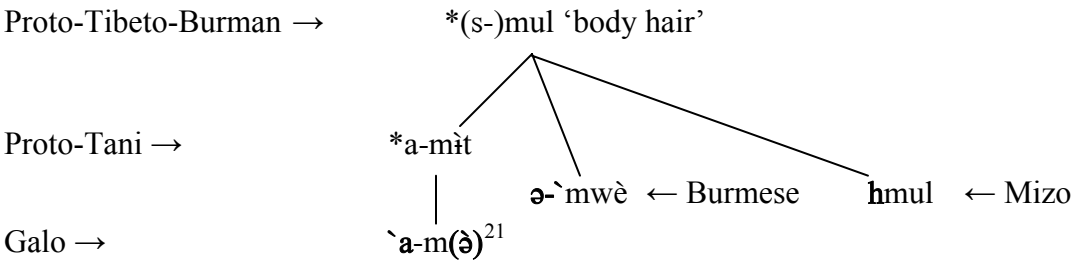


Figure 7 – Rhythm and typological drift in Tibeto-Burman (PTB reconstruction by Matisoff (2003))

As is argued in more detail in Post (2007:§2), the historical morphological and phonological facts taken together suggest a scenario in which a previously isolating, analytical language with a basically morphosyllabic typological profile became increasingly synthetic and agglutinating in consort with a rhythmic shift to a trochaic pattern.

The suggestion made here, however, is that prosody has not only *driven* developments in some aspects of Tani grammatical organization (such as the shift from

[21] Although not exhibited here, expected cases of prefixal root-harmonization are also commonly (if irregularly) attested in Tani, as in Lare Galo ***ɨhɨɨ*** 'wood' < PT ****a-*** 'Noun Prefix' + ****sɨŋ*** 'wood').

monosyllabic, simplex root to disyllabic, complex lexeme as the basic lexical unit, and the fusion of free sequences of simplex functional morphemes into complex functional words), it has in effect *stayed one step ahead* of grammatical organization by creating word-level units which are subject to reanalysis by speakers – and which the grammar may eventually adjust to accommodate.

9. The views of Galo native speakers on "words"

A topic on which this paper has been perhaps strangely silent thus far is that of the views of Galo native speaker-writers toward the identification of "words". Indeed, the phonological word-grammatical word "mismatch" in Galo *may* be a problem for linguistic analysis, but surely when Galo native speaker-writers consult their intuitions, they are able to report whether a particular string of morphemes should be divided into one, two or three words, etc.? Well, unfortunately, no. Throughout the early stages of my research into Galo and other neighbouring Tani languages, I felt myself almost perpetually bewildered and confounded by the reticence with which native speaker consultants offered judgments concerning the number of "words" represented by a particular string. Different consultants might privilege relatively larger or smaller words – potentially, observing grammatical or phonological units of analysis – or, the same consultant might be inconsistent in his/her judgments, whether on the same or on different occasions. (38)-(41) illustrate two sentences from a recorded text; (38) and (40) represents this author's transcription while (39) and (41) represent the independent transcription of one of my consultants. Note particularly the transcription of the predicates in **bold**. While, in my attempt to consistently represent (what I took to be) phonological words in my transcription, I represented two "words" in each case, my consultant represented one "word" in (39) and two "words" in (41). Note that the grammatical and phonological conditions in both cases are almost identical.

(38) *taɬkə́...pətûp arúu lokkə̀...nêndə kulà...**kekkáa kù**.* ← **Author's transcription**
tatɬ́k = əə pətùp arúu lokə̀ = əə nèn-dò(o)-kú-là(a) kéK-káa-kú = ̀
frog=TOP container hole LOC.ABL=TOP exit-IPFV-CMPL-NF flee-PF-CMPL=FI
'The frog got out of the container and escaped.'

(39) *tatike petup aru lokke nendo kula **kekkaku**.* ← **Consultant's transcription**

(40) *pətûp lò mɨ̀əm...**cəpkâa kù**.* ← **Author's transcription**
pətùp = lo bɨ̀ɨ = əm cə̀p-káa-kú = ̀
container=LOC 3.SG=ACC pinch-PF-CMPL=FI
'He got stuck inside the container.'

(41) *petublo miem **cebka ku**.* ← **Consultant's transcription**

When my consultant was later asked, he saw the discrepancy immediately, and confessed that he felt that the decision was ultimately arbitrary: both transcriptions were, in his view (and in those of others subsequently asked) "correct". Cases such as this are extremely common in my data, and, in short, the intuitions of Galo native speakers

concerning the identification of "word boundaries" in some general sense applicable to Galo writing are, in my experience, inconsistent at best.

Since Galo has traditionally existed as an unwritten language,[22] this has historically not been a problem. The Galo word for 'word' *agóm* also means 'language', 'speech' and 'point/matter', and might perhaps be more accurately translated as 'item of speech'; there are no native means for referring to more or less large speech items. In fact, the writing systems of most Asian languages traditionally lack strict representation of "word" boundaries in the strict manner observed by most European languages; Thai script, for example, to this day demarcates only *sentence* and *phrasal* boundaries via spaces in written text. However, as the majority of modern Galo settle on a Roman-based medium, the conventional representation of "word" boundaries in Roman-based scripts with which the Galo are familiar – particularly, English – is presenting major challenges to conventionalization, with no clearly principled solution in evidence as of this writing. Time will tell, of course, how this challenge will eventually be addressed.

10. Conclusion

The purpose of this paper has been first, to illustrate a case of "mismatch" among grammatical and phonological words in Galo; second, to illustrate a number of effects of this mismatch in the organization of Galo grammar, and third, to suggest a possible general explanation for these facts in terms of a historical shift in the rhythmic profile of Galo or its ancestral language(s). The facts of Galo reviewed here would appear to support a independent definitions of "word" at phonological and grammatical levels of analysis – neither of which are directly governed by the other, nor by a third, more general type of unit. They would also, however, suggest the existence of a general functional pressure toward consolidation or unification of grammatical and phonological "words" over time. A further suggestion has been made to the effect that, in Galo at least, the driving force behind the diachronic innovation of new word structures may be primarily prosodic, and that the grammar may subsequently restructure itself to accommodate the now-prevailing word shapes.

Abbreviations

References

Aikhenvald, A. Y. (2002). "Typological parameters for the study of clitics, with special reference to Tariana." In R. M. W. Dixon and A. Y. Aikhenvald, Eds. *Word: A Cross-Linguistic Typology*. Cambridge, Cambridge University Press: 42-78.

Campbell, L. (2001). "What's wrong with grammaticalization?" *Language Sciences* 23(2-3): 113-161.

[22] A Roman-based script has been in casual use for some decades at least, called "Padam" script by Galo due to its earlier use in the context of Christian missionary work among the Paadam-Mising tribespeople. However, no standardization of its use or systematic adaptation for Galo has ever been attempted; most Galo accordingly consider Padam script to be inadequate to their purposes, and for the Galo language to thus in principle "lack writing".

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, University of California Press.

Dixon, R. M. W. and A. Y. Aikhenvald (2002). "Word: A typological framework." In R. M. W. Dixon and A. Y. Aikhenvald, Eds. *Word: A Cross-Linguistic Typology*. Cambridge, Cambridge University Press: 1-41.

Donegan, P. J. and D. Stampe (1983). "Rhythm and the holistic organization of linguistic structure." *Chicago Linguistic Society* 19(2: Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax): 337-353.

--- (2004). "Rhythm and the synthetic drift of Munda." In R. Singh, Ed. *The Yearbook of South Asian Languages and Linguistics*. Berlin/New York, Mouton de Gruyter: 3-36.

Enfield, N. J. (2004). "Adjectives in Lao." In R. M. W. Dixon and A. Y. Aikhenvald, Eds. *Adjective Classes: A Cross-Linguistic Typology*. Oxford, Oxford University Press: 323-347.

Matisoff, J. A. (2003). *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley, University of California Press.

Nevis, J. A., B. D. Joseph, et al., Eds. (1994). *Clitics: A Comprehensive Bibliography 1892-1991*. Amsterdam/Philadelphia, John Benjamins.

Plank, F., Ed. (1995). *Double Case*. New York, Oxford University Press.

Post, M. (2006). "Compounding and the Structure of the Tani Lexicon." *Linguistics of the Tibeto-Burman Area* 29(1): 41-60.

--- (2007). *A Grammar of Galo*. PhD Dissertation. Research Centre for Linguistic Typology. Melbourne, La Trobe University.

--- (forthcoming 2008). "Verbs of posture, existence, location and possession and their grammaticalization pathways in the Tani languages." In S. Morey and M. Post, Eds. *North East Indian Linguistics*. New Delhi, Foundation/Cambridge University Press India.

--- (under review). "Predicate derivations in the Tani languages." In S. Morey and M. Post, Eds. *North East Indian Linguistics II: Papers from the Second International Conference of the North East Indian Linguistics Society, Feb. 5-9, 2007*. Gauhati University, Assam, India.

Prince, A. and P. Smolensky (2002). *Optimality Theory: Constraint Interaction in Generative Grammar (ROA Version)*. Rutgers University and Johns Hopkins University.

Sapir, E. (1921). *Language*. New York, Harcourt, Brace.

Traugott, E. C. (2001). *Legitimate counterexamples to unidirectionality*. Paper presented at Freiburg University, October 17th, 2001.

Zwicky, A. M. (1994). "What is a clitic?" In J. A. Nevis, B. D. Joseph, D. Wanner and A. M. Zwicky, Eds. *Clitics: A Comprehensive Bibliography 1892-1991*. Amsterdam/Philadelphia, John Benjamins: xii-xx.