

# John Benjamins Publishing Company



This is a contribution from *Studies in Language* 33:4  
© 2009. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

# The phonology and grammar of Galo “words”

## A case study in benign disunity

Mark W. Post

Research Centre for Linguistic Typology

“Words” may be independently defined and identified in Galo (Tibeto-Burman > Western 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 phonological “words” which consist of two grammatical “words”, and grammatical “words” which consist of two phonological “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, phonological word, grammatical word, syntax/prosody interface, Tibeto-Burman languages, Tani languages

### 1. Introduction

It is by now well-known that the concept of a “word” — taken as a general linguistic object, not restricted to any particular linguistic subdomain — is far from unproblematic (see Dixon and Aikhenvald (2002), among others). 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. Most previous work in this area has focused on relatively

marked cases in which forms which may appear to be a “word” according to certain grammatical criteria behave more like an “affix” (hence no longer a “word”) according to certain phonological criteria; such units are typically (for better or for worse) described as “clitics”, and their host environments as “clitic groups”.<sup>1</sup> Far rarer are works treating cases in which what appears to be an “affix” according to certain grammatical criteria behaves phonologically as though it were a “word”. Some very recent works have begun to broach this topic (see especially Hall and Hildebrandt (2008)), although again usually in terms of a marked subset of a given language’s forms.

Despite these undoubted advances in the problematization of “word” typology, some commonly-held assumptions remain which are seemingly not limited to any particular theoretical or methodological orientation or tradition — first, that a generalized 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 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).<sup>2</sup> 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 (i.e., limited to a particular subdomain of the grammar and/or relatable to a restricted type of morpheme).

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 (Western) 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), among others) *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. For an example of the type of data with which we will mainly be concerned — “words” which occur within the Galo predicate complex — first compare (1)–(2).<sup>3</sup>

- (1) `zabdù  
 záp-dùu  
 talk-IPFV  
 ‘talking’
- (2) `zabráp `duukù  
 záp-ráp-dùu-kú  
 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). 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 the form, 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, again 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 generalized 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, from a general perspective, the various senses of “word” identified by Dixon and Aikhenvald in their well-known (2002) study, in the process clarifying the senses in which terms referring to “words” and related linguistic forms 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; §9 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.<sup>4</sup> This is the sense in which a native speaker of a language may be aware of a given item 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 *triggered processes or rules* (internal and external sandhi). Thus, for example, English *started* [ˈstɑːr.təd] is a single phonological word in that it exhibits a single stress accent, exhibits word-internal assimilation behaviour at the stem-suffix boundary (/t/ → [r]), and so on. While not all linguists assume the basic independence of this type of “word” unit from morphosyntactic structure, works which argue forcefully for a fundamentally

phonological “word” unit (generally, between “foot” and “phrase”) include Nespor and Vogel (1986), papers in Hall and Kleinhenz (1999), and a recent compilation of papers introduced by Hall, Hildebrandt et al (2008).

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 cannot be interrupted, 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 (GMSEA) tradition at least. An “affix” is a morpheme which is an *dependent grammatical constituent* of a word to which it is in some sense 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 genitive *'s* as in *the king's dog* is a clitic since it is a constituent of the genitive phrase *the king's* of which *king* is head (it could just as easily be enclitic to another genitive phrase constituent, as in *the king of France's dog*). For fuller and more general accounts 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 illustrations of the realization and interaction of these concepts in terms of synchronic and diachronic aspects of Galo grammar. First, however, we turn to a brief overview of some relevant linguistic features of Galo. The description is based on the 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

A transitional (Eastern-convergent) member of the Western Tani branch of Tibeto-Burman spoken by around 30,000–40,000 hill-tribespeople in the North East

Indian state of Arunachal Pradesh, Galo is a basically synthetic and agglutinating language, with statistically verb-final constituent order, accusative case-marking and three major lexical classes noun, adjective and verb (adverbs are primarily derived). Noun phrase and predicate structures are quite distinct; most nominal operators are functional words, postpositions and/or phrasal enclitics, while most predicate operators are suffixes. Verbal cross-referencing of core arguments (“agreement”) is not found, nor are there any word-level nominal inflections other than personal pronominal number and case (3).

- (3) *ɲún nám aráa lò indù.*  
 [ɲunù]<sub>NP.S</sub> [námó aráa=lo]<sub>NPOBL</sub> [ín-dùu]<sub>PRED</sub>  
 1.PL [house interior=LOC] [go-IPFV]  
 [N RN=POSP] [ROOT-SFX]  
 ‘We’re going inside the house.’

Galo exhibits a robust structural distinction between independent and dependent predicates/clauses (a.k.a. “finite/non-finite” or “main/subordinate”) (4). Clearcut instances of verb-serialization have not been found in modern Galo; however, what may once have been post-head serialized verbs now occur as a very large and productive set of bound predicate-derivational formatives, as the ‘Exhaustive’ derivation *-ɲám* ‘EXH’ in (4); some additional discussion of predicate structure and constituency will be found in the section on grammatical wordhood in §4 below.

- (4) *ɲún báal donám tó.*  
 [ɲunù]<sub>NP.A</sub> [báa-là(a)]<sub>PRED.NF</sub> [dó-ɲám-tó]<sub>PRED.FIN</sub>  
 1.PL roast-NF eat-EXH-PFV  
 ‘We roasted (it) and (then) ate it up.’

A basic unity *syllable = morpheme* underlies the organization of much of Galo grammar, just as in many other GMSEA languages (Matisoff 1991; Bickel 2003); this tendency prevails among both lexical roots and grammatical suffixes, as *dó-* ‘eat’ and *-dùu* ‘IPFV’. A relatively much smaller number of synchronically noncompositional polysyllabic roots and suffixes also occur, as *kahi-* ‘hide (oneself); conceal (something)’ and *-ɲoohí* ‘PROG’, although these are usually reconstructible to earlier compositions of monosyllabic forms (for example, the preceding two examples contain a lexicalized (noncompositionally-occurring) Reflexive suffix *-hí* ‘REFL’).

Importantly, and unlike in a great many other Sino-Tibetan languages, roots and their dependents alike are always in principle grammatically *bound* in Galo.<sup>5</sup> That is to say, while there are certain functional classes whose members include simplex, monosyllabic free forms (primarily, pronouns, demonstratives, particles, postpositions and articles), it is in principle not possible to deploy a simple, monosyllabic lexical root or any of its dependents as a grammatical word in any function

(including imperatives). Some additional discussion and exemplification of this point, which is quite central to the paper’s main arguments, will be found in the section on grammatical words in §4.

Basic syllable structure in modern Galo is  $(C_i)V(X)$ , in which  $C_i$  is one of the possible initial consonants  $p/t/c/k$ ,  $b/d/z/g$ ,  $m/n/\eta$ ,  $l/r/j$ , or  $s/h$ ,  $V$  is an obligatory vowel  $a/i/u/e/o/\text{ɨ}/\text{ə}$  and  $X$  is either a nucleus-identical (lengthened) vowel  $V$  (i.e.,  $\text{ɨ}$ ) or one of the possible final consonants  $C_p$   $p/t/(c/)k$ ,  $(b/d/z/g)$   $m/n(/n/\eta)$ ,  $(l/r(/j))$ ,  $(s)$ .<sup>6,7</sup> Of the preceding list of  $C_p$  only the non-parenthesized forms occur in the underlying forms of morphemes; parenthesized forms are introduced in word-medial position following application of internal sandhi processes, some of which will be discussed in the section on phonological word in §5 below. Galo is a quantity-sensitive language, meaning that syllables are either light/monomoraic or 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 this section, we discuss some properties of grammatical words in Galo. Primary focus will be on lexical words, although most of the general principles to be discussed below apply to functional word types equally.

Definition of the grammatical word in Galo depends fundamentally on the three-way distinction between *morpheme*, *word* and *phrase*. Morphemes include roots and suffixes, grammatical words are composed of these, and grammatical phrases are in turn composed of words. Since the vast majority of Galo morphemes are morphologically bound, and must undergo an operation of some kind in order to stand as a pronounceable form-meaning unit — in effect, to “move up” to the level of the word — grammatical words can be seen to occupy a critical *intermediate* position in Galo grammar, organizing non-pronounceable morphological material into pronounceable forms which are capable of deployment in phrasal syntax (5).

<p>(5) <i>hodûm dorɲi</i>          ho-dùm          [[PFX:HIGH.ANIMAL-barking.deer]<sub>N</sub>          ‘two barking deer’</p>	<p><i>dór-ɲi</i>          [CLF:HIGH.ANIMAL-TWO]<sub>ENUM</sub>NP</p>



#### 4.1 Morpheme to word

Looking in more detail at the structure of nouns and adjectives, we find that the vast majority (> 99%) are etymologically complex and dimorphemic, and consist either of two compounded (morphologically bound) roots or a single prefixed (morphologically bound) root. A very small number of nouns and adjectives with mixed semantic values (i.e. which do not constitute one or more natural classes) appear to be morphologically simplex, that is consisting of a single unaffixed, uncompounded root. However, phonological changes apply to ensure that such forms meet a Minimal (bimoraic) word constraint (§5.1.5); for example, the monomoraic root *ɲí-* ‘person’ (as in *ɲipàk* ‘non-hill-tribal’ or *ɲimà* ‘wife’) exhibits nuclear lengthening in the simple word form *ɲíí* ‘person; human being’ (Table 1). In this sense, no nominal or adjectival roots can stand as grammatical words *as such*; an operation of some kind (compounding, prefixation or the application of domain-oriented constraints) is required to license their appearance at the word level.<sup>8</sup>

**Table 1.** Structure of most lexical nouns and adjectives

Type	Class	Initial	Gloss	Final	Gloss	Term	Gloss
[ROOT]	N	—	—	<i>ɲí-</i>	‘person’	<i>ɲíí</i>	‘person’
[ROOT]	ADJ	—	—	<i>zèe-</i>	‘green/blue’	<i>zèe</i>	‘green/blue’
[PFX-ROOT]	N	<i>a-</i>	‘PFX’	<i>kíi-</i>	‘guts’	<i>akíi</i>	‘belly/guts’
[PFX-ROOT]	ADJ	<i>a-</i>	‘PFX’	<i>hòo-</i>	‘long/tall’	<i>ahòo</i>	‘long/tall’
[ROOT-ROOT]	N	<i>là-</i>	‘leg/foot’	<i>cáà-</i>	‘digit’	<i>lácàà</i>	‘toe’
[ROOT-ROOT]	ADJ	<i>là-</i>	‘leg/foot’	<i>zìn-</i>	‘stretch’	<i>lázìn</i>	‘outstretched, of legs’

Turning to verbs, we find that the vast majority (>90%) consist of a single morphologically bound root, to which may be productively attached any of a very wide variety of derivations and inflections, according to intended meaning and syntactic function.<sup>9</sup> In Table 2, note that any “Vroot” can freely cross-combine with either of the listed “Dependents” to form a grammatical word (in this case, a derived subject nominal or an inflected final predicate, respectively); no Vroot can stand as a grammatical word in simplex form, under any conditions.

**Table 2.** Structure of most lexical verbs, with a sample set of possible dependents

Vroot	Gloss	Dependent	Gloss
<i>hí-</i>	‘die’		
<i>ín-</i>	‘walk/go’	<i>-nà</i>	‘NZR:SUB’
<i>dám-</i>	‘beat (as with fists)’		
<i>zí-</i>	‘give’	<i>-diuu</i>	‘IPFV’

The above does not cover all possible word-formation types in Galo, for which a full paper of similar length would be required. The main point to note here is that we are able to clearly establish the possible structures of words in terms of their constituent formatives, which are all in principle morphologically bound and unable to themselves qualify as grammatical words.

#### 4.2. Word to phrase

Turning now to grammatical phrases, we find that phrasal structures in Galo are similarly well-defined in terms of constituency and order, and that the constituents referenced by ordering paradigms and rules are in all cases either grammatical words or other phrases which are likewise composed of grammatical words. First consider the maximal structure of a common noun-headed noun phrase (Figure 1); in all cases, the constituents referenced have either grammatical word or phrase status.

GENP — PRHD — RELC — MNOM — NOM — (RELC) — ENUM — RN — QN — PSHD

**Figure 1.** Order of common nominally-headed noun phrase elements (head is underlined; GENP = Genitive phrase, PRHD = Pre-head referential or referential-relational marker (demonstrative or demonstrative postposition), RELC = Relative clause, MNOM = Modifying nominal, NOM = Nominal, ENUM = Enumerative expression (Classifier and/or Numeral), RN = Relator noun, QN = Qualifying noun, PSHD = Post-head referential and/or relational marker(s) (demonstrative, demonstrative postposition, article, postposition, or licensed sequence thereof, in order)

A number of phrasal properties follow from the “word” status of phrasal constituents. First of all, items which are licensed for ellipsis (mainly, phrasal heads) only permit ellipsis of the entire constituent, not any subpart (no matter what its constituent status in the respective word). So, for example in (5), repeated for convenience as (6), *hodùm* ‘barking deer’ can be omitted in conditions of high predictability, but neither *ho-* nor *dùm-* may be independently ellipsed. Similarly, *dorɲi* could be omitted, although with a difference in meaning (i.e., enumerative classification would no longer be a feature of the reference); however, neither *dór-* nor *ɲi-* could be independently removed.<sup>10</sup>



- |     |                                 |                        |
|-----|---------------------------------|------------------------|
| (9) | <i>ηó bàal dorá.</i>            | <i>*ηó báa(-) dorá</i> |
|     | ηó báa-làa. dó-rá               | ηó báa- dó-rá          |
|     | 1.SG roast-NF eat-IRR           | 1.SG roast- eat-IRR    |
|     | ‘I’ll roast (it) and eat (it).’ |                        |

As is also shown in Figure 2, predicate stems may be expanded by one or more of a very large set of *predicate derivations* (> 300 members), which provide an interesting and provocative dimension to the analysis of Galo predicate grammar. While many predicate derivations are homophonous with and clearly relatable to (other) lexical roots — usually though not always, verb roots — in modern Galo they occur as bound predicate stem-expanding formatives (Post 2009). 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 (10). A certain amount of reordering within the predicate complex is possible — subject to semantic compatibility restrictions — however semantic differences always obtain which relate to the leftward scope of predicate operators. In (11), note that the Causative derivation falls under the scope of the Procedural derivation, while the reverse is the case in (10).

- (10) *tíi-ηám-còo-mò-làa=kèe!*  
 [[[imbibe]<sub>ROOT</sub>-EXH-FIRST-CAUS]<sub>STEM</sub>-IPTV.SDIR]<sub>WORD</sub>[=HORT.POL]<sub>PCL</sub>  
 ‘Let them **finish** drinking **first**, will you (before they are to go on to other things)?’
- (11) *tíi-ηám-mò-còo-làa=kèe!*  
 [[[imbibe]<sub>ROOT</sub>-EXH-CAUS-FIRST]<sub>STEM</sub>-IPTV.SDIR]<sub>WORD</sub>[=HORT.POL]<sub>PCL</sub>  
 ‘**First** let them **finish** drinking, will you (before you do whatever it is you plan to do)?’

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

### 4.3. Interim summary

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. As an outcome of their fixed status

at the level of the syntactic phrase, grammatical words are in principle uninter-ruptible by syntactic words, and neither can any of their internal constituents be subjected to syntactic ellipsis.

## 5. Phonological word in Galo

In this section, we adopt a phonological perspective on the Galo “word”. 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*

The domain of stress-assignment in Galo is the phonological phrase, which usually corresponds well to a syntactic phrase of some kind and is typically identifiable in terms of boundary phenomena such as tonal downdrift and intonational pauses. The metrical foot is strongly trochaic (falling), with primary stress accent generally falling on the initial mora of a minimally bimoraic, usually disyllabic, phonological word. Primary phonetic correlates of stress accent are relative amplitude and full (non-reduced) vocalic specification, with the latter being the most reliably-recognizable in analytical practice. 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.<sup>11</sup> In (12), we find seven phonological words, each realizing a single metrical foot bearing a single stress accent. Note here the discontinuity between the number of phonological words expressed on the surface and the number of grammatical words expressed in the interlinearization.

- (12) [<sup>ˈ</sup>abó <sup>ˈ</sup>taníi]...[<sup>ˈ</sup>míikáa <sup>ˈ</sup>nammá]...[<sup>ˈ</sup>gumbòk <sup>ˈ</sup>zaaláa <sup>ˈ</sup>jù].  
 abó-taníí            míi-káa-nam=ə            gùm-bók=záa=làa.            jùu  
 father-mankind char-TENT-NZR:RLS=TOP lean-DOWN/SOUTH=REAL=NF REP  
 ‘Abo Tani [lit. Father 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 (12), the effects of Triggered foot-strengthening are observed in *-nam=ə* ‘NZR:RLS=TOP’, realized [<sup>ˈ</sup>nammá]; another example is *tabə=ə* ‘snake=TOP’, realized [<sup>ˈ</sup>tabbá].<sup>12</sup>

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 (13). For a fuller description of these and other similarly stress-related morphophonological processes, see Post (2007:§4.1.4).

- (13) [*ˈlâb(ə)rəm ˈgeegáp ˈnammá*]...  
 lâbòr=əm                    géé-gáp-nam=ə  
 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 (i.e., they cannot usually stand as a grammatical word; see §4). Rather, the surface Tone-Bearing Unit (TBU) is a *phonological word*; thus, the underlying tones of sub-word morphemes 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 simplex 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 3).<sup>13</sup>

Table 3. Direct projection of underlying root tones in phonological words

Form. 1	Gloss	Form. 2	Gloss	Word	Gloss
<i>ní-</i>	‘person’	—	—	<i>níi</i>	‘person’
<i>ta-</i>	‘MDIM’	<i>bá-</i>	‘snake’	<i>tabá</i>	‘snake’
<i>ta-</i>	‘MDIM’	<i>bà-</i>	‘sugar cane’	<i>tabà</i>	‘sugar cane’
<i>híi-</i>	‘urinate’	<i>-nam</i>	‘NZR:RLS’	<i>híinám</i>	‘to urinate’
<i>hii-</i>	‘plug/clamp’	<i>-nam</i>	‘NZR:RLS’	<i>híinàm</i>	‘to plug/clamp’

When a phonological word consists of multiple underlying TBUS, its surface tone is derived by rule according to the patterns exemplified in Table 4.

**Table 4.** Derivation of surface tones from multiple underlying tones

Form. 1	Gloss	Form. 2	Gloss	Word	Gloss	Pattern
<i>lák-</i>	‘arm/hand’	<i>cáə-</i>	‘digit’	<i>lakcáə</i>	‘finger’	H + H → H
<i>lák-</i>	‘arm/hand’	<i>ci-</i>	‘left’	<i>lakci</i>	‘left hand/arm’	H + L → L
<i>là-</i>	‘leg/foot’	<i>cáə-</i>	‘digit’	<i>lácəə</i>	‘toe’	L + H → L
<i>là-</i>	‘leg/foot’	<i>ci-</i>	‘left’	<i>laci</i>	‘left leg/foot’	L + L → L

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 clear meaning to the phonological word *duukù*, s/he is usually able to identify its tone as Low/Tense just as readily as that of *lácəə* ‘toe’ in Table 4.

### 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 across a phonological word boundary. Underlyingly vowel-initial functional morphemes (including postpositions/enclitics) tend not to exhibit a glottal stop onset; resyllabification across such boundaries is therefore possible (14).<sup>14</sup>

- (14) *porók-luggóm zuud<sup>o</sup>lá...zañigó zal<sup>o</sup>ká*  
 porók-lugó=əm úu-dó(o)-là(a)=’ añi=gə alá:k=ə  
 fowl-crow=ACC awake-STAT-NF=NFI1 self=GEN hand/arm=TOP  
*dollòm zidú...*   
 dolò=əm í-dùu=’  
 paddy=ACC 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 (14) — 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 (15), the speaker hesitates at the word onset,

phonetically 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).

- (15) *nó...gogbooló...tá/...tapâa lammò.*  
 nó gók-boolo tá/ tá-pàa-là(a)-mò  
 2.SG call-COND listen/ listen-ATTN-ABIL-NEG  
 ‘If you call, she...she surely won’t hear.’

Now compare (16), in which the speaker corrects himself “mid-grammatical-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. (17) and (18) simply illustrate that both *-dó(o)* ‘STAT’ and *-dàk* ‘COS’ are indeed grammatical predicate suffixes, in the sense outlined in §4, and not auxiliaries or some other such independent grammatical word type.

- (16) *allô-rôa nè*  
 allò-roò nè  
 tomorrow-day.after.tomorrow IRR.TMP.PUNC  
*caatâr dóo/...dagêe bə...*  
 càa-târ-dó(o)/-dàk-é=bə  
 ascend-TO.END-STAT/-COS-IPFV.DISJ=AVZR  
 ‘After they come up/...come up tomorrow or the next day...’

- (17) *bùl càadoo kú.*  
 bulù càa-dó(o)-kú  
 3.PL ascend-STAT-CMPL  
 ‘They’re starting to come up (as a general practice).’

- (18) *bùl caadâk kú.*  
 bulù càa-dàk-kú  
 3.PL ascend-COS-CMPL  
 ‘Now they’re coming up (in contrast to whatever they were doing before).’

### 5.1.5 *Word minimality*

As in perhaps most other languages of the world, if not indeed all (McCarthy and Prince 1986: 10), phonological words in Galo tend strongly to consist of a single metrical foot, and to, therefore, be minimally bimoraic. The constraint does not seem to apply to most functional words, including articles (such as the Individuator in (7)) and Postpositions (such as the Adverbializing subordinator in (16));



such forms are most often analyzed as clitics by Post (2007), and exempted from word minimality requirements. However, all lexemes appear to be subject to bimoraicity requirements, and exhibit nuclear lengthening in the very few cases in which an underlyingly monomoraic root structure appears to be exhibited. For an example, see again §4.1.

## 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*.<sup>15</sup>

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 *gók-boolo* ‘call-COND’ [gogbooló], as in (15).

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: *nóm-káa* ‘swallow-PF’ ‘swallowed’ is realized [ɲoŋkáa], not \*[ɲomkáa], and *ín-p̃hì* ‘go-REACH’ ‘reach’ is realized [imp̃hì], not \*[inp̃hì]. 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.

### 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 in *híidáa* ‘stick’) and word-medially when following a vowel *and* when standing as onset of a heavy ((C)VX) syllable (as in *ihíi* ‘wood’). Word-medially when following a consonant and/or when standing as onset of a light ((C)V) syllable, *-h* surfaces [s] (as in *namsúu* ‘stinky’ and *isi* ‘water’).

Now also consider the behaviour of the Reflexive suffix *-hí* ‘REFL’ in (19)–(20). In (19), *-hí* ‘REFL’ follows a consonant-final morpheme as the final syllable within a phonological word; its initial fricative is therefore realized [s]. In (20), *-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].

- (19) *zapsí tokè!*  
*záp-hí-tó=kèe*  
 talk-REFL-IPTV.ODIR=HORT.POL  
 ‘Talk to yourself!’

- (20) *zabmín hít<sup>o</sup>kè!*  
 záp-mín-hí-tó=kèè  
 talk-RECP-REFL-IPTV.ODIR=HORT.POL  
 ‘Talk amongst yourselves!’

Next, an underspecified consonant *-K* is found in Galo which reflects a Proto-Tani syllable-final consonant of uncertain form (but which is likely to have been *\*-ɕ*; see Sun (1993:§4.3.2.3)). 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 [cinnám] and *cíK-ùp* ‘throw.spear-SHATTER’ ‘throw a spear such that something shatters’, realized [cigúp]. In (21), 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 (22), 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 on opposite sides of a *phonological word boundary*.

- (21) *plastiká ziddùu kú.*  
 plastik=ə                      zíK-dùu-kú  
 plastic(<Eng)=TOP melt-IPFV-CMPL  
 ‘The plastic is now melting.’
- (22) *dooná plastikám amzík duukù*  
 dooná=ə plastik=əm                      ám-zíK-dùu-kú  
 sun=TOP plastic(<Eng)=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 is believed that grammatical and phonological criteria for (or perhaps “aspects of”) word-hood should necessarily converge upon the same

type of unit, 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”; the phonological criteria? Or the grammatical criteria? If one set of criteria is discounted as a means of identifying “words” in the most basic or general sense, then what sort(s) of unit *do* these other criteria identify?

With reference to (22) above, let us briefly entertain two possible alternatives to the identification of a grammatical word/phonological word “mismatch”, to determine whether a unifying principle can or can not in fact be salvaged: first, one in which phonological factors would be discounted, with grammatical criteria in turn treated as the sole or primary 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 (22) should be re-transcribed as in (23).

- (23) *dooná plastikám amzíkduukù*  
 dooní=ə plastik=əm                      ám-zíK-dùu-kù  
 sun=TOP plastic(<Eng)=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 (24), 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).

- (24) *tapèk perrò*  
 tapèk-**perrò**  
 leech-jungle.leech  
 ‘jungle leech’

In (24), the initial and final terms share a common **bolded** root *pèK*- ‘leech’, seemingly reflecting PT *\*pac*. 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, (24) exhibits the same seemingly word-oriented characteristics as does the predicate in (23), and Galo speakers are quite clear on the potentially independent grammatical and phonological “word” statuses of both *tapèk* and *perrò* in (24). Writing “*tapèkperrò*” 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 the numbers of “words” in a string, we might re-transcribe (23) as in (25).

- (25) *dooná plastikəm amzík duukù*  
 dooní=ə plastik=əm      **ám-zíK**      **dùu-kú**  
 sun=TOP plastic(<Eng)=ACC **roast-MELT AUX.IPFV?-CMPL**  
 ‘The sun is melting the plastic.’

The challenge with respect to (25) 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.<sup>16</sup>

There are two problems with this possible solution. The first is that it would be left to explain why it should be that a form with a single function ‘Imperfective’ should surface as a suffix in a sentence like (21), but as an “auxiliary” in a sentence like (25), when the only apparent difference between the two sentences is that in (25), the predicate stem has been derivationally expanded.<sup>17</sup> Furthermore, analysis of *-dùu* as an auxiliary head in (25) 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 (§4.2). In Table 5, 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’.

**Table 5.** “Suffix/Auxiliary” alternations in the predicate complex

<i>riġlġi dù</i>	‘wanting to wash it’	<i>rikkák liidù</i>	‘wanting to wash it clean’
<i>rikkên dù</i>	‘easy to wash’	<i>rikkák kendù</i>	‘easy to wash clean’
<i>riksí dù</i>	‘washing oneself’	<i>rikkák hidù</i>	‘washing oneself clean’
<i>riġnám dù</i>	‘washing everything’	<i>rikkák ħamdù</i>	‘washing everything clean’
<i>riġbáa dù</i>	‘still washing it’	<i>rikkák bædù</i>	‘still washing it clean’
<i>riġnóo dù</i>	‘habitually wash it’	<i>rikkák ħoodù</i>	‘habitually wash it clean’
(...)		(...)	

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 *-lii* ‘DESD’ should have a different word-level grammatical status in the first column of Table 5 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 morphologically bound predicate formative in both cases. The only difference is that it occupies the second syllabic position in a grammatical predicate in the first column, and the third syllabic position 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 disappear.

## 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 important implications for our understanding of the synchronic and diachronic organization of Galo grammar.

### 7.1 “Functor fusion”

As in a number of other Tibeto-Burman languages, it is very common in Tani languages to encounter disyllabic forms — or forms which were at one time disyllabic — which appear to represent fusions of two or more previously simplex functional morphemes. Dual and plural pronouns usually 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 *lokə* ‘ABL’, which seems to derive historically from the sequence *lo* ‘LOC’ plus *\*kə* ‘GEN/ABL’.<sup>18</sup>

In Galo, we find literally dozens of such forms, reflecting fusions of a very wide variety of morpheme types. Table 6 presents a small selection of disyllabic functional words in Galo which are believed to have arisen via fusion of two previously simplex morphemes.

Table 6. Selection of disyllabic fused forms and their presumed etymologies

Cat.	Word	Gloss	Form. 1	Cat.	Gloss	Form. 2	Cat.	Gloss
PRO	<i>ɲunù</i>	1.PL	* <i>ɲó</i>	PRO	1.SG	* <i>lù(ɲ)</i>	RN?	'group'?
DEM.POSP	<i>təkà</i>	SEMB.UP	<i>tà</i>	DEM	DST.UP	* <i>kà</i>	POSP	GEN
DEM.POSP	<i>higì</i>	SPRX.IND	* <i>hì</i>	DEM	SPRX	* <i>go?</i>	ART	IND
PCL	<i>maad̩i</i>	TAG.2	(-) <i>máa</i>	PINFL?	NEG	<i>d̩i</i>	PCL	WOND
PCL	<i>bərə̀e</i>	CJEC	<i>bà</i>	POSP	SBRD	<i>rè̀e</i>	PCL	PQ
PCL	<i>larè̀e</i>	DUB	<i>là(a)</i>	CNJ	CNJ	<i>rè̀e</i>	PCL	PQ

Space prevents us from entertaining a full description of even the abbreviated list of “functor fusions” in Table 6, each of which has its own associated “story” of source construction, functional extension, and, in many cases, phonological change. Instead we’ll focus here on what may be the most salient forms in terms of number and systematicity of members, frequency of use, clarity of diachronic development and typological relevance, the so-called “demonstrative postpositions” described in Post (2007:§7.4).

Deriving historically from NP-peripheral sequences of the form DEMONSTRATIVE + POSTPOSITION, demonstrative postpositions in Galo count as single grammatical words which retain both the deictic and the relational-marking functionalities of their two simplex source forms. In (26) and (27), we see that both demonstratives and postpositions occur at the right edge of a noun phrase (also cf. Figure 1). While post-head is the most frequent position in which demonstratives in Galo are attested, it is also possible for demonstratives to occur pre-head, as well as simultaneously pre- and post-head, “bracketing” the NP (28); however, neither pre-head nor “bracketing” patterns are permitted in the case of postpositions (29).

- (26) *dolúu tà*  
 doolúu tà  
 village DST.UP  
 [NOM DEM]<sub>NP</sub>  
 ‘that village up there’

- (27) *dolúu ló*  
 doolúu lo  
 village LOC  
 [NOM POSP]<sub>NP</sub>  
 ‘in the village’
- (28) *t̂ doolúu t̂*  
 t̂ doolúu t̂  
 DST.UP village DST.UP  
 [DEM NOM DEM]<sub>NP</sub>  
 ‘that there village up there’
- (29) \**lo doolúu (ló)*  
 lo doolúu (lo)  
 LOC village (LOC)

From an initially compositional sequence DEMONSTRATIVE + POSTPOSITION, Galo demonstrative postpositions would then have arisen via a grammatical fusion of these two morphemes; this hypothetical process is schematized in (30).<sup>19</sup>

- (30) \**dolúu t̂lò* → *doolúu tolò*  
 doolúu t̂=lo doolúu tolò  
 village DST.UP=LOC village DST.UP.LOC  
 [NOM DEM=POSP]<sub>NP</sub> [NOM DEM.POSP]<sub>NP</sub>  
 \*‘in that village up there’ ‘in that village up there’

There are two important things to note here. The first is that the modern Galo form exhibits progressive vowel-harmonization, an irregular but pervasive phonological process found widely in Tani languages and which generally correlates to the word-level lexicalization of previously compositional morpheme sequences (Post 2006).<sup>20</sup> The second is that unlike the simplex postposition illustrated in (29), the new demonstrative postposition is able to occur pre-head or “bracketing” an NP, like a simple demonstrative (31). Also like all simple demonstratives, demonstrative postpositions may be used pronominally (32)–(33).

- (31) *t̂l doolúu tolò*  
 tolò doolúu tolò  
 DST.UP.LOC village DST.UP.LOC  
 [DEM.POSP NOM DEM.POSP]<sub>NP</sub>  
 ‘in that there village up there’
- (32) *t̂̂ nà*  
 t̂ nà  
 DST.UP DECL  
 ‘It’s that one up there.’

- (33) *tól nà.*  
 tolò          nà  
 DST.UP.LOC DECL  
 ‘It’s up there.’

In sum, retaining the relational marking and deictic functionality of both source forms, and taking on the distributional patterning of the demonstrative source form, a new category of grammatical word has arisen with category-specific (non-predictable) internal phonological structure: the demonstrative postposition.

The important thing to note here, though, is that the fusion itself was almost certainly at least partially motivated by the fact that the source collocations would have been frequently uttered as *phonological words*. Patterning phonologically as a unit, the grammar came to treat them as a unit, and the emergence of a novel grammatical category is the result.

“Functor fusions” are perhaps even more noticeable 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’. Largely semantically equivalent, the first form derives from a collocation of a Change of state predicate inflection -*dàk* ‘COS’ plus a seemingly non-native Additive particle *kòm* ‘ADD’.<sup>21</sup> 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’ (34). When suffixed to a disyllabic stem, the Concessive suffix is realized as an independent phonological word (35); this follows the basic behaviour of any relatively lengthy predicate word, such as those exemplified in Table 5 above.

- (34) *âmbə rìdàk kòm, nokkəm zərjâa rá.*  
 əmbə      rì-dakkòm      nò-kə=əm      zər-jàa-rə  
 ANAP.PADV happen-CONC 2.SG-GEN=ACC spin-MORE-IRR  
 ‘That (being the case) **notwithstanding**, my (top) will spin longer than yours (will).’

- (35) *ŋún kaamáa dakkòm, doolúu hígì dookáa hidù!*  
 ŋunù káa-máa-dakkòm      doolúu hígì      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’ (36). Naturally, from this function they can then be extended to marking an argument NP with a concessive sense ‘even’ (37).

- (36) *ə̀g dakkòm, bulù...tukâa bulù, apúk-anág bə̀...caamâa rə̀.*  
 ə̀gə̀      **dakkòm** bulù tukkâa bulù apúk-anák=bə̀ cə̀a-mâa-rə̀  
 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.’
- (37) *hijni...ərək go dakkòm álʳə̀ əmbóo lo...*  
 hijni    ə̀rək=go **dakkòm** alə̀-rə̀    ə̀m-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’ – which seems to have its ultimate source in a positional verb *dàk* ‘stand’ (Post 2008) — has ever been capable of functioning independently as a noun marker; therefore, it seems unlikely that noun-marking uses such as those in (36)–(37) could have arisen compositionally.<sup>22</sup> The second is that, although it is conceivable that (35)–(37) could all represent instances of a single subordinating particle/word *dakkòm*, capable of both predicate and nominal scope, this analysis is untenable in (34) (where *-dakkòm* licenses a grammatical predicate by suffixing directly to a bound verbal root) — and yet, the functional values of *-dakkòm* in (34) and (35) 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 (35) would in fact have *encouraged* reanalysis of *dakkòm* as an independent *grammatical* word, ultimately leading to extensions in its functionality — just as in the case of demonstrative postpositions sketched out above. This hypothetical development is sketched in Table 7; in Table 7, note that while the grammatical status of *(-)dakkòm* is hypothesized to have undergone several developments, its phonological value remains unchanged throughout.

**Table 7.** Development of noun particles from predicate subordinators via reanalysis of independent phonological words as grammatical words (for glosses and translations, cf. (35)–(37))

subordinated inflected predicate	kâa-mâa-dàk=kòm	[kaamâa dakkòm,]	→
subordinated predicate stem	kâa-mâa-dakkòm	[kaamâa dakkòm,]	→
subordinated nominal	ə̀gə̀ dakkòm	[ə̀gə̀ dakkòm,]	→
particle-marked nominal	ərək go dakkòm	[ərək gó dakkòm...]	

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 with a variety of 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 versatile particle (*ə*)<sup>í</sup> ( “’” represents an extra-high tone) occurs three times in (38), 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*)<sup>í</sup> 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 importance.

- (38) *ayni lokkə hikai maan<sup>ə</sup>mə (...)* *ənə bədək lokkə<sup>í</sup>*  
*ayni lokkə hikai-maa-nam=ə anə bəə-dək lokkə=(ə)<sup>í</sup>*  
 bit ABL.SRC teach(<Ind)-NEG=TOP mother bear-COS ABL.SRC=**EMPH**  
*annə<sup>í</sup> abb<sup>w</sup>ə<sup>í</sup> hób<sup>ə</sup>gə*  
*anə=ə=(ə)<sup>í</sup> abó=ə=(ə)<sup>í</sup> hobə=əgə*  
 mother=COP.IPFV=**EMPH** father=COP.IPFV=**EMPH** mithun=ANAP.IND  
*moodii lo...rəho nà zaatə.*  
*moodii=lo rə-rho-nà zaatə=ə*  
 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 (39) (cf. also (12) above). In Post (2007), the non-standard symbol  $\equiv$  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.<sup>23</sup>

- (39) *hodùm hór<sup>o</sup>cìn rəkú ə́í maané.*  
 hodùm-horə=cìn      rə-kú $\equiv$ (ə)í $\equiv$ máa=né  
 barking.deer-boar=ADD live/exist-CMPL $\equiv$ EMPH $\equiv$ NEG=DECL.ADM  
 ‘Even wild game was *nowhere* to be found, see (because a tiger had scared them all away).’

In (39), the full form of the grammatical predicate word is *rə-kú-máa* ‘live/exist-CMPL-NEG’; Admonitive declarative particle *né* is treated by Post (2007) as an enclitic which is structurally outside the predicate word (it can also mark focused noun phrases, for example). That (ə)í ‘EMPH’ occurs within, and not at the boundary of, a grammatical word, is clear from the fact that the right edge of *rəkú\_* is not in fact a grammatical word boundary, inasmuch as Completive suffix *-kú* ‘CMPL’ is incapable of licensing a final predicate (40); to form a grammatical predicate word in this case, a licensing predicate inflection such as Negative *-máa* ‘NEG’ must be suffixed to the morphologically bound predicate stem.

- (40) \**hodùm hór<sup>o</sup>cìn rəkú.*  
 hodùm-horə=cìn      rə-kú  
 barking.deer-boar=ADD live/exist-CMPL

Furthermore, and crucially for our purpose here, Emphatic particle (ə)í ‘EMPH’ can *only* interrupt the predicate at a *phonological word boundary*. Since *rəkú* (as in (39)) constitutes one phonological word, not two (i.e., *rə\_kú* is not a phonological word boundary), the sentence in (41) is unacceptable. (42) simply shows that this is not a problem related to the relative ordering of (ə)í and *-kú* (*rə-máa* ‘live/exist-NEG’ would constitute an acceptable final predicate on its own).

- (41) \**hodùm hór<sup>o</sup>cìn rə́í kumá.*  
 hodùm-horə=cìn      rə $\equiv$ (ə)í $\equiv$ kú-máa  
 barking.deer-boar=ADD live/exist $\equiv$ EMPH $\equiv$ CMPL-NEG

- (42) \**hodùm hór<sup>o</sup>cìn rə́í má.*  
 hodùm-horə=cìn      rə $\equiv$ (ə)í $\equiv$ máa  
 barking.deer-boar=ADD live/exist $\equiv$ EMPH $\equiv$ 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.<sup>24</sup>

### 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 four syllables across 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.<sup>25</sup>

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 8.

**Table 8.** Primary formatives of Multiword predicate constructions

Type	Example	Gloss	F1	Gloss	F2	Gloss
DCV	dó...làa-	‘subsist’	dó-	‘eat’	làa-	‘take’
	pi...pàa-	‘make a living’	pi-	‘craft’	pàa-	‘get’
DPD	-pàa...là(a)	‘AS CAN BE MAN-AGED’	-pàa	‘ATTN’	-là(a)	‘ABIL’
	-kúp...-lék	‘HELTER SKELTER’	-kúp	‘UPSIDE DOWN’	-lék	‘RIGHTSIDE UP’
ESR	-báá...-jáá	‘CONSTANTLY’	-báá	‘DUR’	-jáá	‘RDUP’
	-pèn...-jèn	‘FULLY SEPARATE’	-jèn	‘SEP’	-jèn	‘RDUP’

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 8).

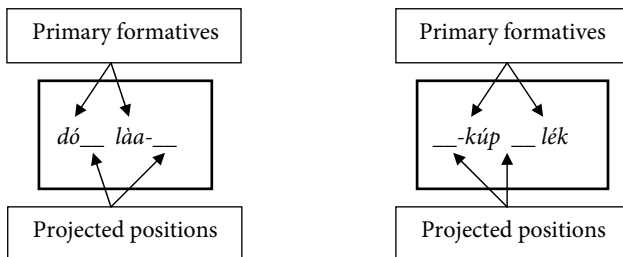


Figure 3. DCV Template

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 (43)–(44).

- (43) *dolâa laalâa kú*  
 dó-là(a) làa-là(a)-kú  
 eat-NF take-NF-CMPL  
 ‘came to make a living’
- (44) *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 a Multiword predicate; minimally, there are two, and in examples such as (43)–(44), 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 (43), 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.

Here again, then, the mismatch between grammatical and phonological words in Galo presents the language with opportunities to develop constructions which, while entirely ordinary in terms of their distinct phonological and underlying

grammatical structures, exhibit a complex interaction which ultimately produces a highly intricate surface form.

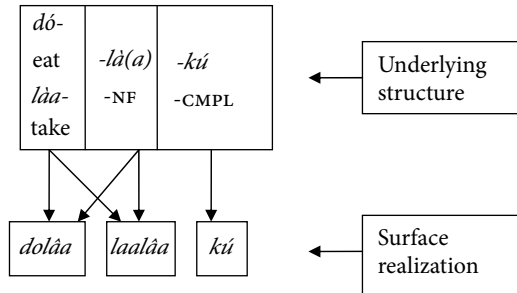


Figure 5. Mapping relation between underlying and surface structures of Multiword predicate (cf. (43))

#### 7.4 “Auxiliation” in polar question responses

GMSEA languages 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 (45)–(46).

- (45) A: *tacên duurè?*  
 tá-cèn-dùu=rèè  
 listen-KNOW-IPFV=PQ  
 A: ‘Do you understand?’

- (46) B: *tacên dù.*  
 tá-cèn-dùu  
 listen-KNOW-IPFV  
 B: ‘Yes, I do.’

Importantly, the response in (46) requires repetition of the predicate head; it is not possible, for example, to simply reply “*dù*” (treating the Imperfective suffix as though it were an auxiliary-like predicate head). However, a very small number of predicate derivations — seemingly, only three out of the hundreds available — appear to license a different kind of response. A question containing either the Desiderative derivation *-lîi* ‘DESD’, Ability derivation *-là(a)* ‘ABIL’ or Reflexive derivation *-hî* ‘REFL’ may be answered by treating the predicate derivation as though it were an auxiliary verb-like predicate head, omitting the predicate root (47)–(48).

- (47) *nó əpàk larəi?*  
 nó əpàk-là(a)-rə=(ə)ì  
 2.SG discard-**ABIL**-IRR=PQ  
 ‘Will you be able to quit (smoking)?’
- (48) *larà.*  
 Ø-là(a)-rə  
 Ø-**ABIL**-IRR  
 ‘Sure I will.’
- (49) *nó əpàk zirəi?*  
 nó əpàk-zí-rə=(ə)ì  
 2.SG discard-**BEN**-IRR=PQ  
 ‘Will you throw it away for him?’
- (50) \**zirə.*  
 Ø-zí-rə  
 Ø-**BEN**-IRR

Other predicate derivations do not license this type of response (49)–(50).<sup>26</sup>

It is currently unknown whether this irregular behaviour in Lare Galo polar question responses is more likely to represent a morphosyntactic conservation from an earlier stage of the 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)* ‘**ABIL**’. Like Non-final suffix *-là(a)* ‘**NF**’ (43) and a handful of other predicate formatives including *-dó(o)* ‘**STAT**’, Ability *-là(a)* ‘**ABIL**’ is subject to the irregular but pervasive process of Phrase-medial truncation. In Phrase-medial truncation, a qualifying morpheme with an underlyingly long rhyme surfaces with a short rhyme when occupying a particular position in a predicate string. The position in which the short rhyme is exhibited is lexically-specified according to the morpheme (second in the case of *-dó(o)* ‘**STAT**’, third in the case of *-là(a)* ‘**ABIL**’). In all other positions, the rhyme surfaces with the etymologically conservative long form (again, compare (43)).<sup>27</sup> Note, then, that (48) exhibits the *truncated* form.<sup>28</sup> In 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 (47) 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.<sup>29</sup>

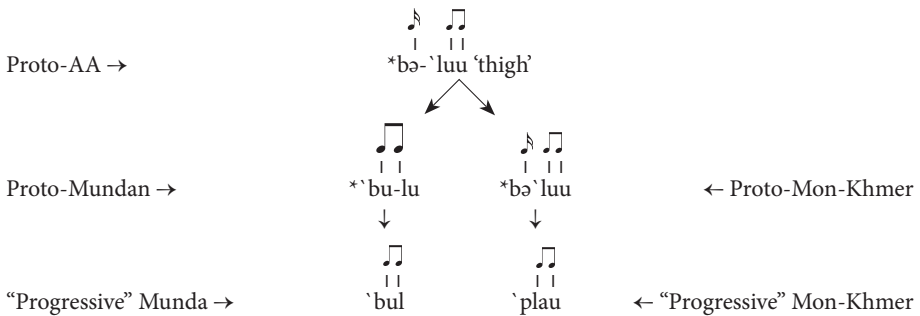
## 8. A diachronic perspective: Rhythm and the synthetic drift of Tani

The preceding subsections have illustrated some seemingly disparate facts about Galo grammar which are all argued to relate in some fundamental way to the existence of a “mismatch” between phonological words and grammatical words in Galo. As we noted in the outset of the paper, absent entirely from traditional grammar, this phenomenon remains rarely-identified across languages (outside of discussions of “clitics”). In a series of very recent papers by Bickel, Hall, Hildebrandt, and Schiering targeting the notion of phonological (or “prosodic”) word more generally, an array of related phenomena are introduced in Kiranti languages (Tibeto-Burman, Nepal), among others (see e.g. Schiering, Hildebrandt et al. (MS), Hildebrandt (2007), Bickel (2007), Hall and Hildebrandt (2008) and Bickel, Hildebrandt et al. (MS)). However, such accounts have not usually attempted a general explanation.<sup>30</sup> My purpose in the present section is to provide an admittedly somewhat speculative account of how the types of phenomena treated in above sections may have come about in Galo.

In two important and far-reaching papers, Donegan and Stampe (1983; 2004) put forth a theory of morphosyntactic change in which overall shifts in morphological typology — which are often assumed by scholars to result in some general and usually not well-specified way from “language contact” — are suggested to be more directly caused by a language-internal shift in *prosodic* organization (which, however, may itself derive ultimately from language contact — potentially, spanning long geographical distances). 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 by Donegan and Stampe 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,

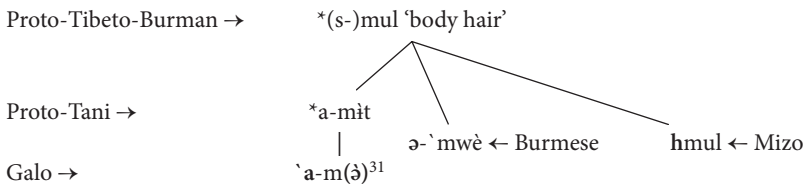


synthetic/agglutinating structures, a (C)V(X) syllable canon, stable, geminate clusters, stable, monophthongal vocalism, harmonic prosodies, and register rather than contour tones. Languages which develop iambic (rising) patterns should undergo the opposite types of developments: prefixes/prepositions, less synthetic structures, complex onset clusters, diphthongal vocalism, and contour rather than register tones. The basic form of the principle is first illustrated in Figure 6.



**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 Tibeto-Burman 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 — whether at the Proto-Tani stage or earlier — 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 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 functional reanalysis by speakers — and which the grammar eventually adjusts itself to accommodate.

## 9. Conclusion

The main purposes of this paper have been, first, to illustrate a case of “mismatch” among grammatical and phonological words in Galo, and second, to illustrate a number of effects of this mismatch in the synchronic and diachronic organizations of Galo grammar. Finally, a general explanation was offered in terms of a historical shift in the rhythmic profile of Galo or its ancestral language(s). The facts of Galo reviewed here would thus appear to support 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. Such facts would thus tend to argue both against the viability of generalized formal constraints which would unify the phonological and grammatical expression of “words”, such as those hypothesized by Prince and Smolensky (2002), as well as against the general assumption in functional linguistics that “words” automatically operate, and may be referenced as, a single type of generalized linguistic object.

The Galo facts would also, however, suggest the existence of a general *functional pressure* toward consolidation or unification of grammatical and phonological “words” over time. Assuming, as was argued in §8, that the innovation of new phonological word structures in Tani languages followed a shift in their prosodic profile, we then have evidence of subsequent restructuring of the lexicon and grammar to accommodate the innovated word shapes: for example, from a proto-lexicon containing mainly monosyllabic grammatical and phonological words such as PT \**lak* ‘arm/hand’ and \**keŋ* ‘finger’, a modern lexicon has arisen which contains mainly disyllabic grammatical and phonological words (reflecting earlier compounds) such as *lakcǎǎ* ‘finger’ (cf. Table 4). Similarly, as was discussed in §7.1, fusions of demonstrative + postposition sequences such as *tǎ* ‘DST.UP’ and

*lo* ‘LOC’ — once sequences of two monosyllabic grammatical and phonological words — have resulted in the emergence of the new, hybrid morphological category of Demonstrative postpositions such as *tolo* ‘DST.LOC.UP’ — a single, disyllabic grammatical and phonological word.

Thus, evidence from Galo would suggest that — like so many other putative linguistic “universals” — the widely-held notion of a unified, general category “word” (to the extent that it exists) must be expressed as a functionally-motivated, diachronically-operating tendency rather than as a synchronic formal constraint.

## Abbreviations

A	Transitive subject	MDIM	Masculine diminutive
ABIL	Ability	MNOM	Modifying nominal
ABL	Ablative	N	Noun
ACC	Accusative	NEG	Negative
ADD	Additive	NF	Non-final
ADJ	Adjective	NFI	Non-final intonation
ADM	Admonitive	NOM	Nominal
ANAP	Anaphoric	NP	Noun phrase
ART	Article	NZR	Nominalizer
ATTN	Attainment	O	Transitive object
AUX	Auxiliary	OBL	Oblique
AVZR	Adverbializer	PADV	Pro-adverbial
BEN	Benefactive	PCL	Particle
C	Consonant	PF	Perfect
CAUS	Causative	PFV	Perfective
CJEC	Conjectural	PFX	Prefix
CLF	Classifier	PL	Plural
CMPL	Completive	POL	Polite
CNJ	Conjunction	POSP	Postposition
CONC	Concessive	PQ	Polar question
COND	Conditional	PRED	Predicate
COP	Copula	PRHD	Pre-head demonstrative
COS	Change-of-state	PSHD	Post-head demonstrative, article, or postposition
DCV	Discontinuous compound verb	PTB	Proto-Tibeto-Burman
DECL	Declarative	PT	Proto-Tani reconstruction by Sun (1993)
DEM	Demonstrative	PUNC	Punctual
DER	Derivation		

DESD	Desiderative	QN	Qualifying noun
DPD	Discontinuous predicate derivation	RDUP	Reduplicant/reduplication
DSJ	Disjunct	REAL	Reality
DST	Distal	RELC	Relative clause
DUB	Dubitative	REP	Reported information
DUR	Durative	RLS	Realis
ENUM	Enumerator	RN	Relator noun
EMPH	Emphatic	S	Intransitive subject
ESR	Expressive semi-reduplication	SDIR	Self-directed
EXH	Exhaustive	SEMB	Semblative
EXIS	Existential	SEP	Separation
FI	Final intonation	SFX	Suffix
FIN	Final	SG	Singular
GEN	Genitive	SPRX	Speaker-proximate
GENP	Genitive phrase	SRC	Source
HORT	Hortative	STAT	Stative
ICEP	Inceptive	SUB	Subject
INAN	Inanimate	TENT	Tentative
IND	Individuator	TMP	Temporal
INFL	Inflection	TOP	Topic
IPFV	Imperfective	TBU	Tone-bearing unit
IPTV	Imperative	V	Vowel
IRR	Irrealis	X	Segment (C or V)
LOC	Locative		

## Notes

1. For an early compilation of the very large literature on clitics see Nevis, Joseph et al. (1994); for a general discussion of clitics see Zwicky (1994), as well as the more updated survey in Aikhenvald (2002).
2. 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 a precise specification of the grammatical-word/phonological-word relationship in the functionalist literature; in practice, however, a fundamental (if prototypical) unity seems to be assumed (as in Givón (2001 [1984])).
3. Transcription follows IPA except where  $c = [tɕ]$  and  $z = [dz]$ . Tones are High/Plain  $\hat{\ }_$ , Low/Tense  $\grave{\ }_$ , or Rising-Falling  $\hat{\ }_$ . The surface (pronounceable) Tone-Bearing Unit is the usually polysyllabic *phonological word* (given in the first line of examples), while the underlying Tone-Bearing Unit is the (often unpronounceable) *morpheme* (given in the second line of examples). That

is, Galo is a “word tone” language, largely in the sense of Mazaudon (in press). Some further discussion of tone patterning in Galo is also found in §5.1.2 below).

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

5. While it is undoubtedly the case that the modern forms of many languages traditionally identified as “extreme isolating” or “monosyllabic” such as Chinese, Thai and Vietnamese in fact contain large numbers of compounds and bound morphemes (Chao 1968; Matisoff 1991) — and thus may not be as “extremely” isolating as had at one time been claimed — it is also the case that in these languages a great number of roots *may* occur as simplex lexemes, whether they also occur as complex word formatives or not. This is *not* the case in Galo, in which the overwhelming majority of morphemes are grammatically bound and unpronounceable in isolation, full stop.

6. *K* is an underlyingly underspecified consonant; its surface realization is discussed in §5.2.2 below.

7. 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 *aii* ‘tooth’ < Proto-Tani \**afii*), although the syllabicity of such forms remains somewhat unclear. They do not impact the present discussion, however, and may be safely disregarded.

8. Note that the basically phonological criterion of moraic lengthening is not being used here to *define* a grammatical word. Rather, a grammatical word is *defined* in terms of its relationship between a grammatical morpheme and a grammatical phrase, and thus *recognised* in terms of its actual deployment in a grammatical phrase. The phonological criterion is invoked here to underscore the fact that simplex lexical words are not simply “roots patterning as words” (i.e. “exceptions” to the general morphological distinction between “morpheme” and “word” levels); rather, they are words which are non-prototypically *specified for* simplex morphological content, and which are subject to word-level phonological constraints as an outcome of their word-level grammatical licensing.

Further to this point, it is worth acknowledging that the putatively “bound” root *zèe-* ‘green/blue’ (Table 1) does not exhibit any observable structural changes when “moving” to the level of the “word”; its “root” and “word” forms thus being homophonous, there is no positive evidence in favour of viewing *zèe-* as morphologically bound (the fact that it alternates with the prefixed form *jazèe* ‘green/blue’ offers little comfort, as there is no reason in principle why a free root could not also be prefixed). That said, there is also no evidence *against* assuming that *zèe-* is subject to the same language-wide principles which govern the surface contents of other non-clitic words: the Minimal word constraint applies, but since the form already qualifies, no audible changes occur. I adopt the assumption that this is the case, while conceding that a skeptical reader might have good reason to view *zèe-* as a legitimate counterexample to the basically bound status of roots in Galo.

9. A relatively much smaller number of basic verbs are morphologically complex, though based on synchronically non-productive formations (or formations with limited productivity). Most such forms also function as nouns or adjectives, and appear to either count as instances of zero-

derivation, or else reflect earlier lexicalization of pre- or post-head dependents (either incorporated nominal roots (in the pre-head case) or following derivational suffixes).

10. Note that this is not simply a feature of the grammatical interdependence of classifiers and numerals in Galo. In functions such as anaphoric reference, a classifier can be used independently of a numeral — in which case, a *prefixed* form of the classifier is used. See Post (2007:§8.2.2.2) for discussion and examples.

11. Certain types of underlyingly trisyllabic sequences exhibit evidence of resolving into two-footed phonological words, although this analysis remains problematic. While somewhat peripheral to the focus of this paper, some additional remarks on the topic may be found in Post (2007: 174).

12. A very similar process of “onset prothesis” in Belhare is described in optimality-theoretic terms by Bickel (MS-1998).

13. Note that tones in Galo are conventionally marked via a single diacritic over the penultimate mora of a TBU; however, the tonal specification is not in fact a property of that mora, but is rather a property of the overall unit in which it occurs (underlying morpheme and/or surface phonological word). For additional information on the phonetic realization and phonological derivation of tones in Galo, see Post (2007: 195).

14. In Northwestern Galo, the glottal stop onset is phonemic, and is underlyingly assigned to some vowel-initial lexemes but not to others. The phonemicity of glottal stop onsets appears to be in the process of breaking down in Lare Galo.

15. In other Galo dialects, such as Pugo Galo, regressive nasal assimilation is additionally observed.

16. 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 un-inflecting post-head auxiliary verbs — in this case, *diiu-* ‘sit; stay; exist (animate); be at (for an item construed as ‘sitting’). Auxiliary-like behaviour of a seemingly cognate form *duj* is still represented in Pagro Mising, a Tani language of the Eastern branch (Post 2008).

17. This assumes, of course, that it is necessary to analyze *-diiu* as a “suffix” in (21), inasmuch as it enables a bound verbal root to stand as a grammatical word (see again §4.1). Continuing to analyze *-diiu* 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 “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.

18. *\*kə̀* does not occur as a postposition in modern Galo, although it does occur as such in Upper Belt Minyong (Tani, Eastern; author’s field notes) and is reconstructed to Proto-Tani as such by Post (2007); probably, it ultimately reflects Proto-Tibeto-Burman *\*ka* (DeLancey 1984). Modern Galo reflexes include the pronominal Genitive suffix *-kə̀*, as well as the Genitive postposition *gə̀* (seemingly reflecting an earlier process of lenition).

19. The reconstructed form *tə̀lə̀* is in fact attested in Upper Belt Minyong, a Galo-bordering Tani language of the Eastern branch with relatively more conservative segmental features at the word

level (author’s field notes); however, synchronic compositionality has not yet been extensively researched in the case of the Minyong form.

20. Areas of the lexicon in which vowel-harmonization is particularly pervasive include erst-while \**a*- and otherwise-prefixed nominal/adjectival roots (cf. §4.1), complex pronouns (cf. Table 6 line 1) and fused functors such as demonstrative postpositions.

21. Not accepted by most modern Lare Galo speakers as a simple Additive particle, the form seems to have an ultimate origin in the Eastern Tani language Minyong, and to have entered Galo via the Minyong-bordering Pugo dialect (together with a good number of other Minyong forms). Despite dispreferring the simplex form, however, most Lare speakers accept and use the complex fused form.

22. Thus a “polygrammaticalization” account would probably be untenable in this case.

23. A reviewer suggests the (also non-standard) notation “ $\leq \geq$ ” as an alternative, viewed as a combination of clitic “=” and infix “<>”. While sympathetic to the suggestion, I hesitate in order to avoid the impression that versatile particles are capable of occurring within morphemes and/or unanalyzable lexical words; this is seemingly not the case in Galo.

24. A reviewer appropriately points out the similarity between this aspect of Versatile particles’ distribution and the patterning of “endoclitics” such as the person-markers described for Udi (North East Caucasian, Georgia) by Harris (2000). While there are some important differences to be noted — Galo Versatile particles have a far wider distribution than do Udi person markers (cf. again (38)), and unlike Udi person markers, Galo Versatile particles are incapable of occurring inside morphemes — similar challenges are posed here to the Lexical Integrity Hypothesis (DiScullio and Williams 1987), in which it is asserted that the internal (morphological and/or prosodic) structures of words cannot be accessed by the rules of syntax. Clearly, the prosodic structure of the predicate is central to the patterning of Galo Versatile particles.

25. While not ideal exemplars in terms of their contemporary grammatical status and functions, Galo multiword predicates can be understood to fall within the broad family of “elaborate expressions” found widely among GMSEA languages (Matisoff 1988: 39, among others).

26. (50) is unacceptable as a response to (49). If interpreted as a sentence headed by the verb root *zí* ‘give’ — the likely historical source form of the Benefactive suffix *-zí* ‘BEN’ — 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 (49), since it would not address the question concerning ‘discarding’.

27. Ability *-là(a)* ‘ABIL’ is reconstructed as PT \**-lan* (Sun 1993); the expected Lare Galo reflex following regular Final velar nasal deletion with compensatory lengthening is *-làa*.

28. 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. For example, the question *nó rì-là(a)-rǎ=(ǎ)ì* ‘2.SGdo-ABIL-IRR=PQ’ ‘Will you be able to do it?’ is realized *nó rilâa rǎì*, exhibiting the long form in second syllable position. However, a positive response remains in the short form *larǎ*.

29. 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 in this case 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.

30. I assume here that “explanation” of any ultimately biological structure must ultimately involve reference to diachrony, i.e. to its evolution, and that positing a series of “constraints” or other variables from which languages may somehow “choose” does not count as explanation in this sense (Bybee 1988; Givón 2002).

31. Although not exhibited here, expected cases of prefixal root-harmonization are also commonly (if irregularly) attested in Tani, as in Lare Galo *ihii* ‘wood’ < PTs \*a- ‘Noun Prefix’ + \**siŋ* ‘wood’).

## References

- Aikhenvald, Alexandra. Y. 2002. Typological parameters for the study of clitics, with special reference to Tariana. *Word: A Cross-Linguistic Typology*, R.M.W. Dixon & Alexandra Y. Aikhenvald (eds.), 42–78. Cambridge, Cambridge University Press.
- Bickel, Balthasar. 2003. Prosodic tautomorphemicity in Sino-Tibetan. *Language variation: Papers on language variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, David Bradley; Randy J. LaPolla; Boyd Michailovsky & Graham Thurgood (eds.), 89–99. Canberra: Pacific Linguistics.
- Bickel, Balthasar. 2007. Free prefix ordering in Chintang. *Language* 83(1): 43–73.
- Bickel, Balthasar. 1998. Rhythm and feet in Belhare morphology. Ms, Rutgers Optimality Archive #287.
- Bickel, Balthasar; Kristine A. Hildebrandt & René Schiering. 2007. The distribution of phonological word domains: A probabilistic typology. Ms., University of Leipzig, University of Manchester & University of Münster.
- Bybee, Joan L. 1988. The diachronic dimension in explanation. *Explaining Language Universals*, John A. Hawkins (ed.), 350–379. Oxford: Blackwell.
- Campbell, Lyle. 2001. What’s wrong with grammaticalization? *Language Sciences* 23(2–3): 113–161.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley CA: University of California Press.
- DeLancey, Scott. 1984. Etymological notes on Tibeto-Burman case particles. *Linguistics of the Tibeto-Burman Area* 8(1): 59–77.
- DiScullio, Anna-Maria & Edwin Williams 1987. *On the definition of word*. Cambridge MA: The MIT Press.



- Dixon, R.M.W. & Alexandra Y. Aikhenvald. 2002. Word: A typological framework. *Word: A cross-linguistic typology*, R.M.W. Dixon & Alexandra Y. Aikhenvald (eds.), 1–41. Cambridge, Cambridge University Press.
- Donegan, Patricia J. & David Stampe. 1983. Rhythm and the holistic organization of linguistic structure. *Chicago Linguistic Society* 19(2): 337–353 (Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax).
- Donegan, Patricia J. & David Stampe. 2004. Rhythm and the synthetic drift of Munda. *The year-book of south asian languages and linguistics*, Rajendra Singh (ed.), 3–36. Berlin: Mouton de Gruyter.
- Enfield, Nick J. 2004. Adjectives in Lao. *Adjective classes: A cross-linguistic typology*, R.M.W. Dixon & Alexandra Y. Aikhenvald (eds.), 323–347. Oxford: Oxford University Press.
- Givón, T. 2001[1984]. *Syntax: An introduction*, Vol. 1. Amsterdam: John Benjamins.
- Givón, T. 2002. *Bio-linguistics: The Santa Barbara lectures*. Amsterdam: John Benjamins.
- Hall, T. Alan & Kristine A. Hildebrandt. 2008. Phonological and morphological domains in Kyrong Tibetan. *Linguistics* 46(2): 215–248.
- Hall, T. Alan; Kristine A. Hildebrandt, et al. 2008. Introduction: Theory and typology of the word. *Linguistics* 46(2): 183–192.
- Hall, T. Alan & Ursula. Kleinhenz (eds.). 1999. *Studies on the phonological word*. Amsterdam: John Benjamins.
- Harris, Alice C. 2000. Where in the word is the Udi clitic? *Language* 76(3): 593–616.
- Hildebrandt, Kristine A. 2007. Prosodic and grammatical domains in Limbu. *Himalayan Linguistics Journal* 8: 1–34.
- Matisoff, James A. 1988. *The dictionary of Lahu*. Berkeley CA: University of California Press.
- Matisoff, James A. 1991. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20: 469–504.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley CA: University of California Press.
- Mazaudon, Martine. In press. On tone of Tamang and some neighbouring languages. *Proceedings of the symposium Cross-Linguistic Studies of Tonal Phenomena*. Tokyo: Tokyo University of Foreign Studies.
- McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Ms, University of Massachusetts & Brandeis.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Nevis, Joel. A.; Brian D. Joseph; Dieter Wanner & Arnold M. Zwicky (eds.). 1994. *Clitics: A comprehensive bibliography 1892–1991*. Amsterdam: John Benjamins.
- Plank, Franz (ed.). 1995. *Double case*. New York NY: Oxford University Press.
- Post, Mark W. 2006. Compounding and the structure of the Tani lexicon. *Linguistics of the Tibeto-Burman Area* 29(1): 41–60.
- Post, Mark W. 2007. A grammar of Galo. Ph.D Diss., La Trobe University, Research Centre for Linguistic Typology.
- Post, Mark W. 2008. Verbs of posture, existence, location and possession and their grammaticalization pathways in the Tani languages. *North East Indian linguistics*, Stephen Morey & Mark W. Post (eds.), 127–152. New Delhi: Foundation/Cambridge University Press India.
- Post, Mark W. 2009. Predicate derivations in the Tani languages: Root, suffix, both or neither? *North East Indian Linguistics Volume 2*. Stephen Morey & Mark W. Post (eds.), 175–197. New Delhi: Cambridge University Press India.

- Prince, Alan & Paul Smolensky. 2002. Optimality theory: Constraint interaction in generative grammar (ROA Version). Rutgers University & Johns Hopkins University.
- Sapir, Edward. 1921. *Language*. New York NY: Harcourt, Brace.
- Schiering, René; Kristine A. Hildebrandt, et al. The prosodic word is not universal. Ms, University of Leipzig & University of Manchester,
- Sun, Jackson Tianshin. 1993. A historical-comparative study of the Tani branch of Tibeto-Burman. Ph.D Diss., University of California, Berkeley.
- Traugott, Elizabeth Closs. 2001. Legitimate counterexamples to unidirectionality. Paper presented at Freiburg University, October 17th, 2001.
- Zwicky, Arnold M. 1994. What is a clitic? *Clitics: A comprehensive bibliography 1892–1991*, Joel A. Nevis; Brian D. Joseph; Dieter Wanner & Arnold M. Zwicky (eds.), xii–xx. Amsterdam: John Benjamins.

*Author's address*

Mark W. Post  
Research Centre for Linguistic Typology  
La Trobe University  
Bundoora, VIC 3086  
Australia  
m.post@latrobe.edu.au