Against Split Morphology

Robert Hepburn-Gray
University at Buffalo (SUNY)

Proceedings of the 24th International Conference on
Head-Driven Phrase Structure Grammar

University of Kentucky, Lexington

Stefan Müller (Editor)

2017

CSLI Publications

pages 199–211

http://csli-publications.stanford.edu/HPSG/2017

Keywords: SBCG, Word and Paradigm morphology, noun classes, Niger-Congo

Hepburn-Gray, Robert. 2017. Against Split Morphology. In Müller, Stefan (Ed.), *Proceedings of the 24th International Conference on Head-Driven Phrase Structure Grammar, University of Kentucky, Lexington*, 199–211. Stanford, CA: CSLI Publications.

Abstract

In this paper I present data from several Niger Congo languages, illustrating how the paradigms which make up the noun class systems of these languages are problematic to analyze within traditional morphosyntactic frameworks. I outline possible solutions to this problem, and argue for the introduction of an exemplar-based Word and Paradigm (Blevins 2006) approach to morphology within SBCG. I then outline the consequences of this approach for the structure of the SBCG lexicon.

1 Introduction

The Niger-Congo family is perhaps most well known for the distinctive noun class systems which can be found in many of its languages. Although there is considerable diversity to be found within the family itself, comparative research on noun classes has revealed something of a distinctive "Niger-Congo type" of noun class. The features of this type make Niger-Congo noun classes quite different from familiar gender and number systems of the Indo-European family. In particular, I argue that there is little evidence to accept the traditional dichotomy of inflectional and derivational morphological processes referred to as "split-morphology."

The remainder of this paper will proceed as follows. Section 2 reviews the assumptions commonly made in morphological theory. Section 3 illustrates the properties typical of Niger-Congo noun class systems. Sections 4 and 5 examine in more detail two features of Niger-Congo noun class systems which are particularly problematic for a split morphological analysis. Section 6 reviews two modern approaches to paradigmatic structure in morphological theory. Finally, Section 7 outlines the formal analysis proposed for Niger-Congo noun class systems.

2 Assumptions of (Split-)Morphological Theory

Most morphological theory, including that which is currently employed within Sign-Based Construction Grammar (Sag, 2012), often differentiates between inflectional and derivational processes. This division is referred to as "split morphology" by Bauer (1997). In SBCG, this distinction is explicitly represented in the hierarchy of construct(ion) types, as shown here:

[†]I would like to thank the anonymous reviewers for their helpful comments on this paper. I would also like to thank Rui Chaves and Jeff Good for their comments on previous versions of this paper. Lastly, I would like to thank the participants of the conference for the stimulating discussion.

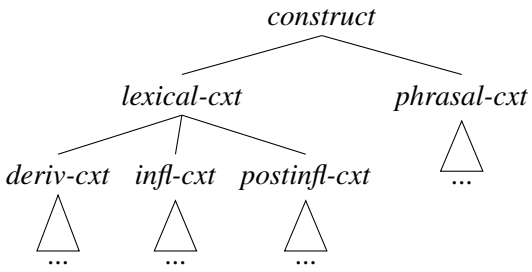

Figure 1: SBCG Construct Types

This traditional division of inflection and derivation has been defined and supported by a number of different criteria, outlined in Stump (2005) and adapted here:

1. Derivation can change part-of-speech class, while inflection cannot
2. Inflection applies to a category without exception; derivation applies sporadically
3. Inflection is semantically regular; derivation is frequently less than fully semantically regular
4. Inflection is syntactically determined; derivation is not
5. Derivational processes apply before inflectional processes

A central assumption of split morphology is the distinction between word and lexeme. Inflectional constructions are said to create words from a lexeme, whereas derivational constructions are said to create new lexemes from old ones. Booij (2012, 5) bases this distinction on what is referred to as an "intuitive difference" between the relationship between words such as *walks* and *walking* on the one hand and *tax* and *taxable* on the other.

This distinction between word and lexeme is also assumed in SBCG, and is represented in the type signatures of the *infl-cxt* and *deriv-cxt*:

(1) Split Morphology in SBCG (Sag, 2012):

$$\textit{infl-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \textit{word} \\ \text{DTR} & \textit{list(lexeme)} \end{bmatrix}$$

$$\textit{deriv-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \textit{lexeme} \\ \text{DTR} & \textit{list(lex-sign)} \end{bmatrix}$$

A final assumption made in most morphological theory is that a theory which is able to describe more language types is superior to one which describes fewer

language types. Here, however, this assumption will not be made, since a an aspect of the morphological theory may be motivated by data from one language, but not for another. Here I assume that languages which present different morphological systems may call for substantially different types of analyses.

3 Typology of Niger-Congo Noun Class Systems

Although noun classes can certainly be found outside of the Niger-Congo family, there are a set of properties common in Niger-Congo noun class languages which give the appearance of distinctive type. These properties, outlined by Kießling (2013, 44-45), are listed below.

1. All nouns are assigned to a limited set of noun classes
2. All nouns control, by virtue of their assignment to a class, a system of concordial agreement which penetrates vast sections of the morphosyntax
3. Class assignment is governed by semantic principles so that classes could be described as semantic networks, but not necessarily synchronically active/cognitively real (Dingemanse, 2006, 22-23)
4. Most noun classes form singular-plural pairs or genders

To illustrate these properties please consider the noun class system of Otoro, a Kordofanian language (Stevenson, 2009), outlined below.

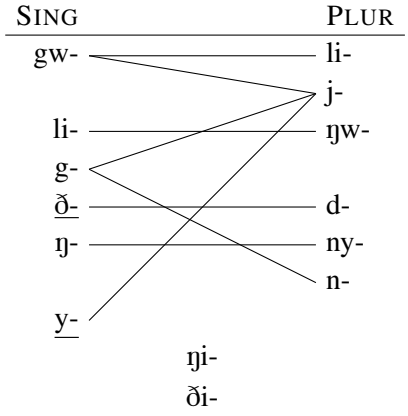

Figure 2: Otoro Noun Class System

GEN	SING	PLUR	GLOSS
gw-/li-	gwiji	liji	‘person’
gw-/j-	gwaʈe	jaʈe	‘tree’
g-/j-	gilöð	jilöð	‘hoe’
ð-/j-	ðimu	jimu	‘scorpion’

Table 1: Example Otoro Paradigms

The diagram in Figure 2 represents the noun class system of Otoro using a format common to Niger-Congo linguistics. As is often the case in Niger-Congo noun class systems, there are classes which participate in multiple ‘genders’, such as *gw-*, *j-*, and *g-*, which form pairs with multiple classes, and *ð-* and *y-*, which participate in single and double class genders, (indicated here by the underlining of the noun class marker). Class markers in the bottom center are never paired. Unpaired classes are typically found to contain mass nouns, abstracts, and liquids in Niger-Congo languages.

A characteristic of many noun class languages of the Niger-Congo family is that number is present semantically, but is not an active morphosyntactic feature (Welmers, 1973). In Indo-European, there exist patterns, such as subject-verb agreement, which are sensitive to number, but not gender. In Niger-Congo, however, systems are often found where there exist no constructions which are sensitive to number distinct from the feature of noun class.

4 Number as a ‘derivational’ process

Lumun (Smits, 2011) represents a particularly irregular number system. It is schematized below:

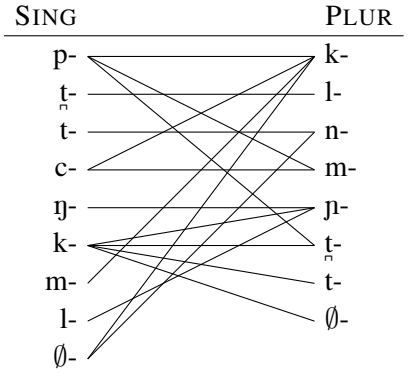

Figure 3: Lumun Noun Class System

It is obvious from the diagram that the minor irregularities seen in the Otoro

noun class system above are far more widespread in Lumun. Note also that many of the noun class markers are phonologically identical to another marking the opposite number category.

Noun class markers have three possible functions in the noun class system: to mark a noun as singular, to mark a noun as plural, or to serve as the class marker of a one-class noun. Inspection of the Lumun noun class system shows that a majority of noun class markers performs all three of these functions in the system.

NCM	SING NCM	PLUR NCM	SINGLE NCM
p-	✓		✓
t̪-	✓	✓	✓
t-	✓	✓	✓
c-	✓		✓
k-	✓	✓	✓
m-	✓		✓
n-		✓	✓
ŋ-	✓		✓
ɲ-		✓	
l-	✓	✓	✓
∅-	✓	✓	✓

Table 2: Lumun Noun Class Markers and their 'Genders'

When these facts are combined with the already noted observation that a singular noun class marker can pair with multiple plural noun class markers (and vice versa), the result is that any given noun class marker is capable of functioning as a marker of many different genders. The most extreme example of this is the noun class marker *k-*, whose genders are listed below:

GEN	SING	PLUR	GLOSS
k-/∅-	kʊmmʊk	ʊmmʊk	'pot/pots'
k-/t̪-	kupú	t̪upú	'peice of k.o wood/k.o wood'
k-/t-	kua	tua	'strand of hair/hair'
k-/ɲ-	kʊkkʊ́	ɲʊkkʊ́	'groundnut/ groundnuts'
p-/k-	pɪra	kɪra	'tree/forest'
c-/k-	cít	kít	'eye/eyes'
∅-/k-	ɪkɛ	kɪkɛ	'giraffe/giraffes'
k-	kəɽɛt		'abusive language'

Table 3: The nine 'genders' of class marker *k-*

This noun class marker participates in nine different genders. Looking at the system as a whole, there are no less than twenty-six genders from only eleven

phonologically distinct noun class markers. Furthermore, the number distinctions in these genders are not semantically regular. Rather than just straightforward singular and plural meaning, many of these noun class markers bear the functions of singulative and collective as well.

Due to this degree of irregularity, Smits argues that number marking in Lumun should be considered a derivational rather than inflectional process. This analysis follows a similar analysis of the noun class system of Swahili by Schadeberg (2001), where it was shown that the noun class system bore more of the properties typical of derivational morphology than those of inflectional morphology. As a consequence of each noun class marker being affixed by a derivational process, the noun class system does *not* involve genders.

The analyses of Smits and Schadeberg have the advantage of eliminating the vast accidental homophony that a gender-based analysis would require in the noun class markers. In addition, the treatment of class rather than gender/number as the base of the system follows the observation by Welmers (1973) that number does not seem to be an active morphosyntactic category.

However, the feature of class is active in the agreement of many nominal dependents, not only in the marking of number on nouns themselves. Agreement is generally considered the inflectional category *par excellence*. The analyses of Smits and Schadeberg would consequently treat class marking as derivational within the nominal domain, but as inflectional within the domain of agreement targets. This type of asymetry is undesirable, and I propose that a superior analysis involves the abandonment of the inflection/derivation assumption for morphological systems such as these.

5 Paradigm Networks

Paradigm networks such as the following can be found throughout the Niger-Congo family (Hepburn-Gray, 2016).

NC Paradigm	*-dooma* 'kaba'	*-taat* 'annona'
si-/mun-	'kaba tree'	'annona tree'
bu-/i-/di-	'kaba fruit'	'annona fruit'
ja-	'leaves of the kaba tree'	'leaves of the annona tree'

Table 4: Botanical Paradigm Network in Baïnounk (Cobbinah, 2013, 319)

In this botanical paradigm network roots referring to specific tree species enter into different noun class paradigms depending on what part of the plant is being referenced. One paradigm refers to the tree itself, a second refers to the fruit of the tree, and a third refers to the leaves of that tree. A second type of paradigm network is shown below.

Class	Acipu	Karishen	Kadonho	Hausa	Gloss
8	*c-cípù*	*∅-rísìnô*	*d-dípó*	k-kɔ́gɔ́	Person
2	*à-cípù*	*ò-rísìnô*	*ò-dípó*	ò-kɔ́gɔ́	People
1		*kò-rísìnô*	*kò-dípó*		Town/Area
6	*cì-cípù*	*tì-rísìnô*	*tì-dípó*	tì-kɔ́gɔ́	Language

Table 5: Ethnic Group Paradigm in Cicipu (McGill, 2007, 61)

This data from Cicipu, a Kainji language, is an example of an ethnic group paradigm network. Here, a root referring to a certain ethnicity can be associated with different paradigms to create different "words", whether this word is referring to a person of this ethnic group, the language spoken by this group, or the area inhabited by this group.

The principle problem of class networks such as these for formal models of morphosyntax is that there is no principled way to identify which of these nouns should correspond the base lexeme from which the other words are derived. A possible solution would involve a separate lexeme, from which all of these words are derived. The problem with this analysis, however, is that this lexeme would somehow have to be barred from entering into any inflectional construction, since none of the above words could correspond to this lexeme without first undergoing a derivational construction.

Koenig (1999, 150) discusses a related example in English, where there appears to be a lexeme which undergoes a mandatory derivational process. This example involves the sets *regress/regressive/regression* vs. **agress/agressive/agression*. The absence of the verb *agress* is explained as a missing root, which is only constructionally introduced in the *agressive/agression* constructions. However, it is not a root that is missing, but rather a fully inflected word which occupies the verb cell in a derivational paradigm.

6 Paradigms in Morphological Theory

With respect to the modeling of paradigmatic knowledge, Stump & Finkel (2013) distinguish between the canonical extremes of the PURE WORD-AND-PARADIGM MORPHOLOGY (PWPM) hypothesis and the PURE EXPONENCE-BASED MORPHOLOGY (PEM) hypothesis. These hypotheses differ with respect to the way two features are represented: inflectional class membership and morphological rules. These differences are represented in the following table.

CRITERION	PWPM	PEM
IC membership	represented by means of a set of lexically listed principal parts	represented by means of a diacritic+one or more stems
Rules	implicative rules formulated in terms of realized cells	rules of exponence formulated in terms of stems

Table 6: Differences between the PWPM and PEM hypotheses (Stump & Finkel, 2013, 265)

Stump and Finkel argue that the first distinction, between the representation of inflectional class with principal parts or with diacritics, is a false one. They proceed to use diacritics in their formalism.

As for rules, Stump and Finkel propose a hybrid model in which rules of exponence are the primary method of representing inflectional relations and implicative rules of referral are used to represent syncretism between paradigm cells. The choice of rules of exponence as primary is based primarily on stem variation in Sanskrit. Here, however, it is assumed that the diversity of morphological systems can correspond to a diversity in types of morphological representation, rules of exponence are not motivated for the Niger-Congo languages discussed so far.

What Stump and Finkel refer to as "the obvious advantage of compactness" (Stump & Finkel, 2013, 266) of the *diacritic+stem* approach is less obvious for these Niger Congo languages. A inflection class diacritic is certainly compact in languages with large inflectional paradigms, but for these Niger Congo languages the paradigm diacritic only specifies two cells, both of which can already be specified by a single morphosyntactic feature (CLASS).

Furthermore, the introduction of paradigm diacritics would be akin to introducing the notion of gender to these languages. As was mentioned above, Schadeberg (2001) and Smits (2011) argue that 'gender-based' analyses are inappropriate for Niger-Congo languages such as Swahili and Lumun. The following section will illustrate the alternate analysis proposed here.

7 Formalism

Based on the issues proposed in the previous sections, I propose here a WORD AND PARADIGM (Blevins, 2006) model of morphology for these Niger-Congo languages. In the theory proposed here, a lexeme is *not* a type of sign. A "lexeme" is simply the knowledge that a set of words is paradigmatically related, as well as whatever information is shared among these words. This captures the notion of the 'abstractive' lexemes introduced in Blevins (2006), where notions such as stems and lexemes have no status within the model, but rather are abstractions over sets

of fully inflected word forms.

The type hierarchy of construction types I propose for Niger-Congo noun class morphology is presented in Figure 4 below. The paradigm cells of principle parts are fully stored in the lexicon. Forms of other paradigm cells, be they 'inflectional' or 'derivational' paradigms, are generated via an analogical construction (*an-cxt*). Compound constructions (*comp-cxt*) are the only morphological construction type which remains distinct from the analogical constructions, since they necessarily include multiple words, and therefore represent a hybrid construction between word and phrase.

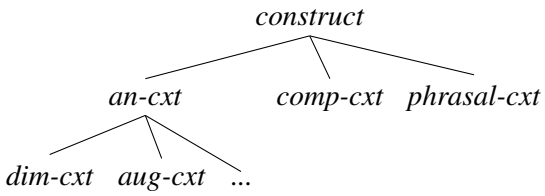

Figure 4: Construct Types

For the representation of morphological information related to a word, I adapt the following type signature of the *head* feature from Koenig (1999).

(2) Head Feature Type Signature:

$$head \Rightarrow \begin{bmatrix} \text{LXM} & \textit{lex-prop} \\ \mu\text{-FEAT(URES)} & \mu\textit{-prop} \end{bmatrix}$$

Information shared between word forms of a lexeme is represented in the LXM attribute. Information that is particular to a specific paradigm cell (or proper subset of paradigm cells) is represented within the attribute μ-FEAT.

The following is the type signature for the sign type *word* of a noun in a Niger-Congo noun class language. The value of the LXM attribute is a valule matrix containing a label (LBL and a semantic frame (FRAME). The LBL feature is shared between all words associated with a single lexeme. Reducing the notion of a lexeme to this feature is meant to mirror the 'abstractive' sense of lexeme taken from Blevins (2006). The value of the FRAME attribute is a semantic frame representing whatever semantics are common to all the words of a lexeme.

(3) Type Signature of Sign type *word*:

$$\begin{bmatrix} \textit{word} & & \\ \text{MORSYN} \mid \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{LXM} & \begin{bmatrix} \text{LBL} & \textit{lxm-lbl} \\ \text{FRAME} & \boxed{1} \end{bmatrix} \\ \mu\text{-FEAT} & \begin{bmatrix} \text{CLASS} & \textit{class} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \boxed{1} \oplus L \end{bmatrix} \end{bmatrix}$$

The only morphological feature necessary for nouns in these languages is CLASS, following the analysis of Schadeberg (2001) in which noun class pairs are not treated as genders, and therefore number is only present as a semantic feature. Here *L* represents the (possibly empty) set of semantic frames which may be associated with a particular paradigm cell.

The figure in (4) illustrates the type signature of the analogical construction. Different types of *an-cxt* take the principle part of a lexeme, and associate the morphosyntactic features of the desired paradigm cell with the new word. The phonology of the word is determined by a function, which takes as input the phonology of the principle part and the features of the new paradigm cell. This function contains the 'implicative rules' of the PURE WORD AND PARADIGM MORPHOLOGY approach, which generates a proportional analogy with the corresponding exemplar paradigm. The exact nature of this phonological function will not be addressed here.

(4) Type Signature of *an-cxt*:

$$\textit{ana-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{F}(\boxed{1},\boxed{2}) \rangle \\ \text{CAT} & \begin{bmatrix} \text{LXM} & \boxed{3} \\ \mu\text{-FEAT} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \begin{bmatrix} \textit{word} & \\ \text{PHON} & \boxed{1} \\ \text{CAT} & \begin{bmatrix} \text{LXM} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

An example of an analogical construction is presented below in (5). This construction illustrates the construction for the *bu-* class of the Baïnounk botanical paradigm network. The daughter of this construction is any word with a frame corresponding to a tree species. The mother of this construction is a word for the fruit of this tree species with the corresponding *class* feature.

(5) The 'fruit' construction of the botanical paradigm network:

$$\textit{bu-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{PHON} & \langle \text{F}(\boxed{1},\boxed{2}) \rangle \\ \text{CAT} & \begin{bmatrix} \mu\text{-FEAT} & \boxed{2}\begin{bmatrix} \text{CLASS} & \textit{bu} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \begin{bmatrix} \textit{fruit-frame} \\ \text{INDEX} & i \\ \text{TREE-SPEC} & j \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{CAT} \mid \text{HEAD} & \begin{bmatrix} \text{LXM} \mid \text{FRAME} & \begin{bmatrix} \text{TREE-SPEC} & j \end{bmatrix} \\ \mu\text{-FEAT} & \begin{bmatrix} \text{CLASS} & si \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

8 Conclusion

This paper has analyzed noun class systems of various Niger-Congo languages. These systems exhibit properties which make them quite different from gender systems of familiar Indo-European languages. The argument was made that the these differences where substantial enough to result in morphological systems where the typical distinction between inflectional and derivational morphological processes is unmotivated. Furthermore it was argued that these systems are most economically modeled using a Word and Paradigm model of morphology. Finally, a formalism incorporating this Word and Paradigm approach into Sign-Based Construction Grammar was briefly sketched. There is plenty of room for further work which will more fully flesh out the formalism, especially the exact nature of the implicative rules and phonological functions which generate non-exemplar word forms.

References

Bauer, Laurie. 1997. Derivational paradigms. In Jaap van Marle & Geert Booij (eds.), *Yearbook of morphology 1996*, 243–256. Dordrecht: Kluwer.

Blevins, James. 2006. Word-based morphology. *Journal of Linguistics* 42. 531–573.

Booij, Geert. 2012. *The grammar of words: Introduction to linguistic morphology*. Oxford: Oxford University Press 3rd edn.

Cobbinah, A.Y. 2013. *Noun classification and verbal nouns in baïnounk gubëeher*. London: dissertation.

Dingemanse, M. 2006. *The semantics of bantu noun classification*. Leiden: Leiden University MA thesis.

Hepburn-Gray, Robert. 2016. A survey of niger-congo noun class agreement systems, Paris.

Kießling, Roland. 2013. On the origin of niger-cong noun classification. In *Historical linguistics 2011: Selected papers from the twentieth international conference on historical linguistics, osaka, 25-30 july 2011*, 43–65. Amsterdam: John Benjamins.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

McGill, Stewart. 2007. The cicipu noun class system. *Journal of West African Languages* 34(2). 51–90.

Sag, Ivan. 2012. Sign-based construction grammar: An informal synopsis. In Ivan Sag & Hans Boas (eds.), *Sign-based construction grammar*, 61–187. Stanford: CSLI Publications.

Schadeberg, Thilo. 2001. Number in swahili grammar. *Afrikanistische Arbeitspapiere* 68. 7–16.

Smits, Heleen. 2011. Lumun noun classes and number. In *Afrikanische sprachen im fokus. linguistische beitrage zum 19. afrikanistentag, mainz, 8.-10. april 2010*, 271–283. Cologne: Rudiger Koppe Verlag.

Stevenson, Roland C. 2009. A grammar of otoro. In Thilo Schadeberg (ed.), *Tira and otoro: Two grammars of kordofanian languages by roland c. stevenson*, Cologne: Rüdiger Köppe Verlag.

Stump, Gregory. 2005. Word formation and inflectional morphology. In *Handbook of word formation*, 49–71. Dordrecht: Springer.

Stump, Gregory & Raphael A. Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.

Welmers, William E. 1973. *African language structures*. Berkeley: UC press.