Reexamining Lexical Integrity: The Case of Verbal Extensions in Bantu

Robert Elwell
Department of Linguistics, the University of Texas at Austin

17 Apr 2006

Abstract. This paper considers how the morphology of verbal extensions in Bantu languages can shed light on the morphology/syntax interface. This is in counter to current treatments based on Bresnan & Mchombo's Lexical Integrity Principle (LIP). I propose using recent alternate methods of analysis in Multimodal Combinatory Categorial Grammar to solve basic shortcomings surrounding the underlying assumptions the LIP makes about these morphemes and how they impact argument selection. The arguments I give are framed both theoretically and computationally, and use data from several Bantu languages. I offer a combinatory analysis of Chichewa, which serves as a vantage point upon which to motivate further evaluation of formal machinery in LFG. Using this approach, we can also argue against wider LFG-based approaches to argument structure, such as Lexical Mapping Theory.

Keywords: syntax, Bantu, Chichewa, lexical mapping theory, lexical integrity, CCG, multi-modal CCG

1. Introduction

Approaches to the morphology-syntax interface have been problematic and fractured. Within mainstream literature, little attention is paid to the computational ramifications in the solutions offered; the Lexical Integrity Principle (LIP) (Bresnan and Mchombo, 1995), the prevalent notion of morphology in Lexical Functional Grammar, states that all morphological permutations of any particular stem exist within the lexicon, and that morphology is on a completely different level than the syntax. This leads to a veritable explosion of forms. Such assumptions exist in spite of the simple, finite-state compositional characteristics of word-level morphology, proven extensively through the use of finite-state tools for morphological analysis (Beesley and Karttunen, 2003). Furthermore, recent evidence given from Turkish (Bozsahin, 2007) brings into great suspicion the initial arguments for Lexical Integrity. The solution given by Bozsahin is in much closer agreement with computational insights regarding morphology; if morphemes can be introduced into a syntactic formalism with minimal computational complication, then this should be a closer model of both competence and performance.

In this paper, I intend to use Bozsahin's analysis of morphemes in Bantu verbal extensions, using data from work which leverages the LIP in its analysis, my own fieldwork (Elwell, 2005), and various other Bantu sources. Doing this will further motivate an argument against lexical integrity in the language family purported to most exemplify its effects. Furthermore, I will use the evidence established from this line of argument to argue against the treatment of verbal argument selection in Lexical Functional Grammar (LFG) known as Lexical Mapping Theory (LMT). This theory has also been supported with evidence from Bantu (Bresnan and Moshi, 1990). However, the cross-linguistic variation which appears to motivate the formalization can be easily accounted for using the tools available in Multi-Modal Combinatory Categorial Grammar (Baldridge, 2002)(Steedman and Baldridge, 2007) using hybrid logic for semantic representation, the utility and speed of which has been shown in chart parsing techniques (White, 2004).

2. Data/Problem

2.1. Lexical Integrity and Verbal Extensions

Verbal stems are a morphological strategy in Bantu languages which append extra arguments to the verb. Given one morpheme or another, the entire argument structure of the sentence could change, or simply be elaborated upon. Examples from Chichewa (Mchombo, 2005) are as follows:

(1) *M-kángo u-na-thyól-á m-pánda*
3-lion 3SM-pst-break-fv 4-chair
'The lion broke the fence'

(2) *M-kángo u-na-thyól-***éts**-*á m-bidzí m-pánda*
3-lion 3SM-pst-break-caus-fv 10-zebra 3-fence
'The lion made the zebras break the fence'

(3) *M-kángo u-na-thyól-* **él**-*á m-bidzí m-pánda*
3-lion 3sm-pst-break-appl-fv 10-zebra 3-fence
'The lion broke the fence for the zebras'

(4) *M-kángo u-na-thyól-* **éts-** **él**-*á m-bidzí mpánda kwa a-lenje*
3-lion 3SM-pst-break-caus-appl-fv 10-zebra 3-fence by 2-hunter
'The lion made the hunters break the fence for the zebras'

(5) *M-bidzí zi-na-thyól-* **éts-** **él-** **édw**-*á mpánda kwá a-lenje (ndí m-kángo)*
10-zebra 1-SM-pst-break-caus-appl-pass-fv 3-fence by 2-hunter (by 3-lion)
'The zebras got the fence broken for (them) by the hunters at the instigation of the lion'

(6) *M-bidzí zi-na-thyól-* **éts-** **él-** **an**-*á mpánda kwá alenje*
10-zebra 10SM-pst-break-caus-appl-recip-fv 3-fence by 2-hunter
'The zebras got the fence broken for each other by the hunters'

These morphemes have been given a thorough treatment in both Lexical Functional and Minimalist literature. Minimalist treatments generally assume that morphemes from several different phrasal nodes raise to the tensed node and combine in a manner which fulfills all necessary fetures. Approaches which utilize movement for morphological composition in Bantu (Henderson, 2005) are insufficient in two ways. First, there is little empirical evidence to explain why the moving morphemes originate in their given phrasal nodes–let alone evidence that these nodes exist. This problem is well portrayed in the "applicative phrase" internal to the verb phrase, but for use only in the case of applicatives–even in languages such as English which require optional oblique representations of applicatives. (McGinnis, 2001) Furthermore, because of the morphological structure of Bantu verbs along with fundamental Minimalist concepts of the hierarchy of phrasal nodes and the VP-internal subject hypothesis, movement to the tense node can result in Turing-complete computational complexity. It would be more desirable to attempt to account for the morphological data using more constrained approaches first, and only increasing computational complexity in necessary cases.

LFG rejects movement as a suitable approach to word- or morpheme-level composition; therefore, the arguments given should be alloted more consideration. However, the exact opposite of the issue seen in Minimalist approaches becomes problematic in LFG. Instead of morphological composition becoming an issue of computational complexity, it becomes an issue of storage of forms: every single possible morphological combination of all verbs are represented in the lexicon. Working under the assumption of no limit on storage, this is unproblematic. However, if it is possible to account for all forms using a minimal amount of computation and thus remove countless extraneous forms

which could easily be syntactically derived, then this would in fact be preferable. Regardless of the actual amount of lexical storage an individual may have, if there are regular processes which act on all of these forms, it is more than probable that these processes are always at work in composing 'word-level' syntactic forms.

Approaches to morphology with regards to argument selection and binding in Chichewa (Mchombo, 2005) (Mchombo, 2007) offer an analysis which leverages the LIP using several combinatory examples, focusing on the clitics *-nso*, meaning 'again' or 'too', and *-be*, meaning 'still':

(7) *M-kángo u-ku-thyól-á-nsó mipando*
3-lion 3sm-pres-'break'-fv-too 4-chair
'The lion is breaking the chairs, too'

(8) *M-kángo u-ku-thyól-á-bé mipando*
3-lion 3sm-pres-'break'-fv-still 4-chair
'The lion is still breaking the chairs'

These clitics "can be attached to a number of hosts" (Mchombo, 2005), but their position at the end of verbs is used to argue for the lexical integrity of the verb stem as an atomic unit that cannot be further decomposed. This analysis is undesirable in a number of ways previously described. The verb stem consists of the root, a final vowel morpheme which agrees for mood and tense, and a number of strictly ordered verbal extensions. Because of the principled ways in which these morphemes interact, the syntax should be able to further decompose these items beyond simple word level. This can be done in a principled, efficient, finite-state manner which can be constrained to reflect the combinatory characteristics of morphemes which originally motivated the Bantu work on the LIP.

Assuming that language learning is an empirical process rather than a sudden event of acquisition, it must be posited that these productive, combinatory forms can only appear because of tacit learning of the functions and behaviors of each morpheme, and not a memorization of any instance of words the speaker is exposed to. I propose a model where an individual deduces morphology rather than memorizes stems. This would better account for produced forms that one may have never been previously exposed to in speech.

Furthermore, by treating verbal stems as lexical, we ignore verbal extensions which, in fact, have become lexicalized, and therefore part of the root. The following data from Ekegusii portrays this as having much looser semantic relation than what is seen with regards to these morphemes and argument selection[1]:

(9) *ógo-sab-a*
inf-ask/pray-fv
'to ask or pray'

(10) *ógo-sab*-**er**-*a*
inf-ask/pray-**app**-fv
'to ask or pray for (someone)'

(11) *ógo-sab*-**erer**-*a*
inf-beg-fv
'to beg'

(12) *ógo-sab***erer**-**er**-*a*
inf-beg-**app**-fv
'to beg for (someone)'

[1] Data from (Elwell, 2005)

In the previous examples, the root *-sab-* shows that it combines with the applicative *-er-* in two ways. It combines inflectionally to produce an applicative reading and require an extra argument. Derivationally, the root combines with the affix to produce a new root with an alternate reading. As seen in 12, both strategies can be utilized simultaneously. Furthermore, while the inflectional suffixes appear only once, and are therefore, like all items in the syntax, resource-sensitive, derived roots containing extensions can be repeated to create new semantic readings in a fairly unprincipled manner:

(13) *ó!kw-éég-á*
inf-imitate-fv
'to imitate'

(14) *ó!kw-éég***ér**-*á*
inf-learn-fv
'to learn'

(15) *ó!kw-éég***érer**-*á*
inf-favor-fv
'to favor'

The above examples show another inherent shortcoming of strictly lexicalizing verbal extensions as part of the verb stem. Lexicalized forms of stem and extension combinations do exist, and the roots to which they attach then acquire a non-compositional semantic reading. Here, we arrive at an issue of spurious lexical ambiguity. By giving both sets of data the same treatment, we again find ourselves at an inadequate model of word-level morphology.

2.2. Extensions and Typological Argument Asymmetries

Treating verbal extensions in accordance with the LIP has also led to the treatment in LFG known as Lexical Mapping Theory (LMT). This is motivated by the following typological differences in Bantu regarding argument symmetry or asymmetry (Bresnan and Moshi, 1990):

(16)
- Passivizability of objects
- Object marker candidacy of each argument
- Unspecified object deletion with object markers
- Co-occurrence of reciprocals with passives
- Co-occurrence of reciprocals with object markers
- Co-occurrence of reciprocals with unspecified object deletion

For any of these items, data has shown that languages such as Kichaga are symmetrical in all of these matters. However, languages such as Chichewa are asymmetric, and therefore either do not allow these phenomena or only allow it for half of the viable arguments to which these could apply.

As in the previous section, evidence will be used from work in the field of LMT. The given CCG analysis proposed above and provided below should properly account for the data with no necessary addition to the theoretical framework. I will discuss the several points on which symmetrical and asymmetrical arguments differ and that each change is combinatory, referring to various points in the analysis below.

3. Analysis of Verbal Morphology

There are several criteria which a proper account of morphology couched in the syntax should satisfy. The lexical categories should properly reflect the rules of composition. They should not allow any forms not seen in the language. In this case, if a category for a morpheme violates the serial ordering of the verbal morphology, the morpheme is inadequately categorized. The semantics of the morphemes should not only account for their impact upon the root, but their impact upon the stem as well.

Overall, when treating the morphology, the goal of the syntactic approach is reversed: rather than using methods which should be extensible cross-linguistically, the categories used should be constrained to each language being treated. This does not mean that there would not exist strong similarities for certain categories between languages of the same family, but rather that there must exist principled reasons for the similarity, and that the analysis must be constrained by empirical data specific to that language, and not because of genetic relations. These formal principles not only serve as a proper guide to account for morphology, but furthermore serve to properly treat the Bantu object asymmetries which have served previously to motivate Lexical Mapping Theory in LFG.

3.1. The Simple Verb

Before treating verbal extensions, it is necessary to first arrive at a proper account for Bantu verbal morphology in general. This must be done to examine in what exact manner these morphemes fit. The Bantu verb is well-known for its incorporated pronouns for both subject and object. These serve to allow complex monolexical sentences where the subject and object are implied by their agreement with one of the dozen or so noun classes to be found in most Bantu languages.

Treatments given in LFG (Bresnan and Mchombo, 1987) assume that incorporated pronouns are anaphoric and serve the grammatical function of the optional noun phrase. This line of argumentation also claims that the coreferential noun phrase, when present, serves a "non-argument function". This means that rather than primarily fulfilling the unification features of the subject, the noun phrase instead can serve to demarcate topic or focus, or to serve as an adjunct to more fully describe the incorporated pronoun which is serving the main position in the functional structure.

This is similar to the radical lexicalist approach taken in CCG to account for languages such as Finnish–a language with a similarly high degree of agglutination and pronoun incorporation (Karttunen, 1989). Here, the work is primarily done by the categories and the semantics, rather than the functional structure. The verb composes to form an `s` category, and other words in the sentence are composed of various `s`-slash-`s` categories which serve to specify the more general semantics found in pronoun incorporation. For Bantu, this would mean that a subject morpheme agreeing with a certain noun class would denote, semantically, the set of all entities within this noun class. The hearer should be able to anaphorically deduce the coreferent. In other cases, where the coreferent is overt, it serves to offer a semantic reading of the intersection with the noun class marker such that the argument of the verb is an intersection of the set of all entities in the given agreeing class with the set of all entities with a denotation that matches that given by the adjunct noun phrase.

Drawing from both of these, my approach treats the verb as a domain of morphological composition which results in an `s` category. However, there are computational gains to be achieved from using hybrid logic as opposed to a lambda calculus semantics, and I will opt for such a semantic treatment in my derivations. Furthermore, there are cases where overt arguments will be necessary, as object marking is optional within the verb.

An intransitive verb in Chichewa appears as follows[2]:

[2] Data from (Watkins, 1937).

(17) *ku-táfun-a*
inf-chew-ind
'To chew'

(18) *ni-táfun-a*
1sg-chew-ind
'I chew'

(19) *n-a-táfun-a*
1sg-pst-chew-ind
'I chewed' (recent past)

(20) *n-dzá-táfun-a*
1sg-fut-chew-ind
'I will chew'

(21) *Sam a-dzá-táfun-a*
sam 3sg-fut-chew-ind
'Sam will chew'

It should be noted here that there is a fundamental issue which bringing morphology into syntax raises–the treatment of zero morphemes. In CCG, zero morphemes have been treated in two major ways, both mostly focused on the efficiency realized via parsing. One approach treats the root as having the underlying feature which the zero morpheme would add to the semantic reading (Hoeksema, 1985). The problem with this approach is the cancellation of this default feature when met by an overt morpheme offering a different semantic reading. One solution offered posits the existence of a zero morpheme that carries the 'default' feature (Aone and Wittenburg, 1990). This is then applied to the correct form via an algorithm which imposes trial and error attempts at unary rules which introduce each zero morpheme in the grammar. While both approaches are satisfactory and serve an excellent purpose, they are both founded on argumentation regarding how to view null morphemes in general. In other work, I show that null morphemes or default features have a direct connection to entropy, and that the learning and understanding of this morphological dichotomy of overt markedness and covert unmarkedness is a deductive process (Elwell, 2007). Here, I will treat present tense forms as containing no valuation for tense under the assumption that the correct form is deduced in through the semantics and shown in the syntax as a unary rule acting upon a droppable slash.

Given the data in examples 17 to 21, the language requires the following categories:

(22) $táfun := (\mathsf{vp}\backslash_{\star}\mathsf{sm})\backslash_{\varnothing}\mathsf{tm} : @_e\mathbf{chew} \wedge @_e\langle\textsc{Actor}\rangle x$

(23) $n := \mathsf{sm} : @_x\langle\textsc{Person}\rangle 1 \wedge @_x\langle\textsc{Number}\rangle sg$[3]

(24) $a := \mathsf{sm} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle sg$

(25) $a := \mathsf{tm} : @_e\langle\textsc{Tense}\rangle recent.past$

(26) $dzá := \mathsf{tm} : @_e\langle\textsc{Tense}\rangle future$

(27) $ku := \mathsf{vp}/_{\star}(\mathsf{vp}\backslash_{\star}\mathsf{sm}\backslash_{\varnothing}\mathsf{tm}) : @_e\langle\textsc{Tense}\rangle infinitive$

(28) $a := \mathsf{s}\backslash_{\star}\mathsf{vp} : @_e\langle\textsc{Mood}\rangle ind.$

(29) $Sam := \mathsf{s}|\mathsf{s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle sg \wedge @_x\mathbf{Sam}$

Note the usage of $_{\varnothing}$ as a modality. This denotes a droppable slash, with all of the restrictions of the $_{\star}$ modality–strict functional application or dropping. The droppable slash for the case of verbs allows the following unary rule:

[3] Note that I will ignore phonological variation in my description of categories, such as that seen in example (**??**)

(30) $\mathsf{X}\backslash_{\varnothing}\mathsf{tm} : @_e \ldots \Rightarrow_{\varnothing_{tm}} \mathsf{X} : @_e \ldots \wedge \langle \textsc{Tense} \rangle \mathit{present}$ $(\varnothing_{tm})$

The categories above will account for the treatment of noun phrases as topicalized coreferents to incorporated pronouns, as seen in LFG and further discussed below. Derivations for examples (18) and (21) are as follows [4]:

(31)

ni	*táfun*	*a*
sm	$(\mathsf{vp}\backslash\star\mathsf{sm})\backslash_{\varnothing}\mathsf{tm}$	$\mathsf{s}\backslash\star\mathsf{vp}$
$@_x\langle \textsc{Person} \rangle 1 \wedge @_x\langle \textsc{Number} \rangle sg$	$@_e\mathbf{chew} \wedge @_e\langle \textsc{Actor} \rangle x$	$@_e\langle \textsc{Mood} \rangle ind.$
	$<\varnothing_{tm}$	
	$\mathsf{vp}\backslash\star\mathsf{sm}$	
	$@_e\mathbf{chew} \wedge @_e\langle \textsc{Actor} \rangle x \wedge \langle \textsc{Tense} \rangle present$	
	$<$	
	vp	
	$@_e\mathbf{chew} \wedge @_e\langle \textsc{Actor} \rangle x \wedge @_x\langle \textsc{Person} \rangle 1 \wedge @_x\langle \textsc{Number} \rangle sg$	
		$<$

$\mathsf{s} : @_{e\,\mathbf{chew}} \wedge @_e\langle \textsc{Actor} \rangle x \wedge @_x\langle \textsc{Person} \rangle 1 \wedge @_x\langle \textsc{Number} \rangle sg \wedge @_e\langle \textsc{Mood} \rangle ind.$

(32)

Sam	*a*	*dzá*	*táfun*	*a*
$\mathsf{s}\vert\mathsf{s}$	$\mathsf{sm}\backslash\star\mathsf{np}$	tm	$(\mathsf{vp}\backslash\star\mathsf{sm})\backslash_{\varnothing}\mathsf{tm}$	$\mathsf{s}\backslash\star\mathsf{vp}$
			$<$	
			$\mathsf{vp}\backslash\star\mathsf{sm}$	
			$<$	
		vp		
		$<$		
	s			
	$>$			
s				

For transitive verbs, the following set of sentences are used to prove the specific characteristics discussed above for the treatment of Bantu subject and object markers in LFG (Bresnan and Mchombo, 1987):

(33) *njûchi zi-ná-lúm-a alenje*
cl7.bees cl7-pst-bite-ind. hunters
'The bees bit the hunters' (S V O)

(34) zi-ná-lúm-a alenje njûchi
'The bees bit the hunters' (V O S)

These are the only two configurations allowed for verbs that do not have an object marker. The following combinations are allowed when the object marker -*wá*- is present:

(35)
1. S V O
2. V O S
3. O V S
4. S O V
5. O S V

Bresnan and Mchombo use this evidence to motivate two major principles: *(1)* object NPs are VP-internal, and *(2)* when both subject and object markers are present, these incorporated pronouns serve the purpose of subject and object grammatical function, while the subject and object noun phrases are topicalized forms. This can be interpreted into CCG using the following categories:

[4] Example (32) omits semantics for the purposes of margins.

(36) *njûchi* := $\mathsf{s|s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl \wedge @_x\mathbf{bees}$

(37) *alenje* := $\mathsf{s|s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle pl \wedge @_x\mathbf{hunters}$

(38) *zi* := $\mathsf{sm} : @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl$

(39) *ná* := $\mathsf{tm} : @_e\langle\textsc{Tense}\rangle past$

(40) *lúm* := $((\mathsf{vp}\backslash\star\mathsf{sm})\backslash_\varnothing\mathsf{tm})\backslash\star\mathsf{om} : @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$

(41) *wá* := $\mathsf{om} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle pl$

For all transitive verbs, there also exists a second category to account for the requirement of a rightward nominal argument if there is no object marker:

(42) For all X := $((\mathsf{vp}\backslash\star\mathsf{sm})\backslash_\varnothing\mathsf{tm})\backslash\star\mathsf{om}$,
there also exists an entry, X := $(((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))\backslash\star\mathsf{sm})\backslash_\varnothing\mathsf{tm}$

This generality will also be valuable during the discussion of verbal extensions. The configurationality alternations seen in examples (33) to (35) are accounted for in Figure 1. Here, it is shown that agreement and selection are conditioned by slash directionality and semantics. For subject and object markers that reflect the same class marker, there will be an amount of ambiguity due to the non-configurational characteristic of the s|s category. This ambiguity would be expected in LFG as well given the treatment of nominals as coreferential adjuncts. By using the category which leverages the om as its source of argument selection, the s|s nominal categories become optional, as seen by the verb derivation in Figure 2. This verb results in an s category without either nominals, which follows with the language data and the analysis from Bresnan & Mchombo.

For the case of the verb, I have shown that assigning the proper categories can result in verbal composition which uses only functional application, reflecting the finite-state nature of morphological composition in general. With incorporated pronouns, the syntax and semantics interact in a manner which makes nominals represented by pronouns optional. Assuming there is a greater inconvenience from the storage of an innumerable explosion of forms rather than the least complex type of computational strategy, I have affirmed the usefulness of Bozsahin's treatment of morphology. From here, were are at a suitable juncture to examine verbal extensions.

3.2. Verbal Extensions

Chichewa displays a series of verbal extensions which are serially ordered and can occur simultaneously. They all alter the argument structure in a constant fashion. In this section, I will treat a number of extensions one at a time, using cumulative combinations of extensions to show exactly how they interact with both the verb and one another.

3.2.1. *The Causative*

The causative marker in Chichewa has a variety of forms: *-its-/-ets-/-iz-/-ez-/-z-*. Essentially, the purpose of this extension is to alter the argument structure in the following way: the subject takes on a causing role rather than an agentive role; the object of the verb root remains the same; a second object in the form of an overt nominal is required as the entity being caused to perform an action. This is evident in the example (2) given previously.

In order to treat this, the following categories must be assigned:

(43) *mkángo* := $\mathsf{s|s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle pl \wedge @_x\mathbf{lion}$

(44) *mpánda* := $\mathsf{s|s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 3 \wedge @_x\mathbf{fence} \wedge @_x\langle\textsc{Number}\rangle pl$

(45) *mbidzí* := $\mathsf{s|s} : @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 10 \wedge @_x\mathbf{zebra} \wedge @_x\langle\textsc{Number}\rangle pl$

(46) *thyól* := $((\mathsf{vp}\backslash\star\mathsf{sm})\backslash_\varnothing\mathsf{tm})\backslash\star\mathsf{om} : @_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$

njûchi	*zi*	*ná*	*lúm*	*a*	*alenje*
$\mathsf{s \vert s}$ $@_x\langle\textsc{Person}\rangle 3$ $\wedge @_x\langle\textsc{Class}\rangle 7$ $\wedge @_x\langle\textsc{Number}\rangle pl$ $\wedge @_x\mathbf{bees}$	sm $@_x\langle\textsc{Class}\rangle 7$ $\wedge @_x\langle\textsc{Number}\rangle pl$	tm $@_e\langle\textsc{Tense}\rangle past$	$(((\mathsf{s}/_\star(\mathsf{s \vert s}))/_\star(\mathsf{s}\backslash_\star\mathsf{vp}))\backslash_\star\mathsf{sm})\backslash_\varnothing\mathsf{tm}$ $@_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$	$\mathsf{s}\backslash_\star\mathsf{vp}$ $@_e\langle\textsc{Mood}\rangle ind.$	$\mathsf{s \vert s}$ $@_y\langle\textsc{Person}\rangle 3 \wedge$ $@_y\langle\textsc{Number}\rangle pl$ $\wedge @_y\mathbf{hunters}$
		$(\mathsf{s}/_\star(\mathsf{s \vert s}))/_\star(\mathsf{s}\backslash_\star\mathsf{vp})\backslash_\star\mathsf{sm}$ $\wedge @_e\langle\textsc{Tense}\rangle past$	$: @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$ $\wedge @_e\langle\textsc{Patient}\rangle y$ $<$		
	$(\mathsf{s}/_\star(\mathsf{s \vert s}))/_\star(\mathsf{s}\backslash_\star\mathsf{vp})$ $\wedge @_e\langle\textsc{Tense}\rangle past$	$: @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$ $\wedge @_e\langle\textsc{Patient}\rangle y$	$\wedge @_x\langle\textsc{Class}\rangle 7$ $\wedge @_x\langle\textsc{Number}\rangle pl$ $<$		
	$\mathsf{s}/(\mathsf{s \vert s})$ $\wedge @_e\langle\textsc{Tense}\rangle past$	$: @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$ $\wedge @_e\langle\textsc{Patient}\rangle y$	$\wedge @_x\langle\textsc{Class}\rangle 7 \wedge$ $\wedge @_x\langle\textsc{Number}\rangle pl$	$@_e\langle\textsc{Mood}\rangle ind.$ $>$	
	s $\wedge @_e\langle\textsc{Tense}\rangle past$	$: @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$ $\wedge @_e\langle\textsc{Patient}\rangle y$	$\wedge @_x\langle\textsc{Class}\rangle 7 \wedge$ $\wedge @_x\langle\textsc{Number}\rangle pl$	$@_e\langle\textsc{Mood}\rangle ind.$ $@_y\langle\textsc{Number}\rangle pl$	$@_y\langle\textsc{Person}\rangle 3 \wedge$ $\wedge @_y\mathbf{hunters}$ $>$
$\wedge @_x\mathbf{bees}$	s $\wedge @_e\langle\textsc{Tense}\rangle past$	$: @_e\mathbf{bite} \wedge @_e\langle\textsc{Actor}\rangle x$ $\wedge @_e\langle\textsc{Patient}\rangle y$	$\wedge @_x\langle\textsc{Class}\rangle 7 \wedge$ $\wedge @_x\langle\textsc{Number}\rangle pl$	$@_e\langle\textsc{Mood}\rangle ind.$ $@_y\langle\textsc{Number}\rangle pl$	$@_y\langle\textsc{Person}\rangle 3 \wedge$ $\wedge @_y\mathbf{hunters}$ $>$

*Figure 1.*Derivation of ‘the bees bite the hunters’ which requires a nominal object.

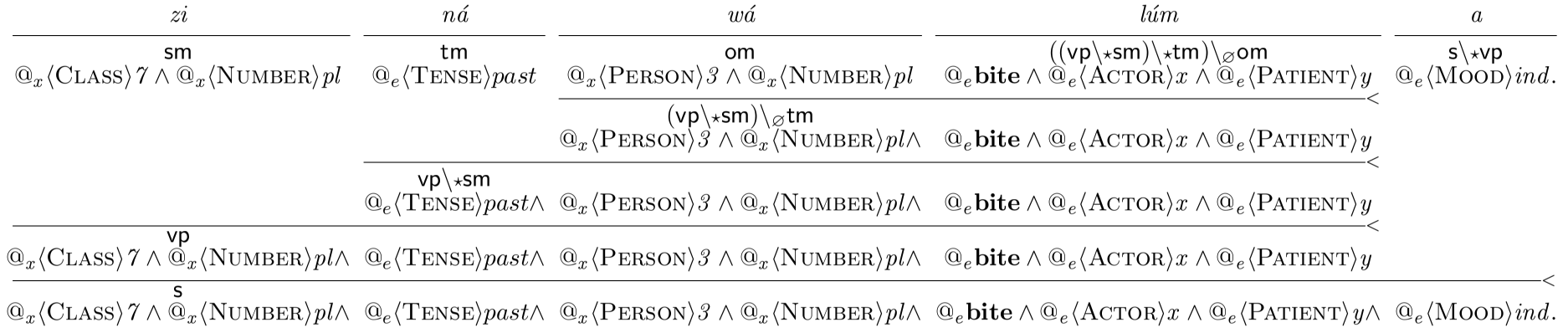

*Figure 2.*Derivation of 'bit' including incorporated pronouns.

(47) $u := \mathsf{sm} : @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl$

(48) $ets := ((\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\mathbf{cause} \wedge @_e\langle\textsc{Causee}\rangle z$

Note the logical form of example (48). This is accompanied with the following rule:

(49) $@_i\mathbf{j} \sqcup @_i\mathbf{cause}$
$\rightarrow @_i\mathbf{cause\text{-}j}$

Essentially, this makes the logical predicate the causation of the event rather than the event itself. The actor and patient associate with the same argument, but the actor takes on a causing role. The causee becomes the entity which actually acts upon the patient, but cannot be referred to strictly as agent because this entity, too, is being acted upon.

Given the entries and rules above, the sentence in example (2) parses in the following manner:

(50)

mkángo	*u*	*na*	*thyól*	*éts*	*á*	*mbidzí*	*mpánda*
s|s	sm	tm	(((s/⋆(s|s))/⋆(s\⋆vp))\⋆sm)\∅tm	((s\⋆vp)/⋆(s|s))/⋆(s\⋆vp)	s\⋆vp	s|s	s|s
			((s/⋆(s|s))/⋆(s\⋆vp))\⋆sm (<)	(s\⋆vp)/⋆(s|s)) (>)			
			(s/⋆(s|s))/⋆(s\⋆vp) (<)	s\⋆vp (>)			
			s/(s|s) (>)				
			s (>)				
			s (>)				

Note that this derivation shows these categories accurately capture the proper dependencies expected from the verbal extension and verb root; these nested dependencies show that the causative extension is seeking the causee, while the verb root is seeking the agent and paient. The following example shows the semantics for this derivation, step by step:

(51)

1. $@_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y,$
$@_e\mathbf{cause} \wedge @_e\langle\textsc{Causee}\rangle z$
2. $@_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge @_e\langle\textsc{Tense}\rangle past,$
$@_e\mathbf{cause} \wedge @_e \wedge @_e\langle\textsc{Causee}\rangle z\langle\textsc{Mood}\rangle ind.$
3. $@_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge @_e\langle\textsc{Tense}\rangle past \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl,$
$@_e\mathbf{cause} \wedge @_e\langle\textsc{Causee}\rangle z \wedge @_e\langle\textsc{Mood}\rangle ind.$
$\wedge @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge @_z\langle\textsc{Number}\rangle pl$
4. $@_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge @_e\langle\textsc{Tense}\rangle past \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl \sqcup$
$@_e\mathbf{cause} \wedge @_e\langle\textsc{Causee}\rangle z \wedge @_e\langle\textsc{Mood}\rangle ind.$
$\wedge @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge @_z\langle\textsc{Number}\rangle pl$
$\rightarrow @_e\mathbf{cause\text{-}break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge @_e\langle\textsc{Tense}\rangle past \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl$
$\wedge @_e\langle\textsc{Causee}\rangle z \wedge @_e\langle\textsc{Mood}\rangle ind.$
$\wedge @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge @_z\langle\textsc{Number}\rangle pl$
5. $@_e\mathbf{cause\text{-}break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge @_e\langle\textsc{Tense}\rangle past \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle pl$

$\wedge\, @_e\langle\textsc{Causee}\rangle z \wedge @_e\langle\textsc{Mood}\rangle \mathit{ind.}$
$\wedge\, @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge\, @_z\langle\textsc{Number}\rangle \mathit{pl} \wedge @_y\langle\textsc{Person}\rangle 3$
$\wedge\, @_y\langle\textsc{Class}\rangle 3 \wedge @_y\mathbf{fence} \wedge @_y\langle\textsc{Number}\rangle \mathit{pl}$

6. $@_e\mathbf{cause\text{-}break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge\, @_e\langle\textsc{Tense}\rangle \mathit{past} \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle \mathit{pl}$
$\wedge\, @_e\langle\textsc{Causee}\rangle z \wedge @_e\langle\textsc{Mood}\rangle \mathit{ind.}$
$\wedge\, @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge\, @_z\langle\textsc{Number}\rangle \mathit{pl} \wedge @_y\langle\textsc{Person}\rangle 3$
$\wedge\, @_y\langle\textsc{Class}\rangle 3 \wedge @_y\mathbf{fence} \wedge @_y\langle\textsc{Number}\rangle \mathit{pl}$
$\wedge\, @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle \mathit{pl} \wedge @_x\mathbf{lion}$

3.2.2. *The Applicative*

The applicative extension is seen in the forms *-ir-/-er-/-il-/-el*. It is much like the causative, in that it requires an extra argument. In this case, however, the role of the argument with respect to the verb is different. In order to account for the applicative as seen in example (3), the following lexical entries for the causative is necessary:

(52) $\mathit{el} := ((\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\langle\textsc{Recipient}\rangle z$

For this entry, there is no necessary logical rule, but rather a simple addition of an argument both syntactically and semantically. Rather than taking part in the action of the verb, this new argument is benefitting in some way by the action. Again, as seen in example (3), this category accounts for the nesting of extension and argument within the verb phrase. The derivation for this example is as follows:

(53)

mkángo	*u*	*na*	*thyól*	*él*	*á*	*mbidzí*	*mpánda*
s|s	sm	tm	$(((\mathsf{s}/\star(\mathsf{s}\|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))\backslash\star\mathsf{sm})\backslash_\varnothing\mathsf{tm}$	$((\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}\|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp})$	$\mathsf{s}\backslash\star\mathsf{vp}$	s|s	s|s
			$((\mathsf{s}/\star(\mathsf{s}\|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))\backslash\star\mathsf{sm}$ (<)	$(\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}\|\mathsf{s}))$ (>)			
			$(\mathsf{s}/\star(\mathsf{s}\|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp})$ (<)	$\mathsf{s}\backslash\star\mathsf{vp}$ (>)			
			$\mathsf{s}/(\mathsf{s}\|\mathsf{s})$ (>)				
			s (>)				
			s (>)				

This derivation is identical to that in example (50), differing only in the semantic output:

(54) $@_e\mathbf{break} \wedge @_e\langle\textsc{Actor}\rangle x \wedge @_e\langle\textsc{Patient}\rangle y$
$\wedge\, @_e\langle\textsc{Tense}\rangle \mathit{past} \wedge @_x\langle\textsc{Class}\rangle 7 \wedge @_x\langle\textsc{Number}\rangle \mathit{pl}$
$\wedge\, @_e\langle\textsc{Recipient}\rangle z \wedge @_e\langle\textsc{Mood}\rangle \mathit{ind.}$
$\wedge\, @_z\langle\textsc{Person}\rangle 3 \wedge @_z\langle\textsc{Class}\rangle 10 \wedge @_z\mathbf{zebra}$
$\wedge\, @_z\langle\textsc{Number}\rangle \mathit{pl} \wedge @_y\langle\textsc{Person}\rangle 3$
$\wedge\, @_y\langle\textsc{Class}\rangle 3 \wedge @_y\mathbf{fence} \wedge @_y\langle\textsc{Number}\rangle \mathit{pl}$
$\wedge\, @_x\langle\textsc{Person}\rangle 3 \wedge @_x\langle\textsc{Class}\rangle 3 \wedge @_x\langle\textsc{Number}\rangle \mathit{pl} \wedge @_x\mathbf{lion}$

Here, there is no alteration to the semantics of the verbal predicate, only the addition of an argument. A combination of causative and applicative is seen in example (4). In this case, it is shown that only one argument is satisfied in the manner seen for both the categories of both causative and applicative. What this implies is that these two extensions in tandem still only require one overt argument. Furthermore, because this required argument is the outermost extension, nesting is still

respected. When these factors fall under consideration with the fact that Bantu verbal extensions are strictly ordered, it is prudent to assume that there is a degree of lexification of these morphemes in combination with one another, but not with the verb. Under this assumption, the veritable explosion of lexical entries is contained; only the number of possible extension combinations become an issue for the lexicon, while verbal roots will receive a more realistic one-to-one representation in the lexicon. The lexical entry for these two extensions in cooperation are therefore as follows:

(55) *étsél* := $((\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\mathbf{cause} \wedge @_e\langle\textsc{Recipient}\rangle z$

Note the lack of the Causee argument in the semantics given here. This is in keeping with the 'by' phrase added to the causee argument in the syntax. This extra word must be used to refer to a semantic role not yet given a value, but inferred semantically. In this case, the category for any of the many class-referential 'by' phrases are as follows, with the pertinent lexeme being used to exemplify the entire class:

(56) *kwá* := $(\mathsf{s}|\mathsf{s})/(\mathsf{s}|\mathsf{s}) : @_x\langle\textsc{Class}\rangle 2 \wedge @_f\langle\textsc{Causee}\rangle x$

Note that Causee is not necessarily the role taken by entity being referred to, but rather this is the eventual deduction of the listener. Because all but one extra argument of the multiple verbal extensions to be accounted for must be referred to in what could be referred to as an oblique manner, there is some deal of ambiguity involved in the semantics. It will be shown for the case of the passive that Agent can just as easily be deduced in a similar manner. Using these lexical entries, example (4) can be derived as seen in Figure 3. Note the dependencies which are captured. Given this sentence, there is only one possible role for the hunter among the many which a 'by' phrase would delineate.

3.2.3. *The Passive*

Passivization takes what would be the most adjacent rightward argument of the verbal root and requires that it be found immediately to the left of the verb rather than on the right. Furthermore, the subject marker maintains agreement with the nominal that has moved to this position. Like other verbs, passivized verbs do not require a nominal subject. As the only extension in question, the category is fairly unproblematic; because of verbal extensions, ditransitives are rarely root-based in Bantu. An example of the passive by itself is as follows (Dubinsky and Simango, 1996):

(57) *nyemba zi-na-phik-idw-a*
10.beans cl.10-pst-cook-pass-fv
'The beans were cooked'

Given its interaction upon the verb, the proper lexical entry for the passive extension would be as follows:

(58) *idw* := $(\mathsf{s}\backslash\star((\mathsf{s}/\star(\mathsf{s}|\mathsf{s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\mathbf{pass}$

While this category seems complicated, it is absolutely necessary to prevent the licensing of an overt nominal object. Given this entry, the following semantic rule is necessary:

(59) $@_i\mathbf{j} \sqcup @_i\mathbf{pass}$
$\rightarrow @_i\mathbf{be\text{-}j\text{-}ed}$

This preserves which entity is the actor, which is important given the morphological agreement found above. This is shown in the following derivation:

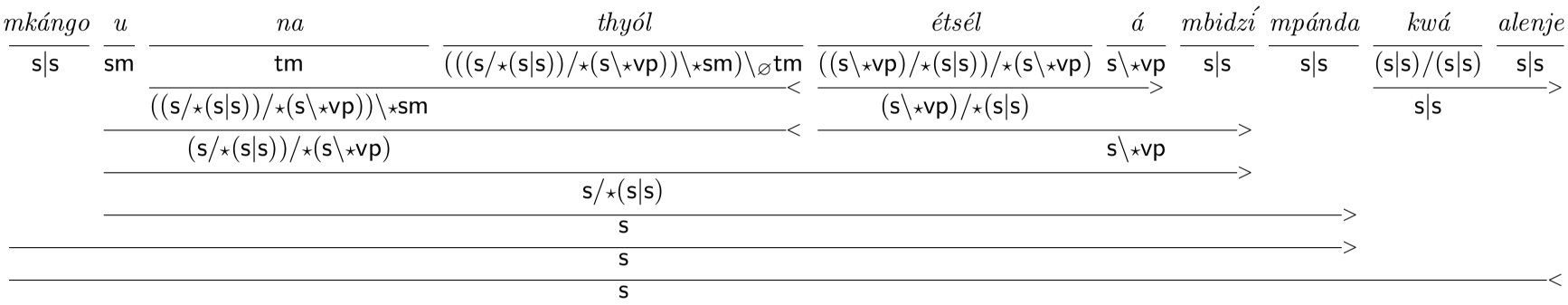

*Figure 3.*Derivation for ‘the lion made the hunters break the fence for the zebras’.

(60)
$$\dfrac{\dfrac{\dfrac{\dfrac{\textit{nyemba}}{\mathsf{s|s}}\ \dfrac{\textit{zi}}{\mathsf{sm}}\ \dfrac{\textit{na}}{\mathsf{tm}}\ \dfrac{\textit{phik}}{(((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))\backslash\star\mathsf{sm})\backslash_{\varnothing}\mathsf{tm}}}{((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))\backslash\star\mathsf{sm}}<\ \dfrac{\textit{idw}}{(\mathsf{s}\backslash\star((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))/\star(\mathsf{s}\backslash\star\mathsf{vp})}\ \dfrac{\textit{a}}{\mathsf{s}\backslash\star\mathsf{vp}}}{\mathsf{s}\backslash\star((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))}>}{(\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp})}>}{\mathsf{s}}<}{\mathsf{s}}>$$

Combinations of verbal extensions will take the verbal root as an argument in order to mitigate nominal argument selection by the verbal root. For instance, example (5) shows the general utility of treating combinations of verbal extensions as lexified items. Here, this allows the proper selection of the passivized argument as the applicative recipient. Because the this argument has saturated the requirement of the fronting of a nominal, there also seems to be some degree of semantic blending among the three extensions. The category for the extension combination is as follows:

(61) *étsélédw* := $(\mathsf{s}/\star\mathsf{vp})/\star(\mathsf{s}/\star\mathsf{vp}) : @_e\mathbf{pass\text{-}app} \wedge \exists @_e\langle\textsc{Causer}\rangle$

This will require the following semantic rule:

(62) $@_e\mathbf{j} \sqcup @_e\mathbf{pass\text{-}app}$
$\rightarrow @_e\mathbf{be\text{-}j\text{-}ed\text{-}for}$

This will preserve the position of agent as the subject, but also provide the correct relation in regards to the primal verb root.

3.2.4. *Reciprocal*
The reciprocal extension *-an-* combines with the applicative extension to denote its namesake. Another argument for lexifying extension combinations lies here; this particular morpheme is generally found with the applicative. In cases where it does not, it serves to affirm the treatment of certain extension combinations as treating the verb root as an argument in order to prevent transitivity, as seen in the following examples (Watkins, 1937):

(63) *adjeremani ni angerezi ba-pon-an-a*
germans and english 3pl-fight-rec-fv
'the germans and the english are fighting each other'

(64) *bantu bamuna s-a-kan-an-a*
people male neg-3pl-refuse-rec-fv
'men do not refuse each other'

The lexical entry for the reciprocal by itself would be as follows:

(65) *an* := $(\mathsf{s}\backslash\star((\mathsf{s}/\star(\mathsf{s|s}))/\star(\mathsf{s}\backslash\star\mathsf{vp}))/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\mathbf{rec}$

This will require the following semantic rule:

(66) $@_e\mathbf{j} \wedge \langle\textsc{Agent}\rangle x \wedge \langle\textsc{Patient}\rangle y \sqcup @_e\mathbf{rec}$
$\rightarrow @_e\mathbf{j} \wedge \langle\textsc{Agent}\rangle x \wedge \langle\textsc{Patient}\rangle x$

In essence, such a rule ties in the syntactic fact that the object of the argument becomes the subjects in a reciprocal manner. To account for the extension combination in example (6), the following category would be necessary:

(67) *étsélán* := $(\mathsf{s}\backslash\star\mathsf{vp})/\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\mathbf{app\text{-}rec} \wedge @_e\mathbf{cause}$

Because applicative and reciprocal combine to make the recipient and agent the same rather than the agent and patient, the following semantic rule is needed:

(68) $@_e\mathbf{j} \wedge \langle \textsc{Agent} \rangle x \sqcup @_e\mathbf{app\text{-}rec}$
$\rightarrow @_e\mathbf{j} \wedge \langle \textsc{Agent} \rangle x \wedge \langle \textsc{Recipient} \rangle x$

Note that for example (6), the 'by' phrase is deduced to be the causee because of the saturation of other semantic roles delineated by the semantics of the extensions combined with the verb.

3.3. Clitics and Extensions

One argument given for lexical integrity of the verb stem in particular is the presence of postverbal clitics (Mchombo, 2005). Purportedly, the clitics seen in examples (7) and (8) would prevent any kind of combinatory treatment of the morphemes which comprise the verb stem. Treated certain manners, this would be true; the wrong category for such clitics may prevent the saturation of arguments required both by the verb root as well as extensions. However, given the semantic information offered by these clitics, it is equally arguable that these clitics attach directly to the final vowel. If they do so and maintain the same category the final vowel does, this would in no way prevent either argument selection or the combinatory characteristic of morphemes. Furthermore, the directionality of slashes and lexification of extension combinations would further reflect the combinatory characteristics of clitics only being found at the end of the verb. Given this, the category for either clitic, using *-be* as an example, is seen in the following lexical entry:

(69) $be := (\mathsf{s}\backslash\star\mathsf{vp})\backslash\star(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\langle \textsc{Aspect} \rangle \mathit{continuous}$

Also, given the argument that a clitic can attach to any "number of hosts", it may be the case that the clitic's category is simple as $\mathsf{x}\backslash\star\mathsf{x}$. In other words, it could imply that it is categorially underspecified, and can simply attach to anything as long as it doesn't change its category.

A second clitic-based argument for lexical integrity of the stem revolves around its role in reduplication. Because the clitic attaches after the reduplicated stem, the argument is that the verbal root, extensions, and final vowel are irreducible and maintain one lexical entry per possible combination of each section. The following examples have been used as evidence for this assertion (Mchombo, 2005):

(70) chinkokomo chija chi-da-**kúlir-a-kúlir-a**-be
'The rumbling steadily increased in loudness'

(71) Ona zinthu zikunka zí-ku-**yípír-a-yípír-a**-tu
'Look, things are progressively worsening'

These examples show the exclusion of clitics from reduplication. Note *-tu* as also having an impact upon the aspect, further affirming the treatment of this clitic as attaching to the final vowel. Despite the previous interpretations on the evidence above, the lexical integrity view continues to be disadvantageous given the presupposition that there is a limit on storage in the lexicon. Current treatments in morphophonology regards the reduplicant as a single morphological affix, Red, which attains its phonological form via interaction of specific constraints (Kager, 1999). Given this view of the reduplicant as a primal underspecified morpheme before its phonological characteristics are achieved, the issue becomes somewhat simpler. Because the reduplicant includes changes the final vowel undergoes, it can be assumed that it is morphologically after the final vowel but before the clitic. In fact, it could be assumed that the reduplicant is similarly a kind of clitic, given that it also has a repetitive aspectual marking, and clitics seem to offer an extra aspectual realization to the verb itself. The reduplicant would have the following lexical entry:

(72) RED := $(\mathsf{s}\backslash\star\mathsf{vp})\backslash\star(\mathsf{s}\backslash\star\mathsf{vp})$: $@_e\langle\text{ASPECT}\rangle$*repetitive*

This accounts for the combinatory aspects of the reduplicant. It can be assumed that there are extrasyntactic constraints which influence the ordering of reduplicants before overt clitics, but because it combines with the final vowel first, the reduplicant does find itself at a stage of combination before the other clitic can combine with the rest of the verb proper. What this affords is an opportunity for the phonology to delineate the proper reduplicated form. Even if the second clitic attaches before this occurs, there may exist constraints which could exclude all rightward syntactic combinators from appearing in the reduplicated form. Finally, work in morphological analysis shows that reduplication can occur at a finite-state speed, and combinatorially as an under-valued primal reduplicant form which is later or concurrently modulated by the phonological module (Beesley and Karttunen, 2003), making a combinatory treatment of the affix a desirable approach.

3.4. Section Summary

Here, I have shown that the verbal morphology of Chichewa can be further reduced lexically. Assuming lexical space limitation and human abilities of deduction would serve as decent incentives to reduce lexification of any forms which can be further analyzed combinatorially, the very low computational cost of finite-state functional application makes this formal approach a useful alternative. This would account for a speaker's ability to synthesize morphemes into a novel usage never before heard by the individual. The lexical integrity principle would either assume that the speaker manages to deduce this morphology to create new lexical entries but would not make these lexical entries purely morphological in nature, or that every form the speaker can use is learned. Either of these presuppositions is inherently flawed. The analysis given here allows for novel inflections as well as novel arrangements of argument structure for a given verbal root.

4. Lexification of Lexical Mapping

Lexical Mapping Theory (LMT) is a sub-theory of Lexical Functional Grammar which accounts for argument selection in a given language. It has been especially useful for languages with verbal affixes that affect argument structure, because these show a clear relation between the syntactic changes and the stimulus for the change.

Here, I will argue that given the analysis in the above section, LMT is an unnecessary addition to syntactic theory. Furthermore, because it builds on abstract categories in LFG, such as SUBJ, OBJ, OBJ-θ, and OBL, which are initially intended to control argument selection, the sub-theory itself can be classified as a "degenerating problemshift" (Lakatos, 1970); rather than examining the combinatory nature of these traditional syntactic roles, an extra level of abstraction became necessary in order to explain how they actually interact with the syntax.

In order to motivate my claims, I will show in a number of ways how what LMT has to offer in the typological analysis of Bantu argument asymmetries (Bresnan and Moshi, 1990) can be easily translated into simple combinatory rules with a fairly uncomplicated semantic interface. The clear benefit here is the removal of an extra layer of abstraction, as well as an extra layer of syntactic rules which apply parametrical restrictions upon the argument structure.

In this section, I will follow the multiple points given by Bresnan and Alsina in example (16) to motivate the lexical characteristic of these asymmetries. Doing so will obviate a rule-based system of argument selection.

4.1. Passives

In Kichaga, both the applicative argument and verbal argument can be passivized (Bresnan and Moshi, 1990):

(73) *n-á-í-lyi-í-a* *mka* *kélya*
foc-1s-pr-eat-app-fv 1.wife 7.food
'He is eating food for/on his wife.'

(74) *mka* *n-á-í-lyi-í-o* *kélya*
1.wife foc-1s-pr-eat-app-pass 7.food
'The wife is being benefitted/adversely affected by someone eating the food'

(75) *kelyá* *k-í-lyi-í-o* *mka*
7.food 7s-pr-eat-app-pass 1.wife
'The food is being eaten for/on the wife'

According to Bresnan & Moshi, Chichewa does not exhibit the passivization seen in example (75). Given the semantics offered for the combination of passive and applicative in example (61) and its following lexical entry, this can be confirmed, given that a combination of passive and applicative must follow the above formal treatment. Conversely, Kichaga would exhibit two different lexical entries for combination of the applicative and passive. It can be assumed that these two entries vary in scope; one scopes applicative over passive, and one scopes passive over applicative:

(76) Passive-Applicative:
ií := $\mathsf{s}\backslash\star(\mathsf{s}/\star(\mathsf{s}|\mathsf{s})) : @_e\mathbf{pass\text{-}app}$

(77) Applicative-Passive:
ií := $\mathsf{s}\backslash\star(\mathsf{s}/\star(\mathsf{s}|\mathsf{s})) : @_e\mathbf{pass} \wedge \langle\textsc{Recipient}\rangle z$

The advantages to this approach are that it can be claimed for certain Bantu languages that only one particular reading for the extension combination exists. This would be the case in Chichewa, but not Kichaga. In fact, the variation in scope is more supporting evidence for treating extension combinations as a single lexical item; ambiguities in scope for morphemes reflect ambiguities in combination. Both readings can be assumed to originate from assumptions of what extension combined with the other before combining with the verb root.

4.2. Object Markers

While symmetrical languages like Kichaga allow the object marker to be any of the verbal arguments, including those added by extensions, Chichewa seems to only allow object markers for what could be considered the innermost nested argument of the verb. For instance, the object of the applicative can be only mentioned by an object marker, but the patient of the applicativized verb cannot. In Kichaga, however, both arguments can be mentioned as object markers concurrently. This suggests the following lexical entry for the applicative in Chichewa:

(78) *el* := $(((\mathsf{s}/\star(\mathsf{s}|\mathsf{s}))\backslash\star\mathsf{vp})/(\mathsf{s}\backslash\star\mathsf{vp}) : @_e\langle\textsc{Recipient}\rangle z$

This entry would treat the verb with its incorporated pronouns and tense marker as an argument. This is affirmed by cases where the infinitive can take a verb root with its object marker as an argument to its right. By doing this, we capture the dependency that the object marker is at this point referring to the object of the extension, as it serves in part to saturate an argument of the extension. The $\mathsf{s}/\star(\mathsf{s}|\mathsf{s})$ becomes the manner in which the argument of the verb is then saturated.

A similar entry for Kichaga would result in s rather than $\mathsf{s}/\star(\mathsf{s}|\mathsf{s})$ at the final step of the derivation for the verb alone.

4.3. Unspecified Object Deletion

Another source of variation between symmetrical and asymmetrical object languages is the grammaticality of verbal constructions which exclude specific root arguments. For instance, in Kichaga, the patient argument of a verb can be deleted, even when there is an overt applicative argument that is not deleted. However, in Chichewa, this is not licensed. This can be just as easily argued as a case of lexical argument structure. Especially given the verb-heavy sentential characteristics of Bantu languages, it is no far cry to assume that certain languages, like Kichaga, will display no need for overt arguments whatsoever, even without an object marker.

The deletion of an unspecified object, information theoretically, can be envisioned as a case where there is such a degree of ambiguity in the object that not overtly referring to it conveys as much information as overtly doing so. Given this, it would be unsurprising that languages which would allow it would be those which would treat properly inflected verb roots as an s category, because while this type of encoding works under constraints other than purely syntactic, it would still be ill-formed if the language required an overt object in some form. A language like Chichewa does require either an object marker or an over nominal, and this is seen in the lexical categories for transitive verbs.

It can further be considered that there are functional benefits in this case for erring on either the symmetrical or asymmetrical side. Forcing certain syntactic well-formedness requirements on a sentence through lexical categories will prevent ambiguity, while a language like Kichaga will be fairly rife in ambiguous phrases. However, it benefits from being able to manipulate language in a slightly freer fashion.

4.4. Reciprocalization

Chichewa displays an asymmetry here in that it cannot reciprocalize the patient in the presence of an applied argument. This is another argument for the treatment of extension combinations as a single lexical item in the language, and reflects the treatment given in example (68). Kichaga, on the other hand, can reciprocalize the patient of the verb even if applicativized in a non-instrumental manner. This is again an instance of asymmetry and clarity versus symmetry and ambiguity. Chichewa uses structural syntactic asymmetry as a cue for its semantic interface, while Kichaga takes advantage of a greater degree of combinatory freedom at the expense of increased ambiguity. The following Kichaga example is disallowed in Chichewa (Bresnan and Moshi, 1990):

(79) *wachaka wá-í-wˈágh-i-an-a mangi*
2.chaga 2s-pres-kill-app-rec-fv 1.chief
'The Chagas are killing each other for the chief'

However, if Kichaga does display true symmetry, then this would also have the reading 'The Chagas are killing the chief for each other'. Knowing the intended reading would become a discourse-specific–and therefore more complex–undertaking. Because of the one-to-one relational categories verbal extensions and their semantics display in Chichewa, disambiguation of argument structure becomes significantly more syntactic in nature.

4.5. Interactions of Object Properties

Symmetrical and asymmetrical object types in Bantu also differ in the number of objects which can be affected by verbal extensions. In Chichewa, there are much stricter requirements on what

can be deleted or passivized, while Kichaga allows a degree of freedom which, as previously stated, leads to some deal of ambiguity.

4.5.1. *Co-occurrence of Passives with Object Markers*
Kichaga allows object markers to co-occur with passives, while Chichewa does not. This is an extension of the morphosyntax of Kichaga, which, as established, is significantly freer than Chichewa. In Chichewa, "the object marker never appears on a passive verb" (Bresnan and Moshi, 1990). This can be attributed to the passive extension taking as an argument the lexical entry for a verb root which seeks an overt rightward nominal. Kichaga would have no such constraints, and would therefore be able to passivize one argument and use the object marker position to refer to a second argument, such as an applicative.

4.5.2. *Unspecified Object Deletion with Passives*
Symmetrical Bantu languages allow the deletion of lexical arguments of the verb root, while asymmetrical languages such as Chichewa disallow such deletion. This would follow from the general restriction on unspecified object deletion, and is explained by the verbal categories seen in the analysis of Chichewa above. An analysis of Kichaga would result in much freer word order and syntactic arrangement at the expense of the ambiguity of many dependencies which are properly attained through the Chichewa analysis.

4.5.3. *Unspecified Object Deletion with Object Markers*
This is another case where the general issue accounts for the specific issue. Here, languages like Kichaga can delete a verb root argument with an object marker present for the argument of an extension, while Chichewa cannot. For Kichaga, this is a case where a droppable slash exists for either transitive verb lexical entry. This would be controlled in the same manner as the generality given in example (42) for Chichewa. The fact that Chichewa maintains its requirement for an overt nominal argument given an object marker for a different argument suggests that the previous lexical entry given in example (78) is sufficient to account for the data given.

4.5.4. *Co-occurence of Repiprocals with Passives*
The following examples show a grammatical construction in Kichaga and an ungrammatical construction in Kichaga:

(80) Kichaga:

shímíí *sh-í-kor-í-an-o* *(na) wachaka*
8.firebrands 8s-pres-burn-app-rec-pass (by) 2.chaga

'Firebrands are being used by the Chagas to burn each other'

(81) Chichewa:

**mikôndo* *i-na-mény-ér-an-ídw-á* *ndí alenje*
4.spears 4s-recpst-hit-app-rec-pass-fv by hunter

'Spears were used by the hunters to hit each other'

This is a result of the previously described one-to-one proclivity in Chichewa for verbal extension categories that refer to a specific kind of argument structure. The combination of applicative and reciprocal seen in example (6) suggests that the argument which is applicativized must also be

the argument which is achieving the reciprocation arrangement. This would be a ramification of lexification of verbal extensions; certain combinations of units within a lexical entry will only carry a specific semantic reading, despite the opportunity for ambiguity.

4.5.5. *Co-occurrence of Reciprocals with Unspecified Object Deletion*
Reciprocalized applicatives require an overt nominal object for the verb root in Chichewa, while this is not necessary in Kichaga. This asymmetry is accounted for in the above description of Chichewa morphosyntax. The droppable slash generality for Kichaga would account for its argument optionality.

4.6. Section Summary

In this section, I have described how differences between symmetrical and asymmetrical languages with respect to the verbal object can be accounted for via categories. This is a practical and advantageous approach because it reduces the amount of theoretical abstraction necessary to account for the capabilities of verbal object selection. By making the issue strictly combinatory, I have shown that verbal extensions and verbal roots interact to either allow or prevent certain behaviors which Kichaga and Chichewa differ in.

The necessary first step to illuminating this as a combinatory issue was to confront and question the Lexical Integrity Principle. Using the LIP as a precept, LMT does become necessary to account for the various behaviors seen above. These are combinatory in nature, but in a manner which was heretofore opaque due to the treatment of morphology given in Lexical Functional Grammar. However, with the assumptions of the LIP dispelled, we find ourselves at an proper vantage point to pursue alternate explanations for the patterns noted between Kichaga and Chichewa.

5. Summary

In this article, I have discussed several important points regarding the Lexical Integrity Principle and how it relates to syntactic analysis. I have argued that from a processing perspective, it is disadvantageous. Furthermore, I have shown that treating morphemes in the syntax allows for proper argument selection and relation of dependencies. Given this, Lexical Mapping Theory becomes an unnecessary mechanism for analysis of changes to verbal argument structure both within a single language as well as typologically. I have identified that LMT and the LIP are irreducibly intertwined; if we must assume such strict lexical integrity, LMT is necessary for argument selection, but it is questionable in its theoretical assumptions, especially given the treatment of Chichewa offered above.

Future work is necessary to further affirm that LMT is totally unnecessary. A CCG treatment of applicativization in Indonesian would reaffirm the approach given here by both confirming that applicative verbal morphemes do have a combinatory effect on argument structure and at the same time achieve the same analytical insights on argument structure that Lexical Mapping Theory provides. However, without this, the work given here does stand as a viable alternative to both the LIP and LMT without any formal cost and multiple computational and theoretical gains.

References

Aone, C. and K. Wittenburg: 1990, 'Zero morphemes in Unification-Based Combinatory Categorial Grammar'. In: *28th Annual Meeting of the Association for Computational Linguistics*. University of Pittsburgh, Pittsburgh, pp. 188–193.

Baldridge, J.: 2002, 'Lexically Specified Derivational Control in Combinatory Categorial Grammar'. Ph.D. thesis, University of Edinburgh.

Beesley, K. R. and L. Karttunen: 2003, *Finite State Morphology*. Stanford, CA: CLSI Publications.

Bozsahin, C.: 2007, 'Lexical Integrity as a Corollary of Type Dependence of Morphology and Syntax'. Draft of a forthcoming paper.

Bresnan, J. and S. A. Mchombo: 1987, 'Topic, Pronoun, and Agreement in Chichewa'. *Language: Journal of the Linguistic Society of America* **63**, 741–82.

Bresnan, J. and S. A. Mchombo: 1995, 'The Lexical Integrity Principle: Evidence from Bantu'. *Natural Language and Linguistic Theory* **13**, 181–254.

Bresnan, J. and L. Moshi: 1990, 'Object Asymmetries in Comparative Bantu Syntax'. *Linguistic Inquiry* **21**, 147–185.

Dubinsky, S. and S. R. Simango: 1996, 'Passive and Stative in Chichewa: Evidence for Modular Distinctions in Grammar'. *Language* **72**(4), 749–781.

Elwell, R.: 2005, 'A Morphosyntactic Analysis of the Ekegusii Verb'. Undergraduate honors thesis presented in April, 2005 at the State University of New York at Albany, Albany, New York.

Elwell, R.: 2007, 'The Role of Frequency in Historical Change'. Master's thesis, University of Texas at Austin.

Henderson, B.: 2005, 'The Syntax of Agreement in Bantu Relatives'. Presented given at the Ninth Annual Meeting of the Texas Linguistics Society.

Hoeksema, J.: 1985, *Categorial Morphology*. New York and London: Garland Publishing, Inc.

Kager, R.: 1999, *Optimality Theory*. Cambridge University Press.

Karttunen, L.: 1989, 'Radical Lexicalism'. In: M. Baltin and A. Kroch (eds.): *Alternative Conceptions of Phrase Structure*. Chicago University Press, Chicago, pp. 43–65.

Lakatos, I.: 1970, 'Falsification and the methodology of scientific research programmes'. In: I. Lakatos and A. Musgrave (eds.): *Criticism and the Growth of Knowledge*. Cambridge University Press, pp. 91–195.

McGinnis, M.: 2001, 'Phases and the syntax of applicatives'. In: *Proceedings of NELS*. pp. 1–17.

Mchombo, S. A.: 2005, 'Argument Binding and Morphology in Chichewa'. Presentation given at the Ninth Annual Meeting of the Texas Linguistics Society.

Mchombo, S. A.: 2007, 'On the Causative in Bantu'. Presentation given at the 38th Annual Meeting of the American Conference of African Linguistics.

Steedman, M. and J. Baldridge: 2007, 'Combinatory Categorial Grammar'. Unpublished manuscript to appear.

Watkins, M. H.: 1937, *A Grammar of Chichewa, a Bantu Language of British Central Africa*. Millwood, NY: Krause Reprint Company.

White, M.: 2004, 'Efficient Realization of Coordinate Structures in Combinatory Categorial Grammar'. *Research on Language and Computation* **4**, 39–75.