Using LFG a-structure to determine the subcategorization frames of Setswana verbs

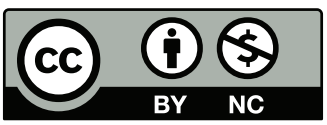
Ansu Berg
School of Languages, Subject group Setswana
North-West University, South Africa
ansu.berg@nwu.ac.za

Laurette Pretorius
College of Graduate Studies
University of South Africa (UNISA), South Africa
laurette@acm.org

Rigardt Pretorius
School of Languages, Subject group Setswana
North-West University, South Africa
rigardt.pretorius@nwu.ac.za

ABSTRACT

The verb lexicon, and more specifically verb subcategorization frames, forms an integral part of any computational grammar. In building a Lexical Functional Grammar (LFG) computational grammar for Setswana, it is therefore necessary to determine, in a principled way, the subcategorization frames for both basic verb roots as well as extended verb roots since verbal extension is a productive process in Setswana. Sequences of verbal extensions are not uncommon. In this article we investigate how four common extensions, viz. the causative, applicative, reciprocal and passive, each change the LFG argument structure of a basic verb root that they extend. We also consider what the subcategorization frame of the newly extended verb root is by applying lexical mapping theory (LMT). Due to the compositionality of the meaning of verbal extensions, we are able to generalise our results to sequences of extensions with the purpose of determining their subcategorization frames.



Keywords: LFG, argument structure, subcategorization frames, Setswana verb morphology, verbal extensions, lexical mapping theory (LMT)

Disciplinary field of study: Setswana computational linguistics

About the authors

Ansu Berg is a senior lecturer in the department of Setswana at the North-West University. The focus of her research is Setswana morphology and syntax and the computational analysis thereof.

Laurette Pretorius is emeritus professor and extraordinary professor of Computer Science at the University of South Africa. Her research in natural language processing and computational linguistics is focussed on the technological development of the South African languages.

Rigardt Pretorius is an associate professor of Setswana at the North-West University. His research on Setswana linguistics is focussed on the morphosyntax of Setswana.

1 Introduction

In studying human language, "[k]nowing about grammar ... helps us understand what makes sentences and paragraphs clear and interesting and precise".[1] Taking this position one step further, building and using computational grammars for human languages help computers understand sentences and paragraphs better. This task keeps growing in importance as more and more communication takes place in digital space in and across increasing numbers of languages. Many formalisms have been developed and are used for this purpose. Examples of well-known formalisms are Definite Clause Grammar (DCG) (Pereira and Shieber 2002), Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994), Tree-adjoining grammar (TAG) (Joshi 1985), Combinatory Categorial Grammar (CCG) (Steedman 2001), Grammatical Framework (GF) (Ranta 2011) and Lexical Functional Grammar (LFG) (Dalrymple 2001). Of specific interest here is the use of LFG for the development of a computational grammar for Setswana, a member of the Bantu family[2] of languages and one of the official, albeit less-resourced (De Pauw et al. 2011), languages of South Africa.

Conceptually, a (computational) grammar requires a set of *rules* that systematically describes the structure, more specifically the syntax and morphology, of a specific language, and a (ideally comprehensive) *lexicon* of words that may be used in building correct and meaningful phrases and sentences in this language using the mentioned grammar rules. Creating a comprehensive computational grammar lexicon is an ambitious task, with verbs being arguably one of the most complex word categories in any language (VerbNet[3], WordNet[4], FrameNet[5], PropBank[6], Palmer et al. 2017), also in Setswana.

Verbs usually express the semantics of an event being described as well as the relational information among participants. They also exhibit morphosyntactic structures that encode that information, all while displaying a rich range of semantic and syntactic behaviour. More specifically, *subcategorization* denotes the tendency of verbs to have restrictions on the arguments that they can take, and we may think of a *subcategorization frame* as a formalisation of the notion of subcategorization. In other words, subcategorization frames are the syntactic environments in which any given verb can occur, and they form part of the essential information included in a lexical entry for a given verb. Therefore, accurately representing verb subcategorization frames "constitutes a central part of any grammar development effort" (Butt et al. 1999). For example, a grammar that is able to parse the English sentence "John buys Mary flowers" has to provide in its lexicon the verb "buy" together with a subcategorization frame that allows "buy" to take as arguments a subject (John), a direct object (flowers) and an indirect object (Mary).

In LFG, argument structure (*a-structure*) deals with predicate-argument information such as the number and types of arguments of a predicator, usually a verb. This structure represents the essential interface between the semantics and syntax of predicators. Semantically, "a-structure represents the core participants in events/states/actions designated by the predicator" while, syntactically, this structure "represents the minimal information needed to characterize

[1] http://www2.ncte.org/statement/qandaaboutgrammar/, Position statement "Some Questions and Answers about Grammar", July 1, 2002, NCTE.
[2] https://www.ethnologue.com/subgroups/sotho-tswana-s31
[3] https://verbs.colorado.edu/verbnet/
[4] https://wordnet.princeton.edu/
[5] https://framenet.icsi.berkeley.edu/fndrupal/
[6] https://verbs.colorado.edu/~mpalmer/projects/ace.html

the syntactic dependents of an argument-taking head" (Bresnan et al. 2016, 326). Therefore, a-structure is important in the context of lexical items and their subcategorization frames as it facilitates the mapping between thematic (semantic) roles and grammatical functions. Indeed, in LFG, a-structures "are lexical syntactic structures" (Bresnan et al. 2016, 329) that allow the representation of subcategorization frames of verbs. More specifically, it is by means of the LFG sub-theory of Lexical Mapping Theory (LMT) that the lexicon and syntax are connected and the subcategorization frames of verbs determined (Bresnan et al. 2016, 329). Thus, by employing LMT, a systematic approach may be followed to determine these frames for subsequent use in the lexicon - in our case that of the Setswana computational grammar.

The Setswana verb has a complex (agglutinative) morphology where various prefixes and suffixes can be added to the verb root. These prefixes and suffixes have an influence on tense, aspect, mood and polarity (TAMP), as well as, crucially, on meaning. Moreover, the productive verb extensions (suffixes) i.e. the causative, applicative, reciprocal and passive extensions also have an important influence on the valency of a verb, the capacity of a verb to take a specific number and type of arguments (Mchombo 2007, 203-204; Khumalo 2007, 13; Khumalo 2014, 145; Chavula 2016, 16).

In this article we show how LMT may be employed to determine verb subcategorization frames for Setswana verbs in a principled way. We systematically consider basic roots (roots with no verbal extensions), so-called extended verb roots with a single verbal extension, and also extended verb roots with multiple verbal extensions with each newly added extension possibly changing the valency of the verb root that it extends (Berg 2018).

The structure of the article is as follows: in section 2, we present, by way of context, a brief description of LFG c- and f-structure and explain LFG a-structure and LMT. Section 3 contains an exposition of the morphological structure of the Setswana verb, with specific emphasis on the root, its prefixes and its (sequences of) valency changing productive verbal extensions. Sections 4 and 5 constitute the main research contribution. They are devoted to the application of LMT to determine the subcategorization frames (SCFs) of verbs with a single verbal extension and multiple extensions, respectively. Section 6 concludes the article and proposes a number of ideas for future work.

2 Lexical Functional Grammar

The LFG grammar architecture consists of several parallel levels of representation, i.e. the c(onstituent)-structure, m(orphology)-structure, a(rgument)-structure, f(unctional)-structure, s(emantic)-structure, p(honological)-structure and i(nformation)-structure (Kaplan 1995, 23-24). These levels are mutually constraining through so-called functional projections (correspondence functions). For the purposes of grammar development in LFG, sentences are analysed in terms of c- and f-structure (Kaplan & Bresnan et al. 1995, 175). The *c-structure* is defined by language-specific context-free production rules that model word order and phrase structure while the *f-structure* represents functional or syntactic information about the internal structure of the sentence (Dalrymple 2001, 7). The *f-structure* contains surface grammatical functions (GFs) (cf. §2.2) as well as features that represent the morphosyntactic properties of constituents. In Setswana, these features are linguistic notions such as noun class, person, number, tense, aspect, and mood. The representation of the f-structure is formalised through

an attribute-value matrix (AVM), a set of pairs where the first member of the pair indicates the attribute while the second member expresses the value of that attribute (Dalrymple 2001, 30). For an example of such an AVM, see the f-structure in Diagram 2.

2.1 Argument structure

We reiterate that a-structure deals with predicate-argument information such as the number and types of arguments of a predicator. In terms of semantics, a-structure represents the core participants in actions designated by the predicator. In terms of syntax, it represents the nucleus of information required to characterise the syntactic dependents of an argument-taking head (Bresnan 2001, 304). In LFG, these two representations are linked through LMT. Before explaining this theory briefly, we introduce the central notion of grammatical functions (GFs).

2.2 Grammatical functions

In LFG, GFs are considered primitives of syntactic representation. The LFG inventory of GFs consists of the following: subject (SUBJ), object (OBJ), secondary object (OBJ_θ), oblique (OBL_θ), complement (COMP), open complement (XCOMP), adjunct (ADJUNCT) and open adjunct (XADJ). The subscript θ in OBJ_θ always refers to a secondary object while the subscript θ in OBL_θ refers to a semantic role (cf. Table 1), for example the subscript *theme* in OBL_{theme} and the subscript *loc* in OBL_{loc} refer to the semantic roles associated with the respective values of θ. These GFs may be cross-classified according to certain properties (Bresnan 2001, 94–98, 307–309; Dalrymple 2001, 8–27):

- A *governable* (argument) function is a GF that is subcategorised for or governed by the predicate: SUBJ, OBJ, OBJ_θ, OBL_θ, COMP and XCOMP. A GF that is not governable is often referred to as a *modifier* (nonargument function): ADJUNCT and XADJ. Governable GFs satisfy the so-called uniqueness requirement, i.e. each predicator may subcategorise only for a single GF of each type. Modifiers are not subject to this requirement.
- The *core* GFs (terms) are SUBJ, OBJ and OBJ_θ while *noncore* GFs (nonterms) are OBL_θ, COMP and XCOMP.
- A *semantically restricted* GF can only be associated with a limited set of semantic roles: OBJ_θ and OBL_θ. A s*emantically unrestricted* GF may be connected to any semantic role or even more than one semantic role: SUBJ and OBJ.
- An *open* GF does not contain an internal subject phrase and the SUBJ must be specified externally to its phrase: XCOMP and XADJ. A *closed* GF contains an internal subject phrase: SUBJ, OBJ, OBJ_θ, OBL_θ, COMP and ADJUNCT.
- The *objective* GFs are OBJ and OBJ_θ while the *nonobjective* GFs are SUBJ and OBL_θ.
- LFG also allows for grammaticalized *discourse* functions: topic (TOP), focus (FOC) and SUBJ. The SUBJ has the unique property of being both an argument function and a grammaticalized discourse function (Bresnan 2001, 98, 308).
- For the copulative verb, the semantic relationship between the SUBJ and the phrase following the verb necessitates a specific subcategorization frame. Traditionally, this phrase was

assigned an XCOMP function. A more recent development in LFG has been the introduction of the so-called PREDLINK function (Butt et al. 1999, 70; Attia 2008, 141–171; Sulger 2009, 32). Butt et al. (1999, 69) explain that predicative constructions include a copulative verb (linking verb) that takes a subject and another argument. PREDLINK is a closed category and "there is no control equation between the SUBJ and the PREDLINK and hence no need for NPs, APs, and PPs to have subject arguments" (Butt et al. 1999, 70).

2.3 Lexical Mapping Theory

In LFG, LMT lies at the heart of connecting the lexicon and the syntax, more specifically, it facilitates the formalisation of SCFs for lexical categories (Alsina & Mchombo 1993, 24–27; Butt 1995, 30–32; Bresnan 2001, 307; Dalrymple 2001, 195–213; Falk 2001, 93–114; Ackerman & Moore 2013, 9–20; Her 2013, 47–57; Bresnan et al. 2016, 324–344). A number of variations of or modifications to LMT have been suggested and LMT remains an active field of research, also in Bantu linguistics (see, for example, Findlay 2016; Szűcs 2018; Li 2019; Jerro 2015). However, a detailed discussion of the field of LMT falls outside the scope of this article. We opted for LMT as in Bresnan et al. (2016) and return to this topic in section 3.2.2.

The basic principles of LMT are as follows:

(i) A predicator with its argument roles is stated as an ordering that represents the decreasing relative prominence of the roles. The predicator is usually a verb and the argument roles are selected from a hierarchy of semantic roles in accordance with the lexical semantics of the predicator. The hierarchy of the thematic (semantic) roles is *agent* > *beneficiary* > *experiencer/goal* > *instrument* > *patient/theme* > *locative*. The various roles[7] are described in Table 1.

Table 1: Description of the thematic (semantic) roles

Thematic role (θ)	**Description**
Agent (ag)	The initiator of the action, capable of volition
Beneficiary (ben)	The entity for whose benefit the action is performed
Experiencer (exp)	The living entity that experiences the action or event
Goal (goal)	The location or entity in the direction of which something moves
Instrument (instr)	The medium by which the action or event is carried out
Patient (pat)	The entity affected by the action, undergoes change of state
Theme (theme)	The entity that is moved by or directly receives the action or event
Locative (loc)	The specification of the place where the action or event is situated

[7] Adapted from http://elies.rediris.es/elies11/cap5111.htm, http://www.linguisticsnetwork.com/semantics-thematic-roles/ and http://folk.uio.no/liljao/inf5830/inf5830_semanticroles.pdf

In (1) the a-structure of the Setswana example includes the predicator and its argument roles.

(1) *rekel-* 'buy for' < agent, beneficiary, theme >

(ii) A syntactic classification of each role is indicated by a feature. These features are [+o], [-o], [+r] and [-r], respectively, referring to objective, non-objective, restricted and unrestricted GFs. Assigning features to thematic (semantic) roles in the a-structure is based on correspondences shown in Diagram 1.

According to the Intrinsic Role Classification (IRC), patientlike roles map to [-r], secondary patientlike roles map to [+o], and all other thematic roles map to [-o]. The most prominent thematic role (often agent) will map to [-o] by virtue of the above principle and to [-r] due to its leftmost (highest) position in the hierarchy of thematic roles. The Default Role Classification (DRC) assigns [-r] to the most prominent thematic role and [+r] to all other roles. No Default assignment is allowed to a thematic role that has already been mapped to [-r].

	-r	+r
-o	SUBJ	OBL_{θ}
+o	OBJ	OBJ_{θ}

Diagram 1: Correspondences between features of argument functions

(iii) The IRC and DRC are applied, as in (2).

(2)	*rekel-*	'buy for'	<	agent	beneficiary	theme >
	IRC:			[-o]	[+o]	[-r]
	DRC:			[-r]	[+r]	

Next, the correspondence matrix in Diagram 1 is applied, obtaining (3).

(3)	*rekel-*	'buy for'	<	agent	beneficiary	theme >
	IRC:			[-o]	[+o]	[-r]
	DRC:			[-r]	[+r]	|
				|	|	|
				SUBJ	OBJ_{θ}	SUBJ/OBJ

(iv) The mapping is completed when the intrinsic and default argument role classifications are further constrained by following the subject and bi-uniqueness conditions that are well-formedness conditions on the relation between thematic roles and GFs:

***Subject condition**:* Every predicator must have a SUBJ. The most prominent thematic role ([-o])

is the agent, which is realised as the SUBJ. If an agent is not available, the thematic role that follows ([-r]) is realised as the SUBJ.

Bi-uniqueness condition: A thematic role in the a-structure must be associated with a unique function, and vice versa.

Finally, as shown in (4), a thematic role is associated with only one GF and the GFs are not associated with more than one thematic role.

(4)	*rekel-*	'buy for'	<	agent	beneficiary	theme	>
	IRC:			[-o]	[+o]	[-r]	
	DRC:			[-r]	[+r]	|	
				|	|		
				SUBJ	OBJ_θ	SUBJ/OBJ	
	bi-uniqueness:			SUBJ	OBJ_θ	OBJ	

We have therefore shown that by starting with the lexical semantics of a predicator (here the extended verb root *rekel-*), we are able to obtain its SCF. Indeed, *rekel-* subcategorises for a SUBJ, OBJ_θ and OBJ.

3 Setswana verbs

3.1 Introduction

Setswana is classified as a Bantu language and, as such, has an extensive noun class system, a system of grammatical agreement and an agglutinative morphology. Setswana verbs are subdivided into main, auxiliary and copulative verbs. Main verbs, the focus of this article, can be intransitive, transitive or ditransitive. A detailed exposition of Setswana morphology and syntax falls outside the scope of this article, but may be found in Krüger (2006).

The formation of Setswana verbs is governed by a set of linguistic rules (morphotactics) where prefixes and suffixes may be sequenced and combined to form valid verb forms and by phonological and orthographical alternation (morphophonological alternation rules) that model the sound changes that occur at morpheme boundaries (Pretorius, L., et al. 2010, 133–134). In terms of morphotactics the prefixes and suffixes have a fixed order in the verb structure and are also structural elements that execute a process of adapting or extending the meaning of a word. They may also have an influence on tense, aspect, mood and polarity (TAMP) (Krüger 2006, 268). The presence of the productive extensions in the morphological structure of a verb not only extends and alters the meaning of the verb root, but may also affect its valency. In order to determine the SCF of the (extended) verb root systematically, we need to consider these valency changes and their effect on the a-structure of the extended verb root.

3.2 The morphological structure of Setswana verbs

The core of the main verb is the root. The main verb may include inflectional prefixes as well as derivational and inflectional suffixes.

3.2.1 The root

The root is usually a bound morpheme that carries the basic lexical meaning of the word and requires one or more affixes to form a complete word (Kosch 2006, 7). The Setswana verb *ba a lela* 'they cry' consists of the subject agreement morpheme *ba*, the present tense morpheme *a*, the root *lel-* and the verbal ending *-a*. The valency of such a verb root is determined by the meaning of the root (Pretorius, R., et al. 2012, 208–209). For example, in (5) the verb root *lel-* 'cry' is intransitive, based on its meaning. It should therefore only subcategorise for a SUBJ *bana* 'children'.

(5) *Bana **ba a lela**.*[8]
ba-ana ba-a-lel-a
NPRE2-child AGRSUBJ2-PRESPRE-cry-VEND
'The children are crying.'

In (6) the verb root *rek-* 'buy' is transitive (as before based on its meaning) and should therefore subcategorise for a SUBJ *basadi* 'women' and an OBJ *ditlhako* 'shoes'.

(6) *Basadi **ba reka** ditlhako.*

ba-sadi	ba-rek-a	di-tlhako
NPRE2-woman	AGRSUBJ2-buy-VEND	NPRE8-shoe

'The women buy shoes.'

Based on its meaning, the verb root *f-* 'give' in (7) is ditransitive and therefore should subcategorise for a SUBJ *basadi* 'women', an OBJ_{θ} *bana* 'children' and an OBJ *ditlhako* 'shoes'.

(7) *Basadi **ba fa** bana ditlhako.*

ba-sadi	ba-f-a	ba-ana di-tlhako
NPRE2-woman	AGRSUBJ2-give-VEND	NPRE2-child NPRE8-shoe

'The women give the children shoes.'

[8] Note on glossing in the Setswana examples: The approach taken in presenting the examples is based as far as possible on the Leipzig glossing rules and the guidelines of this journal. However, we deviate slightly from these to make our examples clear. Due to the disjunctive orthography of Setswana, we distinguish between an *orthographic word* and a *linguistic word* (Kosch 2006, 3-4). For this purpose, we have four lines instead of the usual three. In line one, the surface form of the Setswana example is given. The verb is given in bold face, and, in this case, consists of three orthographic words, but one linguistic word. In line two, the linguistic words are shown. The hyphen is used to separate the morphemes in each linguistic word. This allows for an accurate and succinct representation of the morphology of linguistic words in line three. For the sake of consistency, we use our own tagset, provided at the end of the article.

When a causative, applicative, reciprocal or passive extension is suffixed to a basic verb root an extended verb root is formed. The extended root denotes a polymorphemic structure that consists of the basic root plus at least one suffix, excluding the verbal ending (Kosch 2006, 7). For example, the extended verb root *rekel-* 'buy for' consists of the basic root *rek-* 'buy' and the applicative extension *-el-*.

3.2.2 The subject and object agreement morphemes

The verb root can be preceded by several prefixes i.e. a subject agreement morpheme (obligatory, except in the imperative mood), a negative morpheme (occupying one of two possible slots), the present tense, progressive, potential and temporal morphemes, one or more object agreement morphemes (see (9) and (10) below), and the reflexive morpheme.

The subject agreement morpheme and the object agreement morpheme indicate class gender, person and number. A subject agreement morpheme usually appears with an overt SUBJ, for example, in (5) the subject agreement morpheme *ba* in the verb *ba a lela* 'they cry' shows agreement with the SUBJ *bana* 'children'. However, if the SUBJ is known the subject agreement morpheme can also represent this GF, for example in (8) the subject agreement morpheme *ba* refers to a known, but omitted SUBJ.

(8) ***Ba a lela***.
ba-a-lel-a
AGRSUBJ2-PRESPRE-cry-VEND
'They are crying.'

Setswana is therefore a so-called pro-drop language as it is not obligatory for an overt SUBJ to be included in a sentence. The subject agreement morpheme not only exhibits the agreement between a verb and its SUBJ, but also performs an anaphoric function rendering a pronominal interpretation of the missing argument. It therefore enables the reconstruction of the missing SUBJ and behaves as an incorporated pronoun that functions as the SUBJ in the sentence (Krüger 2006, 32, 53). In LFG the special PRED attribute with a 'PRO' value is distinguished in the f-structure (Butt et al. 1999, 13). The f-structure in Diagram 2 shows that the subject agreement morpheme *ba* in (8) has a PRED attribute where the value is pronominal ('PRO').

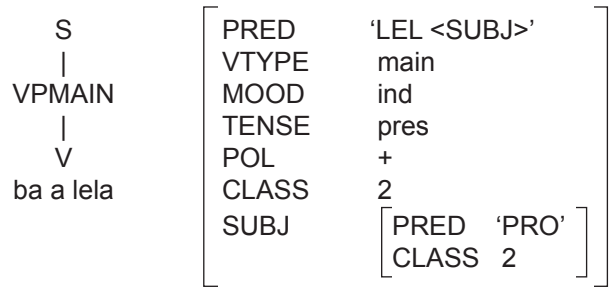


Diagram 2: The c- and f-structure of *Ba a lela* in (8)

An object agreement morpheme represents an omitted OBJ and an anaphoric relationship exists

between the object agreement morpheme and the omitted OBJ (Louwrens 1994, 11). Similar to the subject agreement morpheme, the object agreement morpheme also behaves as an incorporated pronoun where the PRED attribute includes a 'PRO' value and this is also included in the f-structure (Mchombo 2004, 20–22). In (10) the object agreement morpheme *di* in the verb *ba a di reka* 'they buy it' refers to an omitted object. The f-structure in Diagram 3 shows that the object agreement morpheme *di* in (9) has a PRED attribute where the value is pronominal ('PRO').

(9) *Basadi* ***ba a di reka***.
ba-sadi ba-a-di-rek-a
NPRE2-woman AGRSUBJ2-PRESPRE-AGROBJ8-buy-VEND
'The women buy them.'

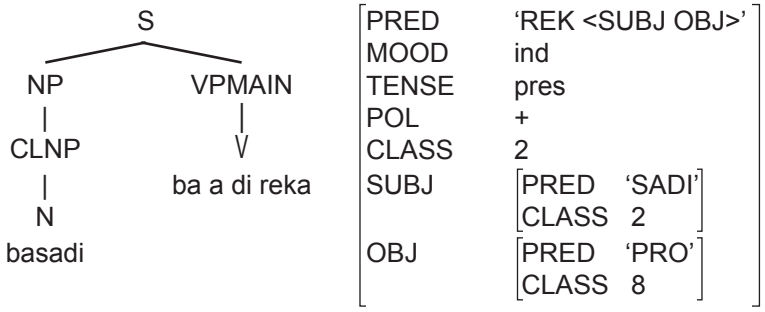


Diagram 3: The c- and f-structure of *Basadi ba a di reka* in (9)

Moreover, it is important to note that the SCF of the verb root occurs in the f-structure. In our examples (8) and (9) they are 'TSAMAY <SUBJ>' and 'REK <SUBJ OBJ>'.

By way of illustration we also consider the example (10). The f-structure in Diagram 4 shows that both the object agreement morphemes *di* and *mo* in (10) have a PRED attribute where the value is pronominal ('PRO'). Moreover, the SCF of the verb *rekel-* 'buy for' is REKEL<SUBJ, OBJ_θ OBJ>, in accordance with (4).

(10) *Basadi* ***ba a di mo rekela***.
ba-sadi ba-a-di-mo-rek-el-a
NPRE2-woman AGRSUBJ2-PRESPRE- AGROBJ8-AGROBJ1-buy-VEND
'The women buy them for her.'

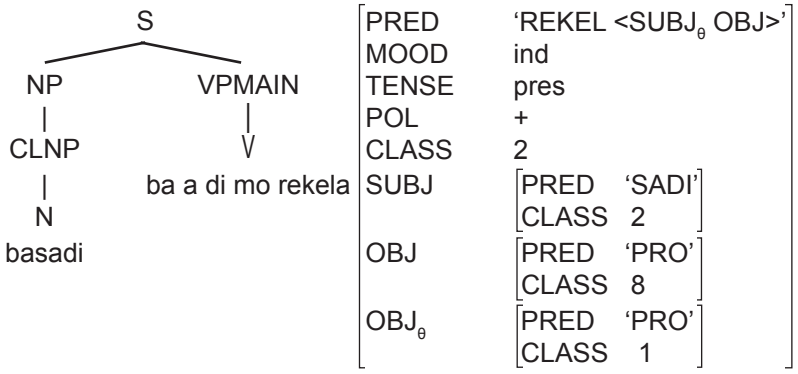


Diagram 4: The c- and f-structure of *Basadi ba a di mo rekela* in (10)

We now return to our statement that Setswana can have "one or more object agreement morphemes" (Pretorius, R., et al. 2012), shown in the above examples. Marten and Kula (2012) provide a detailed exposition of morphosyntactic variation in object marking in Bantu and found that Bantu languages exhibit notable variation in this regard. In a typology of object symmetry, Marten et al. (2007) classified Setswana as symmetrical with respect to word order, passivisation and object marking (Jerro 2015). In Pretorius et al. (2012) we confirm that Setswana exhibit, amongst others, the following characteristics:

1. "There may be more than one object agreement morpheme per verb in Setswana";

2. "Two object agreement morphemes may be used in Setswana in certain linguistic instances. These instances include a number of verbs which allow indirect objects based on their meaning, certain verbs with the causative suffixes, certain possessive constructions, and some verbs with the applied suffixes ... The use of two object agreement morphemes appears infrequently in Setswana (Cole 1955)."

3. "Marten et al. (2007, 267) indicate that the order of multiple object agreement morphemes in Setswana is free. Technically this is true, but it is our contention that there may be preferences based on discourse pragmatics." Pretorius et al. (2012) provide attested examples, but conclude that "[t]he reason for this phenomenon is not clear and more (corpus-based) research is necessary in this context."[9]

4. "Marten et al. (2007, 269, Example 27) indicate that either object may follow the verb in Setswana. Without any scientific proof to the contrary, this to us seems possible, but rather irregular. Object agreement morphemes may be used in any order (our Example 22), but it is doubtful whether this is possible with objects themselves. Since the focus in this article is on object agreement morphemes, the matter of the order of objects falls outside its scope."

5. "Either object can be the subject of the sentence in the passive."

6. "Either object can be expressed by an object agreement morpheme."

It has been shown that the LMT of Bresnan et al. (2016) presents problems in catering for two object symmetry, and modifications to LMT have been proposed (see, for example, (Her 2013) and (Jerro 2015)). The question now arises: why did we not use one of these modified LMT's in our work? The explanation is as follows: from the findings 1–6 above it is clear that full object symmetry in Setswana has not been shown. By "full" we mean that in each occurrence of two objects, both usages are equally acceptable and equally common. In general, there still seems to be a preferred usage in most cases where such symmetry can linguistically occur. In our work we adopted the pragmatic view of focussing on the usages that are commonly used and accepted. An investigation into the use of variants of LMT for capturing relevant aspects of object symmetry in Setswana forms part of our future work.

[9] "Applied and causative forms of transitive verbs may have two, and sometimes even three, non-coordinate objects … When there are two such objects, the one is the object of the simple form of the verb, the other the object of the derivative form. The latter termed the principle object, is almost invariably expressed or represented by its OC, the former, termed the secondary of subsidiary object may be merely implied. Though both objects may be either alternatively or simultaneously represented by OCs, it is more usual for the principal object to be so treated. When both objects are expressed or represented by their OCs, the principle object, or its concord, has precedence of order and is placed nearest the verb" (Cole 1955, 430).

3.2.3 The productive verbal extensions

Various derivational or inflectional morphemes may be suffixed to the verb root i.e. the causative, reciprocal, applicative[10], passive and perfect suffixes, a verbal ending and the relative or imperative suffixes. The causative, reciprocal, applicative and passive suffixes are so-called productive verbal extensions and are derivational. The perfect suffix is an inflectional morpheme and specifies either past tense or perfective aspect. The verbal endings are inflectional morphemes and the specific form of the verbal ending is determined by mood, tense and polarity information (Krüger 2006, 268).

The causative extension

The causative extension adds a meaning of *cause to do*, *make to do*, *help to do* or *assist in doing* to the verb. The two phonologically related causative extensions in Setswana are *-is-* and *-y-* (Cole 1955, 203). These extensions change intransitive verbs to transitive verbs and transitive verbs to ditransitive ones (Krüger 2006, 227). For example, the intransitive verb *o a taboga* 'he runs' in (11) changes to a transitive verb *o tabogisa* 'she lets run' in (12) and the transitive verb *ba kwala* 'they write' in (13) changes to a ditransitive verb *o kwadisa* 'she lets write' in (14). The verb in (12) requires the inclusion of an OBJ *ngwana* 'child' while the verb in (14) requires the inclusion of both an OBJ_{θ} *bana* 'children' and an OBJ *teko* 'test'.

(11) *Ngwana* ***o a taboga****.*
mo-ana o-a-tabog-a
NPRE1-child AGRSUBJ1-PRESPRE-run-VEND
'The child is running.'

(12) *Mosadi* ***o tabogisa*** *ngwana.*
mo-sadi o-tabog-is-a mo-ana
NPRE1-woman AGRSUBJ1-run-CAUSSUF-VEND NPRE1-child
'The woman lets the child run.'

(13) *Bana* ***ba kwala*** *teko.*
ba-ana ba-kwal-a (ne)-teko
NPRE2-child AGRSUBJ2-write-VEND NPRE9-test
'The children are writing a test.'

(14) *Mosadi* ***o kwadisa*** *bana teko.*
mo-sadi o-kwal-is-a ba-ana (ne)-teko
NPRE1-woman AGRSUBJ1-write-CAUSSUF-VEND NPRE2-child NPRE9-test
'The woman lets the children write a test.'

The applicative extension

An applicative extension indicates that the action of the verb is carried out *on behalf of*, *to the*

[10] The applicative extension is often also referred to as the applied extension.

advantage of or *with respect to something or some place* (Cole 1955, 199). The applicative extension in Setswana is *-el-* with the phonological variants *-ets-* and *-lets-*. It changes an intransitive verb to a transitive verb and a transitive verb to a ditransitive verb (Pretorius, R., et al. 2012, 211). For example, the intransitive verb *o a tshameka* 'he plays' in (15) changes to a transitive verb *o tshamekela* 'he plays for' in (16) and the transitive verb *o reka* 'she buys' in (17) changes to a ditransitive verb *o rekela* 'she buys for' in (18). The verb in (16) requires the inclusion of an OBJ *setlhopha sa ntlha (sa sekolo sa rona)* 'first team (of our school)' while the verb in (18) requires the inclusion of both an OBJ_{θ} *bana* 'children' and an OBJ *dijo* 'food'.

(15) *Katlego **o a tshameka**.*
(-)-Katlego o-a-tshamek-a
NPRE1A-Katlego AGRSUBJ1A-PRESPRE-play-VEND
'Katlego is playing.'

(16) *Katlego **o tshamekela** setlhopha sa ntlha (sa sekolo sa rona).*
(-)-Katlego o-tshamek-el-a se-tlhoph-a
NPRE1A-Katlego AGRSUBJ1A-play-APPLSUF-VEND NPRE7-choose-DEVSUF
sa ntlha (sa se-kolo sa
POSSPART7 ADV (POSSPART7 NPRE7-school POSSPART7
rona)
POSSPROP1PL)
'Katlego plays for the first team (of our school).'

(17) *Mosadi **o reka** dijo.*
mo-sadi o-rek-a di-j-o
NPRE1-woman AGRSUBJ1-buy-VEND NPRE10-eat-DEVSUF
'The woman buys food.'

(18) *Mosadi **o rekela** bana dijo.*
mo-sadi o-rek-el-a ba-ana
NPRE1-woman AGRSUBJ1-buy-APPLSUF-VEND NPRE2-child
di-j-o
NPRE10-eat-DEVSUF
'The woman buys food for the children.'

The reciprocal extension

The reciprocal extension *-an-* expresses a meaning of *each other* or *one another* and suggests at least two participants (Cole 1955, 210). These two participants can be represented in Setswana by three different syntactic realisations i.e. a plural noun (*bana* 'children' in (19)) or plural pronoun, a coordinate phrase (*mosimane le mosetsana* 'boy and girl' in (20)) or a discontinuous reciprocal construction in which the one participant is the SUBJ and the other participant is represented by means of an associative particle phrase (*le wena* 'with you' in (21)).

(19) *Bana **ba a ratana**.*
ba-ana ba-a-rat-an-a
NPRE2-children AGRSUBJ2-PRESPRE-love-RECSUF-VEND
'The children love each other.'

(20) *Mosimane le mosetsana **ba a ratana**.*
mo-simane le mo-setsana ba-a-rat-an-a
NPRE1-boy CONJ NPRE1-girl AGRSUBJ2-PRESPRE-love-RECSUF-VEND
'The boy and girl love each other.'

(21) *Mosimane **o dumelana** le wena.*
mo-simane o-dumel-an-a le wena
NPRE1-boy AGRSUBJ1-agree-RECSUF-VEND ASSPART PERSPROP2SG
'The boy agrees with you.'

The passive extension

The passive extension *-iw-* has a variant form *-w-* and can be suffixed to both intransitive and transitive verbs. In Setswana, the SUBJ of an active sentence functions as the complement of the agentative particle *ke* 'by' in an agentative particle phrase which functions as an ADJUNCT (Cole 1955, 192–196; Krüger 2006, 254–259). The transitive verb *o roma* 'she sends' in (22) changes to the intransitive passive verb *ba romiwa* 'they are sent' in (23).

(22) *Mosadi **o roma** bana.*
mo-sadi o-rom-a ba-ana
NPRE1-woman AGRSUBJ1-send-VEND NPRE2-child
'The woman sends the children.'

(23) *Bana **ba romiwa** ke mosadi.*
ba-ana ba rom-iw-a ke mo-sadi
NPRE2-child AGRSUBJ2 send-PASSSUF-VEND AGPART NPRE1-woman
'The children are sent by the woman.'

Now that we have explained the valency and meaning changes that occur due to the various verbal extensions, we are ready to use these insights regarding the lexical semantics of Setswana (extended) verb roots as predicators to obtain their SCFs, by applying LMT.

4 Subcategorization frames of Setswana verbs

As described in section 2.3, the application of LMT requires (i) identifying thematic roles, making use of the descriptions in Table 1, (ii) assigning the semantic features [-o], [+o], [-r] and [+r] to the thematic roles according to the IRC and the DRC, (iii) mapping semantic features to the appropriate argument functions by using the correspondence matrix in Diagram 1 and (iv) invoking the subject and bi-uniqueness conditions, if applicable. In the following four sub-

sections we use specific examples to show how the different verbal extensions manifest in the SCFs of the specific extended verb roots, and then formulate general patterns.

4.1 The causative extension

In (11) the only thematic role present is that of agent viz. *ngwana* 'child'. According to the IRC and DRC, the semantic features [-o] and [-r] are assigned to the agent and via the correspondence matrix (Diagram 1) map the agent to SUBJ (24). This mapping is valid for and generalises to any verb root that has agent as thematic role.

In (12) the causative extension to the basic verb root *tabog-* 'run' results in a change of meaning that requires not only an agent (*mosadi* 'woman'), but also an obligatory patient role (*ngwana* 'child') in the a-structure of the extended verb root *tabogis-* 'let run'. Applying LMT, the agent maps to SUBJ by the same reasoning as in the previous paragraph, while according to the IRC, the patient obtains the [-r] feature. No DRC is necessary since [-r] has already been assigned. In this instance, the correspondence matrix (Diagram 1) maps [-r] to both SUBJ and OBJ as indicated in the [-r] column. The bi-uniqueness condition is then invoked to disambiguate this mapping and yields OBJ since SUBJ has already been assigned (25).

(24)

tabog- 'run' <	agent >
IRC:	[-o]
DRC:	[-r]
	|
	SUBJ

(25)

tabogis- 'let run' <	agent	patient >
IRC:	[-o]	[-r]
DRC:	[-r]	|
	|	|
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

In (13), the agent is *bana* 'children' and the transitive basic verb root *kwal-* 'write' has *teko* 'test' as theme. According to the IRC, the semantic feature [-r] is assigned to this role. By means of the DRC and the bi-uniqueness condition, it maps to OBJ (26). In (13), the extended verb root *kwadis-* 'let write' has *mosadi* 'woman' as agent, *bana* 'children' as patient and *teko* 'test' as theme (27). As before, the agent role maps to SUBJ, and the theme role to OBJ. By applying the IRC, the DRC and the correspondence matrix, the patient is assigned the [+o] and [+r] semantic features that maps to OBJ_{θ}, where θ has the value of patient.

(26) *kwal-* 'write' < agent theme >

IRC:	[-o]	[-r]
DRC:	[-r]	|
	|	
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

(27) *kwadis-* 'let write' < agent patient theme >

IRC:	[-o]	[+o]	[-r]
DRC:	[-r]	[+r]	|
	|	|	
	SUBJ	OBJ_{θ}	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ_{θ}	OBJ

In example (28) we see that the causative extension can also influence a-structure as follows: When the basic verb root *tsen-* 'enter' is extended to *tseny-* 'put' ('let enter' is lexicalised as 'put'), the a-structure not only (semantically) requires an agent and a patient, but also an obligatory locative role. In (29) the agent is *mosadi* 'woman', the theme *buka* 'book' and the locative is the phrase *mo kgetsing* 'in the bag'. The agent and theme roles are mapped as SUBJ an OBJ respectively. By applying the IRC and DRC, the respective semantic features of the locative phrase are [-o] and [+r] and it maps to OBL_{loc}.

(28) *Mosadi* ***o tsenya*** *buka mo kgetsing.*

mo-sadi	o-tsen-y-a	(ne)-buka
NPRE1-woman	AGRSUBJ1-enter-CAUSSUF-VEND	NPRE9-book
mo	(ne)-kgetsi-ing	
$LOCPART_{mo}$	NPre9-bag-LOCSUF	

'The woman puts the book in the bag.'

(29) *tseny-* 'put' < agent theme locative >

IRC:	[-o]	[-r]	[-o]
DRC:	[-r]	[+r]	|
	|	|	
	SUBJ	SUBJ/OBJ	OBL_{loc}
bi-uniqueness:	SUBJ	OBJ	OBL_{loc}

Thus, in f-structure notation the SCFs of our examples are:
'TABOG<SUBJ>',but
'TABOG**IS**<SUBJ, **OBJ**>';
'KWAL<SUBJ, OBJ>', but
'KWAD**IS**<SUBJ, OBJ, **OBJ$_\theta$**>'; and
'TSEN**Y**<SUBJ, OBJ, **OBL$_{loc}$**>'.

In general, the Setswana causative extension is a *valency increasing extension* as it allows the inclusion of (i) a patient, (ii) both a patient and theme or (iii) both a theme and a locative in the a-structure of an extended verb root. A verb root with a causative extension may, in addition to the obligatory SUBJ, therefore subcategorise for an OBJ, both an OBJ$_\theta$ and an OBJ, or both an OBJ and an OBL$_{loc}$.

4.2 The applicative extension

In (15) and (16), the basic verb root *tshamek-* 'play' is extended to *tshamekel-* 'play for', the a-structure of which not only (semantically) requires an agent but also an obligatory beneficiary role. As before, the agent (*Katlego*) maps to the SUBJ. In (16), the beneficiary *setlhopa sa ntlha* 'first team' is classified as [-r] and by applying the correspondence matrix and the bi-uniqueness condition it is then mapped to OBJ (30).

(30)

tshamekel-	'play for'	<	agent	beneficiary	>
IRC:			[-o]	[-r]	
DRC:			[-r]	|	
			|	|	
			SUBJ	SUBJ/OBJ	
bi-uniqueness:			SUBJ	OBJ	

The extended verb root *rekel-* 'buy for' in (18) requires not only the agent role (*mosadi* 'woman'), but also the obligatory roles of beneficiary (*bana* 'children') and theme (dijo 'food') in its a-structure. As usual, the agent maps to SUBJ. The theme in (18) is classified as [-r] and by applying the correspondence matrix and the bi-uniqueness condition it is mapped to OBJ (31). The beneficiary takes the semantic features [+o] and [+r] in terms of the IRC and DRC. By applying the correspondence matrix, it is mapped to OBJ$_\theta$.

(31)

rekel-	'buy for'	<	agent	beneficiary	theme	>
IRC:			[-o]	[+o]	[-r]	
DRC:			[-r]	[+r]	|	
			|	|	|	
			SUBJ	OBJ$_\theta$	SUBJ/OBJ	
bi-uniqueness:			SUBJ	OBJ$_\theta$	OBJ	

The applicative extension changes the intransitive verb root *tsamay-* 'walk' in (32) to a transitive extended verb root *tsamael-* 'walk towards' in (33). As usual, the agent *batho* 'people' maps to SUBJ (33). The phrase *kwa nokeng* 'to the river' in (34) has a locative role. By applying the IRC and DRC, the respective semantic features of this phrase are [-o] and [+r] and it is then mapped to OBL_{loc}.

(32) *Batho **ba a tsamaya.***
ba-tho ba-a-tsamay-a
NPRE2-person AGRSUBJ2-PRESPRE-walk-VEND
'The people are walking.'

(33) *Batho **ba tsamaela** kwa nokeng.*
ba-tho ba-tsamay-el-a kwa
NPRE2-person AGRSUBJ2 walk-APPLSUF-VEND LOCPART$_{kwa}$
(ne)-noka-ing
NPRE9-river-LOCSUF
'The people walk towards the river.'

(34)

tsamael-	'walk towards	<	agent	locative	>
IRC:			[-o]	[-o]	
DRC:			[-r]	|	
			|	|	
			SUBJ	OBL_{loc}	

The applicative extension also changes the intransitive verb root *tsamay-* 'walk' in (35) to a ditransitive extended verb root *tsamael-* 'walks to' in (36), the a-structure of which obligates an agent, a beneficiary and a locative role. By following the previously described procedure, the agent (*ngwana* 'child') is mapped to SUBJ, while the beneficiary (*mmaagwe* 'his mother') and the locative (*kwa toropong* 'to town') are mapped to OBJ and OBL_{loc} respectively (37).

(35) *Ngwana **o a tsamaya**.*
mo-ana o-a-tsamay-a
NPRE1-child AGRSUBJ1-PRESPRE-walk-VEND
'The child is walking.'

(36) *Ngwana **o tsamaela** mmaagwe kwa toropong.*
mo-ana o-tsamay-el-a (-)-mma
NPRE1-child AGRSUBJ1-walk-APPLSUF-VEND NPRE1A-mother
wa gagwe kwa (ne)-toropo-ing
POSSPART1A POSSPROP1SG LOCPART$_{kwa}$ NPRE9-town-LOCSUF
'The child walks to town for his mother.'

(37)	*tsamael-*	'walk for'	<	agent	beneficiary	locative	>
	IRC:			[-o]	[-r]	[-o]	
	DRC:			[-r]	[+r]	|	
				|	|	|	
				SUBJ	SUBJ/OBJ	OBL_{loc}	
	bi-uniqueness:			SUBJ	OBJ	OBL_{loc}	

Thus, in f-structure notation the SCFs of our examples are:
'TSHAMEK<SUBJ>', but
'TSHAMEK**EL**<SUBJ, **OBJ**>';
'REK<SUBJ, OBJ>', but
'REK**EL**< SUBJ, $\mathbf{OBJ_\theta}$, OBJ>';
'TSAMAY<SUBJ>', but
'TSAMA**EL**<SUBJ, $\mathbf{OBL_{loc}}$>'; and
TSAMA**EL**<SUBJ, **OBJ**, $\mathbf{OBL_{loc}}$>'.

In general, the Setswana applicative extension is thus also a *valency increasing extension* as it allows the inclusion of (i) a beneficiary, (ii) a beneficiary and a theme, (iii) a locative, or (iv) a beneficiary and a locative in the a-structure of an extended verb root. Consequently, a verb root extended with an applicative extension may, in addition to the obligatory SUBJ, subcategorise for OBJ, both OBJ and OBJ_θ, OBL_{loc}, or both OBJ and OBL_{loc}.

4.3 The reciprocal extension

In determining the subcategorization frames of a verb root with a reciprocal extension, we need to take a closer look at the agent role in the a-structures of the three different syntactic realisations mentioned in section 3.2.3.

The plural noun or pronoun (19)

The a-structure of the extended verb root *ratan-* 'love each other' requires only an agent role *bana* 'children', which, according to the IRC and the DRC, are assigned the semantic features [-o] and [-r] and mapped to SUBJ, as before.

The cases of a coordinate phrase and a discontinuous reciprocal construction are somewhat more complex and different approaches have been proposed to determine the SCFs of the inflected verb (Hurst 2006; Hurst 2010; Khumalo 2014).

The coordinate phrase (20)

Mainly two valency reducing approaches and one valency preserving approach have been proposed (Hurst 2006; Hurst 2010). For completeness we summarise them briefly. In the valency reducing approaches only one GF is assigned to the two thematic roles (agent and patient). In the first approach, so-called *suppression* is used and in the second approach *argument unification*.

Suppression (38) prevents the patient role to be mapped to a GF (Hurst 2006, 258), thereby ensuring that the bi-uniqueness condition is satisfied. The Ø notation is used to indicate the suppression of a thematic role. The role of agent (pertaining to both participants, *mosimane le mosetsana* 'the boy and the girl'), is mapped to SUBJ. The advantage of this approach for Setswana is that it has already been applied to Chichewa, also a Bantu language, where the participants stated in a coordinate phrase also fulfil the role of agent and is mapped to SUBJ (Mchombo 1991, 16).

(38) *ratan-* 'love each other' < agent patient >

IRC:	[-o]	[-r]
DRC:	[-r]	|
	|	
Subject condition:	SUBJ	Ø

In the second valency reducing approach, both the agent and patient (or beneficiary) are mapped simultaneously to one GF, i.e. SUBJ. The participants are both agents and patients (or beneficiaries) of the action (Alsina 1996, 260–263; Hurst 2006, 258). Hurst (2010, 315) refers to this process as *argument unification* because the two arguments are allocated to a single slot. This argument slot is then mapped to an f-structure by means of the standard mapping principles (39). However, a disadvantage is that strictly speaking, argument unification does not form part of LMT.

(39) *ratan-* 'love each other' < agent patient >

Argument unification:	⟍ ⟋
IRC:	[-o]
DRC:	[-r]
	|
Subject condition:	SUBJ

In the third approach, the reciprocal morpheme gives rise to a so-called reciprocal pronoun (PRO_{recip}) (Hurst 2006, 258–259; Khumalo 2014, 156). This pronoun is then assigned the patient role that maps to OBJ (40). In the Bantu languages, however, there is a noted incompatibility between reciprocal verbs and OBJs or object agreement morphemes (Khumalo 2014, 158). Therefore, although this approach is technically possible, it is not considered suitable for Setswana.

(40) *ratan-* 'love each other' < agent patient >

	(PRO_{recip})	
IRC:	[-o]	[-r]
DRC:	[-r]	|
	|	
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

In our work, we opted for the suppression approach, primarily because it falls within the LMT framework and has already been used for other Bantu languages. A more detailed investigation into this matter forms part of our future work.

The discontinuous reciprocal construction (21)

The participants are *mosimane* 'boy' and *wena* 'you' - a complement in an obligatory comitative phrase *le wena* 'with you'. The agent (*mosimane* 'boy') maps to SUBJ while, following Hurst (2010, 319–324), the comitative phrase *le wena* 'with you' is considered a thematically underspecified argument because it is not assigned a thematic role. According to the IRC and DRC, the semantic features [-o] and [+r] are assigned to this thematically underspecified argument (indicated in (41) by [ ]) and it maps to OBL_{ass}.

(41)	*dumelan-*	'agree with each other'	<	agent	>	[ ]
	IRC:			[-o]		[-o]
	DRC:			[-r]		[+r]
				|		|
				SUBJ		OBL_{ass}

Thus, in f-structure notation the SCFs of our examples are:
'RAT<SUBJ, OBJ>', but
'RAT**AN**<SUBJ>'; and
'DUMEL<SUBJ>', but
'DUMAL**AN**< SUBJ, $\mathbf{OBL_{ass}}$>'.

In general, the Setswana reciprocal extension is a *valency reducing extension*. It changes the a-structure of the inflected verb by suppressing the patient role. A verb that includes a reciprocal extension therefore does not subcategorise for a direct object and/or secondary object but may subcategorise for an OBL_{ass} when a discontinuous construction is used with the reciprocal.

4.4 The passive extension

In the active sentence in (22) the agent is *mosadi* 'woman' and the patient *bana* 'children'. These roles are mapped to SUBJ and OBJ, as before (42). In the passive sentence the patient (*bana* 'children') in (22) is mapped to SUBJ and the agent role is suppressed (43) (Dalrymple 2001, 208–209; Bresnan et al. 2016, 339–442). In Setswana, the agent is the complement of the agentative particle *ke* 'by'. An ADJUNCT function is assigned to agentative particle phrases such as *ke mosadi* 'by the woman'.

(42) *rom-* 'send' < agent patient >
IRC: [-o] [-r]
DRC: [-r] |
| |
SUBJ SUBJ/OBJ
bi-uniqueness: SUBJ OBJ

(43) *romiw-* 'being sent' < agent patient >
IRC: [-o] [-r]
DRC: [-r] |
| |
Ø SUBJ/OBJ
Subject condition: SUBJ

Thus, in f-structure notation the SCFs of our examples are:
'ROM<SUBJ, OBJ>', but
'ROM**IW**<SUBJ>'.

The Setswana passive extension is a *valency reducing extension*. It changes the a-structure of the inflected verb by suppressing the agent role, while the patient maps to SUBJ as discussed in section 3.2.3.

Setswana verbs may also include more than one extension and in the following section we consider the influence of multiple extensions on a-structure and SCFs.

5 Multiple extensions and subcategorization frames

Extended roots can also include more than one extension as extensions can be sequenced in Setswana verbs. This is true for Bantu languages in general. For example, Bosch & Pretorius (2017) provide an extensive analysis of the phenomenon in Zulu. Two common approaches to the study and analysis of extension sequences are the so-called CARP default template of the Bantu languages and compositionality. In the CARP template, the order is fixed (causative > applicative > reciprocal > passive) although not all slots have to be filled (Hyman 2002). The compositionality approach is governed by semantic scope in which different orders produce different meanings (Manova & Aronoff 2010; Rice 2011, Chavula 2016, 203). In Setswana, adherence to the CARP template has been the dominant approach, although compositionality is also relevant since extension sequences that do not adhere to CARP may occur in rare instances (Pretorius 2014). Pretorius (2014) also reports on a corpus-based investigation of attested verb extension sequences. Examples of common sequences are as follows:

- Causative + Applicative
- Causative + Reciprocal
- Causative + Passive

- Applicative + Reciprocal
- Applicative + Passive
- Applicative + Reciprocal + Passive
- Causative + Applicative + Reciprocal
- Causative + Applicative + Reciprocal + Passive

The following important question now arises: How can we employ the insights that we have gained in the previous section to determine the SCFs of extended verb roots with *multiple* extensions? The answer may be found in observing the compositional nature of the meaning of such a verb root as being extended one extension at a time to which LMT can then be applied repeatedly. For example, for the root *tsen-* 'enter' we could have:

tsen- '*enter* or *attend'* as in (44);
tsen-y- '*put in*, *cause to* enter or *let enter'*, as in (28);
tseny-ets- '*put in for*, *cause to enter for'* as in (45);
tsenyets-an- '*put in for each other*, *cause to enter for each other'* as in (46);
tsenyetsan-w- '*be caused to enter for each other'*, as in (47); and
tsenyets-w- '*be deposited for'* as in (48).

(44) *Rre **o tsena** kopano.*
(-)-rre o-tsen-a kopano
NPRE1A-father AGRSUBJ1A-enter-VEND NPRE9-meeting
'Father attends a meeting.'

(45) *Rre **o tsenyetsa** mme madi mo bankeng.*
(-)-rre o-tsen-is-el-a
NPRE1A-father AGRSUBJ1A-enter-CAUSSUF-APPLSUF-VEND
(-)-mme ma-di mo (ne)-banka-ing
NPRE1A-mother NPRE6-money LOCPARTmo NPRE10-bank-LOCSUF
'Father puts money in the bank for mother.'

(46) *Batho **ba tsenyetsana** madi mo bankeng.*
ba-tho ba-tsen-is-el-an-a
NPRE2-person AGRSUBJ2-enter-CAUSSUF-APPLSUF-RECSUF-VEND
ma-di mo (ne)-banka-ing
NPRE6-money LOCPARTmo NPRE9-bank-LOCSUF
'The people put money in the bank for each other.'

(47) *Madi **a tsenyetsanwa** ke batho mo bankeng.*
ma-di a-tsen-is-el-an-w-a
NPRE6-money AGRSUBJ6-enter-CAUSSUF-APPLSUF-RECSUF-PASSSUF-VEND
ke ba-tho mo (ne)-banka-ing
AGPART NPRE2-person LOCPART_{mo} NPRE9-bank-LOCSUF
'Money is put in (deposited into) the bank for each other by people.'

(48) *Madi **a tsenyetswa** mme mo bankeng ke ntate.*

ma-di	a-tsen-is-el-w-a		
NPRE6-money	AGRSUBJ6-enter-CAUSSUF-APPLSUF-PASSSUF-VEND		
(-)-mme	mo	(ne)-banka-ing	ke
NPRE1A-mother	LOCPART_{mo}	NPRE9-bank-LOCSUF	AGPART
(-)-ntate			
NPRE1A-father			

'Money is deposited into the bank for mother by father.'

The SCFs obtained by applying LMT repeatedly to the basic root and the extended roots, are:

'TSEN<SUBJ, OBJ>'	basic root, transitive
TSEN**Y**<SUBJ, OBJ, $\mathbf{OBL_{loc}}$>'	extended root, increased valency
'TSENY**ETS**<SUBJ, $\mathbf{OBJ_{\theta}}$, OBJ, OBL_{loc}>'	extended root, further increased valency
'TSENYETS**AN**<SUBJ, OBJ, OBL_{loc}>'	extended root, reduced valency
'TSENYETSAN**W**<SUBJ, OBL_{loc}>'	extended root, further reduced valency
'TSENYETS**W**< OBJ_{θ}, SUBJ>'	extended root, reduced valency

Two final comments are relevant:

(i) From the compositionality of the meaning of the extended root as extensions are added, it is clear that the order (CARP or otherwise) in which extensions occur, poses no difficulty for the application of LMT and can be handled as above.

(ii) We focussed on common verbal extensions. Rare extensions such as the "reduplicated" causative (*-isis-*) and applicative (*-elel-*) that semantically play the roles of intensification (Cole 1955, 209) and completion (Cole 1955, 203), respectively, are not considered since they do not change the argument structure of the verb roots to which they are suffixed.

6 Conclusion

The causative, applicative, reciprocal and passive extensions inflect meaning onto the verb root in Setswana that results in a change in the valency of the verb root; causative and applicative extensions increase valency while reciprocal and passive extensions decrease valency in systematic ways. This change in valency has an influence on the argument structure of the verb root, both basic and extended. Moreover, the occurrence of verb roots with multiple extensions is common in Setswana. The study in the context of LFG a-structure and LMT for SCFs (Berg 2018), as reported here, is novel.

As we explained at the outset, the lexicon, and more specifically verb subcategorization frames, forms an integral part of any computational grammar. By considering the LFG a-structure of Setswana verb roots and showing how LMT can be used to obtain SCFs for (extended) verb roots in a principled way, we are a step closer to *semi-automatically* generating SCFs for extended verb roots from known basic verb roots. This is important for building a comprehensive computational verb lexicon for a broad coverage LFG grammar for Setswana.

Three topics for future work immediately come to mind: firstly, the use of the insights and

results of this article towards the semi-automated building of such a comprehensive computational verb lexicon for Setswana. Secondly, an investigation into the suitability of variants of LMT for capturing relevant aspects of object symmetry in Setswana, and thirdly, the broadening of the approach to other Bantu languages.

References

Ackerman, Farrell & John Moore. 2013. "Proto-Properties in a comprehensive theory of argument realization." In *From quirky case to representing space: papers in honour of Annie Zaenen*, edited by Tracy Holloway King & Valeria De Paiva, 9–20. Stanford: CSLI.

Alsina, Alex & Sam A. Mchombo. 1993. "Object asymmetries and the Chichewa applicative construction." In *Theoretical aspects of Bantu grammar, CSLI Lecture Notes,* edited by Sam A. Mchombo, no. 38: 17–45. Stanford: CSLI.

Alsina, Alex. 1996. *The role of argument structure in grammar: evidence from Romance.* Stanford: CSLI.

Attia, Mohammed A. 2008. "Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation." PhD diss., University of Manchester.

Berg, Anna Susanna. 2018. "Computational syntactic analysis of Setswana." PhD diss., North-West University.

Bosch, Sonja E. & Laurette Pretorius. 2017. "A computational approach to Zulu verb morphology within the context of lexical semantics." *Lexikos* 27: 152–182.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan., Ash Asudeh, Ida Toivonen & Stephan Wechsler. 2016. *Lexical-Functional Syntax*. Oxford: Wiley-Blackwell.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu*. California: CSLI.

Butt, Miriam, Tracy Holloway King, María-Eugenia Nino & Frédérique Segond. 1999. *A grammar writer's cookbook*. Stanford: CSLI.

Chavula, Jean Josephine. 2016. "Verbal derivation and valency in Citimbuka." PhD diss., University of Leiden.

Cole, Desmond. T. 1955. *An introduction to Tswana grammar*. Cape Town: Longman.

Dalrymple, Mary. 2001. *Lexical Functional Grammar. Syntax and Semantics, Volume 34.* California: Academic Press.

De Pauw, Guy, Gilles-Maurice De Schryver, Laurette Pretorius & Lori Levin. 2011. "Introduction to the special issue on African Language Technology." *Language Resources and Evaluation* 45, no. 3: 263–269.

Falk, Yehuda N. 2001. *Lexical-Functional grammar: an introduction to parallel constraint-based syntax*. Stanford: CSLI.

Findlay, Jamie Yates. 2016. "Mapping theory without argument structure." *Journal of Language Modelling* 4, no. 2: 293–338.

Her, One-Soon. 2013. "Lexical Mapping Theory revisited." In *From Quirky Case to Representing Space: Papers in Honour of Annie Zaenen*, edited by Tracy Holloway King & Valeria De Paiva, 47–60. Stanford: CSLI.

Hurst, Peter. 2006. "The syntax of the Malagasy reciprocal construction: an LFG account." *Proceedings of LFG06*, edited by Miriam Butt & Tracy Holloway King, 256–274. Stanford: CSLI.

Hurst, Peter. 2010. "The syntax of lexical reciprocal constructions." *Proceedings of LFG10*, edited by Miriam Butt & Tracy Holloway King, 311–331. Stanford: CSLI.

Hyman, Larry. 2002. "Suffix ordering in Bantu: A morphocentric approach." In *Yearbook of morphology*, edited by Geert Booij & Jaap Van Marle, 245–281. Netherlands: Springer.

Jerro, Kyle. 2015. "Revisiting object symmetry in Bantu." *Selected Proceedings of the 44th Annual Conference on African Linguistics*, edited by Ruth Kramer, Elizabeth C Zsiga & One Tlale Boyer., 130–145. Somerville, MA.

Joshi, Aravind K. 1985. "Tree-adjoining grammars: how much context sensitivity is required to provide reasonable structural descriptions?" In *Natural Language Processing*, edited by David Dowty, Lauri Karttunen and Arnold Zwicky, 206–250. Cambridge: Cambridge University Press.

Kaplan, Ronald M. & Joan Bresnan. 1995. "A formal system for grammatical representation." In *Formal issues in lexical-functional grammar*, edited by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell & Annie Zaenen, 28–130. Stanford: Stanford University.

Kaplan, Ronald M. 1995. "The formal architecture of lexical-functional grammar." In *Formal issues in lexical-functional grammar*, edited by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell & Annie Zaenen, 7–27. Stanford: Stanford University.

Khumalo, Langa. 2007. "An analysis of the Ndebele passive construction." PhD diss., University of Oslo.

Khumalo, Langa. 2014. "On the reciprocal in Ndebele." *Nordic Journal of African Studies* 23 no. 3: 140–161.

Kosch, Inge M. 2006. *Topics in morphology in the African language context*. Pretoria: University of South Africa.

Krüger, Casper Jan Hendrik. 2006. *Introduction to the morphology of Setswana*. München: Lincom Europe.

Li, Fengqi. 2019. "The Mapping of Thematic Roles to Grammatical Functions in Mandarin BA Construction Denoting Causation: From the Perspective of Lexical Mapping Theory." *Journal of Language Teaching and Research* 10, no. 3: 618–626.

Louwrens. Louis J. 1994. *Dictionary of Northern Sotho grammatical terms.* Pretoria: Via Afrika.

Manova, Stela & Aronoff, Mark. 2010. "Modelling affix order." *Morphology* no. 20: 109–131.

Marten, Lutz, Nancy C. Kula & Thwala Nhlanhla. 2007. "Parameters of morpho-syntactic variation in Bantu." *Transactions of the Philological Society* 105, no. 3: 253–338.

Marten, Lutz & Nancy C. Kula. 2012. "Object marking and morphosyntactic variation in Bantu." *Southern African Linguistics and Applied Language Studies* 30, no. 2: 237–253.

Mchombo, Sam A. 1991. "Reciprocalization in Chichewa: a lexical account." *Linguistic Analysis* 21, no. 1–2: 3–22.

Mchombo, Sam A. 2004. *The syntax of Chichewa*. Cambridge: Cambridge University.

Mchombo, Sam A. 2007. "Argument binding and morphology in Chichewa." In *Texas linguistics society 9: morphosyntax of underrepresented languages*, edited by Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica. White. CSLI.

Palmer, Martha, Claire Bonial & Jena D. Hwang. 2017. "VerbNet: Capturing English Verb Behaviour, Meaning, and Usage." In *The Oxford Handbook of Cognitive Science,* edited by Susan E.F. Chipman, 315–336. New York: Oxford University Press.

Pereira, Fernando C.N. & Stuart M. Shieber. 2002. *Prolog and natural-language analysis*. Massachusetts: Microtome Publishing.

Pollard, Carl. & Ivan A. Sag. 1994. *Head-driven phrase-structure grammar*. Chicago: University of Chicago Press.

Pretorius, Laurette, Biffie Viljoen, Rigardt Pretorius & Ansu Berg. 2010. "A finite-state approach to Setswana verb morphology." In *Finite-State Methods and Natural Language Processing. Revised Selected Papers, Lecture Notes in Artificial Intelligence*, edited by Yli-Jyrä Anssi, Kornai András, Jacques Sakarovitch & Bruce Watson, 131–138. Berlin: Springer.

Pretorius, Rigardt Samuel. 2014. "The sequence and productivity of Setswana verbal suffixes." *Stellenbosch Papers in Linguistics Plus* 44: 1–23.

Pretorius, Rigardt., Berg, Ansu. & Pretorius, Laurette. 2012. "Multiple objects in Setswana: a computational approach." *Southern African Linguistics and Applied Language Studies* 30, no. 2: 201–218.

Ranta, Aarne. 2011. *Grammatical Framework: Programming with multilingual grammars*. Stanford, California: CSLI Publications.

Rice, K. 2011. "Principles of affix ordering: an overview." *Word Structure* 4, no. 2: 169–200.

Steedman, Mark. 2001. *The syntactic process*. Cambridge, Massachusetts: MIT Press.

Sulger, Sebastian. 2009. "Irish clefting: the LFG perspective." MA diss., University of Konstanz.

Szűcs, P**é**ter. 2018. "A COMP-less approach to Hungarian complement clauses." *Proceedings of LFG18*, edited by Miriam Butt & Tracy Holloway King, 325–342. Stanford: CSLI.

List of tags

ADV	adverb	P1	first person
AGPART	agentative particle	P2	second person
AGROBJ	object agreement morpheme	PASSSUF	passive suffix
AGRSUBJ	subject agreement morpheme	PERSPRO	personal pronoun
APPLSUF	applicative suffix	PL	plural
ASSPART	associative particle	POSSPART	possessive particle
CAUSSUF	causative suffix	POSSPRO	possessive particle
CONJ	conjunction	PRESPRE	present tense morpheme a
DEVSUF	deverbative suffix	RECSUF	reciprocal suffix
$\text{LOCPART}_{\text{kwa}}$	locative particle kwa	SG	singular
$\text{LOCPART}_{\text{mo}}$	locative particle mo	VEND	verbal ending
LOCSUF	locative suffix	1 1A 2 2A ... 20	noun classes 1 to 20
NPRE	noun prefix		