

Computational syntactic analysis of Setswana

Anna Susanna Berg



orcid.org/0000-0001-7596-4558

Thesis submitted for the degree *Doctor of Philosophy in
Setswana* at the North-West University

Promoter: Prof. RS Pretorius

Co-promoter: Prof. L Pretorius

Graduation May 2018



Acknowledgements

I express my appreciation and thanks to the following people who helped and supported me during this study:

- Prof. Rigardt Pretorius, my promoter, for his patient guidance, helpful comments, encouragement and for sharing his outstanding knowledge of Setswana with me.
- Prof. Laurette Pretorius, my co-promoter, for her time, critical comments, suggestions and for the privilege of learning from her experience.
- Prof. Wannie Carstens (School Director), Prof. Justus Roux (Research Director) and Prof. Attie de Lange (Research Director) for their support and encouragement.
- Marcel Hanekom, Nikki Ludwig and Rigardt Pretorius for relieving me of my official workload.
- My parents, sister, brothers and other family members for their support and encouragement.
- Colleagues and friends for their interest.
- FransJohan and Laurette Pretorius for their hospitality and friendship.
- Mrs Margaret Collins for the language editing.

I also thank the Subject Group Setswana (Potchefstroom) and the Research Unit: Languages and Literature in the South African context at the North-West University for financial assistance.

Soli Deo gloria

Summary

The main aim of this study is the computational syntactic analysis of the Setswana simple sentence, using Lexical Functional Grammar (LFG) as framework and XLE as the associated grammar development platform. LFG consists of several parallel levels of representation, but for syntactic analysis the focus is on constituent (c-) and functional (f-) structure as parallel mutually constraining levels of syntactic representation.

We provide a detailed exposition of Setswana grammar in terms of word categories, phrases and the simple sentence, with specific emphasis on nominal classification and concordial agreement, as well as the verb as the morphologically most complex word category. We apply Lexical Mapping Theory (LMT), a sub-theory within LFG, to analyse the argument (a-) structure of the main verb, including the root and its extensions, in order to obtain the subcategorisation frames of the verb roots, as required in the XLE computational grammar lexicon. We also identify and analyse the immediate constituents of the simple sentence in terms of its phrasal structure and their grammatical functions. We use the rich XLE user interface to implement linguistic rules that model this grammar and constitute the XLE parser.

We test the scope, coverage and accuracy of the parser with a systematically hand-crafted test suite that includes both grammatical and ungrammatical test items. We ensure alignment between the linguistic structure of the Setswana simple sentence and phrases and the test suite in order to demonstrate the correctness of our grammar. Finally, we create a treebank, annotated with deep syntactic information, using the XLE interface. The treebank is the first of its kind for Setswana and could serve as a gold standard for testing and evaluating future Setswana parsers. Both our test suite and the treebank, available in `.lfg`, `.SExp` and `.pl` (prolog) format, are freely available.

Key terms

Setswana, grammar, syntax, syntactical analysis, LFG, HLT, parser, parsing, XLE, test suite, treebank

CONTENTS

Acknowledgements	i
Summary	ii
List of Tables.....	vii
List of Figures.....	viii
List of Diagrams	ix
CHAPTER 1 INTRODUCTION.....	1
1.1 CONTEXTUALISATION.....	1
1.1.1 Setswana language.....	1
1.1.2 Technological development of Setswana	5
1.1.3 Lexical Functional Grammar.....	7
1.1.4 XLE	8
1.2 PROBLEM STATEMENT	10
1.2.1 Research questions	10
1.3 AIM	10
1.4 SIGNIFICANCE OF THE STUDY.....	11
1.5 DELINEATIONS AND LIMITATIONS	11
1.6 CHAPTER OUTLINE.....	12
CHAPTER 2 LITERATURE REVIEW.....	14
2.1 INTRODUCTION.....	14
2.2 LEXICAL FUNCTIONAL GRAMMAR	14
2.3 SYNTACTIC STRUCTURE OF SETSWANA	18
2.4 XLE PLATFORM.....	22
2.5 HLT PROFILE OF SETSWANA	23
2.6 SUMMARY.....	25
CHAPTER 3 LEXICAL FUNCTIONAL GRAMMAR	26
3.1 INTRODUCTION.....	26
3.2 CONSTITUENT STRUCTURE.....	27
3.2.1 Phrase structure rules	27
3.2.2 Phrase structure trees	30
3.3 ARGUMENT STRUCTURE.....	31
3.3.1 Grammatical functions.....	32
3.3.2 Lexical Mapping Theory	33
3.4 FUNCTIONAL STRUCTURE	35
3.4.1 Representation of the functional structure	35
3.4.2 Conditions of well-formedness	37

3.4.3	FUNCTIONAL DESCRIPTIONS.....	39
3.5	RELATING CONSTITUENT AND FUNCTIONAL STRUCTURE	41
3.6	SUMMARY.....	44
CHAPTER 4 SETSWANA WORDS		45
4.1	INTRODUCTION.....	45
4.2	HISTORIC DEVELOPMENT OF SETSWANA GRAMMAR	45
4.3	WORD CATEGORIES AND THEIR FEATURES.....	47
4.4	NOUN	47
4.5	PRONOUN.....	52
4.6	VERB	55
4.6.1	Main verb	55
4.6.1.1	Mood.....	57
4.6.1.2	Tense.....	61
4.6.1.3	Aspect.....	65
4.6.1.4	Polarity.....	66
4.6.1.5	Verbal extensions and argument structure	67
4.6.2	Auxiliary verb	92
4.6.3	Copulative verbs	93
4.7	PARTICLES	95
4.8	CONJUNCTION.....	100
4.9	ADVERB	101
4.10	INTERJECTION.....	102
4.11	IDEOPHONE	102
4.12	SUMMARY.....	103
CHAPTER 5 SETSWANA PHRASES.....		104
5.1	INTRODUCTION.....	104
5.2	CLASS NOUN AND LOCATIVE NOUN PHRASE	105
5.3	LOCATIVE CLASS NOUN PHRASE	108
5.4	PRONOUN PHRASE	108
5.5	PARTICLE PHRASES.....	110
5.5.1	Possessive particle phrase.....	110
5.5.2	Qualificative particle phrase	112
5.5.3	Instrumental particle phrase	113
5.5.4	Locative particle phrase	113
5.5.5	Temporal particle phrase.....	115
5.5.6	Agentative particle phrase.....	115

5.5.7	Associative and comparative particle phrases.....	116
5.6	ADVERB PHRASE.....	116
5.7	INTERJECTION AND IDEOPHONE PHRASES.....	116
5.8	VERB PHRASES.....	117
5.8.1	Verb phrase with a main verb.....	117
5.8.2	Verb phrase with an identifying copulative verb.....	118
5.8.3	Verb phrase with a describing copulative verb.....	118
5.8.4	Verb phrase with an associative copulative verb.....	119
5.8.5	Verb phrase with an auxiliary verb.....	119
5.9	COORDINATE PHRASES.....	120
5.10	SEQUENCING OF MODIFIERS.....	122
5.10.1	Juxtaposition.....	124
5.10.2	Nesting.....	124
5.11	SUMMARY.....	125
CHAPTER 6 THE SETSWANA SIMPLE SENTENCE.....		126
6.1	INTRODUCTION.....	126
6.2	IMMEDIATE CONSTITUENTS OF THE SIMPLE SENTENCE.....	126
6.2.1	First constituent.....	127
6.2.2	Second constituent.....	127
6.2.2.1	Main verb phrase.....	127
6.2.2.2	Identifying, describing and associative copulative verb phrases.....	136
6.2.2.3	Auxiliary verb phrase.....	138
6.2.2.4	Inclusion of adjuncts in the verbal phrases.....	139
6.3	WORD ORDER IN THE SIMPLE SENTENCE.....	143
6.4	SUBJECT-VERB AGREEMENT.....	144
6.5	STATUS OF THE SUBJECT AND OBJECT AGREEMENT MORPHEMES.....	148
6.6	SUMMARY.....	152
CHAPTER 7 XLE IMPLEMENTATION OF THE SYNTACTIC STRUCTURE OF THE SETSWANA SIMPLE SENTENCE.....		154
7.1	INTRODUCTION.....	154
7.2	USER INTERFACE.....	154
7.3	GRAMMAR FILE.....	158
7.4	CONFIGURATION AND FEATURES.....	158
7.5	MORPHOLOGY.....	160
7.6	LEXICON.....	162
7.6.1	-unknown lexical entry.....	163

7.6.2	Lexical entries for main and auxiliary verbs	165
7.6.3	Lexical entries for morphological tags	167
7.7	TEMPLATES.....	167
7.8	RULES.....	168
7.8.1	Simple sentence.....	168
7.8.2	Noun phrases.....	169
7.8.2.1	CLNP	169
7.8.2.2	PROP.....	170
7.8.2.3	POSSPARTP	170
7.8.2.4	QUALPARTP	170
7.8.3	Verb phrases.....	171
7.8.3.1	VPMAIN	171
7.8.3.2	VPAUX.....	174
7.8.3.3	VPIDCOP, VPDESCOP and VPASSCOP	174
7.8.3.4	Phrases functioning as obliques and adjuncts.....	175
7.8.4	Sublexical rules	176
7.9	SUMMARY.....	177
CHAPTER 8 TESTING THE GRAMMAR.....		178
8.1	INTRODUCTION.....	178
8.2	TESTING A COMPUTATIONAL GRAMMAR	178
8.3	GRAMMAR SCOPE AND COVERAGE.....	179
8.4	TEST SUITE	180
8.5	RESULTS	183
8.6	TREEBANK.....	184
8.7	SUMMARY.....	185
CHAPTER 9 CONCLUSION		186
9.1	INTRODUCTION.....	186
9.2	ADDRESSING THE RESEARCH PROBLEM	186
9.3	CRITICAL REFLECTIONS.....	187
9.4	RESEARCH CONTRIBUTIONS.....	187
9.5	FUTURE WORK	188
APPENDIX A: Tables from Chapter 1 and Chapter 4		189
APPENDIX B: Tables and treebank formats from Chapter 8		199
APPENDIX C: Morphological tags.....		204
REFERENCE LIST.....		206

List of Tables

Table 1-1:	Setswana noun classes.....	189
Table 4-1:	Pronouns for noun classes.....	190
Table 4-2:	Pronouns for persons.....	191
Table 4-3:	Schematic representation of the morphological structure of verbs.....	58
Table 4-4:	Subject agreement morphemes of noun classes.....	191
Table 4-5:	Subject agreement morphemes of personal pronouns.....	192
Table 4-6:	Consecutive subject agreement morphemes of noun classes.....	192
Table 4-7:	Consecutive subject agreement morphemes of personal pronouns.....	192
Table 4-8:	Object agreement morphemes of noun classes.....	193
Table 4-9:	Object agreement morphemes of personal pronouns.....	193
Table 4-10:	Examples of copulative verbs in sentences.....	194
Table 4-11:	Identifying copulative verbs of personal pronouns.....	196
Table 4-12:	Describing copulative verbs of noun classes.....	196
Table 4-13:	Describing copulative verbs of personal pronouns.....	196
Table 4-14:	Possessive particles.....	197
Table 4-15:	Qualificative particles.....	1980
Table 5-1:	Word categories, subcategories and corresponding Setswana phrases....	105
Table 5-2:	CLNP and LOCNP structure and agreement.....	107
Table 5-3:	PROP structure and agreement with absolute pronoun as head.....	108
Table 5-4:	PROP structure and agreement with demonstrative pronoun as head....	109
Table 5-5:	PROP structure and agreement with inclusive quantitative pronoun as head.....	109
Table 5-6:	PROP structure and agreement with exclusive quantitative pronoun as head.....	109
Table 5-7:	POSSPARTP structure and agreement.....	111
Table 5-8:	QUALPARTP structure and agreement.....	111
Table 5-9:	INSTRPARTP structure and agreement.....	112
Table 6-1:	The syntactic structure of the Setswana simple sentence.....	153
Table 8-1:	Number of test items for various linguistic characteristics in the test suite.....	181
Table 8-2:	Number of lexical entries in the test suite.....	182
Table 8-3:	Number of test items per word length in test suite.....	182
Table 8-4:	Nouns and locative class nouns.....	199
Table 8-5:	Verbs (basic and extended verb roots).....	200
Table 8-6:	Test items and number of valid parses.....	184

List of Figures

Figure 1-1:	The Correspondence Architecture of LFG.....	7
Figure 1-2:	A parsed Setswana simple sentence.....	9
Figure 3-1:	Annotated phrase structure tree and its correspondence function ϕ	42
Figure 7-1:	One solution for a sentence.....	158
Figure 7-2:	Two solutions for a sentence.....	157
Figure 7-3:	Terminal node of sentence o a o reka (she buys it).....	162
Figure 7-4:	Verb subcategorises for a subject, an indirect object and a direct object.....	166
Figure 7-5:	The incorrect use of the present tense morpheme a	172
Figure 7-6:	One object agreement morpheme.....	173
Figure 7-7:	Two object agreement morphemes.....	173
Figure 7-8:	Sentence with three adjuncts.....	174
Figure 7-9:	Inclusion of three adjuncts with a different order.....	174
Figure 7-10:	Expanded display mode of tree showing sublexical information of example (7-1).....	177

List of Diagrams

Diagram 3-1: Mother and daughter nodes.....	27
Diagram 3-2: Correspondences between features of argument functions.....	34
Diagram 4-1: Absolute tenses.....	61
Diagram 4-2: Relative tense.....	63
Diagram 7-1: Macro structure of the grammar file.....	158
Diagram 7-2: Setswana grammar features and their values.....	160

CHAPTER 1

INTRODUCTION

1.1 CONTEXTUALISATION

The main aim of this study is a rule-based computational syntactic analysis of Setswana with a specific focus on the Setswana simple sentence. In recent years, enabling technologies for Natural Language Processing (NLP) in Setswana were developed but, as one of the core technologies, a parser is still needed for the computational processing of Setswana. In order to develop a parser for Setswana we employ Lexical Functional Grammar (LFG) to frame the description of Setswana grammar in a modern linguistic theory. For the purposes of this study a *grammar* is “a representation of the rules for combining words together to form larger syntactic units, and for combining these units to make sentences” (Farghaly, 2003:10) (cf. §8.2). Subsequently, we implement the Setswana grammar in an existing parser development software, i.e. the XLE parser software.

1.1.1 SETSWANA LANGUAGE

Setswana¹ (ISO 639-3 tsn), a language spoken in southern Africa, is one of the official languages of the Republic of South Africa (RSA) where approximately 8% (4 067 248) of the population are first language Setswana speakers (Statistics South Africa, 2011). It is also the national language in the neighbouring country Botswana, where it is estimated that 79.06% (1 070 000) of the population are first language Setswana speakers (Botswana Central Statistics Office, 2009:14, 339). Furthermore, an estimated 30 000 people in Namibia are first language Setswana speakers (Census Namibia, 2011:67).

Setswana belongs to the Bantu language family and is classified in the South-Eastern Zone of Bantu languages. The South-Eastern Bantu languages are grouped together in language groups based on their similar grammatical structure and vocabulary (Poulos & Louwrens, 1994:2; Krüger, 2006:3). The South Eastern Zone comprises the Sotho language group, the Nguni language group, XiTsonga, and Tshivenda. The Sotho language group consists of Setswana (Tswana), Sesotho sa Leboa (Northern Sotho) and Sesotho (Southern Sotho), whereas the Nguni language group consists of siSwati (Swazi), isiXhosa (Xhosa), isiZulu (Zulu) and isiNdebele (Ndebele) (Cole, 1961:88; Van Wyk, 1967:21–25, 37–38; Lombard *et al.*, 1985:5). In Guthrie’s (1971)

¹. Setswana is also commonly known as Tswana. In earlier publications Setswana is also referred to as Western Sotho. Sepedi is also used instead of Sesotho sa Leboa.

classification, the Sotho languages are placed in group S.30 and the Nguni languages are included in group S.40.

Bantu languages are structurally closely related in terms of typology, as they share certain general characteristics such as a noun class system, a system of grammatical (concordial) agreement, and an agglutinative morphology (Louwrens, 1994a:18). However, the Bantu languages differ with respect to orthography. Whereas the Nguni languages have a conjunctive orthography in which affixes are conjoined with the root, the Sotho languages employ a disjunctive orthography in which the prefixes of the verb are generally written disjunctively. This requires a distinction between a so-called orthographic and linguistic word. An *orthographic word* is a unit that is separated by spaces from other units in the sentence, while a *linguistic word* denotes units that function as members of a word category and has its own particular meaning (Kosch, 2006:3). For example, the sentence in (1-1) contains four orthographic words, but three linguistic words. The accurate modelling of these characteristics are imperative in the development of human language technologies (HLTs) for Setswana.

The Bantu languages are characterised by a grammatical gender, so-called class gender, where nouns are grouped together in classes in a grammatically significant way (Kosch, 2006:89-90). The nouns are grouped in classes (Appendix A: Table 1-1, p.189) by means of their class prefixes which are correspondingly referred to as gender number prefixes (Kosch, 2006:90). Moreover, Setswana noun classes have semantic significance (Cole, 1955:68–105; Krüger, 2006:57–98). Each Setswana noun belongs to one of 20 noun classes and are numbered systematically. Classes 1 to 14 consist of singular-plural pairs, noun classes 1 and 2, 3 and 4, 5 and 6, 7 and 8, and 9 and 10 are pairs where the odd numbers indicate the singular and the even numbers the plural. Nouns in class 11 are singular and their plural forms conform to class 10. The nouns in class 14 are singular but their plural forms conform to class 6. Classes 1 and 2 each have a sub class, i.e. classes 1a and 2a. Nouns in class 1a are singular and their plural counterparts appear in class 2a. Classes 15 to 20 do not denote singular or plural. Class 15 contains infinitive nouns. Classes 16 to 20 contain the locative classes (Krüger, 2006:92–98). For the purposes of this study, we distinguish classes 19 and 20; these classes are often either referred to as classes X and Y or the **ga-** and **N-**locative classes (Poulos & Louwrens, 1994:47).

Grammatical (concordial) agreement in the Bantu languages is based on the noun class system (Lombard *et al.*, 1985:54; Rose *et al.*, 2002:4) and is also governed by person and number features. In a Setswana sentence, agreement between a noun and the main verb is expressed by affixes such as the subject agreement morpheme and object agreement morpheme which are

prefixed to the verb root. For example, in (1-1) **ba** is a class 2 subject agreement morpheme and it agrees with the class 2 noun **batho** (people) which contains a class prefix **ba-**.

(1-1) **Batho ba** boile maabane.²
 people they returned yesterday
ba-tho *ba-bo-il-e* *maabane*
 NPre2-person AgrSubj2-return-PerfSuf-VEnd Adv³
 The people returned yesterday⁴.

Noun modifier agreement is established using class-specific words (Louwrens, 1994a:10). For example, in (1-2) **ba** indicates a class 2 demonstrative pronoun which agrees with the class of the noun **batho** (people).

(1-2) **batho ba**
 people these
ba-tho *ba*
 NPre2-person DemPro2D1
 these people

It is well known that the two central phenomena in morphology are *word formation* (also referred to as morpheme sequencing or morphotactics) and *phonological and orthographical alternation* (also referred to as morphophonological alternation) – the sound and spelling changes that occur due to the environment in which a morpheme occurs. In Setswana, as an agglutinative language, both these phenomena play an important role. Affixes are sequenced as structural elements in a word to execute a process of adapting or extending the meaning of a word (Kosch, 2006:133–139). This phenomenon of affixation is particularly prevalent in the formation of nouns (cf. §4.2.1) and verbs (cf. §4.2.3). The meaning of a noun can be extended by a diminutive, feminine, augmentative and locative suffix (Krüger, 2006:73–96). Inflection in verb morphology is expressed by prefixes that indicate class gender, person and number, mood, tense, aspect, and polarity⁵ (Cole, 1955:242–267; Krüger, 2006:198–243), whereas derivation is expressed by causative, applicative, reciprocal, perfect and passive suffixes and an obligatory verbal ending (Cole,

² Note on glossing: The glossing of examples applied in this study is based on the Leipzig Glossing Rules (University of Leipzig, 2015). <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>

³ An explanation of all the morphological tags which are used in the examples, is presented in Appendix C

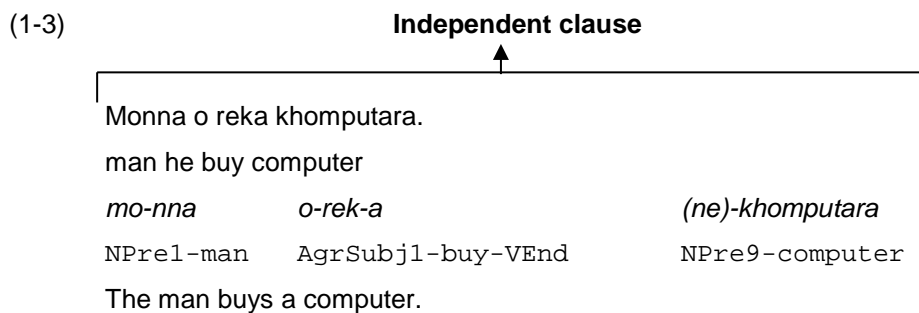
⁴ The determiners *the* and *a* do not have translated equivalents in Setswana. For example, **batho** is translated as *the people* and **motho** as *the person* or *a person*.

⁵ The sequence of verbal prefixes and suffixes is presented in Krüger (2006:257) and Pretorius (2014:1–23).

1955:192–211; Krüger, 2006:257). A detailed exposition of the morphophonological alternation that occurs in Setswana is provided in Krüger (2006). We return to this topic in Chapter 7.

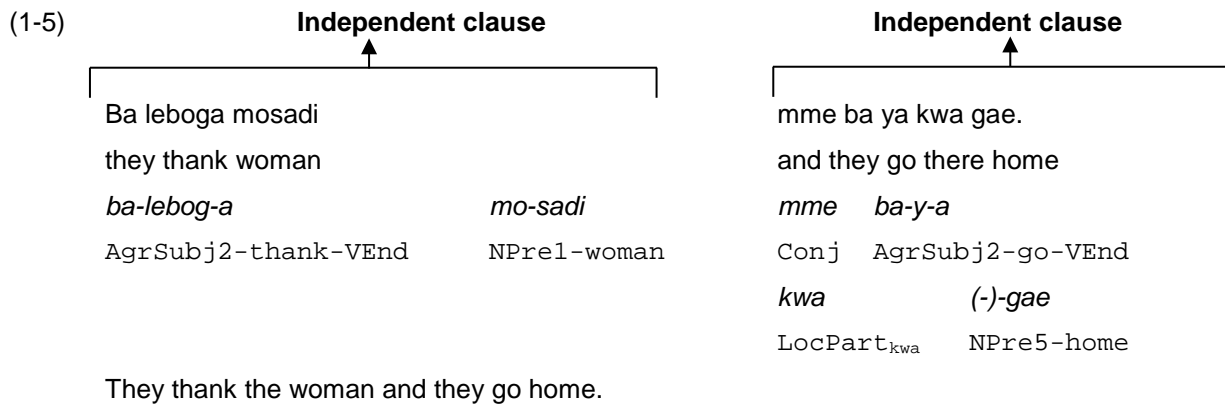
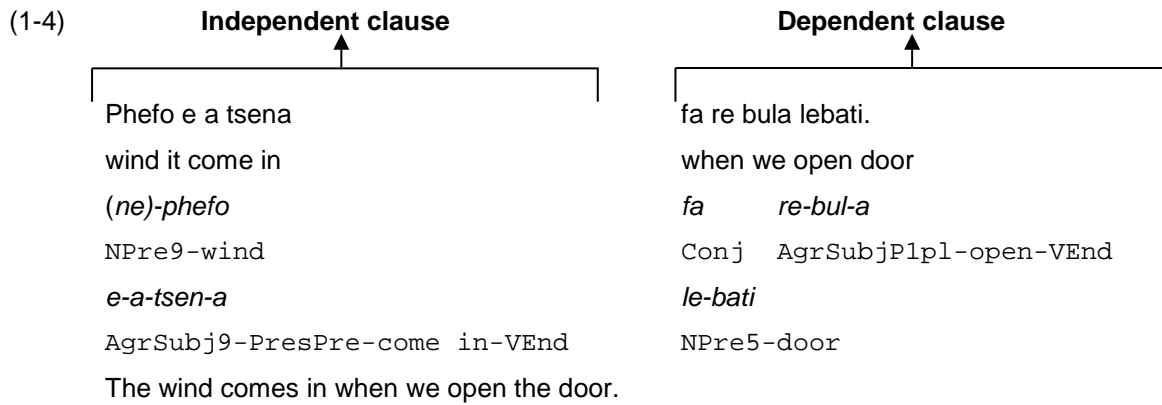
Setswana sentences can be categorised as simple, complex and compound sentences where the division is based on the composition or grammatical structure of sentences (Louwrens, 1994a:178). Setswana sentences also have a specific clausal structure. Independent clauses and dependent clauses are distinguished. The independent clause is a main clause that functions on its own, as it does not depend on another clause (Louwrens, 1994a:84). The dependent clause is a subordinate clause, as it is dependent on an independent clause for its existence (Louwrens, 1994a:28).

The Setswana simple sentence, as an independent clause, consists of a single verbal element (Louwrens, 1991:17). Apart from the verbal element, the sentence also includes various other constituents (Louwrens, 1991:13) such as a subject, objects, obliques, and adjuncts. The structure of the simple sentence is discussed in Chapter 6. The simple sentence in (1-3) consists of only one independent clause where a main verb **o reka** (he buys) is included in the structure.



The complex sentence consists of an independent clause and at least one dependent clause (Watters, 2000:217). As explained by Louwrens (1991:30), the complex sentence consists of two or more verbal elements. One of the verbs is included in the independent clause and one in the dependent clause as illustrated in (1-4). The verb in the dependent clause in (1-4) denotes the participial mood.

A compound sentence consists of two or more independent clauses connected by a conjunction. A compound sentence is formed through coordination in which independent clauses are combined into a single sentence. Both clauses in this sentence are of equal ranking as they have an equal syntactic status (Louwrens, 1994a:29; Watters, 2000:217). A sentence consisting of equally ranked clauses is also called a co-ordinate sentence. In (1-5), the two independent clauses are connected by the conjunction **mme** (and).



1.1.2 TECHNOLOGICAL DEVELOPMENT OF SETSWANA

Broadly speaking, the technological development of a language essentially requires both basic language resources (language data of various kinds) and core technologies for processing these data. Krauwer (2003) mentions basic language resources such as written language corpora, spoken language corpora, mono- and bilingual dictionaries, terminology collections and grammars, core technologies such as taggers, morphological analysers, and **parsers**. He also proposes the notion of a Basic Language Resource Kit (BLARK) ("a minimal set of language resources required to do precompetitive language and speech technology research") as a framework for assessing the technological development of a language. The BLARK has become a *de facto* standard for assessing the technological status of a language, specifically for languages that are considered less-studied or under-resourced (see, for example, Strik *et al.*, 2002⁶; Daelemans *et al.*, 2003⁷; Maegaard *et al.*, 2006⁸; Prys, 2006⁹; Streiter *et al.*, 2006¹⁰; Borin *et al.*, 2008¹¹; Borin *et al.*, 2010¹²; and Anon., 2011¹³). The Setswana BLARK currently includes a

⁶ <http://hstrik.ruhosting.nl/wordpress/wp-content/uploads/2013/04/a92.pdf>.

⁷ <http://www.cnts.ua.ac.be/papers/2005/dbd05.pdf>.

⁸ https://www.researchgate.net/profile/Bente_Maegaard/publication/228379950_The_BLARK_concept_and_BLARK_for_Arabic/links/02e7e517b7f20f11f3000000.pdf

⁹ <http://mt-archive.info/LREC-2006-SALTMIL-WS.pdf#page=37>

¹⁰ <http://mt-archive.info/LREC-2006-SALTMIL-WS.pdf#page=37>

¹¹ <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.148.5870&rep=rep1&type=pdf#page=21>

¹² http://lrec-conf.org/proceedings/lrec2010/pdf/156_Paper.pdf

¹³ <http://www.meta-net.eu/projects/meta-nord/summary>

lemmatiser, tokeniser, morphological analyser, part of speech tagger, sentence separator and a phrase chunker (Brits *et al.*, 2005; Brits, 2006; Pretorius, L., *et al.*, 2008, 2010, 2015; Pretorius, R., *et al.*, 2005, 2009, 2012; Eiselen & Puttkammer, 2014; Eiselen, 2016). A parser for Setswana has not been developed yet.

Indeed, the development of a parser, a technology for computational syntactic analysis, is an important contribution to and a necessary tool in the development of various Human Language Technology (HLT) applications working with natural language data, such as grammar checkers, machine translation systems, manuscript recognition systems, automatic summarising systems, and question answering systems (Babarczy *et al.*, 2007:1). For example:

a prerequisite for building a good machine translation system is a thorough knowledge of how natural language works, and the availability of formalisms and computational tools for the effective modelling of natural language processes and phenomena and the implementation thereof (Butt & King, 2003:132).

Two main approaches for the development of language technologies, including parsers, may be distinguished, viz. symbolic (also referred to as rule-based) and stochastic (also often referred to as data-driven or statistical) approaches (Jurafsky & Martin, 2009:10). A detailed discussion of these approaches falls outside the scope of this study. We adopt a rule-based approach to developing a parser for Setswana and do not cover the growing field of data-driven approaches to parsing.

Accordingly, a rule-based parser is a computer program that processes linguistic units, such as sentences and phrases, to produce syntactic representations of these units based on the grammatical rules that it incorporates (Farghaly, 2003:10; Butt & King, 2003:130; Kübler, 2004:1). These grammatical rules are usually formulated in terms of a specific theoretical approach using an associated formalism, resulting in a formal grammar (Kübler, 2004:2; Forst, 2011:2). Examples of such approaches are Lexical Functional Grammar (LFG) (Kaplan & Bresnan, 1982; Dalrymple, 2001), Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag, 1994), Categorical Grammar (CG) (Karlsson *et al.*, 1995) and Tree-adjoining Grammar (TAG) (Joshi & Schabes, 1997). The formal grammar is then implemented using an appropriate computational platform (Butt & King, 2003:130). Examples of such platforms are the rule-based English CG parser (EngCG) (Samuelsson & Voutilainen, 1997) and XLE (Crouch *et al.*, 2015), which is a platform to develop parsers for various languages making use of LFG. In this study, we provide a detailed exposition of the Setswana simple sentence in terms of LFG and employ the XLE software (Crouch *et al.*, 2015) for the development of a parser.

1.1.3 LEXICAL FUNCTIONAL GRAMMAR

LFG is an approach to linguistic analysis and theory building that has been used since the late 1970s (Dalrymple *et al.*, 1995:1). Its mathematical basis and simplicity has rendered LFG particularly suitable for computational modelling and implementation towards the analysis and understanding of human language (Austin, 2001:26). Kroeger (2004:1) expresses the usefulness of LFG as follows:

LFG has a number of features that make it an attractive and useful framework for grammatical description, and for translation. These include the modular design of the system, the literal representation of word order and constituency in c-structure, a typologically realistic approach to universals (avoiding dogmatic assertions which make the descriptive task more difficult), and a tradition of taking grammatical details seriously.

LFG has been applied to a wide range of languages and is continuously actively developed by an international scientific community. The grammar architecture of LFG consists of several parallel levels of representation (Figure 1-1), 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; Asudeh, 2006:363–387; Asudeh & Toivonen, 2015:400). These levels are mutually constraining through functional projections (correspondence functions). For example, all nodes in the phrase structure tree relate to corresponding elements in the f-structure (cf. §3.5), and this relation is defined by the so-called many-to-one ϕ (phi) function (Falk, 2001:64). The ϕ function is defined as the composition of the μ , α and λ functions. The μ function specifies the mapping from the c-structure tree to the m-structure, the α function specifies the mapping from the m-structure to the a-structure while the λ function specifies the mapping from the a-structure to the f-structure (Asudeh & Toivonen, 2015:400–401).

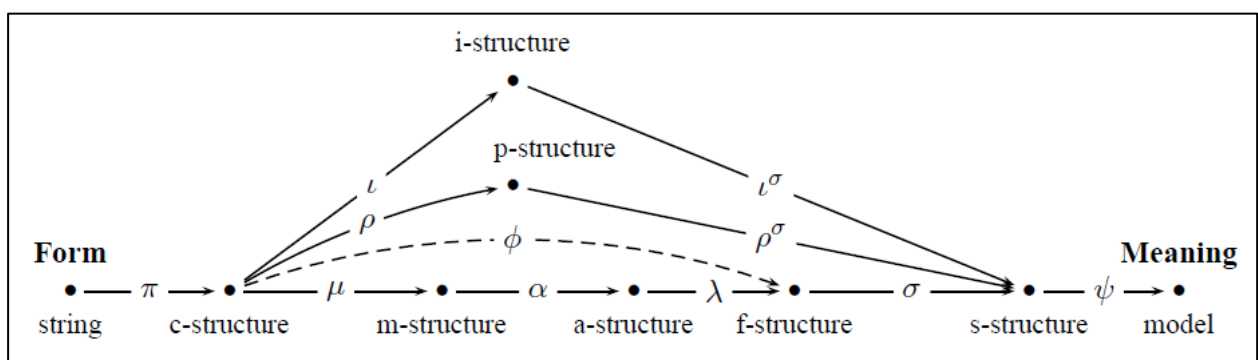


Figure 1-1: Correspondence Architecture of LFG (cf. Asudeh & Toivonen, 2015:400)

However, for the purposes of grammar development in LFG, our focus is mainly on the modelling of c- and f-structure and sentences are therefore analysed in terms of these two structures (Kaplan & Bresnan, 1995:175). The c-structure expresses the order and grouping of constituents. The f-structure expresses the functional roles of these constituents and we employ an a-structure description in this study to determine the subcategorisation frames of main verbs¹⁴. The c-structure has the form of a context-free phrase structure tree and is defined by language-specific constraints on the word order and phrase structure (Kaplan & Bresnan, 1995:175; Dalrymple, 2006:82). The f-structure represents the functional or syntactic information of the internal structure of the sentence (Dalrymple, 2001:7; Forst, 2011:2). The f-structure contains surface grammatical functions, such as subject and object, as well as features which represent the morphosyntactic properties of constituents. These morphosyntactic properties represent linguistic categories such as class, person, number, tense, aspect, and mood. The representation of the f-structure is formalised through an attribute-value matrix (AVM). The AVM is 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). This is discussed in detail in Chapter 3.

1.1.4 XLE

XLE¹⁵ is a grammar development platform, used to parse and generate text using computational grammars couched within the framework of LFG (Butt *et al.*, 1999:172; Crouch *et al.*, 2015). The LFG grammar of a language is presented to the XLE system in:

a priority-ordered sequence of files containing phrase-structure rules, lexical entries, abbreviatory macros and templates, feature declarations, and finite state transducers for tokenization and morphological analysis (Kaplan *et al.*, 2002:29).

XLE requires a tokeniser and a morphological analyser, developed with the Xerox Finite-State tools (XFST) (Beesley & Karttunen, 2003; Kaplan *et al.*, 2004). As such computational tools have already been developed for Setswana (Pretorius, L., *et al.*, 2008, 2010, 2015; Pretorius, R., *et al.*, 2005, 2009, 2012), this study can thus focus on the computational syntactic analysis of Setswana by making use of the XLE parser software as well as the previously mentioned tokeniser and morphological analyser.

The XLE parser parses text into an LFG representation of c- and f-structure and is designed to take advantage of context-freeness in the grammar of a natural language automatically so that it normally parses in cubic time and generates in linear time (Crouch *et al.*, 2015). An example of a

¹⁴ Chapter 3 presents an overview of LFG summarising the basic ideas of constituent structure (c-structure), argument structure (a-structure) and functional structure (f-structure) and the correspondence between these structures.

¹⁵ The XLE documentation is available at: http://www2.parc.com/isl/groups/nlft/xle/doc/xle_toc.html

parsed Setswana simple sentence (1-6) using XLE is shown in Figure 1-2. This figure shows the c- and f-structure of the simple sentence in (1-6). A single valid c-structure is presented as there is only one parse for this sentence, and the f-structure shows one solution.

The representation of the c- and f-structure of the parsed sentence (1-6) is displayed in four windows (cf. §7.2).

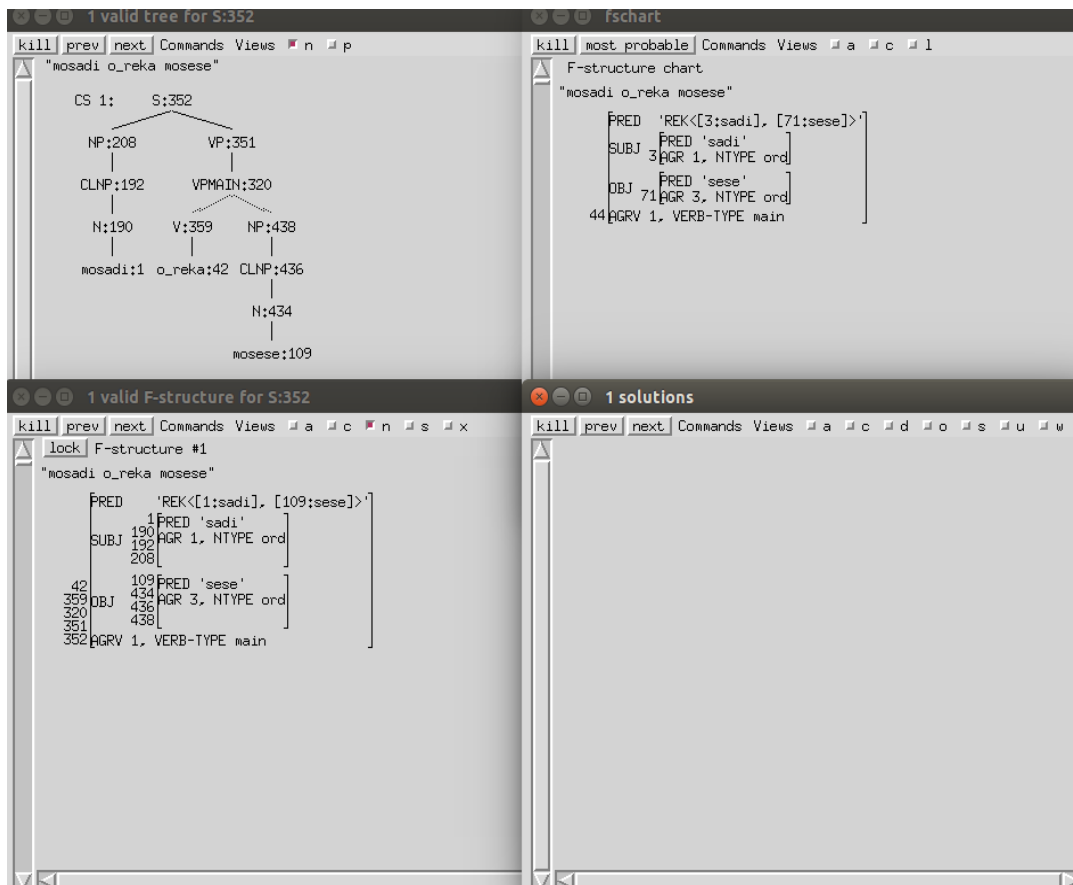


Figure 1-2: A parsed Setswana simple sentence

These four windows represent specific information (Crouch *et al.*, 2015). The upper left window (c-structure window) shows the phrase structure tree. The tree is displayed with the root (sentence) at the top and the leaves (lexical items) at the bottom. The lower left window shows the functional structure (f-structure window), which is displayed as an attribute value matrix structure (AVM) in the standard LFG format. The upper right window (fschart window) shows the f-structure chart that indexes the packed solutions by their constraints. Each constraint appears once in an f-structure, which is annotated by all of the choices where that constraint holds. The lower right window shows the f-structure chart choices (fschartchoices window), which indexes the packed solutions by the alternative choices. If there is only one solution applicable to a sentence, the chart does not show any information.

(1-6) Mosadi o reka mosese.
 woman she buys dress
mo-sadi *o-rek-a* *mo-sese*
 NPre1-woman AgrSubj1-buy-VEnd NPre3-dress
 The woman buys a dress.

1.2 PROBLEM STATEMENT

This study presents a rule-based computational syntactic analysis of Setswana. The main research problem constitutes the accurate formulation of rules representing the structure of the Setswana simple sentence in the LFG framework, the implementation of these rules to develop a computational grammar, the testing of this grammar and the development of a treebank, annotated with deep syntactic information.

1.2.1 RESEARCH QUESTIONS

The following questions, emanating from the problem statement, are addressed in this study:

- How can Setswana syntactic structure be couched in the LFG framework with specific reference to the word categories, the phrasal structure and the simple sentence?
- How is the XLE platform used to implement the LFG representation of the Setswana syntax?
- How can this implementation be tested?
- How can a treebank be developed?

These research questions are addressed in Chapters 3 to 8.

1.3 AIM

The main aim of this study is to present a computational syntactic analysis of the Setswana simple sentence, which could serve as a basis for an extended broad-coverage parser for Setswana in the future. The specific aims of this study are to develop a:

- first LFG grammar for Setswana by describing the syntactic structure of the Setswana simple sentence according to this approach;
- first parser for Setswana by implementing the LFG grammar on the XLE grammar development platform;
- test suite to test the accuracy of the implementation in XLE;
- novel treebank for Setswana, annotated with deep syntactic information.

1.4 SIGNIFICANCE OF THE STUDY

This study constitutes a novel contribution to the broader study of Setswana syntax and its computational modelling and implementation. We employ LFG, a lexicalist, non-transformational, constraint-based theory of generative grammar, and the XLE parser software for this purpose. The development of a Setswana LFG grammar and the development of a novel parser using the XLE platform contribute not only to improving Setswana's HLT profile, but arguably, also provide an accurate formal grammar of the Setswana simple sentence. The development of a novel treebank for Setswana can also serve as a gold standard for future grammar testing and evaluation. This study could also form the basis for extending the grammar to include complex and compound Setswana sentences. Furthermore, the contribution of a computational syntactic analysis of the Setswana simple sentence will benefit the development of various HLT applications for this language.

Owing to the structural similarity of the Bantu languages, and specifically the Sotho group of languages, the contribution of this study may also be used to bootstrap similar (novel) grammars for these languages.

1.5 DELINEATIONS AND LIMITATIONS

This study focusses on a fragment of the Setswana grammar i.e. the *simple sentence*. We emphasise that the Setswana simple sentence is not "simple" since it includes the full complexity of the verb as the most complex word category in Setswana. The expectation is that this study will lead to the future implementation of a broad-coverage LFG grammar for Setswana that would include the Setswana complex and compound sentences. The description of the structure of the simple sentence in this study can be used as the foundation to describe the structure of compound and complex sentences.

While the Setswana *lexicon* in XLE, needed to implement and test the parser, is restricted, it is nevertheless carefully crafted to include all the salient features of a comprehensive Setswana lexicon. It also ensures that no ambiguities in the finite-state Setswana tokeniser and morphological analyser will influence the accuracy of the Setswana grammar as only valid tokens and morphological analyses are presented. Although the focus is only on the simple sentence, the core notions of lexical mapping and subcategorisation are relevant here as the complexity of the verb, its argument structure, and its subcategorisation have important implications for the lexicon. It should also be noted that the parser is tested with a Setswana test suite that is not extracted from a corpus, as a corpus of simple Setswana sentences is not available.

Finally, we point out that the focus of this study in developing an LFG grammar for Setswana is mainly on the *accurate syntactic modelling of the Setswana simple sentence*. An important future initiative would be to investigate the extent to which our current grammar aligns with the current standards and frameworks of ParGram (cf. §2.3), a project in which the emphasis is on *parallel computational grammar* development as support for applications such as machine translation.

1.6 CHAPTER OUTLINE

The structure of the thesis is as follows:

Chapter 1 is an introduction to this study. Geographical information and the distinctive typological characteristics of the Setswana language are presented. The technological development of Setswana is briefly covered and an appropriate theoretical approach and tool for the computational syntactic analysis of Setswana are introduced. The problem statement, aims, significance, delineation, and limitations of the study are presented.

Chapter 2 contains a survey of the literature on topics such as the HLT profile of Setswana; the syntactic structure of the Setswana simple sentence; the LFG framework; the Lexical Mapping Theory (LMT), the sub-theory within LFG, which is concerned with argument-function mapping; the development of an LFG grammar for Setswana; and the use of the XLE platform to develop a parser for Setswana and to execute a computational syntactic analysis. The purpose of this chapter is to contextualise the contribution of this study.

Chapters 3 to 8 contain the main contribution of this study and systematically answer the four research questions.

In *Chapter 3*, an overview of the LFG framework is presented. The c-, a- and f-structure as well as the correspondence between these structures are explained. LMT, a theory of correspondence between semantic roles and grammatical functions, is also summarised.

In *Chapter 4*, a detailed exposition of the features of the word categories (lexical categories) of Setswana is presented. The suffixing of the productive verbal extensions (causative, applicative, reciprocal and passive) to the main verb in Setswana is described and LMT is applied to explore the implications for the argument structure of the resulting Setswana verbs.

In *Chapter 5*, Setswana phrases are proposed and each one of these phrases is described in terms of its head, obligatory complements, possible modifiers, and the agreement phenomena that govern the phrase. Coordination and the sequencing of modifiers are also addressed.

In *Chapter 6*, the syntactic structure of the Setswana simple sentence is examined with respect to its immediate constituents, the order and grammatical functions of these constituents, subcategorisation frames and subject-verb agreement.

In *Chapter 7*, the architecture of XLE is described and the XLE implementation of the LFG model of Setswana syntax, explored in Chapters 4, 5 and 6, is discussed.

Chapter 8 concerns the testing of the computational grammar as covered in Chapter 7. For this purpose, a hand-crafted (manually constructed) test suite is introduced, motivated and then used to test all the salient features of the implemented grammar. The accuracy of the implemented rules concerning the grammar is shown. A novel treebank is developed for Setswana by storing the preferred valid analyses of the XLE output in a user defined folder.

Chapter 9 concludes this study with a short overview of the content of each chapter, a critical assessment of the contribution of the study and an indication of future work concerning Setswana syntax and the computational implementation thereof.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

As the computational syntactic analysis of the Setswana simple sentence is the main aim of this study, an LFG grammar is developed and this grammar is implemented in the XLE parser software. In this chapter, we contextualise this study by reviewing related topics. We specifically focus on the Lexical Functional Grammar (LFG) approach from an historical perspective as well as its use for the linguistic description of a number of Bantu languages including Setswana, the Lexical Mapping Theory (LMT), the grammar of Setswana, the XLE parser software, and the human language technology applications for Setswana.

2.2 LEXICAL FUNCTIONAL GRAMMAR

Kaplan and Bresnan pioneered LFG during the 1970s. Numerous articles and edited volumes focussing on LFG are available. In 1982, a seminal collection of papers concerning the theory of LFG was published (Bresnan, 1982). These papers deal with various linguistic phenomena in various languages. A paper by Kaplan and Bresnan (1982:173–281) containing a first detailed description of LFG is included in this collection of papers. According to Kaplan and Bresnan (1995:7), this paper covers a description of the "basic architectural concepts that underlie the formal theory of Lexical-Functional Grammar" in terms of c- and f- structure as two levels of syntactic description. Moreover, they present a comprehensive account of functional descriptions, the functional well-formedness conditions, the correspondence between c-structure nodes and the f-structures and long distance dependencies. They conclude their paper by presenting an overview of the generative capacity of LFG and explain that LFG can indeed be used to present a suitable linguistic description of a language.

The classical 1995 publication of Dalrymple *et al.* covers a range of topics on LFG theory from 1982 to 1994. It includes an historical overview of the development of the LFG theory (Dalrymple *et al.*, 1995:1–5), an introduction to the formal architecture of LFG (Kaplan, 1995:7–27) as well as a reproduction of the 1982 paper of Kaplan and Bresnan (1995:29–130). This book also includes a collection of papers focussing on nonlocal dependencies, word order, semantics and translation, and mathematical and computational issues.

A comprehensive account of LFG is presented in the books of Dalrymple (2001) and Bresnan (2001)¹⁶. These publications have also become classical references, in which c- and f-structure are discussed in detail and extensively illustrated with examples. Dalrymple (2001) covers linguistic phenomena such as modification, control, anaphora, coordination and long distance dependencies. She illustrates these phenomena with in-depth syntactic analyses. Bresnan (2001) discusses a wide range of syntactic phenomena from typologically diverse languages and shows that these phenomena can be modelled in LFG. This work, and specifically her treatment of certain grammatical aspects of the Bantu language, Chichewa, contributed to our choice of LFG as a suitable theoretical approach for the Setswana syntactic analysis of this study.

The development of the Lexical Mapping Theory (LMT) as a sub-theory within the LFG framework is a significant development for the understanding of the principles and constraints that govern the mapping of arguments to their respective grammatical functions. Dalrymple (2001) and Bresnan (2011) both present a historical overview of the development of LMT and discuss the theory of a-structures focussing on semantic roles, the feature decomposition of argument functions and the mapping of a-structures to grammatical functions. They present a detailed discussion of the intrinsic and default argument classifications as well as the subject and function-argument bi-uniqueness conditions. Dalrymple (2001) illustrates LMT by considering the active and passive versions of the verb, locative inversion and complex predicates. Bresnan (2011) focusses on the analyses of unaccusatives, resultatives, ditransitives and passives. She also discusses and illustrates the morphology of verbs that add or suppress a-structure roles.

Two standard textbooks on LFG theory are Falk (2001) and Kroeger (2004). Falk (2001) presents an introduction to LFG in which the basic concepts of c- and f-structure are described. He applies these structures to several English constructions and compares the LFG theory with the theory of Government and Binding (Haegeman, 1994) and the Minimalist Program (Chomsky, 1995). He also describes the development of LMT and shows how a-structure mediates the mapping between semantic roles and grammatical functions. He uses LMT to analyse passives, unergatives and unaccusatives.

Kroeger (2004) describes topics such as tests for constituency, passivisation and other relation-changing processes, reflexive pronouns, the control relation, topic and focus, relative clauses and Wh-questions, causative constructions, serial verbs, case phenomena and ergativity from an LFG point of view and addresses various unique features of individual European and non-European languages.

¹⁶ A second edition of this book is published in 2016 (Bresnan *et al.*, 2016). This book includes a synthesis of major theoretical developments in lexical-functional syntax over the past few decades.

The International Lexical Functional Grammar Association (ILFGA) is the official organisation for the LFG¹⁷ scholarly community. The Essex LFG website¹⁸, the Google+ ILFGA website¹⁹, and the LFG Facebook website²⁰ can be consulted for technical and theoretical information on LFG. This information includes a comprehensive bibliography of published and unpublished works written in the LFG framework²¹. The proceedings of the LFG conferences are published from 1996 online by CSLI Publications²². These publications confirm that LFG has been successfully applied to the analysis of various languages and a wide range of syntactic constructions. Moreover, some Bantu languages such as Chichewa, Swahili, Zimbabwean Ndebele, Kikongo, Sesotho sa Leboa and Setswana, are also described using the LFG framework.

For Chichewa, the applicative constructions (Alsina & Mchombo, 1988; Alsina & Mchombo, 1990b; Lam, 2007), object asymmetries (Alsina & Mchombo, 1990a, 1993), locative inversion (Bresnan, 1987; Bresnan & Kanerva, 1989; Schachter, 1992), topic, pronoun and agreement (Bresnan, 1997; Bresnan & Mchombo, 1985, 1986, 1987) and the lexical integrity principle (Bresnan & Mchombo, 1995), to name but a few, have been described in LFG. The so-called pro-drop phenomenon in Chichewa and the status of the subject and object agreement morphemes as pronominal or incorporated pronouns have also been studied (Bresnan & Mchombo, 1995:276–284; Bresnan, 2001:148–160; Mchombo, 2001:229–230; Mchombo, 2004:19–22). All the phenomena for Chichewa are insightful for Setswana as many of the typological features are comparable with the description of Setswana. Furthermore, Mchombo (2007) uses LMT to describe a-structure and verbal suffixation. He focusses on argument binding and the reciprocal in Chichewa. He demonstrates that the reciprocal is a detransitivising morpheme and that it reduces by one the arrangement of arguments associated with the non-reciprocalised predicate (cf. §4.6.1.5).

For Swahili, Olejarnik (2009) applies the LFG approach to the analysis of complex predicates, more specifically light verb (V) + noun (N) constructions. The study of Olejarnik (2009) does not present a comprehensive account of the syntactic structure of Swahili using LFG, as her focus is on the description of only one phenomenon. Lipps (2011) also focusses on one phenomenon in Swahili, i.e. the relative clause. He describes the structure of relative constructions using the LFG approach and provides an outline of three relativisation categories. He then provides an LFG

¹⁷ The official website for ILFGA is <https://sites.google.com/site/ilfgalg/home/>

¹⁸ <http://www.essex.ac.uk/linguistics/external/LFG/index.html>

¹⁹ Google+ ILFGA page: <https://plus.google.com/109464318749972104499>

²⁰ The LFG Facebook page: <https://www.facebook.com/lfgpage>

²¹ The LFG bibliography is available at <http://www.essex.ac.uk/linguistics/external/LFG/Bibliography/bibliography.html>

²² <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/>

analysis of the relative constructions supporting his analysis by a computational grammar developed in XLE.

Faaß (2010) presents a novel contribution in developing Sesotho sa Leboa by providing a first morphosyntactic description and implementation of a fragment of this language focussing on the verbal phrase. She presents the morphemes in the morphological structure of the verb as elements of a syntactic constituent structure rather than components of the morphological structure of the verb. She therefore does not adhere to the Lexical Integrity Principle that is one of the key notions of LFG. This principle specifies that no syntactic rule can refer to elements of morphological structure (Dalrymple, 2001:84).

Khumalo (2007:132–161) describes LFG and LMT and is of the opinion that morphological phenomena in Bantu languages lend themselves better to a surface-oriented lexical analysis like LFG. He applies LFG and LMT to present an analysis of the Zimbabwean Ndebele passive construction (Khumalo, 2007:183–213). Furthermore, Khumalo (2014) presents an analysis of the reciprocal in Zimbabwean Ndebele. He uses LFG and LMT to show that the reciprocal in Ndebele is an argument changing verbal extension and it can subcategorise for a direct object. He furthermore shows that the reciprocal in Ndebele can co-occur with the passive.

Fernando (2008) presents a first analysis of Kikongo verbal affixes couched under LMT. He describes the possible affix ordering in Kikongo and then describes the form and function of six verbal affixes (applicative, causative, reciprocal, reflexive, passive and stative). This description leads to a division of the affixes into valency increasing affixes and valency decreasing affixes. He also describes double objects and the sequencing of verbal affixes in Kikongo and its influence on a-structure.

Berg *et al.* (2012, 2013) constitute a first description of certain aspects of Setswana syntax using LFG. They present an LFG description of agreement between a subject and the proper verb and describe noun phrase internal agreement where a noun is modified by a demonstrative pronoun, a possessive phrase and an adjectival phrase. Furthermore, the syntactic structure of Setswana sentences with double objects is described. They show that these objects can be replaced by object agreement morphemes and discuss the pronominal value of these morphemes.

2.3 SYNTACTIC STRUCTURE OF SETSWANA

The syntactic structure of the Setswana simple sentence is described in Chapter 6. In this section, the literature on Setswana grammar as well as the grammatical description of the related Sotho languages is reviewed.

In his seminal publication describing Setswana grammar, Cole (1955) classifies Setswana words into thirteen word (lexical) categories²³ (noun, pronoun, adjective, enumerative, quantitative, possessive, relative, verb, copulative, adverb, ideophone, conjunctive and interjection) (Cole, 1955:59). He presents a systematic description of the morphological structure of these words and describes their function in various syntactic structures. Moreover, he includes introductory notes on the syntax of Setswana and briefly describes the syntactic structure of the substantive, qualificative, predicative and descriptive (Cole, 1955:452–460). He incorporates extensive examples in his description of the morphological and syntactic structures.

A textbook, published by the Department of African Languages and Literature, University of Botswana (2000), provides a notable contribution in its description of Setswana syntax in terms of linguistic rules. The description is based on Generative Grammar and asserts that the basic Setswana sentence is primarily made up of a subject and a predicate. It is emphasised that all Setswana sentences are analysable as $S \rightarrow NP VP$ and a brief overview of the composition of the NP is given (Department of African Languages and Literature, University of Botswana, 2000:3–5). The composition of the VP includes a verb group²⁴ and complements. The verb group is described as "a set of inflectional elements ... followed by the Verb stem" (Department of African Languages and Literature, University of Botswana, 2000:11). A number of phrase structure rules that "provide a framework for the analysis of Setswana simple sentences" is provided, and that these rules generate "simple, declarative, affirmative and active sentences" is added (Department of African Languages and Literature, University of Botswana, 2000:14–16). These rules do not represent an exhaustive discussion of Setswana sentence structure and only selected examples are supplied.

Three significant volumes on Setswana linguistics by Krüger (2006) and Krüger (2013a, 2013b) have been published. Krüger (2006) gives a detailed exposition of the morphological structure of Setswana word categories while the two later volumes (2013a and 2013b) are devoted to the syntactic structure of Setswana word groups from a structural approach. This approach was

²³ The notion of "word (lexical) category" is referred to as "word class" by Krüger (2006, 2013a, 2013b).

²⁴ In this instance the use of the so-called "verb group" refer to the morphological structure of a verb. Note that Krüger (2013a, 2013b), amongst others, use the term "verbal group" to refer to a syntactic structure in which a verb acts as head followed by a complement.

pioneered, and further developed and applied by Van Wyk (1958, 1962, 1964, 1966, 1967) for Sesotho sa Leboa. Krüger (2006, 2013a, 2013b) followed the Van Wykian structural approach in his description and analysis of Setswana. This approach has its origin in the principles of the structural syntax established by Dutch scholars such as De Groot, Uhlenbeck, Paardekooper and Reichling. According to Kosch (1991:49), it "was essentially a classificatory or taxonomic exercise [with] the main aim to list elements and classes of linguistic units". This structural approach "was the acknowledgement that language had a structure which manifested itself in regularities, patterns or rules which had to be discovered methodologically" (Kosch, 1991:49).

In describing the Setswana morphology, Krüger (2006) follows the same principles that Van Wyk (1967) used to classify Sesotho sa Leboa word categories. Krüger (2006) distinguishes between nouns, pronouns, verbs, adverbs, particles, conjunctions, ideophones and interjections; presents a concise exposition of the classification of morphemes; and describes the morphological structure of Setswana words, providing extensive examples. Krüger (2006:293–310) also includes introductory notes on so-called word group formation, briefly discusses the "relevant lexical and functional layers", and presents a summary of guidelines that can be followed in word group formation.

Krüger (2013a) describes the syntactic structure of nominal, pronominal and various particle groups. Krüger (2013b) is devoted to the syntactic structure of verbal groups, word groups including copulatives (identifying, describing, associative), auxiliary verbal groups and conjunction groups. He describes word groups in terms of their "internal structure" and "external function". The internal structure may be thought of as a kind of constituent structure, which he describes in terms of the word categories and how they may be combined. He portrays the external function of a word group as "how it may combine as a member of other more comprehensive structures [and] what its new function(s) is/are in the newly formed internal structure" (Krüger, 2013a:viii).

Krüger (2013b:320) defines a sentence "as a word group with a verbal element as head member". He distinguishes between three types of verbal groups:

- The *active verbal group* minimally contains a predicate that consists of a verb, and the predicate can combine with a subject, an object or a primary descriptive to form a "minimum verbal group" (Krüger, 2013b:52).
- The *copulative verbal group* minimally consists of a copulative verb and a complement. He includes subjects and primary descriptives in his presentation of the structure of copulative groups (Krüger, 2013b:133–186).

- The *auxiliary verbal group* minimally consists of an auxiliary verb and a complement such as an active verbal group. Similarly, a subject, objects and primary descriptives can form part of this group (Krüger, 2013b:192–254).

In his exposition of the external function of a verbal group, he states that this group "can provide the lexical content of an independent sentence" (Krüger, 2013b:104). Krüger (2013b:328–347) presents a concise exposition of five sentence types in Setswana, viz. statements, interrogatives, commands, interjections (exclamations) and vocatives (addresses) without explicitly distinguishing between simple, complex and compound sentences, as discussed in §1.1.1.

While Cole (1955) and Krüger (2006, 2013a, 2013b) are standard references for Setswana grammatical description, Lombard *et al.* (1985), Louwrens (1991), Poulos and Louwrens (1994), Louwrens *et al.* (1995) and Kosch (2006) are also relevant references, since they address the grammatical description of Sesotho sa Leboa, a related language.

Lombard *et al.* (1985) present a comprehensive account of major phenomena in Sesotho sa Leboa such as the word categories. They describe the morphological structure of nouns and pronouns and present the function of these words in word groups and sentences; describe the morphological structure of verbs; make a distinction between transitive and intransitive verbs, auxiliary verbs, and copulative verbs; include information on morphosyntax; and state that there are sub-categories within the verb as a word category (Lombard *et al.*, 1985:139). They distinguish between mood, tense and aspect; explain these subcategories in detail; and describe the morphological structure, meaning and function of particles, conjunctions, ideophones and interjections. While they present introductory notes on the structure of auxiliary verb groups and copulative verb groups, they do not describe the syntactic structure of sentences.

Louwrens (1991) presents an overview of Van Wyk's word identification and classification. He discusses the most characteristic features of simple and complex sentences in Sesotho sa Leboa and shows that the simple sentence consists of at least a subject and a verb (main verb, copulative verb or auxiliary verbal group). The syntactic structure of the complex sentence in Sesotho sa Leboa is described regarding the "modal relationships" that exist between the verb in the main clause and the verb in the subordinate clause (Louwrens, 1991:30–48). Pronominalisation, locative structures, and the use of interrogatives in sentences are also discussed.

Poulos and Louwrens (1994) form part of a three volume series on the linguistic analysis of three South African Bantu languages, viz. Tshivenda, Sesotho sa Leboa and isiZulu, respectively (see

also Poulos (1990) and Poulos and Msimang (1998)). They describe the morphological structure of nouns, pronouns, main verbs and copulas in Sesotho sa Leboa; the indicative, participial, subjunctive and habitual moods of the verb; its different tenses; the auxiliary verb and its complements; and the use of the adverb, ideophone, interjection, conjunction and interrogative in Sesotho sa Leboa.

Louwrens *et al.* (1995) describe the morphological structure of nouns and verbs in Sesotho sa Leboa and present an overview of adverbs, conjunctions and interrogatives. They furthermore include a section on syntax in which they describe the word order in Sesotho sa Leboa. They discuss the structure of the verb in terms of eight moods, absolute and relative tenses and aspect and furthermore describe the transitivity of verbs.

An important contribution by Kosch (2006) addresses topics in morphology such as the word and the morpheme, the nature and environment of the morpheme, suppletion, Sandhi, inflection, derivation, typology and exponence. She follows an "eclectic approach" and illustrates theoretical principles using examples from Sesotho sa Leboa and isiZulu.

The syntax of the related language, Sesotho, is described by Du Plessis and Visser (1992a) and Machobane (2010). This work by Du Plessis and Visser (1992a) forms part of a four volume textbook series on the syntax of four languages, viz. Sesotho, isiXhosa, Tshivenda and XiTsonga, respectively (see also Du Plessis & Visser (1992b), Du Plessis *et al.* (1992) and Du Plessis *et al.* (1995)). Du Plessis and Visser (1995) follow a transformational generative grammar (TGG) approach and cover the morphological and syntactic structures of certain Sesotho phenomena, the properties of the argument structure of verbs, "adjunct clauses", and constructions that include "deficient verbs and copulative verbs". They also present categories that may be used as modifiers in the internal structure of noun phrases.

Machobane (2010) applies Chomsky's theory of Government and Binding (Haegeman, 1994) in her textbook on Sesotho syntax. She lists the noun, verb, preposition and adverb as the word categories of Sesotho and submits a brief overview of the structure of the Sesotho noun phrase, verb phrase, adverbial phrase and the prepositional phrase. While she makes a distinction between simple, complex and compound sentences, she does not present a thorough description of the syntactic structure of these sentence types. She presents a syntactic analysis of one simple sentence and one compound sentence and does not include any description of the structure of simple and compound sentences. She also gives a brief overview of the structure of complex sentences stating that a complex sentence consists of a main clause and a subordinate clause. She pays specific attention to subordinate clauses, distinguishing between "noun clauses,

adverbial clauses, locative clauses, temporal clauses and clauses of reason, clauses of condition, clauses of purpose and clauses of concession".

Notable articles published in various scientific journals on aspects of Setswana grammar focus on topics such as the morphological structure of Setswana (Krüger, 1994), absolute tenses (Pretorius, 2003), verb morphology and the lexical integrity principle (Creissels, 2006), adverbials (Le Roux, 2011) and the noun phrase (Letsholo & Matlhaku, 2014).

A number of MA and PhD studies on various grammatical topics in Setswana provided important insights. These topics include:

- the structure of word groups and simple sentences (Krüger, 1961, 1967);
- auxiliary verbs and deficient verbs (Setshedi, 1974);
- conjunctions (Vermeulen, 1984);
- ideophones (Ras, 1991);
- interrogatives (Khoali, 1994);
- the grammatical description of word categories (Moyane, 1995);
- auxiliary verbs (Pretorius, 1997); and
- adverbials (Le Roux, 2007).

2.4 XLE PLATFORM

The XLE platform and the implementation of the Setswana grammar in XLE are discussed in Chapter 7. Crouch *et al.* (2015) are the main source for XLE²⁵. The development of the XLE platform is a joint project between the NLTT group at PARC and the MLTT group in Grenoble that commenced in October 1993. XLE was specifically designed to facilitate a computational realisation of grammars couched within the LFG framework and is considered as one of the best available grammar development systems, taking into account particular criteria such as depth of analysis and linguistic motivation. It consists of a parser, a generator and a graphical user interface for writing and debugging such grammars (Butt *et al.*, 1999:172; Crouch *et al.*, 2015). This platform has a rich phrase structure rule notation and various kinds of abbreviatory devices such as parameterised templates, macros, and complex categories. A graphical user interface (emacs, tcl/tk) is included in XLE, it is written in C and runs on Linux, Solaris and Mac OS X machines (Crouch *et al.*, 2015). A free educational license may be obtained from the NLTT group at PARC²⁶.

²⁵ XLE documentation available at http://www2.parc.com/isl/groups/nlitt/xle/doc/xle_toc.html

²⁶ The license is available at <http://www2.parc.com/isl/groups/nlitt/xle/XLE-Non-Commercial-License.pdf>. The latest release date of XLE is March 17, 2017.

The XLE documentation (Crouch *et al.*, 2015) includes comprehensive information on the installation of XLE, the loading of a new grammar, the use of the XLE interface, grammatical notations, transfer, and translation. Moreover, the XLE documentation includes documentation on the implementation of a parallel grammar in the Parallel Grammar (ParGram) project. This project is an international collaboration aimed at producing broad-coverage computational grammars for a variety of languages (Butt, *et al.*, 1999, 2002). These grammars are written in the LFG framework and are constructed using XLE. The ParGram project comprises grammars for Arabic, Chinese, English, French, German, Georgian, Hungarian, Indonesian, Irish, Japanese, Malagasy, Murrinh-Patha, Norwegian, Polish, Spanish, Tigrinya, Turkish, Urdu, Welsh and Wolof (Sulger *et al.*, 2013:551).

A Grammar Writer's Cookbook (Butt *et al.*, 1999) provides an excellent and accessible exposition of the various core aspects of computational grammar development. Their decision to couch their exposition in the LFG framework made this book an invaluable resource for this study. Their focus on developing parallel grammars for English, French and German in the ParGram project, using LFG and XLE, further demonstrates the applicability of LFG/XLE for Setswana. In particular, Butt *et al.* (1995:15–52) consider the structure of the clause, verbal elements, nominal elements, determiners and adjectives, prepositional phrases, adverbial elements, coordination as well as constructions that include tag questions, parentheticals and headers and provide the relevant analyses of these structures for English, French and German. Included is a section on language engineering, an overview of the architecture and interface of XLE, the use of finite state tools (Butt *et al.*, 1999:175–183), testing procedures based on treebanks, and annotated test files (Butt *et al.*, 1999:204–209), which are important aspects for the research reported on in this thesis.

Faaß (2010) describes the verbal phrase in Sesotho sa Leboa from a morphosyntactic perspective and implements the structure of this phrase in XLE. Moreover, Faaß and Prinsloo (2011) describe the computational implementation of the infinitive in Sesotho sa Leboa using XLE to model this structure.

2.5 HLT PROFILE OF SETSWANA

Chapter 1 (cf. §1.1.2) provides a broad perspective on HLT for Setswana. In this section, the focus is on the body of literature that directly relates to grammar development for Setswana, viz. a rule-based lemmatiser, tokeniser and morphological analyser, as well as the use of the Grammatical Framework (GF) for the development of a Setswana GF resource grammar. The overview of Eiselen and Puttkammer (2014) concerning the development of language resources for ten South African languages, including Setswana, report on the development of part of speech

taggers, lemmatisers, morphological analysers, tokenisers and sentence separators, developed by means of statistical methods. These resources are available at the Resource Management Agency (RMA)²⁷.

The orthography and agglutinative morphology of Setswana, together with the specific computational challenges that these typological features present for Setswana HLT, have already been addressed through the development of a lemmatiser (Brits *et al.*, 2005; Brits, 2006), a tokeniser and a finite state morphological analyser (see, for example, Pretorius, L., *et al.*, 2015). These so-called enabling technologies are required for, amongst others, rule-based syntactic parsing – as presented in this study.

Brits *et al.* (2005) and Brits (2006) define and identify the Setswana lemma to execute automatic lemma identification. They follow a rule-based linguistic approach using the FSA 6 software for developing finite state automata and transducers for linguistic purposes. Although the lemmatiser was not used in this study, the careful exposition of the relevant linguistic rules by Brits *et al.* (2006) assisted in the development of the morphological analyser used in this study.

Finite-state approaches to computational morphology still represent the state of the art for under-resourced languages (Lindén *et al.*, 2009). The two central problems of morphology, viz. word formation and morphophonological alternation, are modelled using finite state networks and then compiled into a single network that constitutes the morphological analyser (Beesley & Karttunen, 2003:xvi). Pretorius, L., *et al.* (2008, 2010, 2015) and Pretorius, R., *et al.* (2005, 2009, 2012) report on the development of a finite state tokeniser and morphological analyser for Setswana, using the Xerox finite state tools (Beesley & Karttunen, 2003). Both the tokeniser and the analyser were also successfully compiled with the Foma finite state toolkit (Hulden, 2009).

Grammatical Framework (GF)²⁸ is a special-purpose functional programming language for grammar development, specifically well suited to multilinguality (Ranta, 2011). Central to GF grammars are the notions of abstract and concrete syntax. Abstract syntax captures the semantically relevant structure of a language while the concrete syntax facilitates its linear string representations. More than one concrete grammar can be developed for a given abstract grammar, allowing semantically robust translation between the concrete grammars for different languages (Ranta, 2011). Pretorius, L., *et al.* (2017) describe the development of a GF miniature resource grammar for Setswana in which the proper verb in simple declarative and imperative

²⁷ These resources are available at <http://rma.nwu.ac.za>. The RMA will be incorporated in the South African Centre for Digital Language Resources (SADiLaR) (<https://rma.nwu.ac.za/index.php/about-sadilar/>).

²⁸ Grammatical Framework (GF) is described in detail by Ranta (2011) and more information is available at <http://www.grammaticalframework.org/>.

sentences is modelled. The contribution of this development is twofold: firstly, a GF miniature resource grammar, covering a fragment of Setswana grammar, that can form the basis for Setswana CNL applications, and secondly, a parameter system, suitable for the envisaged full Setswana RGL²⁹ (Pretorius, L., *et al.*, 2017).

2.6 SUMMARY

While publications on LFG and the XLE platform abound, our choice of literature in this case was guided by the significance and direct relevance for our study. Various seminal, classical and standard references were discussed. We also included available publications on LFG/XLE for other Bantu languages. These references will be employed in detail in Chapters 3 and 7. Owing to the limited body of available Setswana linguistics literature, we were able to give an extensive account of the work that has been done. In particular, it was shown that just a small number of studies focussing on the Setswana syntax are available. We also considered similar work on Sesotho sa Leboa and Sesotho as related languages. We return to these references in Chapters 4, 5 and 6 where we present an in-depth account of the word categories, phrases and c- and f-structure of the Setswana simple sentence. Our review of the literature on Setswana HLT focussed on developments that directly influenced this study. We return to tokenisation and computational morphological analysis in Chapter 7.

²⁹ The RGL is an available resource for GF and is the standard library of GF which functions as a software library and includes the abstract and concrete syntaxes of various languages (Ranta, 2011:21)

CHAPTER 3

LEXICAL FUNCTIONAL GRAMMAR

3.1 INTRODUCTION

In this chapter, we present an overview of Lexical Functional Grammar (LFG). The focus is on the basic notions of constituent structure (c-structure), functional structure (f-structure) and the relation between a c-structure and its corresponding f-structure as the basis for syntactic analysis in LFG. We briefly discuss the notion of argument structure (a-structure) in the context of the lexicon and subcategorisation frames.

LFG is a grammar framework in theoretical linguistics, a variety of generative grammar. In generative linguistics, the focus is "on modelling the syntactic component, the component of language that deals with the combination of words into phrases, clauses and sentences" (Falk, 2001:1). Bresnan and Kaplan developed LFG during the 1970s in reaction to the transformational grammar (Chomsky, 1957). This development emerged because a need was identified for a theory allowing simultaneous expression of both the phrasal constituency of a sentence and the more abstract functional syntactic organisation thereof. The foundations of the LFG theory are motivated by linguistic, computational, and psycholinguistic considerations (Kaplan & Bresnan, 1982:173–177, Dalrymple, 2001:1–2).

LFG is a non-derivational, constraint based theory of grammar, which distinguishes between two levels of syntactic analysis of a natural language utterance, i.e. a c- and f-structure. These two structures exist in parallel (Kaplan & Bresnan, 1982:175; Dalrymple, 2001:7; Falk, 2001:1–4, 9). Dalrymple (2006:82) states that:

LFG assumes that two syntactic levels are important in the analysis of linguistic structure. F(unctional)-structure represents abstract grammatical functions such as subject and object as well as abstract features such as tense and case. Another level, c(onstituent)-structure, represents the concrete phrasal expression of these relations, governed by language-particular constraints on word order and phrase structure. This duality of syntactic representation is motivated by the different natures of these two structures both within and across languages.

C-structure expresses the order and grouping of constituents, whereas f-structure expresses the functional information of these constituents. As a result, c-structures are language specific and may vary across languages, whereas f-structures in different languages are expected to be largely invariant across languages (Bresnan, 2001:44–45; Austin, 2001:12). Butt *et al.* (1999:7, 12)

maintain that more language universal properties are expressed at the f-structure, and at the c-structure, more language particular properties are encoded. They also explain that the level of c-structure helps to encode certain properties such as linear order, position, and constituent order, and may therefore be different in various languages, while the level of f-structure encodes analyses in terms of the predicate-argument structure and the relationships between heads and their modifiers, which are taken to hold at the language universal level. Dalrymple (2006:82) summarises the theory of LFG as follows:

As the name implies the theory is lexical: the lexicon is richly structured, with lexical relations rather than transformations or operations on phrase structure trees as a means of capturing linguistic generalizations. It is also functional: grammatical functions like subject and object are primitives of the theory, not defined in terms of phrase structure configuration or semantic roles.

The remainder of this chapter consists of a brief exposition of c-structure, f-structure and a-structure, as well as the relationships between a- and f-structure and between the mutually constraining c- and f-structure.

3.2 CONSTITUENT STRUCTURE

A context-free grammar determines the c-structure. The context-free grammar is a set of phrase structure rules that characterise all possible surface structures for a language. These rules encode constituency (dominance) and surface order (precedence) and are language specific (Falk, 2001:33; Bresnan, 2001:113).

3.2.1 PHRASE STRUCTURE RULES

The standard notation for a phrase structure rule is $M \rightarrow p$ where M is the left-hand side of the rule (also called the mother node) and p is the right-hand side consisting of a set of possible daughter nodes for M and it is a regular expression (Dalrymple, 2001:92). An example of a phrase structure rule is presented in (3-1) and explained in Diagram 3-1.

(3-1) $S \rightarrow NP VP$

Mother node	Symbol	Daughter nodes
S	→	NP VP

Diagram 3-1: Mother and daughter nodes

This phrase structure rule is applicable to the structure of a Setswana sentence such as (3-2), which consists of a noun phrase (NP) **mosadi** (woman) and a verb phrase (VP) **o reka mosese** (she buys a dress).

(3-2) Mosadi o reka mosese.

woman she buys dress

mo-sadi

o-rek-a

mo-sese

NPre1-woman

AgrSubj1-buy-VEnd

NPre3-dress

The woman buys a dress.

The right-hand side of a phrase structure rule is formed by means of regular expression operations such as disjunction, optionality and repetition, as well as additional notions (syntactic sugar) such as immediate dominance (ID), linear precedence (LP) and metacategories, as explained below.

LFG allows disjunction over daughter nodes. Disjunction specifies the various possibilities for the right-hand side of a phrase structure rule (Dalrymple, 2001:93, 97). The rule in (3-3) indicates that in this instance the CLNP (cf. §5.2) in Setswana can be either a noun (N) or a noun followed by an absolute pronoun (AbsPro).

(3-3) $CLNP \rightarrow \{N \mid N \text{ AbsPro}\}$

Notation: Marked with curly brackets and the possibilities are separated by a vertical bar |.

Optionality indicates that a specific daughter node in a phrase structure rule may or may not be present (Dalrymple, 2001:94). The rule in (3-4) indicates that the NP is optional. A Setswana sentence can thus include an NP followed by a VP or the sentence can only consist of a VP. The VP is therefore obligatory (cf. §6.2, §6.5).

(3-4) $S \rightarrow (NP) VP$

Notation: A node is represented by enclosing it with round brackets.

A daughter node in a phrase structure rule may be repeated (Dalrymple, 2001: 94). Repetition is encoded by using a Kleene star (*) or a Kleene plus (+) operator. The rules in (3-5) and (3-6) include repetition regarding the use of an adverb phrase (ADV_P) in a Setswana VP.

(3-5) $VP \rightarrow V (NP) ADVP^*$

(3-6) $VP \rightarrow V (NP) ADVP^+$

Notation: The Kleene star (*) operator in (3-5) indicates zero or more occurrences of an *ADVP*, whereas the Kleene Plus (+) operator in (3-6) implies one or more (but not zero) occurrences of this phrase.

Immediate dominance (ID) expresses all the possible word orders of daughter nodes in a rule (Dalrymple, 2001:96). The rule for the Setswana *PROP* (cf. §5.4) in (3-7) is an abbreviation of (syntactic sugar for) two rules. It states that the order of the absolute pronoun and the demonstrative pronoun is not specified³⁰. Either the absolute pronoun can be followed by a demonstrative pronoun (3-8) or it can be preceded by the demonstrative pronoun (3-9).

(3-7) $PROP \rightarrow AbsPro, DemPro$

(3-8) $PROP \rightarrow AbsPro DemPro$

(3-9) $PROP \rightarrow DemPro AbsPro$

Notation: Commas separate the relevant daughter nodes.

Linear precedence (LP) indicates that one daughter node must precede another daughter node in a particular rule (Dalrymple, 2001:96). The rule for an instrumental particle phrase (*INSTRPARTP*) in Setswana in (3-10) shows that an instrumental particle (*InstrPart*) precedes a noun.

(3-10) $INSTRPARTP \rightarrow InstrPart < N$

Notation: Indicated with a < symbol between nodes involved in the precedence³¹.

A metacategory represents several different sets of daughter nodes (Dalrymple, 2001: 94). The metacategory *XP* in (3-11) shows that a phrase of any category can appear as the second

³⁰ Cole (1955:128) explains that the absolute pronoun may stand in “apposition, and usually following, but sometimes preceding a noun or another pronoun”.

³¹ The LP rule is “equivalent to the standard ordered phrase structure rule” (Dalrymple, 2001:96).
In this case $INSTRPARTP \rightarrow InstrPart N$

daughter node. The XP in this instance represents an NP, an adverb phrase (ADVP), an INSTRPARTP or a temporal particle phrase (TEMPPARTP) and disjunction specifies the various possibilities (3-12) that can be included in a Setswana VP.

(3-11) $VP \rightarrow V \text{ XP}$

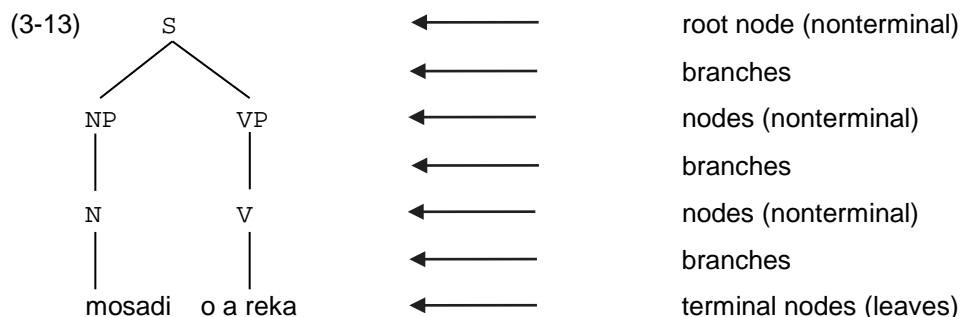
(3-12) $XP \equiv \{NP \mid ADVP \mid INSTRPARTP \mid TEMPPARTP\}$

Notation: The \equiv symbol is used to connect two expressions that are defined to be equivalent.

Any phrase structure rule can also be represented as a phrase structure tree. Phrase structure rules may be interpreted as node admissibility conditions; these rules are descriptions of admissible trees in this sense.

3.2.2 PHRASE STRUCTURE TREES

It follows from §3.2.1 that at the c-structure level, word order and phrasal groupings are represented by means of phrase structure (c-structure) trees. A tree consists of nodes and branches, where the nodes are either nonterminals or terminals (3-13).



Example (3-13) shows that words are atomic units in the syntactic structure of the sentence. Terminal nodes typically represent the lexicon of the grammar.

As a lexical theory, LFG adheres to the Lexical Integrity Principle. This principle implies that no syntactic rule can refer to elements of morphological structure. The internal structure of words is referred to by morphological and not syntactic rules. Morphologically complete words are leaves of the phrase structure tree and each leaf corresponds to one and only one c-structure node. Fully inflected words will therefore occupy the terminal nodes in such a tree – morphemes cannot correspond to terminal nodes (Bresnan, 2001:92; Dalrymple, 2001:84; Falk, 2001:26; Asudeh &

3.3.1 GRAMMATICAL FUNCTIONS

In LFG, grammatical functions (GFs) are considered primitives of syntactic representation. The LFG inventory of GFs consists of the following: subject (SUBJ), object (OBJ), secondary object (OBJ_θ)³², oblique (OBL_θ), complement (COMP), open complement (XCOMP), adjunct (ADJUNCT) and open adjunct (XADJ). The subscript θ of a particular argument refers to the semantic role associated with the argument, for example, in OBJ_{theme} θ refers to the theme semantic role, whereas θ refers to the locative semantic role in the argument OBL_{loc}. 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): ADJUNCT³³ 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 (cf. §3.3.2): OBJ_θ and OBL_θ. A *semantically 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 grammaticalised *discourse* functions: topic (TOP), focus (FOC) and SUBJ. The SUBJ has the unique property of being both an argument function and a grammaticalised discourse function (Bresnan, 2001:98, 308).

Finally, the semantic relationship between the SUBJ and the phrase following the copulative verb necessitates a specific sub-categorisation frame (cf. §6.2.2.2). 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

³² We refer to the secondary object or the indirect object as OBJ-TH in the XLE implementation (cf. Chapter 7).

³³ Quoting from Bresnan *et al.* (2015:97): “The ADJ(UNCT) function binds to a PRED rather than to one of its arguments: ADJ satisfies the completeness and coherence conditions by occurring in the same f-structure as the PRED it modifies.”

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).

3.3.2 LEXICAL MAPPING THEORY

In LFG, the Lexical Mapping Theory (LMT) lies at the heart of connecting the lexicon and the syntax, more specifically, subcategorisation frames for lexical categories (Bresnan & Moshi, 1990:147–185; 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). The basic idea behind LMT is explained in (i) to (iv).

- (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 semantic roles is *agent > beneficiary > experiencer/goal > instrument > patient/theme > locative*. In (3-16) the a-structure of the Setswana example includes the predicator and its argument roles.

(3-16) rekel- (buy for) < agent, patient, 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 (cf. §3.3.1). Basic principles for assigning features to thematic (semantic) roles in the a-structure are based on correspondences shown in Diagram 3-2 (p.34) and may be summarised as follows:

- 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 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 3-2: Correspondences between features of argument functions

The IRC and DRC is applied in (3-17).

(3-17) **rekel-** (buy for) < agent beneficiary theme >
 IRC: [-o] [+o] [-r]
 DRC: [-r] [+r]

(iii) The correspondence matrix in Diagram 3-2 is applied, for example (3-18).

(3-18) **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 (w.f.³⁴) on the relation between thematic roles and GFs:

- Subject condition

Every verb 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. This means that the patient/theme thematic role also maps to a SUBJ.

- Bi-uniqueness condition

This condition implies that a thematic role in the a-structure is associated with a unique function. Conversely, a unique function is also associated with a thematic role in the a-structure. As shown in (3-19), a thematic role is associated with only one GF and the GFs are not associated with more than one thematic role.

³⁴ Similarly, Dalrymple (2001:207) invokes the well-formedness condition (w.f.) for this purpose

(3-19) **rekel-** (buy for) < agent beneficiary theme >

IRC:	[-o]	[+o]	[-r]
DRC:	[-r]	[+r]	
	SUBJ	OBJ _θ	SUBJ/OBJ
w.f.:	SUBJ	OBJ _θ	OBJ

We have shown that by starting with the lexical semantics of a predicator (the extended verb root) (cf. §4.6.1.5), we are able to obtain its subcategorisation frame that is an essential part of the f-structure. We return to this topic in Chapters 4 and 6.

3.4 FUNCTIONAL STRUCTURE

The f-structure is an abstract level that includes the functional or syntactic information of the internal structure of a sentence (Dalrymple, 2001:7; Falk, 2001:11; Forst, 2011:2). An f-structure contains surface GFs, as discussed in §3.3.1, as well as features. Features in LFG refer to morphosyntactic characteristics such as person, number, tense, aspect, and pronoun type, amongst others³⁵. Only the features that have a role in functional syntactic constraints are presented in the f-structure, including grammatical features such as agreement, features specifying morphological structure of an argument such as the verb, and tense and aspect information (Dalrymple, 2001:27). An exposition of the functions and features applicable to the structure of the Setswana simple sentence is presented in Chapters 5 and 6.

3.4.1 REPRESENTATION OF THE FUNCTIONAL STRUCTURE

Formally, an f-structure is a finite set of attribute-value pairs (Bresnan, 2001:47). It is customary to represent these attributes and values as an attribute-value matrix (AVM), as shown in example (3-20).

(3-20) $f:M$ where $M = \begin{bmatrix} \text{ATTRIBUTE_1} & \text{VALUE_1} \\ \text{ATTRIBUTE_2} & \text{VALUE_2} \end{bmatrix}$

In this example, the function f maps the f-structure given by the AVM M as follows (Dalrymple, 2001:30): $(f \text{ attribute_1}) = \text{value_1}$ and $(f \text{ attribute_2}) = \text{value_2}$ ³⁶.

³⁵ Dalrymple (2001:28) presents a list of more commonly assumed features together with their values. However, this list does not define a set of features or values that must be included in the f-structure of a specific language.

³⁶ In LFG the notation $(fa)=v$ is used instead of the standard mathematical notation $f(a)=v$ (Bresnan, 2001:49).

Such a finite set of equations is also referred to as a functional description (f-description). We return to f-descriptions in §3.4.3.

An attribute refers to a GF or feature, as discussed in the previous sections. A value can be represented by a symbol, a semantic form, a set or another AVM (another f-structure) (Falk, 2001: 13–13; Nordlinger & Bresnan, 2011:113). For example, the symbol *sg* in (3-21) represents a singular value for the *NUM* (number) attribute.

(3-21) $\left[\text{NUM } \textit{sg} \right]$

A semantic form refers to the semantic content of a specific word and is represented by the *PRED* (predicate) attribute (Dalrymple, 2001:31, 104–105; Falk, 2001:13; Nordlinger & Bresnan, 2011:113). Dalrymple (2001:104) explains that the semantic form is instantiated to a unique value for each use of the word with which it is associated. This value is presented in single quotes. The semantic form of the Setswana noun **mosadi** (woman) is the noun root **-sadi**³⁷ in (3-22).

(3-22) $\left[\text{PRED } \textit{'SADI'} \right]$

Sub-categorisation requirements are also encoded in the value of the *PRED* attribute. For example, the sentence in (3-2) includes a *SUBJ* **mosadi** (woman) and an *OBJ* **mosese** (dress). The verb **o reka** (she buys) in this sentence subcategorises for these two arguments. This subcategorisation frame is illustrated in (3-23) where the arguments are enclosed in angle brackets (< >) and presented within the value of the *PRED* attribute.

(3-23) $\left[\text{PRED } \textit{'REK <SUBJ OBJ>'} \right]$

Sets are used to represent the f-structure of an argument that may have an arbitrary number of elements. Moreover, these elements can be f-structures. Sets are similarly used to represent coordination, as there is no limit to the number of conjuncts in coordinated structures (Dalrymple, 2001:33–35, 153–158). The f-structure representation in (3-25) of the adjuncts (*ADJUNCTS*) in (3-24) shows that the pronoun **gone** (there) and the adverb **jaanong** (now) can be presented as a set in the *ADJUNCT* function.

³⁷ The semantic form of each word category in Setswana is presented in §4.3 to §4.11 in Chapter 4. Setswana nouns and verbs are often morphologically complex. Their roots are considered as the semantic forms as shown in (3-19) and (3-20).

Dalrymple (2001:37) defines completeness as follows:

An f-structure is *locally complete* if and only if it contains all the governable grammatical functions that its predicate governs. An f-structure is *complete* if and only if it and all its subsidiary f-structures are locally complete.

The completeness condition ensures that all the arguments that a predicate subcategorises for are present in the f-structure. A sentence is ungrammatical if some essential or required material is missing (Dalrymple, 2001:35–36). The f-structure (3-29) of the sentence in (3-28) is incomplete. The transitive verb **o rekela** (buy for) must subcategorise for an OBJ_θ (indirect object) and an OBJ (direct object). The f-structure of this sentence will thus be complete if an OBJ, such as **dijo** (food), is included.

*(3-28) Mosadi o rekela mosetsana.

woman buy for girl

mo-sadi

o-rek-el-a

mo-setsana

NPre1-woman

AgrSubj1-buy-ApplSuf-VEnd

NPre1-girl

The woman buys for the girl.

(3-29) $f:$ $\left[\begin{array}{l} \text{PRED} \quad \text{'REK <SUBJ OBJ}_{\theta} \text{ OBJ}>' \\ \text{CLASS} \quad 1 \\ \text{TENSE} \quad \text{pres} \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'SADI'} \\ \text{CLASS} \quad 1 \end{array} \right] \\ \text{OBJ}_{\theta} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'SETSANA'} \\ \text{CLASS} \quad 1 \end{array} \right] \\ \text{OBJ} \end{array} \right]$

Dalrymple (2001:39) defines coherence as follows:

An f-structure is *locally coherent* if and only if all the governable grammatical functions that it contains are governed by a local predicate. An f-structure is *coherent* if and only if it and all its subsidiary f-structures are locally coherent.

The coherence condition ensures that only the arguments subcategorised for by the predicate are present in the f-structure. An f-structure with extra governable GFs that are not contained in the argument list of its semantic form is disallowed (Dalrymple, 2001:37; Falk: 2001:63–64). The f-structure (3-31) of the sentence in (3-30) is incoherent as the extra governable GF OBJ for **mosetsana** (girl) is included. The verb **o lela** (she cries) cannot subcategorise for an OBJ, as it is an intransitive verb.

(3-30) *Mosadi o lela mosetsana.

woman she cries girl

mo-sadi

o-lel-a

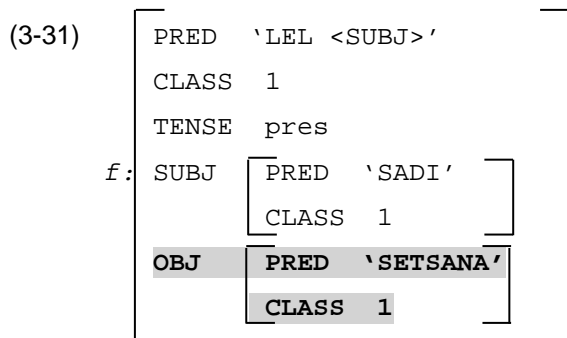
mo-setsana

NPre1-woman

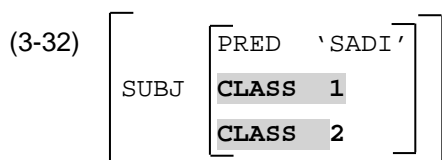
AgrSubj1-cry-VEnd

NPre1-girl

The woman cries the girl.



The consistency (uniqueness) condition disallows an f-structure that contains incompatible constraints. Kaplan and Bresnan (1982:203) describe it as the most important of the well-formedness conditions. A prerequisite for this condition is that every attribute in an f-structure may only have a single value (Dalrymple, 2001:39). For example, the value of the CLASS attribute presented in (3-32) is inconsistent as it shows two values. For **mosadi** (woman), the correct value of the CLASS attribute is 1.



3.4.3 FUNCTIONAL DESCRIPTIONS

An f-description plays an important part in encoding the relationship between a c-structure and an f-structure. In this section, we briefly explain the notion of f-description and in the subsequent section, we explicate its use in relating the c- and f-structure. The equations used on the f-description may be divided into two types, i.e. defining equations and constraining equations. It is also possible to place existential and negative existential constraints on an f-structure. Furthermore, an f-description can include a disjunction of two or more descriptions; it may also be optional or negated.

A defining equation defines an f-structure attribute as existing and having a particular value. Defining equations can furthermore be thought of as defining constraints since they define the required properties of an f-structure (Dalrymple, 2006:86). For example, the defining equation for the CLASS attribute in the f-structure in (3-32) to assume the value 1 is shown in (3-33).

(3-33) (SUBJ CLASS) = 1

A constraining equation is used to impose an additional requirement in an f-structure. This equation does not contribute any features to the f-structure, but it checks that the f-structure is well formed (Dalrymple, 2001:115). For example, the constraining equation in (3-34) constrains the value of the CLASS attribute in (3-33).

(3-34) (f CLASS) =_c 1

Notation: The subscript *c* on the equal sign distinguishes a constraint equation from a defining equation.

An existential constraint is an attribute of an f-structure that is required to be present, but does not require any particular value (Dalrymple, 2001:112). For example, in (3-35) the f-structure contains the attribute TENSE, but its value is not constrained.

(3-35) (f TENSE)

The negative existential constraint prohibits an f-structure to contain a specific attribute (Dalrymple, 2001:114). The negative existential constraint in (3-36) ensures that the f-structure may not have a TENSE attribute.

(3-36) ¬(f TENSE)

An f-description can also consist of a disjunction of two or more f-descriptions in which at least one of the disjuncts must be satisfied for the f-description to hold (Dalrymple, 2001:108). For example, the class 2 pronoun **bone** in Setswana (3-37) can be an absolute pronoun (they) or a possessive pronoun (them).

(3-37) {(f PRONTYPE) = AbsPro | (f PRONTYPE) = PossPro}

An optional f-description may, but need not be satisfied (Dalrymple, 2001:109). For example, when there is no overt SUBJ included in a sentence, the verb optionally contributes an f-description that constrains the value of the PRED attribute of its SUBJ (3-38). In certain conditions, an overt SUBJ does not need to be included in the sentence structure of Setswana. In (3-39), the sentence includes a SUBJ **batho** (people) and in (3-40) the subject agreement morpheme **ba**

refers to an overt SUBJ (a class 2 noun). In this instance, the subject agreement morpheme acts as an incorporated pronoun (cf. §6.5). The optional f-description is enclosed in parentheses.

(3-38) ((SUBJ PRED) = 'PRO')

(3-39) Batho ba a taboga.

people they run

ba-tho

ba-a-tabog-a

NPre2-person

AgrSubj2-PresPre-run-VEnd

The people are running.

(3-40) Ba a taboga.

they run

ba-a-tabog-a

AgrSubj2-PresPre-run-VEnd

They are running.

When an f-description is negated, the f-description is not allowed to hold (Dalrymple, 2001:111). In (3-41), the TENSE attribute is not allowed to assume the value PRES.

(3-41) (f TENSE) \neq PRES

or

$\neg(f$ TENSE) = PRES

3.5 RELATING CONSTITUENT AND FUNCTIONAL STRUCTURE

The relation or mapping between c- and f-structure is at "the heart of the descriptive power of LFG", since it relates all nodes in the phrase structure tree to corresponding elements in the f-structure (Falk, 2001:64). This relation is usually defined by the so-called many-to-one ϕ function³⁸.

We saw that for any given sentence the f-description is a set of equations that represents its f-structure. In this section, we relate the (elements of the) f-structure with the nodes in the phrase structure tree. That is, we show how to construct the function ϕ for this sentence. This is best done by means of an example. We consider the sentence in (3-13), **mosadi o a reka** (the woman buys). Its f-description is presented in (3-42) (see also (3-43): Figure 3-1, p.42).

³⁸ The ϕ function is also referred to as the correspondence function, the mapping function or the projection function.

- (3-42) $(f_1 \text{ SUBJ}) = f_2$ (1)
 $(f_1 \text{ SUBJ CLASS}) =_c (f_1 \text{ SUBJ AGR})$ (2)
 $f_2 = f_3$ (3)
 $(f_3 \text{ PRED}) = \text{'SADI'}$ (4)
 $(f_3 \text{ NTYPE}) = \text{ord}$ (5)
 $(f_3 \text{ CLASS}) = 1$ (6)
 $f_1 = f_4$ (7)
 $f_4 = f_5$ (8)
 $(f_5 \text{ PRED}) = \text{'REK<(\uparrow \text{SUBJ})>'}$ (9)
 $(f_5 \text{ SUBJ AGR}) = 1$ (10)
 $(f_5 \text{ TENSE}) = \text{pres}$ (11)

We note the representation of the noun class to which **-sadi** belongs (see equations 4 and 6 in (3-42)), as well as the concordial agreement between the CLASS of the SUBJ and the agreement morpheme in the verb (equation 2).

The function ϕ is illustrated in ((3-43): Figure 3-1) by mapping each node in the tree to an f-structure as follows: $\phi(S)=f_1$, $\phi(NP)=f_3$, $\phi(N)=f_3$, $\phi(VP)=f_4=f_1$, and $\phi(V)=f_5=f_1$.

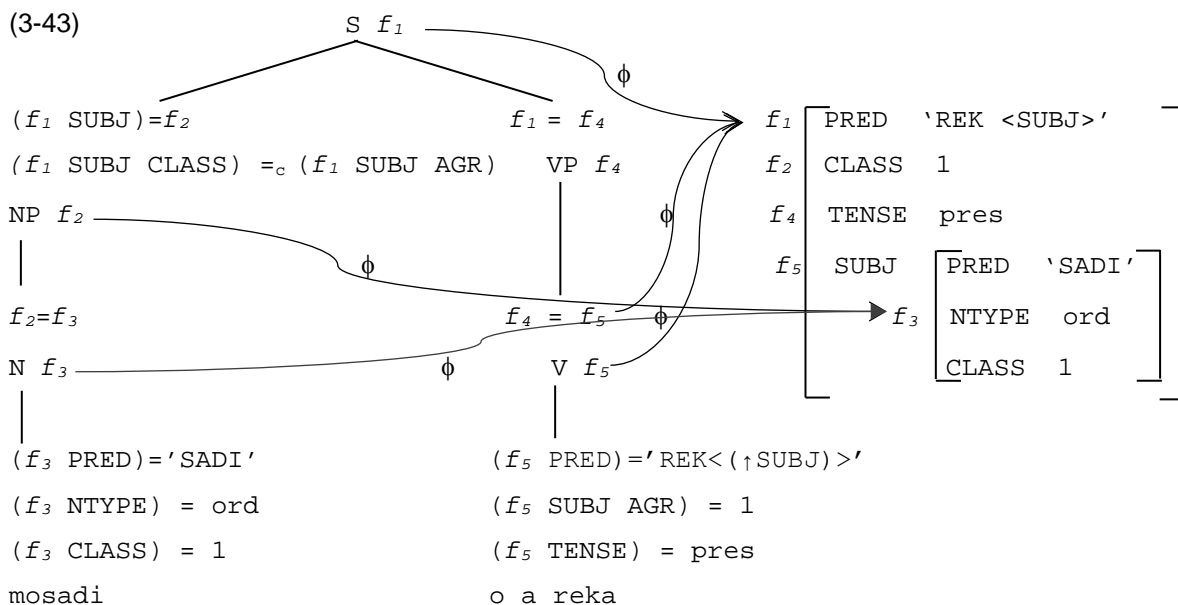


Figure 3-1: Annotated phrase structure tree and its correspondence function ϕ

In order to abstract away from the nodes of a particular tree (as in (3-43)), LFG allows the annotation of phrase structure rules with so-called functional schemata of constraints. These schemata serve as abstractions of f-descriptions. For example, the schemata for $(f_1 \text{ SUBJ}) = f_2$ and $f_2 = f_3$ are $(\uparrow \text{SUBJ}) = \downarrow$ and $\uparrow = \downarrow$ respectively.

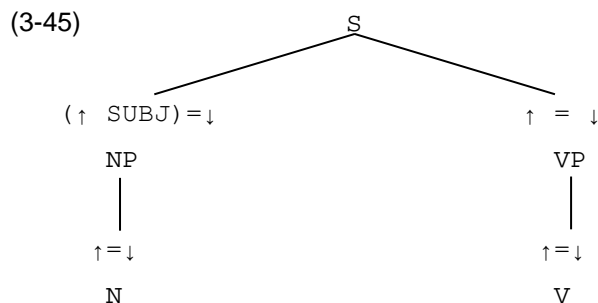
Any node on the right hand side (i.e. a daughter node) of a phrase structure rule may therefore be annotated with constraints that hold between the f-structures of its mother and a daughter node. The notation used in representing these constraints is the symbols \uparrow and \downarrow . The \downarrow refers to the f-structure associated with the daughter node under which the schema containing the \downarrow appears (see (3-44)). The \uparrow refers to the f-structure associated with the mother node of the daughter node under which the schema containing the \uparrow appears.

(3-44) $S \rightarrow NP \quad VP$
 $(\uparrow \text{SUBJ}) = \downarrow \quad \uparrow = \downarrow$

The following information is included in (3-44):

- The annotation $(\uparrow \text{SUBJ}) = \downarrow$ means that the f-structure of the mother node S has a SUBJ attribute the value of which is the f-structure of the NP itself.
- The annotation $\uparrow = \downarrow$ indicates the head relation. The f-structure of the VP node is shared with the mother node S , making all functional information carried by this node also direct information of the mother's f-structure.

Functional schemata can also be included in a phrase structure tree. This is illustrated in (3-45).



A lexical entry in LFG includes a "categorical specification indicating the preterminal category under which the lexical item may be inserted, and a set of schemata to be instantiated" (Kaplan & Bresnan, 1982:185). The lexical entry can thus include the specific word category, the PRED attribute as well as other morphosyntactic properties. As a terminal node, a lexical entry only specifies information about the f-structure of the immediately dominating nonterminal node by making use of \uparrow (Dalrymple, 2001:121). Examples (3-46) and (3-47) show the lexical entries for **-sadi**, the noun root for **mosadi** (woman) and **-rek-** (buy), the verb root for **o a reka** (she buys) and (3-43 (cf. Figure 3-1, p.42) shows how these lexical entries are used in the annotated phrase structure tree for the sentence **Mosadi o a reka** (The woman buys).

(3-46) *sadi* N (\uparrow PRED) = 'SADI'
(\uparrow NTYPE) = ord
(\uparrow CLASS) = 1

(3-47) *rek* V (\uparrow PRED) = 'REK' <SUBJ>
(\uparrow VTYPE) = main
(\uparrow CLASS) = 1
(\uparrow TENSE) = pres

3.6 SUMMARY

In this chapter, we discussed four important aspects of LFG. We introduced the notions of c- and f-structure as parallel mutually constraining levels of syntactic representation³⁹. The c-structure represents the order and grouping of constituents while the f-structure represents the functional information of these constituents. We furthermore briefly addressed the notion of a-structure in the context of the lexicon, GFs and subcategorisation frames, showing how a-structure maps to f-structure by means of LMT. Thereafter we explained and illustrated with a detailed example, the relationship between the c- and f-structure, based on the so-called f-description and mapping function ϕ , which forms the heart of LFG. Finally, we introduced the notion of lexical entry as annotation of the terminal nodes of the c-structure tree to conclude the description of an LFG syntactic representation.

The theoretical background of LFG presented in this chapter serves as the background for Chapters 4, to 6 and where we address Setswana subcategorisation frames, phrase structure and sentence structure, specifically pertaining to the simple sentence. It also provides the background for the grammar development discussed in Chapter 7.

³⁹ The other levels that are distinguished in LFG are m(orphology)-structure, a(rgument)-structure, s(emantic)-structure, p(honological)-structure and i(nformation)-structure (cf. §1.1.3).

CHAPTER 4

SETSWANA WORDS

4.1 INTRODUCTION

The question of what constitutes a word, a word category, a word group, a phrase and a sentence in Setswana is of core importance to this study. It is therefore appropriate first to provide a brief historic overview of how these concepts were studied, formulated and used in Setswana linguistics. We then present an exposition of the word categories of Setswana as basic building blocks for phrases (cf. Chapter 5) and sentences (cf. Chapter 6), with specific reference to the feature structure of each word category.

4.2 HISTORIC DEVELOPMENT OF SETSWANA GRAMMAR

Kosch (1991) provides an excellent exposition of the four major linguistic periods that reflect the development of Bantu grammar from 1876 to 1991, viz. the traditional period, the functional or Dokeian period, the structural or Van Wyk era, and the modern period. While her focus is mainly on Sesotho sa Leboa, her work is also of great relevance for Setswana, which is a sister language of Sesotho sa Leboa, being one of the three Sotho languages. Although the historical development of Setswana grammar falls outside the scope of this study, we highlight those central moments that have played a key role in providing insights into the Setswana word, its word categories, word groups, phrases and sentences.

The traditional period is characterised by the contributions of European missionaries and travellers who presented the early descriptions of Setswana. These contributions included vocabulary lists, lists of phrases, a first Setswana-English dictionary and grammatical descriptions. The classical European languages were mainly used as a basis for the description and the framework set by contemporary European linguistic tradition, was applied (cf. Cole, 1955: xxiii–xxviii).

The more systematic approach to the description of isiZulu by Doke (1927) heralded the functional period. Doke was the first person to develop a new descriptive model focussing on the basic linguistic characteristics of a Bantu language. One of his main aims was to present a classification of words into word categories. He used the sentence as the starting point to identify words and based his word identification on the function of a word in a sentence as well as on the grammatical relationship that exists between words. He classified twelve word categories by incorporating the structural characteristics of words. Doke recognised the distinctive structure of the Bantu

languages and succeeded to move away from the traditional European based views regarding word identification and classification in the Bantu languages (Kosch, 1991:41–54).

Still in the functional tradition, Cole (1955) made a detailed contribution to describe the Setswana grammar. He applied Doke's⁴⁰ model to determine word categories for Setswana and identified thirteen word categories, viz. nouns, pronouns, adjectives, enumeratives, quantitatives, possessives, relatives, verbs, copulatives, adverbs, ideophones, conjunctions and interjections. These categories were presented as six major categories and sub-divisions (Cole, 1955:59). The major categories were distinguished based on a functional grouping according to the function of words as well as the grammatical relationship between words in a sentence. The sub-divisions were distinguished based on the structure of words. He furthermore gave an exposition of the morphological structure of these word categories and described the inclusion of these words in syntactic constructions.

The structural era was ushered in by Van Wyk (1958, 1962, 1964, 1966 and 1967) who did not apply Doke's model, but based his description of Sesotho sa Leboa and isiZulu on insights gained from European structuralists such as De Groot, Uhlenbeck, Paardekooper and Reichling. His main contribution concerned word division, identification and classification (Van Wyk, 1958). He described word *division* as an orthographic problem and considered the semi-conjunctive method of word division as the most scientifically motivated approach. He considered Sesotho sa Leboa and isiZulu word *identification* as a linguistic problem. For this purpose, he formulated and applied four identification tests, i.e. isolatability, separability, transposability and replaceability. He also developed a comprehensive *classification* system consisting of both principles and procedures that need to be applied. It is based on morphological, syntactical, phonological, as well as semantic criteria. This work culminated in a classification of eight word categories for Sesotho sa Leboa (Van Wyk, 1966:251–257). He made a significant contribution to advancing Bantu grammatical study as a scientific discipline and is considered a pioneer in the scientific study of word identification and classification for the Bantu languages (Kosch, 1991:68).

Following the Van Wyk tradition, Krüger (1967) pioneered the classification of eight word categories for Setswana, viz. nouns, pronouns, verbs, particles, conjunctions, adverbs, interjections and ideophones. He is also noted for his novel application of the word classification technique of Van Wyk to distinguish so-called word groups in Setswana, which culminated in recent significant publications (Krüger, 2006, 2013a, 2013b).

⁴⁰ Cole was Doke's student.

Although numerous studies⁴¹ on the use of other, more recent, linguistic theories have been published, no clear trend that suggests a new modern period in Setswana grammatical description can be identified at this stage. The emphasis remains on the word, the word group, and the sentence as a word group.

In subsequent sections, we discuss the *status quo* in Setswana grammatical description with respect to the Setswana words and its word categories (this chapter), word groups and phrases, (Chapter 5) and the simple sentence (Chapter 6); we also employ LFG to frame a Setswana grammatical description. This study may therefore be seen as a contribution towards the description of this language in a modern linguistic theory.

4.3 WORD CATEGORIES AND THEIR FEATURES

Setswana words are classified into eight word categories (parts of speech, lexical categories, word classes) viz. nouns, pronouns, verbs, particles, conjunctions, adverbs, interjections and ideophones⁴². Some of these word categories can in turn, be divided into subcategories (Van Wyk, 1966; Krüger, 2006). It is customary to distinguish between open and closed word categories in the Bantu languages. Nouns and verbs are open (morphologically productive) categories and the pronouns, particles, conjunctions, adverbs, interjections and ideophones constitute the closed (morphologically unproductive) categories (Pretorius, L., *et al.*, 2008:2; Pretorius, 2014:49).

We define the notion of *root* as the inomissible morpheme that carries the basic lexical meaning (principal semantic load) of a word (Kosch, 2006:7). Since, in LFG, the `PRED` feature carries the principal semantic load of a word and has a unique value (Dalrymple, 2001:31), the root constitutes the value of this feature in the case of a noun or a verb. For the other word categories, the value of the `PRED` feature is the word itself.

4.4 NOUN

Each noun in Setswana belongs to a specific noun class (cf. §1.1.1). There are twenty noun classes (Appendix A: Table 1-1, p.189). Nouns belong to a specific class based on the form of their class prefixes and to a lesser extent, their meaning (Louwrens, 1994a:126). These noun

⁴¹ Refer to §2.2 for a number of examples.

⁴² Note on interrogatives: Interrogatives in Setswana do not constitute a word class on their own. Krüger (2013b:349-371) describes the word status and function of various Setswana interrogatives. For example, the interrogatives **eng?** (what?), **kae?** (where?), **leng?** (when?) and **jang?** (how?) are usually classified as adverbs in Setswana because of their usage in typical adverbial position. Khoali (1994) gives an exposition of the use of the different Setswana interrogatives.

classes are divided into the non-infinitive noun classes (classes 1 to 14), the infinitive noun class (class 15), and the locative class nouns (class 16-20) (Krüger, 2006: 57–98).

The nouns in classes 1 to 14 are *ordinary* (basic or common) nouns and their morphological structure consists of a class prefix and a root. In (4-1) the noun **molao** (law) contains a class prefix **mo-** and a root **-lao**. These noun classes typically have singular-plural pairings. The odd class numbers express the singular and the corresponding even class numbers, the plural. In example (4-2) the class prefix **mo-** of class 3 denotes the singular and the class prefix **me-** of class 4 the plural.

(4-1) molao
 law
mo-lao
 NPre3-law
 the law

(4-2)	molao	>	melao
	law		laws
	<i>mo-lao</i>		<i>me-lao</i>
	NPre3-law		NPre4-law
	the law		the laws

A *deverbative noun* is a noun that is derived from an infinitive verb by substituting the infinitive prefix **go** (to) with a noun class prefix and, concomitantly, by substituting the verbal ending **-a** with a deverbative suffix (**-i**, **-e**, **-o** or **-a**) (Krüger, 2006:110). In (4-3) the deverbative noun **lebone** (light) is formed from the infinitive verb **go bona** (to see) by replacing the infinitive prefix with the class prefix **le-** and the verbal ending **-a** with the deverbative suffix **-e**.

(4-3) lebone
 light
le-bon-e
 NPre5-see-DevSuf
 a light

An ordinary noun has valence for four suffixes. This implies that a noun can be inflected with a locative suffix, a diminutive suffix and, in a smaller number of instances, a feminine suffix or an augmentative suffix (Krüger, 2006: 73–92). These suffixes can also occur in combinations if they are semantically plausible.

A *locative noun* is formed by adding a locative suffix to the noun root (Krüger, 2006:87-91). In example (4-4) a locative noun **ditaseleng** (in the roads) is formed from the noun **ditsetla** (roads).

(4-4)	ditsetla	>	mo ditaseleng ⁴³
	roads		in roads locality
	<i>di-tsetla</i>		<i>mo</i> <i>di-tsetla-ing</i>
	NPre10-road		LocPart _{mo} NPre10-road-LocSuf
	the roads		in the roads

A *diminutive noun* is formed by adding a diminutive suffix to the root (Krüger, 2006:74–87). In example (4-5) the diminutive noun **ditsetlana** (small roads) is formed from the noun **ditsetla** (roads).

(4-5)	ditsetla	>	ditsetlana
	roads		roads small
	<i>di-tsetla</i>		<i>di-tsetla-ana</i>
	NPre10-road		NPre10-road-DimSuf
	the roads		the small roads

A *feminine noun* includes the feminine suffix that is added to its root (Krüger, 2006:73–74). The feminine noun **kgosigadi** (queen, chieftainness) in (4-6) is formed from the noun **kgosi** (chief).

(4-6)	kgosi	>	kgosigadi
	king		chief feminine
	<i>(ne)-kgosi</i>		<i>(ne)-kgosi-gadi</i>
	NPre9-king		NPre9-king-FemSuf
	the king		the queen

The suffix corresponding to **-gadi** is also used with augmentative significance (Cole, 1955:110–111). An *augmentative noun* is formed when this suffix is added to a root, for example **podigadi** (a very big goat) (4-7) is formed from the noun **podu** (goat).

⁴³ Owing to morphophonological alternation, the morpheme sequence **di-tsetla-ing** results in the surface form **ditaseleng**. This phenomenon is common in our examples.

(4-7)	podu	>	podigadi
	goat		goat big
	(ne)-podi		(ne)-podi-gadi
	NPre9-goat		NPre9-goat-AugSuf
	a goat		a very big goat

In Setswana, the *adjective*⁴⁴ is classified as a so-called describing noun and is considered a subcategory of the noun. An adjective can be an ordinary adjective or a changeable adjective (Krüger, 2013a:127–136). The *ordinary adjective* is a complete or independent word. For example, the adjective **tsididi** (cold) in (4-8) and (4-9) does not change to show agreement with a preceding describing copulative verb.

(4-8)	mašwi a tsididi.		
	milk is cold		
	ma-šwi	a	tsididi
	NPre6-milk	DesCopV6	adjective
	the milk is cold		

(4-9)	kofi e tsididi		
	coffee is cold		
	(ne)-kofi	e	tsididi
	NPre9-coffee	DesCopV9	adjective
	the coffee is cold		

The *changeable adjective* includes an agreement morpheme (the class prefix of the qualified noun) and an adjectival root. For example, in (4-10) the adjective root **-golo** (big) describes the class 1 noun **motho** (person) and in (4-11), it is used to describe the class 2 noun **batho** (people).

(4-10)	motho yo mogolo		
	person that is big		
	mo-tho	yo	mo-golo
	NPre1-person	QualPart1	NPre1-adjective
	a big person		

⁴⁴ Posthumus (2000:148–158) presents an overview of the use of adjectives in isiZulu.

(4-11) batho ba bagolo
 people that are big
ba-tho *ba* *ba-golo*
 NPre2-person QualPart2 NPre2-adjective
 big people

Infinitives may function as a noun or as a verb. Infinitive nouns belong to class 15. They do not denote singular or plural (Krüger 2006:92–93). The morphological structure of an infinitive noun includes a class prefix **go** as well as a verb root and an ending **-a**. The verb root in (4-12) is **tshamek-** (play). The infinitive noun has valence for locative and/or diminutive suffixes (Krüger 2006:93).

(4-12) go tshameka
 to play
go-tshamek-a
 NPre15-play-VEnd
 the playing

Nouns in classes 16 to 20 constitute the *locative class nouns* in Setswana. They are not marked for singularity or plurality (Krüger, 2006:92–98). These nouns consist of a class prefix followed by a root (4-13). It is not common for these nouns to take suffixes. Krüger (2006:96) provides examples of only a few nouns in classes 16 to 20 that have valence for locative and diminutive suffixes.

(4-13) morago
 behind
mo-rago
 NPre18-behind
 behind

The class and subcategory information presented in this section manifest in the feature structure of the Setswana noun as follows:

- The class information is provided by the class prefix and is essential for modelling agreement in syntactic structures. The value of the CLASS feature is the class number, .that is one of {1, 1a, 2, 2a, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20}.
- Inflected nouns (diminutive, deverbative, augmentative, feminine and infinitive nouns) syntactically behave the same as nouns with no extensions. However, this is not true for

locative nouns, locative class nouns and adjectives since they occur in specific syntactic contexts. We return to this topic in Chapter 5 (cf. §5.2, §5.3 and §5.5.2). The value of the noun type (NTYPE) feature is therefore one of {ord, adjective, locn, loccln}.

The relevant features are applied in an f-structure of the noun **mosadi** (woman) as indicated in (4-14).

(4-14) **mosadi** $\left[\begin{array}{ll} \text{PRED} & \text{'SADI'} \\ \text{NTYPE} & \text{ord} \\ \text{CLASS} & 1^{45} \end{array} \right]$

4.5 PRONOUN

A fixed number of pronouns appear in Setswana (Krüger, 2006:127–144). They are the different types of pronouns for the noun classes (Appendix A: Table 4-1, p.190) and personal pronouns (Appendix A: Table 4-2, p.191).

The *pronouns for the noun classes* agree with the classes and are used to modify or replace a noun. A noun can be replaced when it is regarded as given or old information (Cole, 1955:127). Absolute, demonstrative, quantitative and possessive pronouns are distinguished for the noun classes (Appendix A: Table 4-1, p.190).

An *absolute pronoun* is used to emphasise or modify a noun or to replace a noun that is previously mentioned (Krüger, 2006:131). In (4-15), the absolute pronoun **bone** emphasises the noun and has the meaning of *specific* or *particular*. In (4-16) the absolute pronoun **bone** emphasises the preceding class 2 noun.

(4-15) batho bone
 people specific
ba-tho *bone*
 NPre2-person AbsPro2
 specific people

⁴⁵ Singular and plural are included in the class information and is therefore not indicated explicitly in the f-structure of nouns.

(4-16) Bone ba a dira.
 they they work
bone ba-a-dir-a
 AbsPro2 AgrSubj2-PresPre-work-VEnd
 They are working.⁴⁶

A *demonstrative pronoun* is used to indicate one of three possible distances: distance 1 is a distal position (this, these), distance 2 shows a proximal position (that, those) and distance 3 is a post-distal position (over there) (Krüger, 2006:134–135). In (4-17), the noun **batho** (people) is modified by the demonstrative pronoun **ba** that indicates a distance nearest to the speaker.

(4-17) batho ba
 people these
ba-tho ba
 NPre2-person DemPro2D1
 these people

Inclusive, exclusive, selective, separative and interrogative quantitative pronouns are distinguished for Setswana (Cole, 1955:154–158; Krüger: 2006:137–139). The inclusive quantitative pronoun denotes a meaning of *the whole* or *all* and the exclusive quantitative pronoun has a meaning of *no* or *none*. Selective quantitative pronouns denote the meaning of *only*, while the separative quantitative pronoun denote the meaning of *unknown*. The interrogative quantitative pronoun denotes a meaning of *which* relating to the noun it modifies. In (4-18) the noun **batho** (people) is modified by an inclusive quantitative pronoun **botlhe** (all).

(4-18) batho botlhe
 people all
ba-tho botlhe
 NPre2-person IncQPro2
 all the people

The *possessive pronoun* has "a referential lexical meaning of 'possessor'" (Krüger, 2006:145). In (4-19) the possessive pronoun **gagwe** (his) is the possessor of the noun **mafoko** (words).

⁴⁶ In Setswana, the simple present tense and the present continuous tense are not distinguished, as in English (cf. §4.6.1). Cole (1955:244) states: "The two forms of the present tense positive are virtually identical in significance, thus **Ke a ruta** (I teach, I am teaching), and **Ke ruta bana** (I teach children, I am teaching children)."

(4-19) mafoko a gagwe

words of him

ma-foko

a

gagwe

NPre6-word

PossPart6

PossPro1

his words

Personal pronouns are person and number bound and they have corresponding *possessive pronouns*. Only first and second person pronouns are alluded to here. Third person pronouns are identical to and treated as the absolute pronouns of classes 1, 1a, 2 and 2a. In (4-20), the possessive pronoun **ka** (me) is used as the possessor of the noun **mafoko** (words).

(4-20) mafoko a ka

words of me

ma-foko

a

ka

NPre6-word

PossPart6

PossProPlsg

my words

The possessive pronouns **etsho** (**getsho**), **eno** (**gaeno**) and **gabo**, occurring only in plural form, indicate *communal possession* respectively and not individual possession that is indicated by the ordinary possessive pronoun of the persons. The communal possession expresses the idea that a community (mainly a family) owns something (Cole, 1955:163–164). Example (4-21) is one of just but a few examples mentioned by Cole (1955:164) that illustrate communal possession.

(4-21) bana ba gaeno

children of our place/village

ba-ana

ba

gaeno

NPre2-child

PossPart2

PossProP2pl

children of our place/village

The feature structure of Setswana pronouns can include the following:

- The subcategory information of the pronoun determines the values of the pronoun type (PRONTYPE) feature, that is one of {PersPro, AbsPro, DemPro, PossPro, IncQPro, ExcQPro, SelQPro, SepQPro, IntQPro}.
- Since a personal pronoun indicates first or second person and number, it has the features person (PERS) and number (NUM): PERS as either the value first person (P1) or second person (P2) and NUM either singular (sg) or plural (pl).

- The absolute, demonstrative, possessive and all the quantitative pronouns exhibit a CLASS feature with the value one of {1, 1a, 2, 2a, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20}.

The f-structure of the personal pronoun **rona** (we) is represented in (4-22) while the f-structure of the demonstrative pronoun **yo** (this), which belongs to class 1, is presented in (4-23).

(4-22) **rona**

PRED	'RONA'
FRONTTYPE	PersPro
PERS	p1
NUM	pl

(4-23) **yo**

PRED	'YO'
FRONTTYPE	DemPro
CLASS	1

4.6 VERB

Lexically the verb in Setswana has three sub-categories, viz. auxiliary verbs, main (independent) verbs and copulative verbs (Pretorius, 2003:14). These categories are distinguished based on semantic, morphological and syntactic features (Krüger, 1983:34; Pretorius, 1997:76–81; Kruger (2013b:320) (cf. §2.3).

4.6.1 MAIN VERB

Setswana has an agglutinative morphology (cf. §1.1.1) with the main verb being the most complex word category. In terms of morphotactics (cf. also §1.1.1), each morpheme occupies a specific slot in a verb. The core of the main verb is 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 modify its meaning in order to form a complete word (Kosch, 2006:7).

The main verb may include inflectional prefixes as well as derivational and inflectional suffixes as presented in Table 4-3 (p.58). This table is based on the verbal slot system presented by Krüger (2006:268) that includes the broad structure and the fixed order in which the affix slots are filled.

A subject agreement morpheme (Appendix A: Table 4-4 to Table 4-7, p.191-192) is obligatory in all modal forms of the verb, except for the imperative mood. Other possible prefixes are a negative morpheme (occupying one of two possible slots), the present tense, progressive, potential and

temporal morphemes, as well as one or more object agreement morphemes (including the reflexive morpheme) (Appendix A: Table 4-8 and Table 4-9, p.193). The negative morphemes, subject agreement morphemes, present tense morpheme, progressive morpheme and the future tense morpheme contribute to tense, aspect, mood and polarity (TAMP). The object agreement morpheme, the reflexive morpheme and the potential morpheme do not contribute to TAMP. The productive suffixes⁴⁷ include the verbal extensions and the perfect morpheme. We distinguish between the causative, reciprocal, applied and passive extensions (cf. §4.6.1.5). The perfect is an inflectional morpheme and specifies either past tense or perfective aspect. The final two slots in Table 4-3 (p.58) are filled by the verbal endings and the relative and imperative suffixes. The verbal endings are inflectional morphemes and their form is determined by mood, tense and polarity information.

The terms verb root and verb stem are applied in different ways in Bantu languages (Posthumus, 1994; Pretorius, 2000; Krüger, 2006). A root is that part of the word that does not include any grammatical morpheme, i.e. prefixes or suffixes, and it cannot occur independently. The root establishes the lexical meaning of a word (Pretorius, 2000:61; Krüger, 2006:35). A stem is that part of a word that may include grammatical morphemes and it operates independently. The lexical meaning of a word is established by the stem and the stem has a word correlate outside the structure in which it occurs (Pretorius, 2000:60; Krüger, 2006:35).

In order to accommodate the PRED value in the f-structure of verbs we adopt the following definitions by Kosch (2006:7-12), and therefore distinguish between a basic verb root, an extended verb root, a basic verb stem and an extended verb stem. The basic verb root is a monomorphemic unit and does not include any prefixes or suffixes. An extended verb root denotes a polymorphemic structure that consists of the basic root plus at least one suffix, excluding the verbal ending (Kosch, 2006:7). In (4-24), the basic verb root is **tabog-** (run) and in (4-25) the extended verb root **tabogis-** (let run) consists of the root **tabog-** (run) and the causative extension **-is-**.

Kosch (2006:10) describes the stem of a verb as the verb root plus all the suffixal morphemes in the word but distinguishes between a basic stem and an extended stem. A basic stem consists of a basic verb root followed by a verbal ending. An extended stem consists of a basic root followed by one or more suffixes as well as the verbal ending (Kosch, 2006:10). The basic stem in (4-24) is **taboga** (run) and consists of the verb root **tabog-** and the verbal ending **-a**. The

⁴⁷ The Setswana verbal suffixes are classified into unproductive, semi-productive and productive suffixes (Krüger, 2006:209). However, only the productive suffixes are included here as the unproductive and semi-productive are included as part of the verb root in this study.

extended stem in (4-25) is **tabogisa** (let run) and consists of the verb root **tabog-**, the causative extension **-is-** and the verbal ending **-a**.

(4-24) ba a taboga

they run

ba-tabog-a

AgrSubj2-run-VEnd

they run

(4-25) ba tabogisa

they run let

ba-tabog-is-a

AgrSubj2-run-CausSuf-VEnd

they let run

On the other hand we adhere to the *Lexical Integrity Principle* (cf. §3.2.2) as a result of the way in which verbs are treated in the morphological analyser (cf. §7.5) where disjunctively written prefixes are included as part of verb stems. This ensures that verbs occupy terminal node positions in the c-structure and do not allow for verbal prefixes to be terminal nodes.

The significance of the inflectional prefixes and suffixes in the main verb regarding mood, tense, aspect and polarity is discussed in §4.6.1.1 to §4.6.1.4. The inclusion of the verbal extensions in the main verb and their influence on the meaning and argument structure of a verb are explained in §4.6.1.5.

4.6.1.1 Mood

We distinguish eight verbal moods in Setswana, viz. the indicative, consecutive, imperative, habitual, participial, subjunctive, infinitive and relative moods⁴⁸. The independent moods are the indicative, consecutive, imperative and habitual moods, as they are semantically not dependent on other known, given information. The participial, subjunctive, infinitive and relative moods are semantically and syntactically dependent because they act as subordinate members in complex sentences (Louwrens, 1994a:139–140; Krüger, 2013b:111–128). Since this study concerns the structure of the simple sentence, only the indicative, consecutive, imperative and habitual moods are relevant.

⁴⁸ There are different opinions on the modal classification. See also, for example, Louwrens, *et al.* (1995).

Prefixes						Root	Suffixes						
							Productive suffixes					VEnd	RelSuf
NegPre	AgrSubj	NegPre	PresPre ProgPre PotPre	FutPre	AgrObj RefIPre		CausSuf	ApplSuf	RecSuf	PerfSuf	PassSuf	Mood Tense Polarity	ImpSuf
B	ga	Person Class	se sa a	a sa ka	tla	Person Class i- (reflexive)	-is- -y-	-el-	-an-	-il-	-iw- -w-	-a -e (e or ê)	-ng (relative) -ng (imperative)

A. Linear arrangement

B. Morphological items

Table 4-3: Schematic representation of the morphological structure of verbs (adapted from Krüger, 2006:268)

The *indicative mood* expresses factual statements and the inclusion of a subject agreement morpheme is required in the morphological structure of the verb (Louwrens, 1994a:84–85). The verb **ba ruta** (they teach) in (4-26) includes a subject agreement morpheme **ba** of class 2 that agrees with the noun **basadi** (women).

(4-26) Basadi ba ruta bana dipalo.

women they teach children mathematics

ba-sadi

ba-rut-a

ba-ana

di-bal-o

NPre2-woman AgrSubj2-teach-VEnd

NPre2-child NPre10-count-DevSuf

The women teaches the children mathematics.

The *consecutive mood* indicates that an action follows another action in time. It occurs in a temporal succession to a previous action and thus conveys a subsequent action (Louwrens, 1994a:36). Krüger (2013b:117) explains that although the action in the consecutive mood is chronologically dependent on previous actions, it does not imply syntactic dependence. Therefore, a simple sentence may appear in the consecutive mood. The consecutive mood is marked morphologically in the verbal structure by including a consecutive subject agreement morpheme (Appendix A: Table 4-6 and Table 4-7, p.192). The verb **a kopa** (he then sought) in (4-27) includes a consecutive subject agreement morpheme **a** of class 1 that agrees with the noun **monna** (man).

(4-27) Monna a kopa thuso.

man he (then) seek help

mo-nna

a-kop-a

(ne)-thuso

NPre1-man

ConsAgrSubj1-seeK-VEnd

NPre9-help-DevSuf

The man then sought help.

The *imperative mood* expresses a command, an order or a warning to the second person, singular or plural (Krüger, 2013b:112). A verb in the imperative mood (4-28) may include morphemes such as a negative morpheme, a subject agreement morpheme and an object agreement morpheme. In (4-29), the subject agreement morpheme **o** of the second person singular and the object agreement morpheme **di** (it) of class 10 are included in the verb.

(4-28) Dira!
 work
dir-a
 work-VEnd
 Work!

(4-29) O di tlogele!
 you it/them leave
o-di-tlogel-e
 AgrSubjP2sg-AgrObj10-leave-VEnd
 You must leave it/them!

The *habitual mood* expresses sequential actions, which are performed as a *habit* (Krüger, 2013b:119). There are various ways of marking the habitual mood. Verbs in the habitual mood exhibit the verbal ending **-e** in the positive and negative (4-30).

(4-30) Ba ye kwa tirong ka bese.
 they go to work with bus
ba-y-e *kwa* *(ne)-dir-o-ing*
 AgrSubj2-go-VEnd LocPart_{kwa} NPre10-work-DevSuf-LocSuf
ka *(ne)-bese*
 InstrPart NPre9-bus
 They usually went to work by bus.

Verbs in the habitual mood frequently appear with the auxiliary verb **tle**. This auxiliary verb has alternative forms, i.e. **tla**, **etle** or **etla** and these auxiliary verbs indicate a meaning of *usually* or *frequently* or it indicates *a matter of habit or regular custom* (Cole, 1955:304). Examples from Cole (1955:304) are included to exemplify the use of **-tle** ((4-31) and (4-32)). In (4-31), a long form is used as a present tense morpheme **a** is included. In (4-32), a short form is used as the present tense morpheme **a** is not included in the structure. Cole (1955:304) explains that the use of the long form "is more common and seems to be more definite in significance". The notions of long and short form are discussed in §4.6.1.2.

(4-31) Ke a tle ke reke.
 I usually I buy
ke-a-tle *ke-rek-e*
 AgrSubjP1sg-PresPre-Aux AgrSubjP1sg-buy-VEnd
 I usually buy.

(4-32) Ke tle ke reke.
 I usually I buy
ke-tle *ke-rek-e*
 AgrSubjPlsg-Aux AgrSubjPlsg-buy-VEnd
 I usually buy.

4.6.1.2 Tense

The indicative, participial and relative moods distinguish tense. The consecutive, imperative, habitual, subjunctive and infinitive moods do not distinguish tense. Tense forms in Setswana are divided in terms of absolute and relative tenses.

The *absolute tenses* are the past, present and future and are marked morphologically in the structure of the verb. Nurse (2008, 120) declares that absolute tenses "relate the situation to the present moment". The time of the action in the verb is thus defined in relation to the moment of speech, i.e. coding time (Diagram 4-1). Coding time is the point in time that is used to make an interpretation of time and it is referred to as the deictic centre (Posthumus, 1990:22; Pretorius, 1997:162).

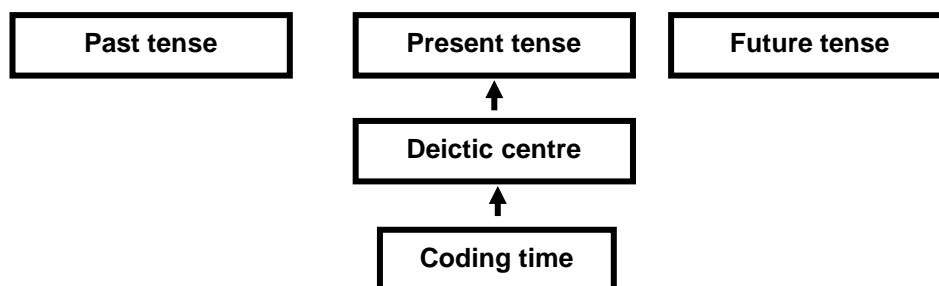


Diagram 4-1: Absolute tenses (cf. Posthumus, 1990:22; Pretorius, 1997:162)

The *present tense* is "generally used to express a habitual or customary action in the present time" (Poulus & Louwrens, 1994:208). In the *positive* of the main verb, two distinct forms, i.e. a long form and a short form, are distinguished. The long form is characterised by the present tense morpheme **a** in the verbal structure (Krüger, 2013b:379). This morpheme does not occur in the short form. The long form is used when an OBJ_θ, OBJ, OBL_θ, or ADJUNCT is not included in a sentence. The short form is used when the main verb in the positive is followed by an OBJ or an OBL_θ and/or an ADJUNCT, or when the main verb is in the negative. The verb in (4-33) and the negative verb in (4-34) are in the short form of the present tense while the positive verb in (4-34) has the long form of the present tense.

(4-33) Re bua Seesimane.
 we speak English
re-bu-a *Se-esimane*
 AgrSubjP1pl-speak-VEnd NPre7-english
 We are speaking English.

<p>(4-34) Positive Re a bua. we talk <i>re-a-bu-a</i> AgrSubjP1pl-PresPre-talk-VEnd We are talking.</p>	<p>Negative Ga re bue. not we talk <i>ga-re-bu-e</i> NegPrega-AgrSubjP1pl-talk-VEnd We are not talking.</p>
---	---

The *future tense positive* is marked by the verbal prefix **tla** (4-35), which conveys an "idea of a promise or definiteness" (Krüger, 2013b:398)⁴⁹.

(4-35) Re tla bua Seesimane.
 we shall speak English
re-tla-bu-a *Se-esimane*
 AgrSubjP1pl-FutPre-speak-VEnd NPre7-english
 We shall speak English.

The *future tense negative* is formed by means of the negative morpheme **ga** and an auxiliary verb **ketla** or **nke** (4-36).

(4-36) Ga re ketla re bua Seesimane.
 we shall not we speak English
ga-re-ketla *re-bu-a* *Se-esimane*
 NegPrega-AgrSubjP1pl-Aux AgrSubjP1pl-speak-VEnd NPre7-english
 We shall not speak English.

The *past tense positive* is specified by using the perfect suffix **-il-**. This tense indicates that an action that started in the past was completed in the past (Pretorius, 2003:21)⁵⁰ as is the case in (4-37).

⁴⁹ The future tense in Setswana is referred to as an immediate future tense by Krüger (2013b:381). He states that it "signifies that an action will take place in the immediate future or that it is intended to be executed soon".

⁵⁰ Krüger (2013b:383) refers to the past tense as the immediate past tense.

(4-37) Re buile Seesimane.

we spoke English

re-bu-il-e

Se-esimane

AgrSubjP1pl-speak-PerfSuf-VEnd

NPre7-english

We spoke English.

The *negative* form of a verb in the *past tense* includes two negative prefixes, i.e. a negative morpheme **ga** and its concomitive negative morpheme **a** (4-38).

(4-38) Ga re a bua Seesimane.

did not we did not speak English

ga-re-a-bu-a

Se-esimane

NegPrega-AgrSubjP1pl-NegPrea-VEnd

NPre7-english

We did not speak English.

The *relative tenses* are the relative past and relative future tense. In the relative tenses the action in the verb is defined to a selected and marked reference point which is interpreted to the time that the utterance is made (coding time) (Posthumus, 1990:22; Pretorius, 1997:164). Relative tenses relate the situation to a time that is already established (Nurse, 2008:120). The reference point may be prior to coding time or subsequent to coding time (Diagram 4-2).

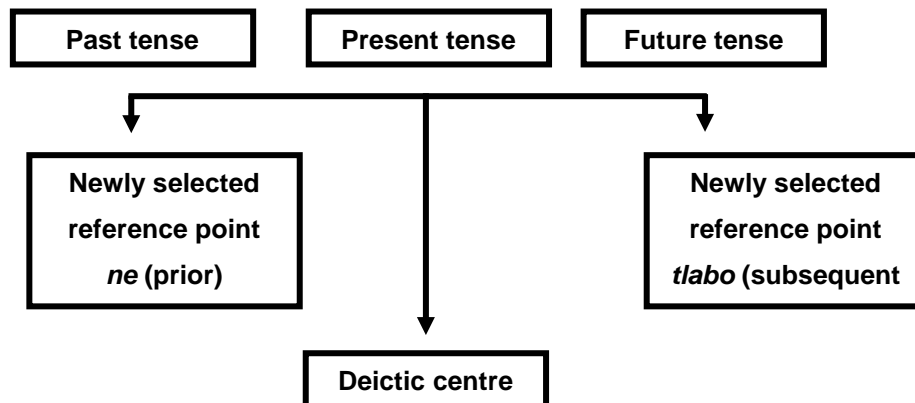


Diagram 4-2: Relative tense (cf. Posthumus, 1990:22; Pretorius, 2003:24)

The *relative past tense* indicates an action that is incomplete and continuing at a certain moment in the past (Krüger, 2013b:396)⁵¹. This tense is formed in the *positive* with the auxiliary verb **ne** (4-39).

⁵¹The relative past tense is also referred to as the remote past tense, the imperfect relative past tense or the past infinitive tense (Pretorius, 1997:249; Krüger, 2013b:396).

(4-39) Re ne re bua Seesimane.
 we were we speak English
re-ne *re-bu-a* *Se-esimane*
 AgrSubjP1pl-Aux AgrSubjP1pl-speak-VEnd NPre7-english
 We were speaking English.

The *negative* of the *relative past tense* is formed by means of the negative morpheme **sa** and the verbal ending **-e** (4-40).

(4-40) Re ne re sa bue Seesimane.
 we were we not speak English
re-ne *re-sa-bu-e* *Se-esimane*
 AgrSubjP1pl-Aux AgrSubjP1pl-NegPresa-speak-VEnd NPre7-english
 We were not speaking English.

The *relative future tense* indicates an action that will be executed or continuing at a given point at some time in the not too near future (Krüger, 2013b:398)⁵². This tense is formed in the *positive* with the auxiliary verb **tlabo** (4-41), which conveys an idea of some more remoteness in the execution of the action (Krüger, 2013b:398).

(4-41) Re tlabo re bua Seesimane.
 we shall we speak English
re-tlabo *re-bu-a* *Se-esimane*
 AgrSubjP1pl-Aux AgrSubjP1pl-speak-VEnd NPre7-english
 We shall be speaking English.

The *negative* of the *relative future tense* is formed by means of the negative morpheme **sa** and the verbal ending **-e** (4-42).

(4-42) Re tlabo re sa bue Seesimane.
 we shall we not speak English
re-tlabo *re-sa-bu-e* *Se-esimane*
 AgrSubjP1pl-Aux AgrSubjP1pl-NegPresa-speak-VEnd NPre7-english
 We shall not be speaking English.

⁵² The relative future tense is also named the remote future tense, the imperfect relative future tense or the future indefinite tense in Setswana (Pretorius, 1997:249; Krüger, 2013b:396).

4.6.1.3 Aspect

Only verbs that appear in the tensed moods include aspectual distinctions in Setswana. Aspect conveys the internal temporal structure of actions or it indicates the duration of an action in a verb (Posthumus, 1993:188; Louwrens, 1994b:139, Pretorius, 1997:170). We distinguish between perfective, progressive and persistive aspect for Setswana.

The relative past tense conveys *perfective aspect* when the action of the complementary verb has been completed with regard to the particular time in the past (Pretorius, 1997:235). The perfective aspect in Setswana is marked morphologically by the verbal extension **-il-**. In (4-43), the action that refers to *write* is a completed action in the past.

(4-43) Mosimane o ne a kwadile teko⁵³.

boy he had he wrote test

mo-simane *o-ne* *a-kwal-il-e*

NPre1-boy AgrSubj1-Aux AgrSubj1-write-PerfSuf-VEnd

(ne)-lek-o

NPre9-try-DevSuf

The boy had written a test.

Progressive aspect represents a situation, as in progress at and around reference time. In other words, it implies that the action is happening at the moment of speech (Pretorius, 1997:230; Nurse, 2008:315). The progressive prefix **sa** (still) expresses this aspect. In (4-44), the action of *playing* is *still being* executed.

(4-44) Bana ba sa tshameka thenese.

children they still play tennis

ba-ana *ba-sa-tshameka-a* *(ne)-thenese*

NPre2-child AgrSubj2-ProgPre-play-VEnd NPre9-tennis

The children still play tennis.

Persistive aspect "affirms that a situation has held continuously since an implicit or explicit point in the past up to the time of speaking" (Nurse, 2008:145). According to Pretorius (1997:230), it is implied that the action has started some time prior to the moment of speech. The persistive in Setswana is marked by **ntse** (have been) or **santse** (still)⁵⁴.

⁵³ Owing to morphophonological alternation, the morpheme sequence **kwal-il-e** results in the surface form **kwadile**.

⁵⁴ The use of **santse** is preferred in Setswana as **sa** is rarely used (Cole, 1955:288),

The auxiliary verb **ntse** (have been) is the past tense of the copulative verb **-nna** (be) (Cole, 1955:288). In (4-45), **ntse** is the auxiliary verb and **thusa** (help) the complementary main verb. In this instance, the meaning indicates an action that has been performed *all along* or *the whole time*.

(4-45) Monna o ntse a ba thusa.

man he has been he them help

mo-nna *o-ntse* *a-ba-thus-a*

NPre1-man AgrSubj1-Aux AgrSubj1-AgrObj2-help-VEnd

The man has been helping them all along (the whole time).

The progressive morpheme **sa** is also used in combination with the auxiliary verb **ntse**. The orthography of this combination is the linguistic word **santse** or **sa ntse**. This auxiliary verb is followed by a complementary main verb. In (4-46), **santse** refers to an action of *playing that is still continuing*.

(4-46) Bana ba santse ba tshameka thenese.

children they still they play tennis

ba-na *ba-sa-ntse* *ba-tshamek-a* *(ne)-thenese*

NPre2-child AgrSubj2-ProgPre-Aux AgrSubj2-play-VEnd NPre9-tennis

The children are still playing tennis.

4.6.1.4 Polarity

Polarity refers to morphological markers in the verb that show that the verb is positive or negative. These markers are the verbal endings and specific prefixes (cf. Table 4-3, p.58). The verbal ending **-a**⁵⁵ indicates positive polarity while the negative prefixes **ga**, **sa**, **se** and **a** as well as the verbal ending **-e** are used to mark negative polarity (cf. §4.6.1.2 and §4.6.3). For example, the present tense positive in (4-47) is indicated with the verbal ending **-a**. In (4-48) the negative is indicated by the negative prefix **ga** and the verbal ending **-e**.

(4-47) Bana ba tshameka thenese.

children they play tennis

ba-ana *ba-tshamek-a* *(ne)-thenese*

NPre2-child AgrSubj2-play-VEnd NPre9-tennis

The children play tennis.

⁵⁵ It should be noted that the root **r-** of the stem **re** and the root **its-** of the stem **itse** also occur with the verbal ending **-e** in the positive. Such stems are labelled as defective stems (Pretorius, 1997:26).

(4-48) Bana ga ba tshameke thenese.
 children they not they play tennis
ba-ana ga-ba-tshamek-e (ne)-thenese
 NPre2-child NegPrega-AgrSubj2-VEnd NPre9-tennis
 The children do not play tennis.

4.6.1.5 Verbal extensions and argument structure

The productive verbal extensions in Setswana are the causative, applicative, reciprocal and passive. These extensions are presented in the so-called CARP default template of the Bantu languages to indicate their fixed order (Hyman, 2002) (cf. (e) below). These inclusions of extensions are suffixed to a verb root, resulting in a new (what we refer to as the extended) verb root. They not only extend and alter the meaning of a verb root, but also affect verb valency. The valency of the inflected verb can be increased, decreased (reduced) or changed (Cole, 1955; Bresnan & Moshi, 1990:147–185; Mchombo, 2007:203–204; Khumalo, 2007:13; Pretorius, R., et al., 2012:203; Khumalo, 2014:145; Chavula, 2016:16). For this reason, it is important to explore the implications of these valency changes of the inflected verb in terms of argument structure to determine their subcategorisation frames systematically. The valency for arguments thus results from the transitivity of the verb root and/or of extension(s) suffixing to it. We apply LMT in the framework of LFG to map a verb stem (lexicon) to its subcategorisation frame (syntax) (cf. §3.3.2). More specifically, we assign the semantic features [-o], [+o], [-r] and [+r] to thematic roles according to the Intrinsic Role Classification (IRC) and the Default Role Classification (DRC). We then use the resulting correspondence matrix (cf. Diagram 3-2, p.34) and, if applicable, the subject and bi-uniqueness conditions are invoked to map the semantic features to the appropriate argument functions (cf. §3.3.2).

(a) Causative

The causative suffix renders the semantic significance of *cause to do*, *make to do*, *help to do* or *assist in doing* to the verb. The SUBJ in a sentence that includes a causative verb (basic root + causative extension) thus causes or brings about an action (Cole, 1955:203; Khumalo, 2007:107–108; Fernando, 2008:116–117). Two phonologically related causative extensions are distinguished in Setswana, i.e. **-is-** and **-y-**. A causative extension changes intransitive verbs to transitive verbs and transitive verbs to ditransitive ones (Krüger, 2006:227, Pretorius, 2014:55).

The inclusion of the causative extension **-is-** changes the intransitive verb **o a taboga** (he runs) in (4-49) to a transitive verb **o tabogisa** (she lets run) in (4-50).

(4-49) Ngwana o a taboga.

child he run

mo-ana

o-a-tabog-a

NPre1-child

AgrSubj1-PresPre-run-VEnd

The child is running.

(4-50) Mosadi o tabogisa ngwana.

woman she run let child

mo-sadi

o-tabog-is-a

mo-ana

NPre1-woman

AgrSubj1-run-CausSuf-VEnd

NPre1-child

The woman lets the child run.

Ngwana (child) is introduced as an agent in (4-49). According to the IRC and DRC in LMT, the semantic features [-o] and [-r] are assigned to the agent and map to SUBJ in (4-51) (cf. §3.3.2).

This mapping is common to all examples that include a SUBJ.

(4-51) **tabog-** (run) < agent >

IRC: [-o]

DRC: [-r]

|

SUBJ

In (4-50) the causative extension introduces **mosadi** (woman) as agent while the agent role of **ngwana** (child) in (4-49) changes to a patient role here. The agent maps to SUBJ, as in (4-51). According to the IRC, the patient obtains the [-r] feature. No DRC is allowed since [-r] has already been assigned. In this instance, the correspondence matrix allows both SUBJ and OBJ (the [-r] column) and we then invoke the bi-uniqueness condition to yield OBJ since SUBJ has already been assigned. Therefore in (4-52), the patient maps to OBJ.

(4-52) **tabogis-** (let run) < agent patient >

IRC: [-o] [-r]

DRC: [-r] |

|
SUBJ SUBJ/OBJ

bi-uniqueness: SUBJ OBJ

A transitive verb that is inflected with the causative extension becomes ditransitive and requires the inclusion of two objects (an OBJ_θ and OBJ). In this instance, the inclusion of the causative

extension **-is-** changes the transitive verb **ba kwala** (they write) in (4-53) to a ditransitive verb **o kwadisa** (she let write) in (4-54).

(4-53) Bana ba kwala teko.

children they write test

<i>ba-ana</i>	<i>ba-kwal-a</i>	<i>(ne)-teko</i>
NPre2-child	AgrSubj2-write-VEnd	NPre9-test

The children are writing a test.

(4-54) Mosadi o kwadisa bana teko.

woman she write let children test

<i>mo-sadi</i>	<i>o-kwal-is-a</i>	<i>ba-ana</i>	<i>(ne)-teko</i>
NPre1-woman	AgrSubj1-write-CausSuf-VEnd	NPre2-child	NPre9-test

The woman lets the children write a test.

In (4-53), the agent is **bana** (children) and the transitive verb **ba kwala** (they write) introduces **teko** (test) as the 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 (4-55), as in (4-52).

(4-55) **kwala-** (write) < agent theme >

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

In (4-54) the causative verb introduces **mosadi** (woman) as the agent, while the agent role of **bana** (children) in (4-53) changes to patient in (4-54) and **teko** (test) remains the theme. The agent and theme thematic roles are mapped as before. By applying the classification roles (IRC and DRC) and the correspondence matrix in (4-56), the patient is assigned the [+o] and [+r] semantic features that map to OBJ_θ, the secondary (indirect) object with θ has the value *patient*. In terms of word order, the secondary object appears nearest to the causative verb.

(4-56) kwadis- (let write)	< agent	patient	theme >
IRC:	[-o]	[+o]	[-r]
DRC:	[-r]	[+r]	
	SUBJ	OBJ _θ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ _θ	OBJ

An intransitive verb that is inflected with the causative extension becomes transitive. As exemplified in (4-57), a causative verb also permits the inclusion of a locative thematic role (Cole, 1955:207). In this example, the transitive verb **o tsenya** (she puts) is formed from the intransitive verb **o tsena** (she enters). The agent is **mosadi** (woman), **buka** (book) is the theme, and the phrase **mo kgetsing** (in the bag) has a locative thematic role. By applying the IRC and DRC in (4-58), the respective semantic features of this locative phrase are [-o] and [+r] and then mapped to OBL_θ, where *θ* is *locative* in this instance. The agent and theme thematic roles are mapped as before.

(4-57) Mosadi o tsenya buka mo kgetsing.

woman she enter let book in bag

mo-sadi

o-tsen-y-a

(ne)-buka

NPre1-woman

AgrSubj1-enter-CausSuf-VEnd

NPre9-book

mo

(ne)-kgets-i-ing

LocPart_{mo}

NPre9-bag-LocSuf

The woman puts the book in the bag.

(4-58) **tseny-** (put) < agent theme locative >

IRC: [-o] [-r] [-o]

DRC: [-r] | [+r]

| | |

SUBJ SUBJ/OBJ OBL_{loc}

bi-uniqueness: SUBJ OBJ OBL_{loc}

Summarising, the causative extension is a valency increasing extension, as it allows for instance, the inclusion of a patient, both a patient and theme or both a theme and locative in a sentence. The causative extension has the effect that a direct object (cf. 4-50), both a secondary and direct object (cf. 4-54), and both a direct object and an oblique (cf. 4-58) can be included in a sentence.

(b) Applicative

An applicative verb (basic root + applicative extension) indicates that the action is carried out on *behalf of, to the advantage of or with respect to something or some place* (Cole, 1955:199; Khumalo, 2007:106–107; Fernando, 2008:116–117). The applicative extension **-el-** has the phonological variants **-ets-** and **-lets-** and it increases the transitivity of a verb: intransitive to transitive and transitive to ditransitive (Pretorius, 2012:211). The inclusion of the applicative extension **-el-** changes the intransitive verb **o a tshameka** (he is playing) in (4-59) to a transitive verb **o tshamekela** (he is playing for) in (4-60).

(4-59) Katlego o a tshameka.

Katlego he play

(-)-Katlego o-a-tshamek-a

NPre1a-Katlego AgrSubj1a-PresPre-play-VEnd

Katlego is playing.

(4-60) Katlego o tshamekela setlhopha sa ntlha (sa sekolo sa rona).

Katlego he play for team of first (of school of us).

(-)-Katlego

o-tshamek-el-a

se-tlhoph-a

NPre1a-Katlego

AgrSubj1a-play-ApplSuf-VEnd

NPre7-choose-DevSuf

sa

ntlha

(sa

se-kolo

sa

rona)

PossPart7

Adv

(PossPart7

NPre7-school

PossPart7

PossProPlpl)

Katlego plays for the first team (of our school).

In (4-59) **Katlego** (Katlego is a person's name) is the agent and maps to SUBJ in (4-61).

(4-61) **tsamek-** (play) < agent >

IRC: [-o]

DRC: [-r]

|
SUBJ

In (4-60), **setlhopa sa ntlha** (first team) is the beneficiary. The beneficiary is classified as [-r] and applying the correspondence matrix and bi-uniqueness condition it then maps to OBJ in (4-62).

(4-62) tsamekel- (play for)	< agent beneficiary >	
IRC:	[-o]	[-r]
DRC:	[-r]	
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

The inclusion of the applicative extension **-el-** changes the transitive verb **o reka** (she buys) in example (4-63) to a ditransitive verb **o rekela** (she buys for) in (4-64).

(4-63) Mosadi o reka dijo.
 woman she buy food
mo-sadi o-rek-a di-j-o
 NPre1-woman AgrSubj1-buy-VEnd NPre10-eat-DevSuf
 The woman buys food.

(4-64) Mosadi o rekela bana dijo.
 woman she buy for children food
mo-sadi o-rek-el-a ba-ana di-j-o
 NPre1-woman AgrSubj1-buy-ApplSuf-VEnd NPre2-child NPre10-eat-DevSuf
 The woman buys food for the children.

In (4-63), **mosadi** (woman), the agent, and **dijo** (food), the theme, are mapped to SUBJ and OBJ in (4-65).

(4-65) rek- (buy)	< agent theme >	
IRC:	[-o]	[-r]
DRC:	[-r]	
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

In (4-64), **bana** (children) has a beneficiary thematic role and takes the semantic features [+o] and [+r] since it is a secondary object ([+o]) in terms of the IRC. The DRC yields a [+r]. Applying the correspondence matrix, it then maps to OBJ_θ in (4-66).

(4-70) tsamael- (walk towards)	< agent locative >	
IRC:	[-o]	[-o]
DRC:	[-r]	[+r]
	SUBJ	OBL _{loc}

The intransitive verb **o a tsamaya** (he walks) in (4-71) changes to a transitive verb **o tsamaela** (he walks for) in (4-72) through the inclusion of the applicative extension **-el-**.

(4-71) Ngwana o a tsamaya.
 child he walk
mo-ana o-a-tsamay-a
 NPre1-child AgrSubj1-PresPre-walk-VEnd
 The child is walking.

(4-72) Ngwana o tsamaela mmaagwe kwa toropong.
 child he walk for his mother to town
mo-ana o-tsamay-el-a (-)-mma
 NPre1-child AgrSubj1-walk-ApplSuf-VEnd NPre1a-mother
wa gagwe kwa (ne)-toropo-ing
 PossPart1a PossProPlsg LocPart_{kwa} NPre9-town-LocSuf
 The child is walking to town for his mother.

In (4-71), **ngwana** (child) is the agent and it maps to SUBJ (4-73).

(4-73) tsamay- (walk) < agent >
IRC: [-o]
DRC: [-r]
SUBJ

In (4-72), **mmaagwe** (his mother) is the beneficiary and the phrase **kwa toropong** (to town) has a locative thematic role. In (4-74), these roles map to OBJ and OBL_{loc} respectively.

(4-74) tsamael- (walk for)	< agent beneficiary locative >		
IRC:	[-o]	[-r]	[-o]
DRC:	[-r]		[+r]
	SUBJ	SUBJ/OBJ	OBL _{loc}
bi-uniqueness :	SUBJ	OBJ	OBL _{loc}

Summarising, the applicative extension is a valency-increasing suffix that changes the argument structure of a verb by introducing any of the following: beneficiary, beneficiary and theme, locative, or beneficiary and locative. The applicative extension thus allows the inclusion of a direct object (cf. 4-60), both a secondary object and a direct object (cf. 4-64), an oblique (cf. 4-68) and both a direct object and an oblique (cf. 4-72).

(c) Reciprocal

The reciprocal extension **-an-** is suffixed to a verb to express a meaning of *each other* or *one another* (4-75) and suggests at least two participants. It therefore commonly occurs with a plural subject (Cole, 1955:210). In Setswana, we broadly distinguish between three different syntactic realisations of these participants, which we illustrate by means of examples.

(i) Plural noun

The participants are represented by a plural noun (as SUBJ) such as **bana** (children) in (4-75)⁵⁶.

(4-75) Bana ba a ratana.
children they love each other
<i>ba-ana</i> <i>ba-a-rat-an-a</i>
NPre2-children AgrSubj2-love-RecSuf-VEnd
The children love each other.

(ii) Coordinate phrase

In (4-76), the SUBJ of the reciprocal verb **ba a ratana** (they love each other) is a coordinate phrase **mosimane le mosetsana** (the boy and girl). Since both the nouns are in class 1, their coordination requires that the subject agreement morpheme takes the plural form associated with class 1, i.e. class 2 **ba**.

⁵⁶ The plural noun can be replaced by a plural pronoun phrase (cf. §5.4), plural possessive particle phrase (cf. §5.5.1) or a plural qualificative particle phrase (cf. §5.5.2).

(4-76) Mosimane le mosetsana ba a ratana.

boy and girl they love each other

mo-simane le mo-setsana ba-a-rat-an-a

NPre1-boy Conj NPre1-girl AgrSubj2-love-RecSuf-VEnd

The boy and girl love each other.

In cases where the participants belong to different noun classes but still exhibit similar semantic features, for example [+human], a coordinate phrase can still be used with an appropriate plural subject agreement morpheme in the reciprocal verb, as **ba-** in example (4-77). Otherwise, a discontinuous reciprocal construction is used (see (iii) below). The notion of coordination and its implication for subject-verb agreement is discussed in detail in §5.9 and §6.4.

(4-77) Mosimane le lekau ba a bonana.

boy and young man see each other

mo-simane le le-kau ba-a-bon-an-a

NPre1-boy Conj NPre5-youn man AgrSubj2-see-RecSuf-VEnd

The boy and the young man see each other.

(iii) **Discontinuous reciprocal construction**

A discontinuous reciprocal construction, in which the one participant is the SUBJ and the other participant is represented by means of a comitative phrase, can also be employed (Khumalo, 2014:148). The comitative phrase in Setswana is represented by an associative particle phrase such as **le wena** (with you) in (4-78). The associative particle phrase consists of the associative particle **le** (with, and) and takes as the complement the phrase that represents the second participant (Cole, 1955:210).

(4-78) Mosimane o dumalana le wena.

boy he agrees together with you

mo-simane o-dumel-an-a le wena

NPre1-boy AgrSubj1-agree-RecSuf-VEnd AssPart PersProP2sg

The boy agrees with you.

We now explore the argument structure of the three syntactic realisations of the reciprocal verb (basic root + reciprocal extension) in Setswana presented above and how these are mapped onto the GFs in order to determine the subcategorisation frame of this verb.

Plural noun (cf. (c) (i) above): The plural noun that indicates the participants of the action in a reciprocal construction is assigned an agent thematic role. According to the IRC and DRC, the semantic features [-o] and [-r] are assigned to the agent and mapped to SUBJ (4-79).

(4-79) ratan- (love each other) < agent >	
IRC:	[-o]
DRC:	[-r]
Subject condition:	SUBJ

For syntactic realisations (ii) and (iii), we consider various approaches as presented in three relevant articles (Hurst 2006; Hurst 2010; Khumalo 2014).

Coordinate phrase (cf. (c) (ii) above): Of the three different approaches that have been proposed in the literature, the first two are valency reducing, while the third is valency preserving. In the valency reducing approaches only one GF is assigned to the two thematic roles (agent and patient). In the first approach, *suppression* is used and in the second approach so-called *argument unification*.

Suppression (4-81) prevents the patient thematic role (4-80) to be mapped to a GF (Hurst, 2006:258)⁵⁷, thereby ensuring that the bi-uniqueness condition is satisfied. The participants, **mosimane le mosetsana** (the boy and the girl), are the agent that maps to SUBJ. An advantage of this approach is that it has already been applied to Chichewa, another Bantu language, where the participants mentioned in a coordinate phrase are also the agent that maps to SUBJ (Mchombo, 1991:16).

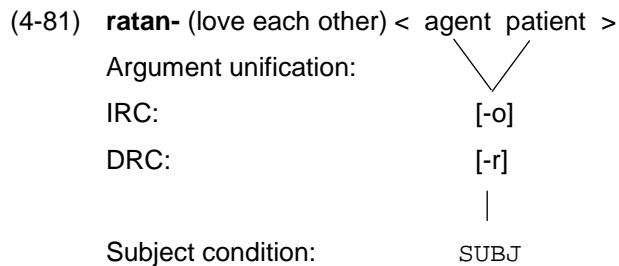
(4-80) ratan- (love each other) < agent patient >		
IRC:	[-o]	[-r]
DRC:	[-r]	
Subject condition:	SUBJ	∅

In the second valency reducing approach, both the agent and patient (or beneficiary)⁵⁸ are mapped simultaneously to one GF, i.e. SUBJ (4-76). The participants are both agents and patients

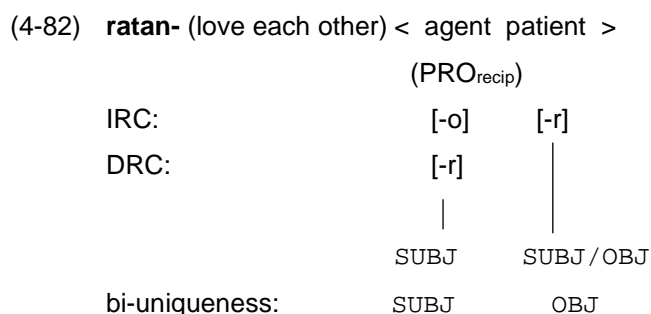
⁵⁷ Falk (2001:111, 2017), Bresnan (2001:339, 341) and Dalrymple (2001:208) describe suppression as a process where a specific argument becomes unavailable for mapping. The ∅ notation is used to indicate the suppression of a thematic role.

⁵⁸ The agent and patient are sometimes referred to as a proto-agent ([P-A]) and a proto-patient (P-P), since some authors prefer to employ Ackerman's (1992) version of LMT. This version of mapping theory incorporates the concept of proto-roles as developed by Dowty (1993).

(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 (4-81). However, a disadvantage is that strictly speaking, argument unification does not form part of LMT.



In the third approach, i.e. a valency preserving approach, the reciprocal morpheme in the morphological structure of the reciprocal verb gives rise to a reciprocal pronoun (PRO_{recip}) (Hurst, 2006:258–259, Khumalo, 2014:156). This pronoun is then assigned the patient thematic role that maps to OBJ (4-82). 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.



In this study, we follow 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.

Discontinuous reciprocal construction (cf. (c) (iii) above): The participants in (4-78) are **mosimane** (boy) and **wena** (you), which occurs as a complement in an obligatory comitative phrase. As before, the agent **mosimane** (boy) maps to SUBJ while, following Hurst (2010:319-324), the comitative phrase 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 and map to OBL_{ass} (4-83).

(4-83) dumalan (agree with each other)	< [agent] >	[]
IRC:	[-o]	[-o]
DRC:	[-r]	[+r]
	SUBJ	OBL _{ass}

(d) Passive

A passive verb is valency reducing, as the OBJ in the active sentence changes to the SUBJ in the passive sentence. In Setswana, the SUBJ of the active sentence functions as the complement of the agentative particle **ke** (by) in an agentative particle phrase (Cole, 1955:192–196; Krüger, 2006:254–259). The passive extension in Setswana is **-iw-** or its variant **-w-**. It can be suffixed to both intransitive and transitive verbs. The transitive verb **o roma** (she sends) in (4-84) changes to the intransitive passive verb **ba romiwa** (they are sent) in (4-85).

(4-84) Mosadi o roma bana.

woman she send children

mo-sadi

o-rom-a

ba-ana

NPre1-woman AgrSubj1-send-VEnd

NPre2-child

The woman sends the children.

(4-85) Bana ba romiwa ke mosadi.

children they send being by woman

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.

In the active sentence in (4-84), **mosadi** (woman) is the agent, and **bana** (children) the patient. These thematic roles are mapped in (4-86) to SUBJ and OBJ respectively.

(4-86) rom- (send)	< agent patient >
IRC:	[-o] [-r]
DRC:	[+r]
	SUBJ SUBJ/OBJ
bi-uniqueness:	SUBJ OBJ

Mosadi (woman) is the agent in (4-85), while **bana** (children) is the patient. The patient maps to SUBJ in (4-87) and the agent thematic role is suppressed (Dalrymple, 2001:208–209; Bresnan, 2016:339–442). The agent in Setswana acts as the complement of the agentative particle **ke** (by) and an ADJUNCT function is assigned to the agentative particle phrase. It is not obligatory to include an agentative particle phrase in a Setswana passive sentence (Cole, 1955: 195–196).

(4-87) romiw- (being send)	< agent patient >
IRC:	[-o] [-r]
DRC:	[-r]
	∅ SUBJ/OBJ
Subject condition:	SUBJ

(e) Sequencing of extensions

In the sections above, we discussed individual extensions (the causative, applicative, reciprocal and passive). It was shown that the individual extensions change the meaning, the argument structure and ultimately the subcategorisation frame of a verb. These individual extensions can also be sequenced. For the purposes of our discussion, we focus on two common approaches to the study and analysis of extension sequences that have been presented in the literature, i.e. the so-called CARP default template of the Bantu languages and compositionality⁵⁹. 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, the CARP template (cf. Table 4-3, p.58) has been the dominant approach, although compositionality is also relevant since extension sequences that do not adhere to CARP may occur in rare cases (Pretorius, 2014).

In the introduction to §4.6.1, we introduced the notion of an extended verb root consisting of a basic verb root and an extension. Generalising this idea, a sequence of two extensions may be

⁵⁹ An in depth discussion of these two approaches falls outside the scope of this study.

seen as an extended root followed by an extension, again resulting in an extended root. For example, **tlisetsa**⁶⁰ (brings for) has a basic root and two extensions, but we may also think of this verb as an extended root **tlis-** (let come) followed by the extension **-el-**. We now explore this phenomenon and its significance for the argument structure resulting from sequences of extensions. We start by applying LMT to explore the argument structure of verbs that include two, three and four verbal extensions. In order to investigate the argument structure of these verbs, we analysed examples, both from a corpus and grammar textbooks, to identify possible trends.

Pretorius (2014) performed a corpus-based analysis of the frequency of use and the sequencing of Setswana verbal suffixes based on statistics, and presents possible combinations concerning the suffixing of the productive suffixes. He used the 67284 orthographic-unit, an annotated NCHLT Setswana corpus that includes 9146 verbs, and identified 1138 verbal suffix combinations in this corpus (Pretorius, 2014:61)⁶¹.

For this first investigation of the argument structure of these verbs in Setswana, we focus on the productive extension sequences, thus sequences formed from productive extensions, viz. the causative, applicative, reciprocal and passive. This indeed reveals important regular behaviour, which we discuss below, using relevant examples.

(i) Causative + Applicative

A causative verb (basic root + causative extension) can in turn be suffixed with an applicative extension resulting in a new extended root. The causative extension adds to the meaning of a *cause to do* to the basic verb root. When an applicative extension is suffixed to the causative verb, a meaning of *doing something for someone is added to the verb* so that a derivation of *cause to do for* is possible (Chavula, 2016:204).

In (4-88), the verb **o tlisa** (she brings) includes the causative extension **-is-** (let). As discussed before, the causative extension requires the inclusion of an OBJ in a sentence. In (4-88), the agent is **mosadi** (woman) and the causative extension introduces a theme, i.e. **dijo** (food). The agent maps to SUBJ and the theme to OBJ (4-89).

⁶⁰ Owing to morphophonological alternation, the morpheme sequence **tl-is-el-a** results in the surface form **tlisetsa**.

⁶¹ Regarding these verbal suffix combinations, 4.42% represents applicative + passive extension sequences, 4.4% causative + passive extension sequences, 0.35% reciprocal + passive extension sequences, 0.7% causative + applicative extension sequences, 0.26% applicative + reciprocal extension sequences and 0.17% applicative + reciprocal + passive extension sequences (Pretorius, 2014:69–70).

(4-88) Mosadi o tliša dijo.

woman she come let food

mo-sadi

o-tl-is-el-a

di-j-o

NPre1-woman

AgrSubj1-come-CausSuf-ApplSuf-VEnd

NPre10-eat-DevSuf

The woman brings the food.

(4-89) **tlis-** (let come) < agent theme >

IRC: [-o] [-r]

DRC: [-r] |
|
SUBJ SUBJ/OBJ

bi-uniqueness: SUBJ OBJ

In (4-90), an applicative extension is suffixed to the causative verb **o tliša** (she brings) resulting in the verb **o tlišetsa** (she brings for). The agent is **mosadi** (woman) and the causative extension introduces a theme, i.e. **dijo** (food), while the applicative extension introduces a beneficiary, **bana** (children) in this case. The agent and theme thematic roles map as before while the beneficiary maps to OBJ_θ in (4-91).

(4-90) Mosadi o tlišetsa bana dijo.

woman she come let for children food

mo-sadi

o-tl-is-el-a

ba-ana

NPre1-woman

AgrSubj1-come-CausSuf-ApplSuf-VEnd

NPre2-child

di-j-o

NPre10-eat-DevSuf

The woman brings food for the children.

(4-91) **tlišets-** (brings for) < agent beneficiary theme >

IRC: [-o] [+o] [-r]

DRC: [-r] [+r] |
| |
SUBJ OBJ_θ SUBJ/OBJ

bi-uniqueness: SUBJ OBJ_θ OBJ

In the following example, (4-92), the verb **o tšenya** (he puts in) includes the causative extension **-y-**. In this example, **rre** (father) is the agent. The causative extension requires both a theme thematic role, i.e. **madi** (money) and a locative thematic role, i.e. the phrase **mo bankeng** (in the bank). The agent maps to SUBJ, the theme to OBJ and the locative to OBJ_{loc} (4-93).

(4-92) Rre o tsenya madi mo bankeng
 father he enter let money in bank
 (-)-rre o-tsen-y-a ma-di
 NPre1a-father AgrSubj1a-enter-CausSuf-VEnd NPre6-money
 mo (ne)-banka-ing
 LocPart_{mo} NPre10-bank-LocSuf
 Father puts money in the bank.

(4-93) **tseny-** (put) < agent theme locative >
 IRC: [-o] [-r] [-o]
 DRC: [-r] | [+r]
 | | |
 SUBJ SUBJ/OBJ OBL_{loc}
 bi-uniqueness: SUBJ OBJ OBL_{loc}

In (4-94), an applicative extension is suffixed to the causative verb resulting in **o tsenyetsa** (he puts for). In this example, the applicative extension necessitates the inclusion of a beneficiary thematic role, i.e. **mme** (mother) and it maps to OBJ_θ (4-95).

(4-94) Rre o tsenyetsa mme madi mo bankeng
 father he enter let for mother money in bank
 (-)-rre o-tsen-is-el-a
 NPre1a-father AgrSubj1a-enter-CausSuf-ApplSuf-VEnd
 (-)-mme ma-di mo (ne)-banka-ing
 NPre1a-mother NPre6-money LocPart_{mo} NPre10-bank-LocSuf
 Father deposits money in the bank for mother.

(4-95) **tsenyets-** (put for) < agent beneficiary theme locative >
 IRC: [-o] [+o] [-r] [-o]
 DRC: [-r] [+r] | [+r]
 | | |
 SUBJ OBJ_θ SUBJ/OBJ OBL_{loc}
 bi-uniqueness: SUBJ OBJ_θ OBJ OBL_{loc}

The causative and applicative extensions are both valency-increasing suffixes. As is evident in (4-90) and (4-94), the causative extension allows the inclusion of a theme (cf. 4-90) or both a theme and a locative (cf. 4-94) while the applicative extension introduces a beneficiary thematic role. The beneficiary then maps to OBJ_θ.

(ii) Causative + Reciprocal

The causative verb may also take a reciprocal extension which adds the meaning of *doing something to each other* (Chavula, 2016:205). In (4-96), the verb **o batlisa** (he seeks) includes a causative extension. In this example, **rre** (father) is the agent, **mme** (mother) the patient, and **dinotlolo tsa koloji** (car keys) the theme. The agent, patient and theme thematic roles are mapped as before (4-97).

(4-96) Rre o batlisa⁶² mme dinotlolo tsa koloji.

father he seek let mother keys of car

(-)-rre	<i>o-batl-is-a</i>	(-)-mme
NPre1a-father	AgrSubj1a- <i>seek-CausSuf-VEnd</i>	NPre1a-mother
<i>di-notlolo</i>	<i>tsa</i>	<i>(ne)-koloji</i>
NPre8-key	PossPart8	NPre10-car

Father helps mother to seek the car keys.

(4-97) batlis- (let seek)	< agent	patient	theme >
IRC:	[-o]	[+o]	[-r]
DRC:	[-r]	[+r]	
	SUBJ	OBJ _θ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ _θ	OBJ

In (4-98), the causative extension allows the inclusion of a theme **dinotlolo tsa koloji** (car keys) in the sentence that maps to OBJ. The reciprocal extension permits a plural denoting the SUBJ, i.e. a coordinate phrase **rre le mme** (father and mother) in this example. They are the agents and this phrase maps to SUBJ (4-99).

(4-98) Rre le mme ba batlisana⁶³ dinotlolo tsa koloji.

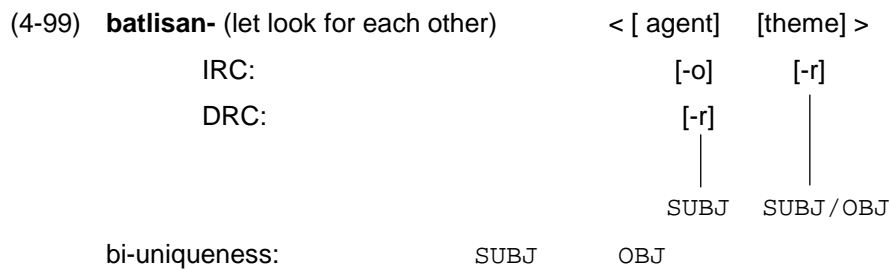
Father and mother they seek let each other keys of car

(-)-rre	<i>le</i>	(-)-mme		
NPre1a-father	Conj	NPre1a-mother		
<i>ba-batl-is-an-a</i>			<i>di-notlolo</i>	<i>tsa</i>
AgrSubj2-want-CausSuf-RecSuf-VEnd			NPre9-key	PossPart9
				<i>(ne)-koloji</i>
				NPre9-car

Father and mother help each other to look for the car keys.

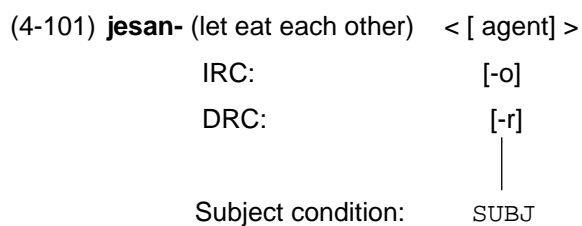
⁶² This is an idiomatic form which means *help to seek* and not the typical *let seek*.

⁶³ This is an idiomatic form which means *help to look for* and not the typical *let look for each other*.



In (4-100), the reciprocal extension requires the inclusion of a plural noun **bana** (children) that maps to SUBJ (101). As was mentioned in 4.6.1.5 (c) above, the reciprocal extension not always requires the inclusion of an OBJ in a sentence.

(4-100) Bana ba a jesana.
 children they eat let each other
ba-ana *ba-a-j-is-an-a*
 NPre2-child AgrSubj2-PresPre-eat-CausSuf-RecSuf-VEnd
 The children feed each other.



The causative extension in (4-98) allows a theme thematic role that maps to OBJ while in (4-100) it is not necessary. In these examples, the reciprocal extension requires a plural denoting SUBJ.

(iii) Causative + Passive

In the active sentence in (4-102), the causative verb **o tabogisa** (he let run) requires the inclusion of a patient thematic role, i.e. **batabogi** (athletes). The agent in the sentence is **mokatisi** (trainer). The agent maps to SUBJ and the patient to OBJ (4-103).

(4-102) Mokatisi o tabogisa batabogi gompieno.
 coach he run let athletes today
mo-katis-i *o-tabog-is-a*
 NPre1-train-DevSuf AgrSubj1-run-CausSuf-VEnd
ba-tabog-i *gompieno*
 NPre2-run-DevSuf Adv
 The coach trains the athletes today.

(4-103) tabogis- (let run)	< agent patient >	
IRC:	[-o]	[-r]
DRC:	[+r]	
	SUBJ	SUBJ/OBJ
bi-uniqueness:	SUBJ	OBJ

The passive sentence in (4-104) includes a causative verb with a passive extension. The patient **batabogi** (athletes) maps to SUBJ (4-105) and the agent thematic role **mokatisi** (coach) is suppressed as before (cf. (e) above). The agent forms the complement of the agentative particle **ke** (by) and, as discussed, the agentative particle phrase is assigned an ADJUNCT function.

(4-104) Batabogi ba tabogisiwa ke mokatisi gompieno.
athletes they run let are by coach today
<i>ba-tabog-i</i> <i>ba-tabog-is-iw-a</i>
NPre2-run-DevSuf AgrSubj2-run-CausSuf-PassSuf-VEnd
<i>ke</i> <i>mo-katis-i</i> <i>gompieno</i>
AgPart NPre1-train-DevSuf Adv
The athletes are trained by the coach today.

(4-105) tabogisiw- (trained)	< agent patient >	
IRC:	[-o]	[-r]
DRC:	[-r]	
	∅	SUBJ/OBJ
Subject condition:		SUBJ

(iv) Applicative + Reciprocal

An applicative verb (basic root + applicative extension) can in turn be suffixed with a reciprocal extension and this combination forms a new extended root. In (4-106), an applicative verb **di kwalela** (they write for) is included. The agent is **ditsala** (friends), **batho** (people) is the beneficiary and **makwalo** (letters) is the theme. These thematic roles map to SUBJ, OBJ_θ and OBJ respectively (4-107).

(4-106) Ditsala di kwalela batho makwalo.

friends write for people letters

di-tsala

di-kwal-el-a

NPre10-friend

AgrSubj10-write-App1Suf-VEnd

ba-tho

ma-kwal-o

NPre2-person

NPre6-write-DevSuf

The friends write letters to the people.

(4-107) **kwalel-** (write for) < agent beneficiary theme >

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

The sequencing of the applicative and reciprocal extensions in **di kwalelana** (they write for each other) (4-108) adds a meaning of *do for each other* to the verb. The reciprocal morpheme does not give rise to a reciprocal pronoun (PRO_{recip}) and therefore a beneficiary thematic role is not allowed. The agent is **ditsala** (friends) and the theme is **makwalo** (letters). The agent maps to SUBJ and the theme to OBJ (4-109).

(4-108) Ditsala di kwalelana makwalo.

friends write for each other letters

di-tsala

di-kwal-el-an-a

NPre10-friend

AgrSubj10-write-App1Suf-RecSuf-VEnd

ma-kwal-o

NPre6-write-DevSuf

The friends write letters to each other.

(4-109) **kwalelan-** (write for each other) < agent theme >

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

(v) **Applicative + Passive**

In the active sentence in (4-110), the applicative verb **o rekela** (she buys for) requires the inclusion of a beneficiary thematic role, i.e. **bana** (children) in the sentence. The agent in the

(4-113) rekelw- (bought for)	< agent beneficiary	theme >
IRC:	[-o] [-r]	[+o]
DRC:	[-r]	[+r]
	∅ SUBJ/OBJ	OBJ
Subject condition:	SUBJ	

In (vi), we explore the argument structure of a verb that includes three verbal extensions and in (vii) a verb that includes four verbal extensions.

(vi) *Applicative + Reciprocal + Passive*

As discussed in (iv) above, a reciprocal extension can be added to an applicative verb (cf. (4-108)). The agent thematic role maps to the SUBJ while the theme thematic role maps to OBJ (cf. (4-109)). A passive extension can in turn be added to this extended root resulting in a new extended root consisting of the basic root, a reciprocal extension as well as the passive extension (4-114). The theme **makwalo** (letters) then maps to the SUBJ while the agent **ditsala** (friends) is suppressed (4-115).

(4-114) Makwalo a kwalelanwa ke ditsala.

letters it write for each are by friends

ma-kwal-o

a-kwal-el-an-w-a

NPre6-write-DevSuf

AgrSubj6-write-ApplSuf-RecSuf-PassSuf-VEnd

ke

di-tsala

AgPart

NPre9-friend

Letters are written to each other by the friends.

(4-115) kwalelanw- (written to each other)	< agent theme >	
IRC:	[-o] [-r]	
DRC:	[-r]	
	∅ SUBJ/OBJ	
Subject condition:	SUBJ	

(vii) *Causative + Applicative + Reciprocal*

In (e) (i) above, an applicative extension is added to a causative verb. A reciprocal extension can in turn be added to this extended root resulting in a new extended root. In (4-94) **rre** (father) is the agent, **mme** (mother) is the beneficiary, **madi** (money) is the theme, and **mo bankeng** (in the bank), the locative. These thematic roles map to SUBJ, OBJ_θ, OBJ and OBL_{loc} (cf. (4-95)). In

example (4-116), a reciprocal extension is added to the extended root. The reciprocal extension necessitates a plural denoting SUBJ and therefore an OBJ_θ is not included in this sentence. **Batho** (people) maps to the SUBJ, **madi** (money) to the OBJ and **mo bankeng** (in the bank) to the OBL_{loc} (4-117).

(4-116) Batho ba tsenyetsana madi mo bankeng.

people they enter let for each other money in bank

ba-tho

ba-tsen-is-el-an-a

NPre2-person

AgrSubj2-enter-CausSuf-App1Suf-RecSuf-VEnd

ma-di

mo

(ne)-banka-ing

NPre6-money

LocPart_{mo}

NPre9-bank-LocSuf

The people put money in the bank for each other.

(4-117) **tsenyetsan-** (put for each other) < agent theme locative >

IRC:

[-o]

[-r]

[-o]

DRC:

[-r]

[+r]

|

|

|

SUBJ

SUBJ/OBJ

OBL_{loc}

bi-uniqueness:

SUBJ

OBJ

OBL_{loc}

(viii) **Causative + Applicative + Reciprocal + Passive**

In (4-116) above, the verb includes a causative, applicative and reciprocal extension. A passive extension can in turn be added to this extended root. In (4-118), **madi** (money) is the theme and **batho** (people) is the complement in the agentive particle phrase **ke batho** (by the people). The causative extension requires the inclusion of a locative thematic role, i.e. **mo bankeng** (in the bank). The theme maps to the SUBJ, the locative to the OBL_{loc}, the agent **batho** (people) is suppressed, and the reciprocal extension does not influence the argument structure (4-1119).

(4-118) Madi a tsenyetsanwa ke batho mo bankeng.

money it enter let for each other by people in bank

ma-di

a-tsen-is-el-an-w-a

NPre6-money

AgrSubj6-enter-CausSuf-App1Suf-RecSuf-PassSuf-VEnd

ke

ba-tho

mo

(ne)-banka-ing

AgPart

NPre2-person

LocPart_{mo}

NPre9-bank-LocSuf

Money is being put in the bank for each other by the people.

(4-119) tseyetsanw- (is being put for each other)	< agent	theme	locative >
IRC:	[-o]	[-r]	[-o]
DRC:	[-r]		[+r]
	∅	SUBJ/OBJ	OBL _{loc}
Subject condition:		SUBJ	OBL _{loc}

(viii) Extension sequences not adhering to CARP

Extension sequences that do not conform to the CARP template are rare. Pretorius (2014) identifies a small number⁶⁵ of examples in the NCHLT corpus with extension sequences that do not conform to the CARP template. However, these examples do not include a reciprocal extension. The suffix **-agan-** that is identified in grammars for Sesotho sa Leboa (Lombard *et al.*, 1985:128, Louwrens *et al.*, 1995:31) as the associative suffix, is also found in the corpus for Setswana. This suffix is semi-productive and is considered as a fossilised part of the verb root. Future corpus-based studies on variations to the known extension sequencing and the associative suffix should be interesting as more corpora become available for Setswana.

(ix) Intensive and completive extensions

Morphologically the causative and applicative extensions appear in reduplicated form in rare cases in Setswana (Cole, 1955:203, 209). The reduplication of these extensions indicates that the action is carried out carefully, intensively or forcefully and it depends on the semantic possibility of the verb to be intensified (Krüger, 2006:243, 246). Semantically and in idiomatic use thus, this reduplicated form no longer indicates the causative or applicative meaning, neither will its influence on argument structure be causative or applicative. Cole (1955:203, 209) refers to the suffix **-isis-** (4-120) as the intensive suffix and to the suffix **-elel-** (4-121) as the perfective suffix (completive suffix).

(4-120) o batla	o batlisisa
he search	he search intensively
<i>o-batl-a</i>	<i>o-batl-isis-a</i>
AgrSubj1-search-VEnd	AgrSubj1-search-IntensSuf-VEnd
he searches	he investigates (he searches carefully)

⁶⁵ Pretorius (2014:61-62, 67-68) identifies seven examples where the order is reciprocal-causative, two examples each of a reciprocal-causative-passive, reciprocal-applicative, and reciprocal-causative-passive order and one example of a reciprocal-causative-applicative order.

(4-121) o gana	o ganelela
he refuse	he refuse completely
<i>o-gan-a</i>	<i>o-gan-elel-a</i>
AgrSubj1-refuse-VEnd	AgrSubj1-refuse-ComplSuf-VEnd
he refuses	he refuses totally (completely)

We summarise: In (a) to (d), specific changes are described regarding the influence of individual extensions on the argument structure of verbs. The same changes are observed in the description of the argument structure of verbs when two, three or four of these individual extensions are sequenced (cf. (e) to (viii)). The causative extension usually introduces a patient, a patient and theme, and a theme and locative thematic roles, and is therefore a valency increasing extension. The applicative extension is also a valency increasing extension as it introduces at least a beneficiary thematic role. However, examples where the applicative also introduces a beneficiary and theme, a locative and a beneficiary, and locative thematic roles, are also observed. The influence of the reciprocal and passive extensions on the argument structure of verbs is the same regarding all the applicable examples, as they are valency decreasing in all the examples.

We conclude that the analysis of the argument structure of verb roots with more than one verbal extension provides evidence that the argument structure of verb roots that have sequences of extensions can be analysed by observing the compositional nature of such a verb root as an extended verb root plus one extension. This approach provides us with a first approximation to the description of the subcategorisation frames of extended Setswana verb roots. However, this topic requires further in-depth investigation.

4.6.2 AUXILIARY VERB

An *auxiliary verb* is dependent on a complement that can be a main verb (cf. §4.6.2), one of the copulative verb phrases (cf. §5.8.2, §5.8.3, §5.8.4) or another auxiliary verb phrase (cf. §5.8.5). The complement expresses the main action (process, event) and the auxiliary verb enriches the meaning of the complement by adding semantic information regarding the progression or completion (Louwrens, 1994a:17). Pretorius⁶⁶ (1997:98) explains that semantically, an auxiliary verb in Setswana expresses a certain type of duration of the action or it expresses the logical time at which the action is executed. He divides auxiliary verbs into proper and improper auxiliary verbs (Pretorius, 1997:259-343). The proper auxiliary verbs usually indicate TAMP information. The improper auxiliary verbs are “proper independent verbs which can also function as auxiliary verbs followed by predicative complements” (Pretorius, 1997:279). They express a semantic value of

⁶⁶ Setshedi (1974) and Pretorius (1997) present a classification and description of the Setswana auxiliary verb as a subcategory of verbs.

continuation or duration, logical time, successive continuation, simultaneous execution or repetition while some of the improper auxiliaries can also have a conjunctive function.

In (4-122) the complement of the auxiliary verb **o ne** (she was busy) is a main verb **a ruta** (she teaches). The auxiliary verb **ne** indicates an action that is incomplete and continuing at a certain moment in the past (Krüger, 2013b:396)⁶⁷.

(4-122) O ne a ruta bana dipalo.

she was she teach children mathematics

o-ne

a-rut-a

ba-ana

AgrSubj1-Aux

AgrSubj1-teach-VEnd

NPre2-child

di-bal-o

NPre10-count-DevSuf

She was teaching the children mathematics.

4.6.3 COPULATIVE VERBS

We distinguish between three copulative verbs, viz. the identifying, describing, and associative copulative verb⁶⁸. These copulative verbs can be used in the same tense forms as those distinguished for main verbs (Krüger, 2013b:422–486).

Identifying copulative verb: The identifying copulative verb has different forms, i.e. **ke**, **-se**, **-le**, **-nna**, **-nne** or **-nnile** (Appendix A: Table 4-10, p.194) which are determined by mood, tense, aspect and polarity (Krüger, 2013b: 140–154). The noun class of the SUBJ has no influence on this form. For example, in (4-123), the identifying copulative verb **ke** is used in the present tense, positive of the indicative mood.

(4-123) Batho ba ke badiri.

people these are workers

ba-tho

ba

ke

ba-dir-i

NPre2-person

DemPro2D1

IdCopV

NPre2-work-DevSuf

These people are workers.

Where the personal pronouns of the first or second person act as the SUBJ, the identifying copulative verb agrees with the person and number of the pronoun (Appendix A: Table 4-11,

⁶⁷ The relative past tense is also named the remote past tense, the imperfect relative past tense or the past indefinite tense (Pretorius, 1997:249; Krüger, 2013b:396).

⁶⁸ Views regarding the identifying and describing copulative verbs in the Sotho languages differ, as some scholars categorise these verbs as particles (Lombard *et al.*, 1985:194; Taljaard & Bosch, 1988; Louwrens, 1991:18; Poulos & Louwrens, 1994:307; Taljard, 2001a, 2001b; Taljard, 2003; Pretorius, 2004:71–73; Krüger 2006:275–278, Krüger, 2013a & 2013b).

p.196). The identifying copulative verb **re** (are) in (4-124) agrees with the personal pronoun **rona** (we).

(4-124) Rona re badiri.

we are workers

<i>rona</i>	<i>re</i>	<i>ba-dir-i</i>
PersProPlpl	IdCopVP1pl	NPre2-work-DevSuf

We are workers.

Describing copulative verb: When a describing copulative verb has a noun as the SUBJ, it exhibits class agreement with that noun (Appendix A: Table 4-12, p.196). The describing copulative verb **o** in example (4-125) agrees with the class 1 noun **mosimane** (boy). Where a personal pronoun is the SUBJ, the verb agrees with it in person and number (Appendix A: Table 4-13, p.196). Describing copulative verbs may also take the form **-se**, **-le**, **-nna**, **-nne** or **-nnile** (Appendix A: Table 4-10, p.194) based on their mood, tense, aspect and/or polarity (Krüger, 2013b:156–173). The describing copulative verbs shows agreement with the SUBJ and these forms therefore take a subject agreement morpheme.

(4-125) Mosimane o mogolo.

boy is big

<i>mo-simane</i>	<i>o</i>	<i>mo-golo</i>
NPre1-boy	DesCopV1	NPre1-adjective

The boy is big.

Associative copulative verb: The associative copulative verb has different forms, i.e. **-na**, **-se**, **-le**, **-nna**, **-nnile** or **-nne** (Appendix A: Table 4-10, p.194) which are determined by mood, tense, aspect and polarity (Krüger, 2013b:178–185). A subject agreement morpheme is prefixed to these forms. In (4-126), the associative copulative verb **ba na** (they have) is in the present tense positive of the indicative mood and the subject agreement morpheme **ba** agrees with the noun **basetsana** (girls).

(4-126) Basetsana ba na le dibuka.

girls they have books

<i>ba-setsana</i>	<i>ba-na</i>	<i>le</i>	<i>di-buka</i>
NPre2-girl	AgrSubj2-AssCopV	AssPart	NPre10-book

The girls have books.

The features that are applicable to Setswana verbs (main, copulative and auxiliary) are presented in §4.6.1, §4.6.2 and §4.6.3. The attributes and values that are applicable in the feature structure of Setswana verbs can therefore include the following information:

- The subcategory information of the verb determines the value of the verb type (VTYPE) feature that is one of {main, idcopv, descopv, asscopv, aux}.
- The agreement feature CLASS takes one of {1, 1a, 2, 2a, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20}.
- The mood feature (MOOD) assumes one of the following values: {ind, cons, imp, hab}.
- The tense feature (TENSE) assumes one of the following values: {pres, past, fut, relpast, relfut}.
- The aspect feature (ASP) has one of the following values: {perf, prog, pers}.
- The polarity feature (POL) has one of the following values: {pos, neg}.

We also include the causative, applicative, reciprocal and passive as features of the main verb, as these verbal extensions have an influence on the argument structure of this verb (cf. §4.6.1). The value of each one of these features is specified with a + notation. The LFG f-structure of the verb **o tabogisa** (he let run) in sentence (4-50) is presented in (4-127).

(4-127) **o tabogisa**

PRED	TABOG	<SUBJ OBJ> '69
VTYPE	main	
CLASS	1	
MOOD	ind	
TENSE	pres	
POL	pos	
CAUS	+	

4.7 PARTICLES

In Setswana no prepositions are distinguished, instead the use of so-called particles expresses the various meanings that are related to prepositions⁷⁰. Particles may or may not exhibit class agreement. The possessive and qualificative particles are class bound while the instrumental, temporal, locative, agentative, associative and comparative particles are not.

The view in this study is that the relative particle distinguished by Krüger (2006:150) is a qualificative particle and that it acts as the head in a specific qualificative particle phrase (cf.

⁶⁹ The subcategorisation frames of main verbs are elaborated on in Chapter 6.

⁷⁰ In Sesotho sa Leboa, prepositions are also not distinguished. Poulus and Louwrens (1994:328) state that the particles in Sesotho sa Leboa express meanings that correlate with the meanings of prepositions.

§5.5.2). Both the *possessive* and *qualificative* particles (Appendix A: Table 4-14, p.197 and Table 4-15, p.198) agree with the preceding noun in the phrase. For example, in (4-128), the class of the possessive particle **la** (of) agrees with the class of the noun **lefatshe** (country) and in (4-129), the qualificative particle **ba** (who are) agrees with the noun **batho** (people).

(4-128) lefatshe la rona

country of us

le-fatshe

la

rona

NPre5-country

PossPart5

PersProPlpl

our country

(4-129) batho ba bantsi

people who are many

ba-tho

ba

ba-ntsi

NPre2-person

QualPart2

NPre2-adjective

many people

The *instrumental particle* **ka** (4-130) conveys the meaning of *with* or *by means of* (Poulos & Louwrens, 1994:328).

(4-130) ka koloji

with car

ka

(ne)-koloji

InstrPart

NPre9-car

with the car

The *temporal particle* **ka** conveys meanings such as *by*, *in*, *at*, *during* or *per* (Poulos & Louwrens, 1994:329). In (4-131), the temporal particle **ka** conveys the meaning of *at* and in (4-132) the meaning of *on*.

(4-131) ka bosigo

at night

ka

bo-sigo

TempPart

NPre14-night

at night

(4-132) ka Matlhatso
 on Saturday
ka *matlhatso*
 TempPart NPre6-wash-DevSuf
 on Saturday

Multiple *locative particles* may be used to convey the meaning of *to*, *at*, *into*, *in*, and *on*. The locative particles **go** and **ga** are used with nouns denoting human beings (Kosch, 2006:120). The locative particle **go** (4-133) expresses the meaning of *with someone personally* or *to*. The locative **ga** (4-134) indicates a meaning of *at someone's place*.

(4-133) go ntate
 to father
go (-)-*ntate*
 LocPart_{go} NPre1a-father
 to father

(4-134) ga ntate
 at father
ga (-)-*ntate*
 LocPart_{ga} NPre1a-father
 at father's place

The *locative particle* **ka** (4-135) is used to convey a meaning of *goes through*, *over*, *out of* or *into* (Krüger, 2006:148).

(4-135) ka motse
 into village
ka *mo-tse*
 LocPart_{ka} NPre3-village
 into the village

We also classify the words **fa**, **mo** and **kwa** as *locative particles* when included in a phrase that indicates locality. This deviates from Krüger (2013a:17, 21, 60, 76), who classifies these words in locative phrases as demonstrative pronouns that modify a subsequent locative noun or locative class noun. A modifier in Setswana usually follows the word that is modified. Our view is in accordance with Poulos and Louwrens (1994:336) and Louwrens (1994a:98-99) who also employ van Wyk's word classification criteria and distinguish **fa**, **mo** and **kwa** as locative particles in Sesotho sa Leboa.

The *locative particle* **fa** (4-136) is used to indicate a meaning of *at (here at)* (Krüger, 2013a:159).

(4-136) fa yunibesiting

at university

fa (ne)-yunibesiti-ing

LocPart_{fa} NPre9-university-LocSuf

at the university

The *locative particle* **mo** (4-137) is used to indicate a meaning of *in, or on (here in, here on)* (Poulus & Louwrens, 1994:336).

(4-137) mo bankeng

in bank

mo (ne)-banka-ing

LocPart_{mo} NPre9-bank-LocSuf

in the bank

The *locative particle* **kwa** (4-138) is used to indicate a meaning of *in or to (there in, there to)* (Poulus & Louwrens, 1994:336).

(4-138) kwa toropong

to town

kwa (ne)-toropo-ing

LocPart_{kwa} NPre9-town-LocSuf

to the town

The *agentative particle* **ke** (4-139) conveys the meaning of *by* in passive sentences (Lombard et al., 1985:173).

(4-139) Mosimane o romilwe ke mosadi.

boy he has been sent by woman

mo-simane *o-rom-il-iw-e* *ke* *mo-sadi*

NPre1-boy AgrSubj1-send-PerfSuf-PassSuf-VEnd AgPart NPre1-woman

The boy has been sent by the woman.

Krüger (2013a:221–223) classifies the word **le** as an *associative particle* and describes the meaning of this word as *together with, with* or *and*. This word is also classified as a conjunction (cf. §4.8). **Le** (with) is classified as an associative particle when it is used as the head in a phrase

that functions as an ADJUNCT in a sentence (4-140) or when it is used as the head of the phrase that follows the associative copulative verb in the positive (4-141).

(4-140) Ke dira le mme

I work with mother

<i>ke-dir-a</i>	<i>le</i>	<i>(-)-mme</i>
AgrSubj1-work-VEnd	AssPart	NPre1a-mother

I work with mother.

(4-141) Mosadi o na le kolo.

woman she is with car

<i>mo-sadi</i>	<i>o-na</i>	<i>le</i>	<i>(ne)-kolo</i>
NPre1-woman	AgrSubj1-AssCopV	AssPart	NPre9-car

The woman has a car.

The word **jaaka** indicates, "that its complement is in a comparative relationship with the action" mentioned in the verb and has a meaning of *like* or *similar to* (Krüger, 2006:149). This word is usually classified as a *comparative particle* in Setswana (Krüger, 2006:149). However, this word can also be used in coordinated structures, in which case it is then classified as a conjunction (cf. §4.8). In (4-142), **jaaka** is used as a comparative particle (Cole, 1955).

(4-142) O bua fela jaaka wena.

he talk just like you

<i>o-bu-a</i>	<i>fela</i>	<i>jaaka</i>	<i>wena</i>
AgrSubj1-talk-VEnd	Adverb	CompPart	PersProP2sg

He talks just like you.

The *hortative particle A* (4-143) is used to express wishes and requests and conveys a meaning of *let* (Krüger, 2013b:115).

(4-143) A nke ke bone!

let me see

<i>a</i>	<i>nke</i>	<i>ke-bon-e</i>
HortPart	Aux	AgrSubjP1sg-see-VEnd

Let me see!

The *interrogative particle A* (4-144) is used to focus the question on the predicate itself (Krüger, 2013b:331, 350).

(4-144) A bana ba tlile?

(question) children they come did

a *ba-ana* *ba-tl-il-e*

IntPart NPre2-child AgrSubj2-come-PerfSuf-VEnd

Did the children come?

The PARTTYPE (particle type) and CLASS features manifest in the feature structure of the Setswana particle as follows:

- A particle has a feature PARTTYPE and can take one of the following values: {PossPart, QualPart, InstrPart, TempPart, LocPart_{go}, LocPart_{ga}, LocPart_{ka}, LocPart_{fa}, LocPart_{mo}, LocPart_{kwa}, AgPart, AssPart, CompPart, HortPart, IntPart}
- The CLASS feature of possessive and qualificative particles is used to indicate agreement phenomena and can take one of the following values: {1, 1a, 2, 2a, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20}.

The LFG f-structure of the possessive particle **la** (of) is presented in (4-145).

(4-145) **la**

PRED	'LA'
PARTTYPE	PossPart
CLASS	5

4.8 CONJUNCTION

A conjunction⁷¹ is used to link words in coordinated structures, as shown for **le** (and), **kgotsa** (or) and **jaaka** (like, such as) in (4-146), (4-147) and (4-148), respectively. Krüger (2013a:219) distinguishes **kgotsa**, **kampo**, **kapa** and **kana** as *alternative particles* and states that they portray a meaning of *or* and *either ... or*. Since these particles are used in coordinated structures, we classify them as conjunctions (Cole, 1955; Le Roux, 2011).

(4-146) rre le mme

father and mother

(-)-rre *le* (-)-mme

NPre1a-father Conj NPre1a-mother

father and mother

⁷¹ Vermeulen (1984) presents a description and exposition of the conjunctions in Setswana.

(4-147) mosetsana kgotsa mosimane

girl or boy

mo-setsana

kgotsa

mo-simane

NPre1-girl

Conj

NPre1-boy

the girl or the boy

(4-148) selo jaaka thipa

thing like knife

se-lo

jaaka

(ne-)thipa

NPre7-thing

CompPart

NPre9-knife

something like a knife

However, a conjunction is also used to link two clauses to form a compound or complex sentence (Louwrens, 1994a:29–33). For example, in (4-149) the conjunction **fa** (when) links the main clause and the subordinate clause.

(4-149) Phefo e a tsen a fa re bula lebati.

wind it come in when we open door

(ne)-phefo

e-a-tsen-a

fa

NPre9-wind

AgrSubj9-PresPre-come in-VEnd

Conj

re-bul-a

le-bati

AgrSubjP1pl-open- VEnd NPre5-door

The wind is coming in when we open the door.

The feature structure of the Setswana conjunction includes a conjunctive form (CONJ-FORM) and the value is the lexical item itself. The LFG f-structure of the conjunction **le** (and) is presented in (4-150).

(4-150) **le** $\left[\begin{array}{ll} \text{PRED} & \text{'LE'} \\ \text{CONJ-FORM} & \text{LE} \end{array} \right]$

4.9 ADVERB

An *adverb*⁷² in Setswana can refer to time, place and manner and is either primitive (basic) or derived (Krüger, 2006:162). An example of a primitive adverb is **jaanong** (now) (4-151). Derived adverbs are formed from nouns, noun roots or verbs by means of an adverbial prefix **ga-**. For example, the adverb **gantsi** (often, many times) is derived from the noun root **-ntsi** (many).

⁷² Le Roux (2007) presents a grammatical analysis of adverbial constructions in Setswana.

(4-151) Re thusa batho jaanong.

we are helping people now

re-thus-a

batho

jaanong

AgrSubjPlpl-help-VEnd

NPre2-person

Adv

We are helping the people now.

The LFG f-structure of the adverb **jaanong** is presented in (4-152).

(4-152) **jaanong** [PRED 'JAANONG']

4.10 INTERJECTION

The interjection is an interposing remark that is used to express reactions and emotions such as joy, approval, disapproval, sorrow, distress, fear, disgust, agreement, disagreement, surprise, disbelief, and to draw attention or to call or urge on animals (Krüger, 2006:165–166). In (4-153); the interjection **heei** (hey) is used to draw attention.

(4-153) Heei, nthuse!

hey, me help

heei

ni-thus-e

Interj

AgrObjPlsg-help-VEnd

Hey, help me!

The LFG f-structure of the interjection **heei** is presented in (4-154).

(4-154) **heei** [PRED 'HEEI']

4.11 IDEOPHONE

An ideophone is a highly expressive word that is often used for sound symbolism regarding colour, sound, smell, taste, feeling, etc. (Louwrens, 1991:79). The ideophone in Setswana functions as an exclamation or as a complement of the verb **-re** (say, mean) (Ras, 1991:108-109). The ideophone **tu** in (4-155) enhances the meaning of *quiet* in the verb **ba didimala** (they kept quiet) to *absolutely quiet*. The sentence in (4-156) contains a main clause as well as a subordinate clause. The ideophone **tuu** in this subordinate clause is a complement of the verb **a re** (he says) (Ras, 1991:111).

(4-155) Ba didimala tu.

they kept quiet absolutely

ba-didimal-a tu

AgrSubj2-quiet-VEnd Ideo

They kept absolutely quiet.

(4-156) A didimala a re tuu.

He kept quiet he said nothing

a-didimal-a a-r-e tuu

AgrSubj1-quiet-VEnd AgrSubj1-say-VEnd Ideo

He kept quit and said absolutely nothing.

The LFG f-structure of the interjection **tuu** is presented in (4-157).

(4-157) **tuu** [PRED 'TUU']

4.12 SUMMARY

In this chapter, we presented a brief historic overview of the study of Setswana linguistic description. We presented a brief exposition of the word categories of Setswana, viz. nouns, pronouns, verbs, particles, conjunctions, interjections and ideophones and followed LFG to summarise and illustrate, with Setswana examples, the distinctive feature structure of each word category. These word categories are the building blocks of the proposed Setswana phrases as presented in Chapter 5. We furthermore applied LMT to discuss the argument structure of the main verb with a causative, applicative, reciprocal, and passive extension as well as the sequencing of these extensions. The analysis of the argument structure of extended verb roots in Setswana (cf. §4.6.1.5) is novel and forms an important contribution to the description of the subcategorisation frames of verbs (cf. Chapter 6) that are necessary for the computational grammar (cf. Chapter 7).

CHAPTER 5

SETSWANA PHRASES

5.1 INTRODUCTION

The phrase structure of Setswana, a head-initial fixed word order language, is based on its word categories and their subcategories, as discussed in Chapter 4. A study of phrase structure is a next step towards explicating the structure of the simple sentence in LFG in Chapter 6. All word categories and their subcategories can function as heads of a phrase. To determine the nature of Setswana phrases, it is important to determine whether a head (word) can function on its own as a phrase or if it needs a specific complement or modifier. In §5.2 to §5.8, we discuss this aspect for each word category and subcategory. We present the head, obligatory complements and possible modifiers for each phrase. We also clarify the agreement phenomena that govern the phrase. In conclusion, we discuss coordination and the sequencing of modifiers. In order to facilitate this discussion, we provide a complete list of proposed Setswana phrases in Table 5-1 (p.105).

Head word		Phrase	Abbreviation
noun		class noun phrase	CLNP
locative noun		locative noun phrase	LOCNP
locative class noun		locative class noun phrase	LOCCLNP
pronoun		pronoun phrase	PROP
adverb		adverb phrase	ADVP
interjection		interjection phrase	INTERJP
ideophone		ideophone phrase	IDEOP
particles	possessive particle	possessive particle phrase	POSSPARTP
	qualificative particle	qualificative particle phrase	QUALPARTP
	locative particle go	locative particle phrase go	LOCPARTP _{go}
	locative particle ga	locative particle phrase ga	LOCPARTP _{ga}
	locative particle ka	locative particle phrase ka	LOCPARTP _{ka}
	locative particle fa	locative particle phrase fa	LOCPARTP _{fa}
	locative particle mo	locative particle phrase mo	LOCPARTP _{mo}
	locative particle kwa	locative particle phrase kwa	LOCPARTP _{kwa}
	instrumental particle	instrumental particle phrase	INSTRPARTP
	temporal particle	temporal particle phrase	TEMPPARTP
	agentative particle	agentative particle phrase	AGENTPARTP
	associative particle	associative particle phrase	ASSPARTP
	comparative particle	comparative particle phrase	COMPPARTP
verbs	main verb	main verb phrase	VPMAIN
	auxiliary verb	auxiliary verb phrase	VPAUX
	identifying copulative verb	identifying copulative verb phrase	VPIDCOP
	describing copulative verb	describing copulative verb phrase	VPDESCOP
	associative copulative verb	associative copulative verb phrase	VPASSCOP

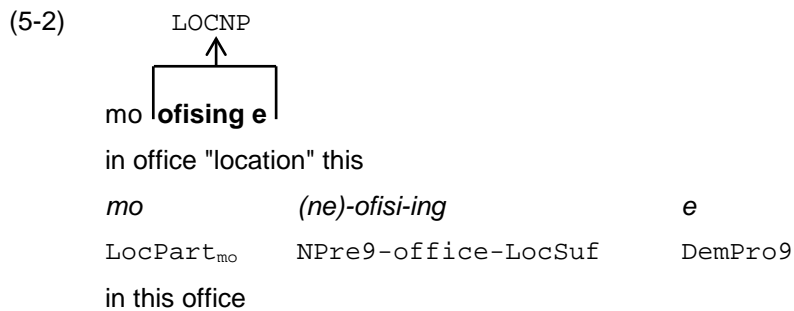
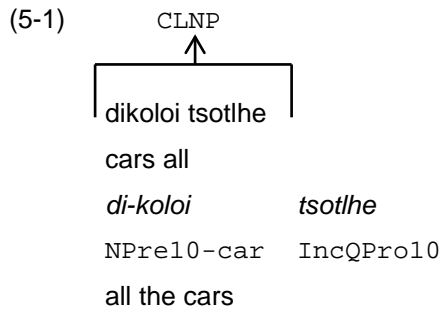
Table 5-1: Word categories, subcategories and corresponding Setswana phrases

5.2 CLASS NOUN AND LOCATIVE NOUN PHRASE

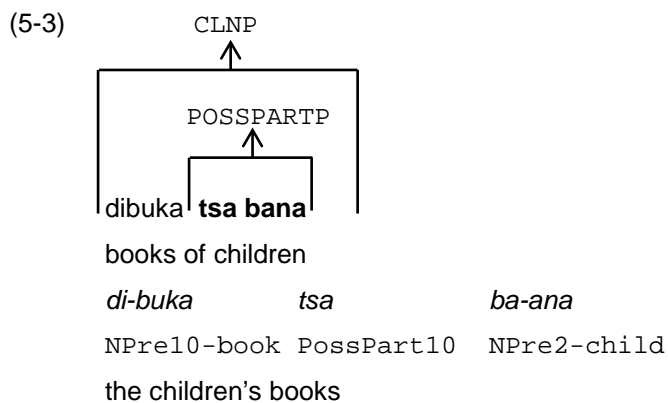
The class noun phrase (CLNP)⁷³ may consist of a (head) noun only or a head noun followed by one or more modifiers. Similarly, the locative noun phrase (LOCNP) consists of a locative noun or a locative noun followed by one or more modifiers. Words or phrases can modify nouns and locative nouns (Table 5-2, p.107). Words that may act as modifiers are the absolute, demonstrative, inclusive quantitative, exclusive quantitative, separative quantitative, selective quantitative or interrogative quantitative pronouns. Class gender agreement between the head

⁷³ This phrase can also be classified as a noun phrase (NP). However, we do not classify this phrase as an NP as the NP in a Setswana sentence is represented by this CLNP (§5.2), a PROP (§5.4), a POSSPARTP (§5.5.1) or a QUALPARTP (§5.5.2).

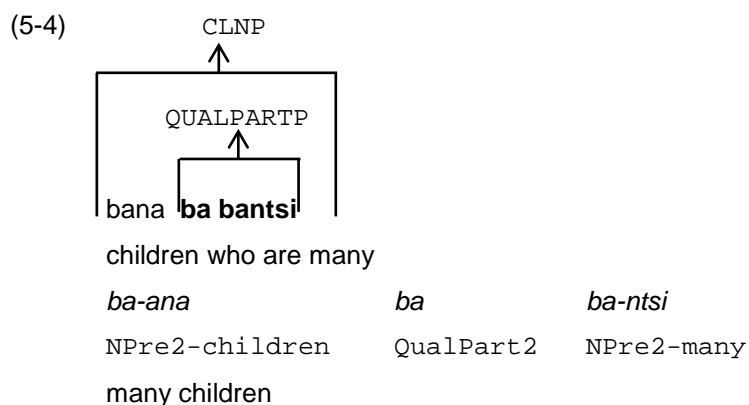
noun and its modifier and between the head locative noun and its modifier is imperative. Example (5-1) shows the class 10 noun **dikoloi** (cars), followed by its modifier, the class 10 inclusive quantitative pronoun **tsotlhe** (all). In (5-2), the LOCNP consists of the class 9 locative noun **ofising** (office "location") and the class 9 demonstrative pronoun **e** (this).



Nouns and locative nouns may be modified by the possessive particle phrase (POSSPARTP) (cf. §5.5.1) or the qualificative particle phrase (QUALPARTP) (cf. §5.5.2). Only the possessive particle in the POSSPARTP agrees with the class of the modified word. In (5-3), the POSSPARTP, modifying the noun **dibuka** (books), consists of the possessive particle **tsa** (of) and the noun **bana** (children). The possessive particle **tsa** exhibits class agreement with the head noun.



Agreement between a noun and the *QUALPARTP* is exemplified in (5-4). The *QUALPARTP* in (5-4) consists of a qualificative particle as head and its complement in the form of a changeable adjective. The qualificative particle **ba** (who are) has as the complement **bantsi** (many) with the root **-ntsi**, and class prefix **ba-** that exhibits class agreement with the qualificative particle. In turn, the qualificative particle **ba** (who are) exhibits class agreement with the head noun **bana** (children). Where the complement is an ordinary adjective, no class agreement occurs (cf. §4.4).



Examples (5-3) and (5-4) show the inclusion of one modifier in a *CLNP*. However, modifiers in *CLNPs* may also be stacked. This is discussed in more detail in §5.10.

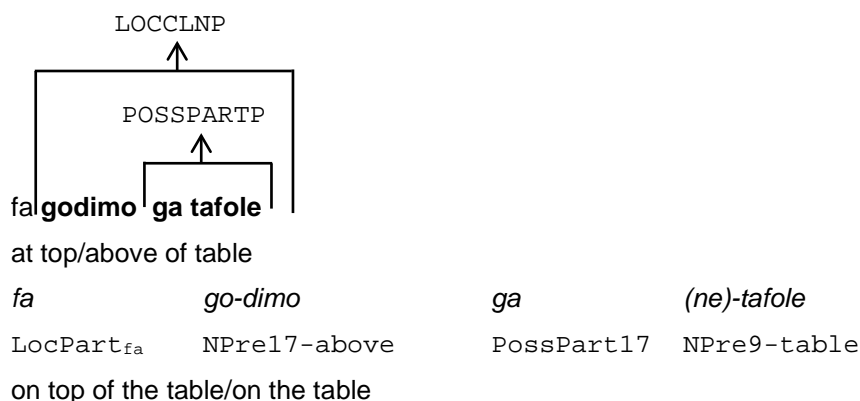
Head	Modifier	Agreement
noun or locative noun	absolute pronoun	The class of the modifying pronoun agrees with the class of the head word.
	demonstrative pronoun	
	inclusive quantitative pronoun	
	exclusive quantitative pronoun	
	separative quantitative pronoun	
	selective quantitative pronoun	
	interrogative quantitative pronoun	
	POSSPARTP	The class of the possessive particle in the <i>POSSPARTP</i> agrees with the class of the head word.
	QUALPARTP	The class of the qualificative particle and the class prefix of the changeable adjective in the <i>QUALPARTP</i> agrees with the class of the head word.

Table 5-2: *CLNP* and *LOCNP* structure and agreement

5.3 LOCATIVE CLASS NOUN PHRASE

The head in the LOCCLNP is a locative class noun (cf. §4.4) and may be followed by a POSSPARTP (cf. §5.5.1) as modifier. The POSSPARTP in this instance always includes the possessive particle **ga** of noun classes 16 to 20, which agrees with the class of the preceding locative class noun. In (5-5) the LOCCLNP consists of a class 17 locative class noun **godimo** (above), modified by the POSSPARTP **ga tafole** (of table), in which the possessive particle **ga** (of) is of class 17.

(5-5)



5.4 PRONOUN PHRASE

A pronoun phrase (PROP) can consist of only a pronoun as head or of a head pronoun with a modifier. An inclusive quantitative pronoun, ADVP, POSSPARTP or a QUALPARTP can modify the personal pronoun. In Table 5-3 to Table 5-6 (p.108-109), the words and phrases that can function as modifiers of absolute, demonstrative, inclusive quantitative and exclusive quantitative pronoun, respectively, are presented.

Head	Modifier	Agreement
absolute pronoun	demonstrative pronoun	The class of the modifying pronoun agrees with the class of the absolute pronoun.
	inclusive quantitative pronoun	
	selective quantitative pronoun	
	separative quantitative pronoun	
	POSSPARTP	The class of the possessive particle agrees with the class of the absolute pronoun.
	QUALPARTP	The class of the qualificative particle and the class prefix of the changeable adjective in the QUALPARTP agree with the class of the absolute pronoun.

Table 5-3: PROP structure and agreement with absolute pronoun as head

Head	Modifier	Agreement
demonstrative pronoun	inclusive quantitative pronoun	The class of the modifying pronouns agree with the class of the demonstrative pronoun.
	selective quantitative pronoun	
	separative quantitative pronoun	
	POSSPARTP	The class of the possessive particle agrees with the class of the demonstrative pronoun.
	QUALPARTP	The class of the qualificative particle and the class prefix of the changeable adjective in the QUALPARTP agrees with the class of the demonstrative pronoun.

Table 5-4: PROP structure and agreement with demonstrative pronoun as head

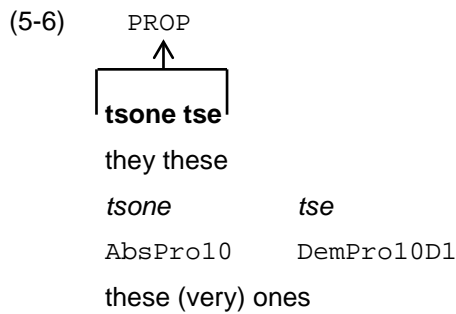
Head	Modifier	Agreement
inclusive quantitative pronoun	POSSPARTP	The class of the possessive particle agrees with the class of the inclusive quantitative pronoun.
	QUALPARTP	The class of the qualificative particle and the class prefix of the changeable adjective in the QUALPARTP agrees with the class of the inclusive quantitative pronoun.

Table 5-5: PROP structure and agreement with inclusive quantitative pronoun as head

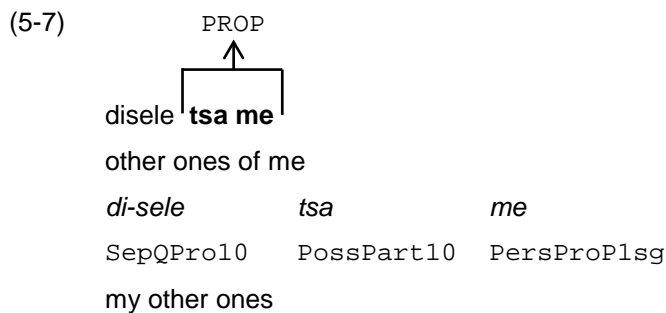
Head	Modifier	Agreement
exclusive quantitative pronoun	POSSPARTP	The class of the possessive particle agrees with the class of the exclusive quantitative pronoun.
	QUALPARTP	The class of the qualificative particle and the class prefix of the changeable adjective in the QUALPARTP agrees with the class of the exclusive quantitative pronoun.

Table 5-6: PROP structure and agreement with exclusive quantitative pronoun as head

By way of example, the PROP in (5-6) consists of the absolute pronoun **tsone** (they), followed by a demonstrative pronoun **tse** (these), as modifier.



A *separative quantitative pronoun* can be modified by a POSSPARTP (cf. §5.5.1). The class of the possessive particle in the POSSPARTP agrees with the class of the separative quantitative pronoun. Similarly, an *interrogative quantitative pronoun* is modified by a POSSPARTP (cf. §5.5.1). The class of the possessive particle agrees with the class of the interrogative quantitative pronoun. The PROP in (5-7) consists of a separative quantitative pronoun **disele** (other ones) as head, modified by the POSSPARTP **tsa me** (of me). The possessive particle **tsa** (of) exhibits class agreement with the separative quantitative pronoun **disele**.



5.5 PARTICLE PHRASES

A particle phrase consists of a specific particle as head word, followed by a complement. We distinguish the following particle phrases: POSSPARTP, QUALPARTP, INSTRPARTP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{fa}, LOCPARTP_{mo}, LOCPARTP_{kwa}, TEMPPARTP, AGPARTP, ASSPARTP and COMPPARTP.

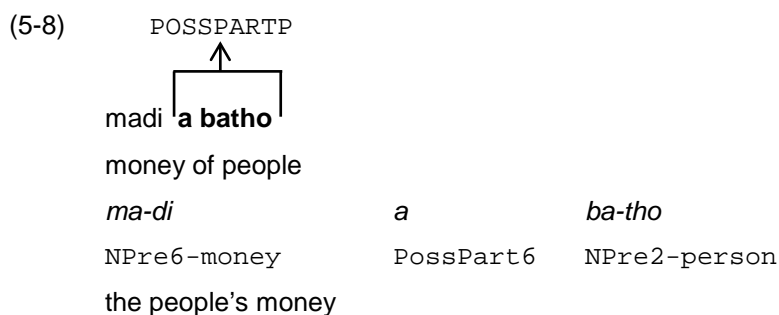
5.5.1 POSSESSIVE PARTICLE PHRASE

The POSSPARTP consists of a possessive particle (Appendix A: Table 4-14, p.197) and its possible complements presented in Table 5-7, p.111). The PROP is not included in this table as a complement of the possessive particle, as personal and absolute pronouns cannot be complements of possessive particles.

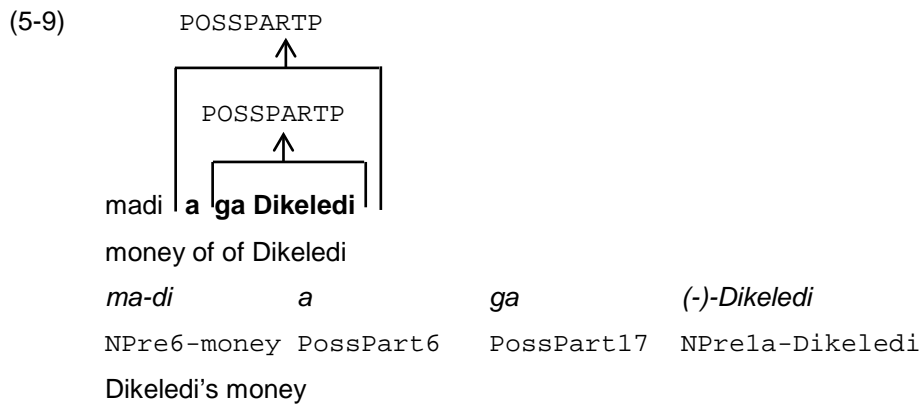
Head	Complement	Agreement
possessive particle	demonstrative pronouns	There is no agreement between the possessive particle and its complement.
	possessive pronouns	
	inclusive quantitative pronouns	
	exclusive quantitative pronouns	
	selective quantitative pronouns	
	separative quantitative pronouns	
	interrogative quantitative pronouns	
	ADVP	
	CLNP	
	QUALPARTP	
POSSPARTP	There is no agreement between the possessive particle in this group and the preceding possessive particle.	

Table 5-7: POSSPARTP structure and agreement

In (5-8), the POSSPARTP **a batho** (of people) consists of the possessive particle **a** (of) and the noun **batho** (people). The POSSPARTP modifies the noun **madi** (money). The possessive particle exhibits class agreement with the preceding noun.



In instances where the possessor refers to kinship nouns, proper names and the interrogatives **mang?** (who? - singular) and **bomang?** (who? - plural), the complement of the possessive particle is another POSSPARTP (Table 5-7, p.111) (Cole, 1955:161; Krüger, 2013a:111). These words function as complements of the class 16 to 20 possessive particle **ga** (of). In (5-9), the complement of the possessive particle **a** (of) is a POSSPARTP consisting of the possessive particle **ga** (of) and the proper name **Dikeledi**.



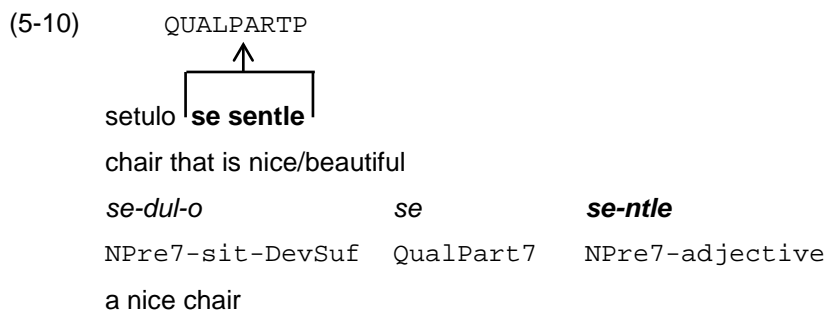
5.5.2 QUALIFICATIVE PARTICLE PHRASE

A *QUALPARTP* consists of a qualificative particle (Appendix A: Table 4-15, p.198) followed by a complement as presented in Table 5-8.

Head	Complement	Agreement
qualificative particle	adjective	There is no agreement between an ordinary adjective and the qualificative particle. The class of the changeable adjective agrees with the class of the qualificative particle.
	relative verb	The class of the complement agrees with the class of the qualificative particle.
	identifying copulative relative verb	
	describing copulative relative verb	
	associative copulative relative verb	
ADVP	There is no agreement between the complement and the qualificative particle.	

Table 5-8: *QUALPARTP* structure and agreement

The *QUALPARTP* in (5-10) consists of a qualificative particle **se** (that is) followed by the adjective **sentle** (nice) as the complement.



5.5.3 INSTRUMENTAL PARTICLE PHRASE

The *INSTRPARTP* includes the instrumental particle **ka** (with) as head followed by a complement, as presented in Table 5-9.

Head	Complement	Agreement
instrumental particle	Interrogative eng?	There is no agreement between the complement and the instrumental particle.
	ADVP	
	CLNP	
	PROP	
	POSSPARTP	
	QUALPARTP	

Table 5-9: *INSTRPARTP* structure and agreement

The *INSTRPARTP* in (5-11) consists of the instrumental particle **ka** (with) followed by the noun **koloji** (car) as the complement.

(5-11) *INSTRPARTP*

↑
ka koloji

with car

ka *(ne)-koloji*

InstrPart NPre9-car

with the car

5.5.4 LOCATIVE PARTICLE PHRASE

The *LOCPARTP* is headed by the locative particles **go**, **ga**, **ka**, **fa**, **kwa** or **mo** each of which is followed by its own distinctive complements. There is no class agreement between the complement and the locative particle.

The *LOCPARTP_{go}* consists of the locative particle **go** (with someone personally, to) as head, followed by a *PROP* or a *CLNP* as complement and, in this instance, the complement belongs to class 1a or 2a. In (5-12), the locative particle **go** (to) is followed by a noun in class 1a.

(5-12) $LOCPARTP_{go}$

↑
┌ **go mme** ─┐

to mother

go (-)-*mme*

$LocPart_{go}$ $NPre1a$ -mother

to mother

The $LOCPARTP_{ga}$ consists of the locative particle **ga** (at someone's place, at) as head, followed by a $PROP$ or a $CLNP$ as complement belonging to class 1a or 2a. In (5-13), the locative particle **ga** (at someone's place) is followed by a noun in class 1a.

(5-13) $LOCPARTP_{ga}$

↑
┌ **ga malome** ─┐

to uncle

ga (-)-*malome*

$LocPart_{ga}$ $NPre1a$ -uncle

to uncle's place

The $LOCPARTP_{ka}$ consists of the locative particle **ka** (goes through, over, out of, into) as head, followed by a $CLNP$, a $LOCPARTP_{fa}$, $LOCPARTP_{mo}$ or $LOCPARTP_{kwa}$ as complement. In (5-14) the locative particle **ka** (into) is followed by a noun.

(5-14) $LOCPARTP_{ka}$

↑
┌ **ka motse** ─┐

into village

ka *mo-tse*

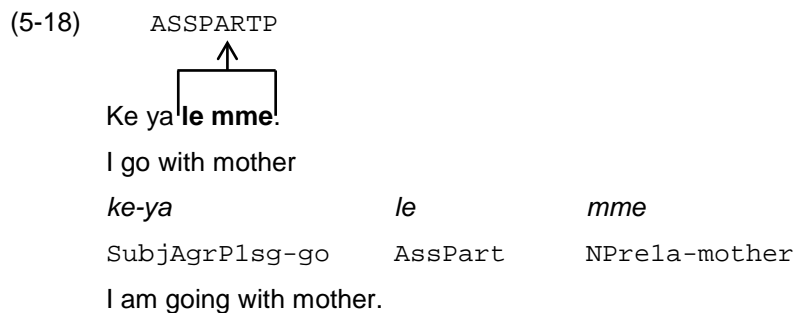
$LocPart_{ka}$ $NPre3$ -village

into the village

The head in a $LOCPARTP_{mo}$, is **mo**, in the $LOCPARTP_{fa}$ **fa** and in the $LOCPARTP_{kwa}$ **kwa**, and these heads can be followed by a $LOCNP$, a $LOCCLNP$, a $LOCPARTP_{go}$ or a $LOCPARTP_{ga}$ as complement. For example, in (5-15), the locative particle **mo** (in) is followed by **lebenkeleng** (shop "location"), a locative noun.

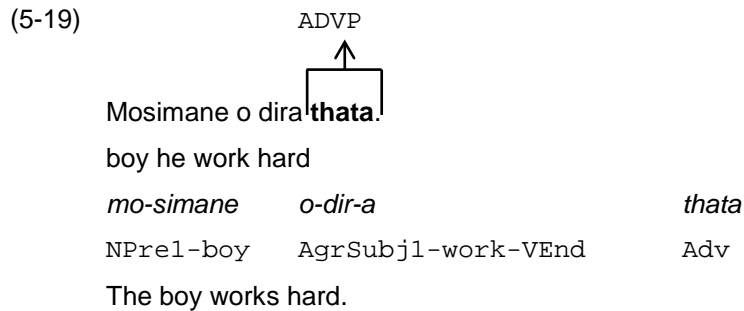
5.5.7 ASSOCIATIVE AND COMPARATIVE PARTICLE PHRASES

The ASSPARTP consists of an associative particle **le** (with) as head, followed by a CLNP, PROP, ADVP, POSSPARTP, QUALPARTP, INSTRPARTP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{mo}, LOCPART_{fa} or LOCPARTP_{kwa} as the complement. There is no agreement between the complement and the preceding associative particle. The structure of the COMPPARTP is similar to that of the ASSPARTP, except that it has the comparative particle **jaaka** (similar to) as head. In the ASSPARTP in (5-18), the complement of the associative particle **le** (with) is the noun **mme** (mother).




5.6 ADVERB PHRASE


As shown in (5-19), an ADVP consists of only an adverb (cf. §4.9).



5.7 INTERJECTION AND IDEOPHONE PHRASES

Interjection phrases (INTERJP) (5-20) and the ideophone phrases (IDEOF) in a simple sentences include only the interjection or ideophone (5-21) (cf. §4.10 and §4.11).

(5-20) INTERJP

Heei! nthuse!
 hey, me help
heei *ni-thus-e*
 Interj AgrObjPlsg-help-VEnd
 Hey, help me!

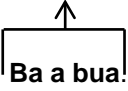
(5-21) IDEOP

 Ba didimala **tu**.
 they kept quiet absolutely
ba-didimal-a tu
 AgrSubj2-quiet-VEnd Ideo
 They kept absolutely quiet.

5.8 VERB PHRASES

As discussed in §4.6, the verb in Setswana is divided into the following subcategories: main verb, identifying copulative verb, describing copulative verb, associative copulative verb and auxiliary verb. Each of these subcategories gives rise to a specific verb phrase structure, i.e. a VP for the main verb, VPIDCOP for the identifying copulative verb, VPDESCOP for the describing copulative verb, VPASSCOP for the associative copulative verb and VPAUX for the auxiliary verb. In subsequent sections, we address the various structures in some detail, focussing on the minimal structure. The full complexity of these structures forms the core of the simple sentence as discussed in Chapter 6.

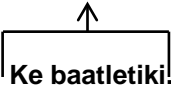
5.8.1 VERB PHRASE WITH A MAIN VERB

The minimal structure of a main verb phrase (VP_{MAIN}) consists of a main verb of which the (complex) morphological structure is presented in §4.6.1. The VP_{MAIN} in (5-22) consists of the main verb **ba a bua** (they talk). This example illustrates that constructions such as **ba a bua** can be considered as a verb, a VP_{MAIN} and even a minimal structure of a sentence. It is not obligatory for an overt SUBJ to be included in a Setswana sentence, but if the SUBJ is omitted, the subject agreement morpheme provides a pronominal interpretation of the missing argument (cf. §6.5).

(5-22) VPMAIN

Ba a bua.
 they talk
ba-a-bu-a
 AgrSubj2-PresPre-talk-VEnd
 They are talking.

5.8.2 VERB PHRASE WITH AN IDENTIFYING COPULATIVE VERB

The minimal identifying copulative verb phrase (VPIDCOPV) consists of an identifying copulative verb (cf. §4.6.3) as head, followed by a non-verbal, non-locative complement in the form of a CLNP, PROP, POSSPARTP or QUALPARTP. The VPIDCOP in (5-23) consists of the identifying copulative verb **ke** (is) and the noun **baatletiki** (athletes).

(5-23) VPIDCOP

Ke baatletiki.
 is athletes
ke *ba-atletiki*
 IdCopV NPre2-athlete
 It is (the) athletes.

5.8.3 VERB PHRASE WITH A DESCRIBING COPULATIVE VERB

The minimal describing copulative verb phrase (VPDESCOP) consists of a describing copulative verb (cf. §4.6.3) as head (Appendix A: Table 4-8 and Table 4-9, p.193), followed by a non-verbal complement. This complement is one of the following:

- Adjectives
- LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{mo}, LOCPARTP_{fa}, LOCPARTP_{kwa}
- LOCCLNP
- Pronouns in classes 16 to 20 (absolute, demonstrative and quantitative pronouns)
- Adverbs of place (place names)
- Interrogatives **kae?**, **jang?** and **go mang?**
- POSSPARTP indicating class 16 to 20.

The VPDESCOP in (5-24) consists of the describing copulative verb **a** (is), followed by a LOCPARTP_{mo} **mo setsidifatsing** (in the refrigerator).

(5-24) VPDESCOP

↑

A mo setsidifatsing.

is in refrigerator "location"

a *mo* *se-tsidifatsi-ing*

DesCopV6 LocPart_{mo} NPre7-fridge-LocSuf

It is in the refrigerator.

5.8.4 VERB PHRASE WITH AN ASSOCIATIVE COPULATIVE VERB

The minimal associative copulative verb phrase (VPASSCOP) consists of an associative copulative verb (cf. §4.6.3) as head, followed in the positive by an ASSPARTP (cf. §5.5.7). In the negative, the complement is a CLNP, PROP, ADVP, POSSPARTP or QUALPARTP. The associative copulative verb **o na** (he has) in (5-25) is positive and is therefore followed by an ASSPARTP **le koloi** (with a car).

(5-25) VPASSCOP

↑

O na le koloi.

he is with car

o-na *le* *(ne)-koloi*

AgrSubj1-AssCopV AssPart NPre9-car

He has a car.

In (5-26), the polarity of the associative copulative verb **ga a na** (he doesn't have) is negative and in this instance it is followed by a noun **koloi** (a car).

(5-26) VPASSCOP

↑

Ga a na koloi.

not he has car

ga-a-na *(ne)-koloi*

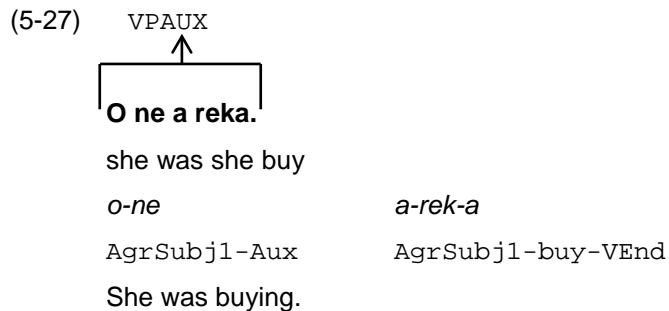
NegPrega-AgrSubj1-AssCopV NPre9-car

He does not have a car.

5.8.5 VERB PHRASE WITH AN AUXILIARY VERB

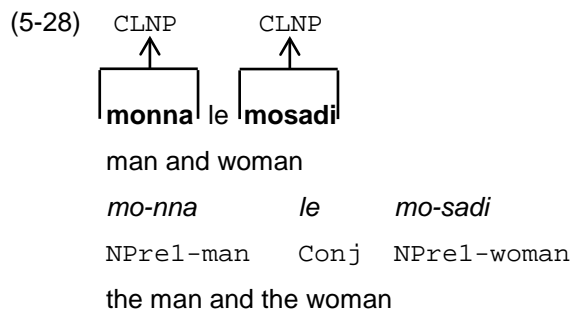
The minimal auxiliary verb phrase (VPAUX) consists of an auxiliary verb (cf. §4.6.2) as head, followed by a verbal complement in the form of a VPMAIN, VPIDCOP, VPDESCOP, VPASSCOP or

another VPAUX. The auxiliary verb **o ne** (she was) in (5-27) is followed by the main verb **a reka** (she buys).

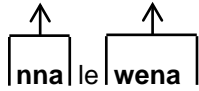


5.9 COORDINATE PHRASES


In this section, we consider the coordinate phrases that can be included as SUBJ and OBJ in the Setswana simple sentence (cf. Chapter 6). A coordinate phrase in Setswana is a phrase that consists of at least two conjuncts linked to each other with a conjunction such as **le** (and). A CLNP (cf. §5.2), PROP (cf. §5.4), POSSPARTP (cf. §5.5.1) and QUALPARTP (cf. §5.5.2) may be used as the conjuncts in a coordinate phrase should it be semantically permissible. The conjuncts can be of the same phrase type or of different phrase types. Conjuncts of the same phrase type are presented in (5-28) to (5-31). In (5-28), the coordinate phrase is structured with two CLNPs.



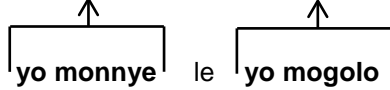
Example (5-29) is a coordinate phrase that is structured with two PROPS, i.e. the personal pronouns **nna** (I) and **wena** (you).

(5-29) PROP PROP

 I and you
nna le wena
 PersProPlsg Conj PersProP2sg
 you and me

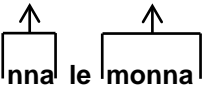
In (5-30), the coordinate phrase is structured with the POSSPARTP **ya me** (of me) and the POSSPARTP **ya gagwe** (of him)

(5-30) POSSPARTP POSSPARTP

 of me and of his
ya me le ya gagwe
 PossPart9 PossProPlsg Conj PossPart9 PossPro1
 his and mine

The coordinate phrase in (5-31) is structured with two QUALPARTPs, i.e. **yo monnye** (the small one) and **yo mogolo** (the big one).

(5-31) QUALPARTP QUALPARTP

 that is small and that is big
yo mo-nnye le yo mo-golo
 QualPart1 NPre1-adjective Conj QualPart1 NPre1-adjective
 the small one and the big one

In (5-32), a PROP and a CLNP is used as the conjuncts in the coordinate phrase.

(5-32) PROP CLNP

 I and man
nna *le* *mo-nna*
 PersProP1sg Conj NPre1-man
 the man and I

If the conjuncts are of different phrase types, a specific order is observed (Cole, 1955:428–430). The personal pronoun of the first person has precedence of word order over the personal pronoun of the second person (5-29) and the nouns and pronouns of the different noun classes (third person) (5-32).

The personal pronoun of the second person has precedence over the pronouns and nouns of the different noun classes (third person) ((5-33) and (5-34)).

(5-33) **wena le ene**
 you and he
wena *le* *ene*
 PersProP2sg Conj AbsPro1
 you and he

(5-34) **wena le monna**
 you and man
wena *le* *mo-nna*
 PersProP2sg Conj NPre1-man
 you and the man

5.10 SEQUENCING OF MODIFIERS

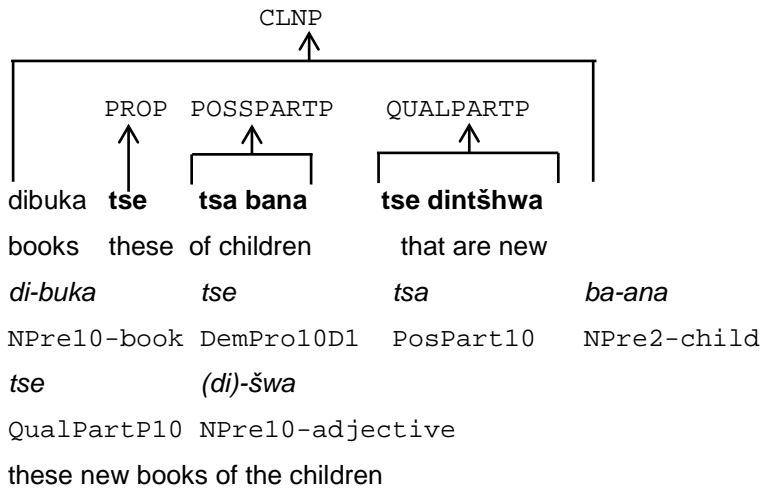
In the discussion of the Setswana phrases above, we consistently considered only one modifier or complement with each head. We now turn our attention to the occurrence of multiple modifiers in Setswana phrases.

Precedence with respect to modifiers is usually determined by the following two orderings of which the first is more commonly used (cf. Cole, 1955:438–440; Department of African Languages and Literature, University of Botswana, 2000:9; Krüger, 2013a:73–92):

PROP > POSSPARTP > QUALPARTP
 or
 PROP > QUALPARTP > POSSPARTP

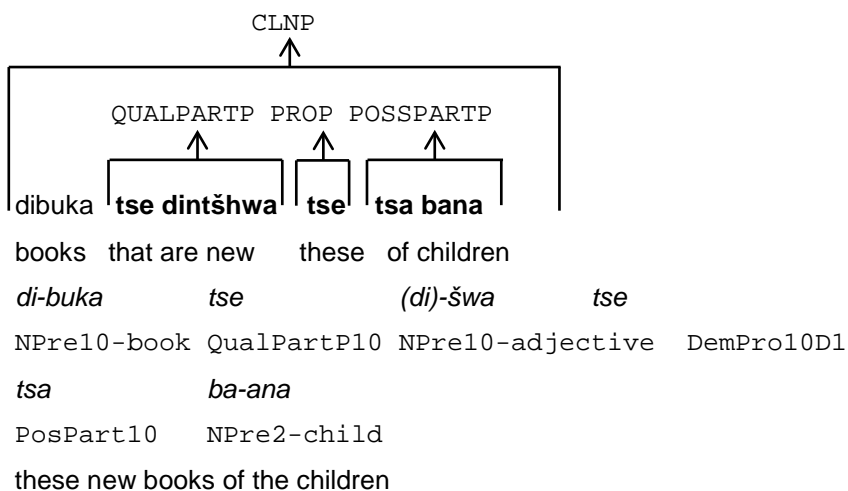
Example (5-35) shows the ordering where the head **dibuka** (books) is modified by a demonstrative pronoun **tse** (these), POSSPARTP **tsa bana** (of the children) and QUALPARTP **tse dintšhwa** (that are new). For this example, the second ordering would also be grammatical.

(5-35)



However, the PROP **tse** (these) can also follow the QUALPARTP **tse dintšhwa** (that are new) (5-36). For this reason, we allow any order of multiple modifiers in Setswana phrases as permissible.

(5-36)

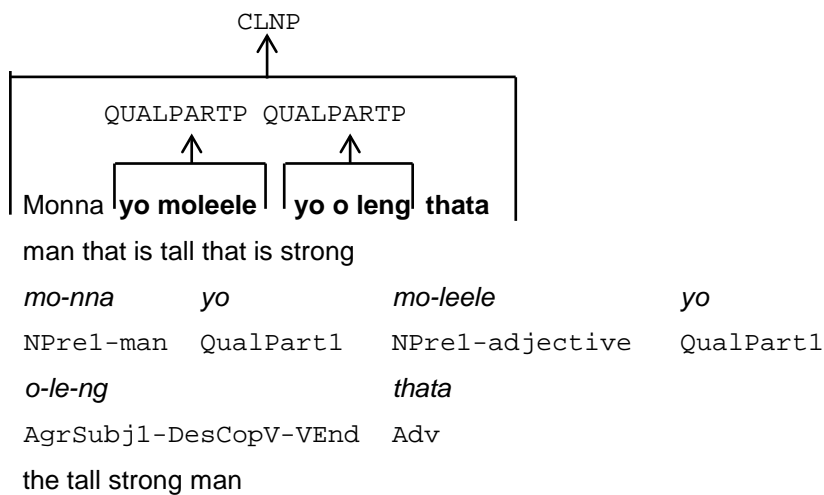


We distinguish between two sources of multiplicity, viz. *juxtaposition* (sequencing or stacking) where a sequence of modifiers modify the same head (§5.10.1), and *nesting* where a modified head occurs inside another modifier or complement of another head (§5.10.2).

5.10.1 JUXTAPOSITION

In examples (5-35) and (5-36) (cf. §5.10), three different modifiers (PROP, POSSPARTP and QUALPARTP) of the head **dibuka** (books) are juxtaposed, while in example (5-37) two similar juxtaposed QUALPARTPs modify the head **monna** (man) (Cole, 1955:438). The QUALPARTP **yo molele** (that is tall) and the QUALPARTP **yo o leng thata** (that is strong) modify the head.

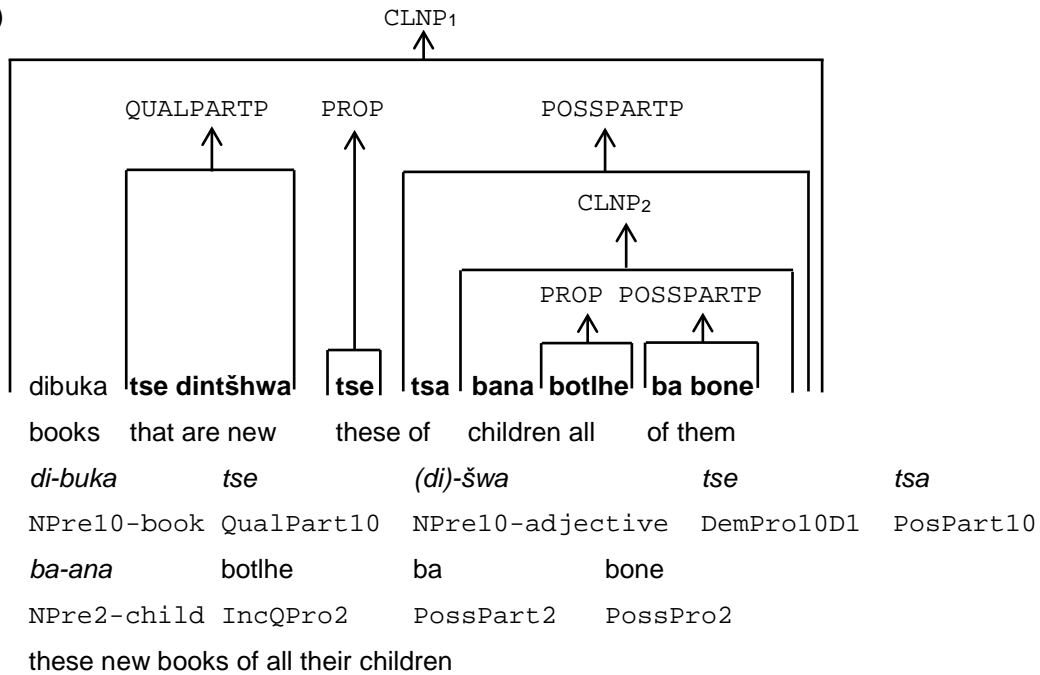
(5-37)



5.10.2 NESTING

Complements and modifiers can in turn, also be modified. For each head that is modified, also for those that occur in these complements and modifiers, the ordering *PROP* > *POSSPARTP* > *QUALPARTP* also applies here. The CLNP₁ in (5-38) consists of the head **dibuka** (books) and its modifiers QUALPARTP, PROP and POSTPARTP. In turn, the POSSPARTP of CLNP₁ consists of the possessive particle **tsa** (of) as head and the CLNP₂ **bana botlhe ba bone** (all their children) as the complement. In this complement, the noun **bana** (children) is the head of the modifying PROP **botlhe** (all) and POSSPARTP **ba bone** (of them).

(5-38)



5.11 SUMMARY

In this chapter, we identified Setswana phrases and presented the syntactic structure of each phrase. Word order and agreement phenomena are essential in structuring grammatical phrases and therefore we described these phenomena for each phrase. The phrases that are presented in this chapter are included in the description of the c- and f-structure of the Setswana simple sentence in Chapter 6.

CHAPTER 6

THE SETSWANA SIMPLE SENTENCE

6.1 INTRODUCTION

In Chapter 4, we identified eight *word categories* for Setswana and summarised the feature structure of each word category in the LFG context. By also describing the a-structure of main verbs, we are now able to present the subcategorisation frames of intransitive, transitive and ditransitive verbs as they occur in sentences. In Chapter 5, we identified Setswana *phrases* and explored the c-structure of each phrase in terms of its head and the possible modifiers or obligatory complements of these heads.

In this chapter, we build on Chapters 4 and 5 to describe the *simple sentence* in terms of its c- and f-structure, the two syntactic levels of representation in LFG. To present the c-structure of the simple sentence, we describe its immediate constituents and present the phrases that can be included in these constituents. We furthermore describe the f-structure of the simple sentence by explaining the GFs of the phrases in the sentences. We make use of the subcategorisation frames of the main verb, as explored in §4.6.1.5.

An important aspect of the c-structure of the Setswana sentence is word order. We therefore describe in some detail the correct ordering of the phrases that make up grammatical simple sentences in Setswana, having already discussed the word order within phrases in Chapter 5. In terms of f-structure, we pay specific attention to Setswana subject-verb agreement as one of the most important characteristics of Setswana syntax. We finally comment on the roles that the subject and object agreement morphemes play in the simple sentence.

Since the general aim of this study is the computational syntactic analysis of the simple sentence, this chapter forms the basis for the computational model and implementation that follows in Chapter 7.

6.2 IMMEDIATE CONSTITUENTS OF THE SIMPLE SENTENCE

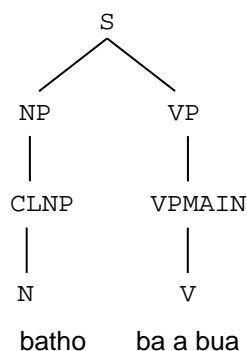
The Setswana simple sentence (S) is an independent clause (cf. §1.1.1). Its immediate (highest level) constituent structure is $S \rightarrow NP VP$ (Department of African Languages and Literature, University of Botswana, 2000:3) where NP (the noun phrase) is the first and VP (the verb phrase) the second constituent.

6.2.1 FIRST CONSTITUENT

The NP, as the first constituent in the Setswana sentence, represents any phrase that performs the SUBJ GF (The Department of African Languages and Literature, University of Botswana, (2000:5). Phrases such as the CLNP (§5.2), PROP (§5.4), POSSPARTP (§5.5.1) or QUALPARTP (§5.5.2) can all act as NPs, as illustrated in terms of the c- (tree) and f-structure (AVM) (6-2) of the sentence in (6-1). SUBJ always forms part of the subcategorisation frame of any verb (cf. §3.3.2). The noun (N) **batho** (people) represents the NP, which performs the SUBJ GF. We return to the second constituent and its c- and f-structure in subsequent sections.

- (6-1) Batho ba a bua.
 people they talk
ba-tho *ba-a-bu-a*
 NPre2-person AgrSubj2-PresPre-talk-VEnd
 The people are talking.

(6-2) c-structure



f-structure

PRED	'BU <SUBJ>'
MOOD	ind
TENSE	pres
POL	+
CLASS	2
VTYP	main
SUBJ	[PRED 'THO' CLASS 2 NTYPE ord]

6.2.2 SECOND CONSTITUENT

Since we distinguish different subcategories of the verb (main, identifying copulative, describing copulative, associative copulative, auxiliary (cf. §4.6.3)), the VP, as second constituent, consists of one of the following phrases: VPMAIN, VPIDCOP, VPDESCOP, VPASSCOP or VPAUX (cf. §5.8). In the remainder of this section, we describe the c- and f-structure of each of these phrases, including subcategorisation frames and ADJUNCTS.

6.2.2.1 Main verb phrase

As discussed in §5.8.1, a VPMAIN can consist of only a main verb (V) **ba a bua** (they talk), as shown in example (5-22). The VPMAIN can also be more complex, containing a variety of phrases

that perform GFs, such as OBJ_θ, OBJ, OBL_θ or ADJUNCT. We note that in the case of the simple sentence, the GFs of COMP (closed complement) and XCOMP (open complement) (cf. §3.3.1) do not occur.

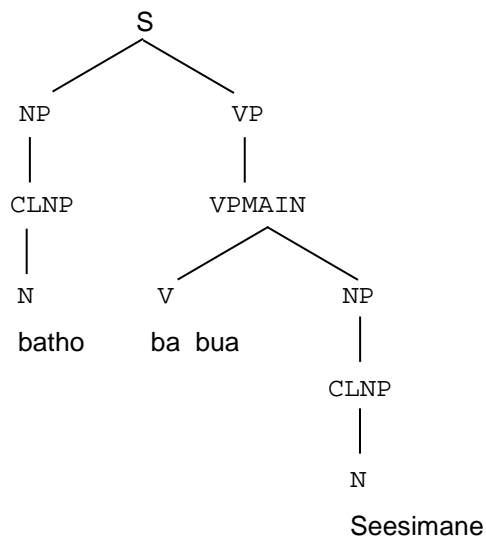
In the remainder of this section we consider the main verb and possible subcategorisation frames, including those resulting from verb extensions (cf. §4.6.1.5) that can make up the VPMAIN phrase.

VPMAIN: a verb (V) that subcategorises for an OBJ

A single NP (a CLNP (§5.2), PROP (§5.4), POSSPARTP (§5.5.1) or QUALPARTP (§5.5.2)) that performs the OBJ GF can follow the verb in the VPMAIN. This valency for an OBJ results from the meaning, tense and mood of the verb. Its meaning is derived from the basic verb root or the extended verb root, formed by suffixing one or more extensions, such as the causative (cf. §4.6.1.5 (a)), applicative (cf. §4.6.1.5 (b)), causative and reciprocal (cf. §4.6.1.5 (e)(ii)) or applicative and reciprocal (cf. §4.6.1.5 (iv)). The short form of the present tense indicative mood requires an OBJ or OBL_{Loc} (cf. §4.6.1.2). For example, in (6-3), the verb **ba bua** (they speak) indicates the short form of the indicative mood and it subcategorises for an OBJ, i.e. the noun **Seesimane** (English) (6-4).

(6-3) Batho ba bua Seesimane.
 people they speak English
ba-tho ba-bu-a se-esimane
 NPre2-person AgrSubj2-speak-VEnd NPre7-English
 The people speak English.

(6-4) **c-structure**



f-structure

PRED	'BU <SUBJ OBJ>'
MOOD	ind
TENSE	pres
POL	+
CLASS	2
VTTYPE	main
SUBJ	[PRED 'THO' CLASS 2 NTYPE ord]
OBJ	[PRED 'ESIMANE' CLASS 7 NTYPE ord]

VPMAIN: a verb (V) that subcategorises for an OBL_{loc}

A LOCCLNP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{fa}, LOCPARTP_{mo} or LOCPARTP_{kwa} can follow the main verb in the VPMAIN. As in the previous case, this valency for an OBL_{loc} GF⁷⁴ results from the meaning, tense and mood of the verb. Its meaning is derived from the basic verb root or the extended verb root, formed by suffixing the applicative extension (cf. 4.6.1.5 (b)). For example, in (6-5), the short form present tense indicative mood of the verb requires a phrase to follow, but it is the meaning of the verb root that requires an OBL_{loc} GF (not an OBJ), i.e. the LOCPARTP_{kwa} **kwa toropong** (to town) in (6-6).

(6-5) Batho ba ya kwa toropong.

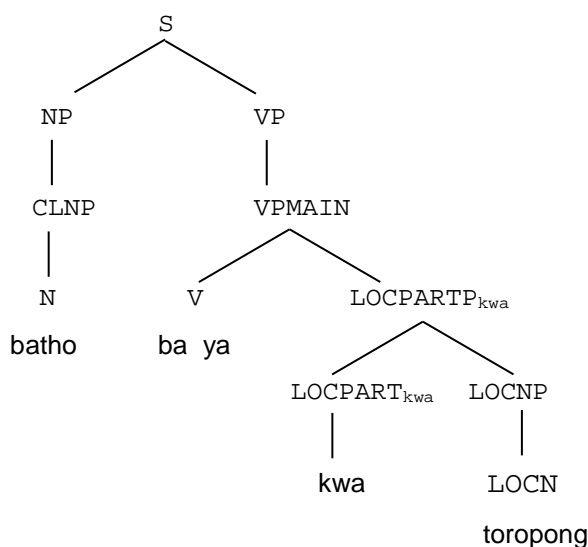
people they go to town locality

ba-tho ba-y-a kwa (ne)-toropo-ing

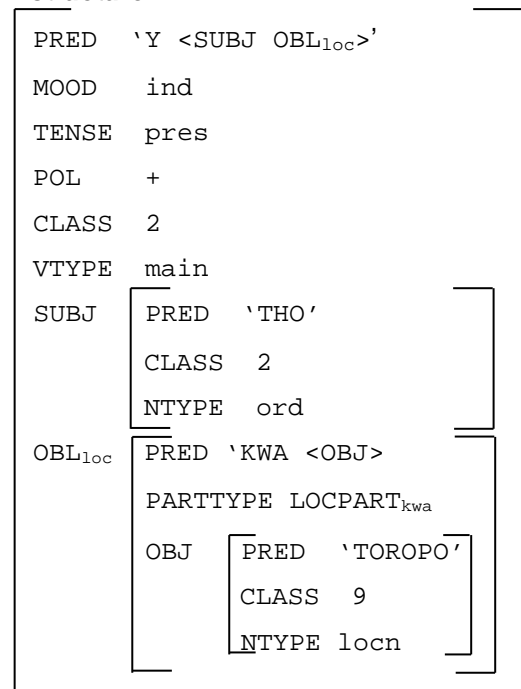
NPre2-person AgrSubj2-go-VEnd LocPart_{kwa} NPre9-town-LocSuf

The people go to town.

(6-6) **c-structure**



f-structure



VPMAIN: a verb (V) that subcategorises for an OBL_{ass}

An extended verb root with a reciprocal extension may require the inclusion of an ASSPARTP (cf. §.4.6.1.5(c) (iii)) in the structure of the VPMAIN. This phrase then functions as an OBL_{ass}. For example, in (6-7), the ASSPARTP **le wena** (with you) is the OBL_{ass} (6-8).

⁷⁴ The OBL_o often resembles ADJUNCTS in form and can be distinguished from the ADJUNCTS only by whether they are required by the predicate or not (Butt *et al.*, 1999:51)

(6-7) Mosimane o dumalana le wena.

boy he agrees together with you

mo-simane o-dumel-an-a

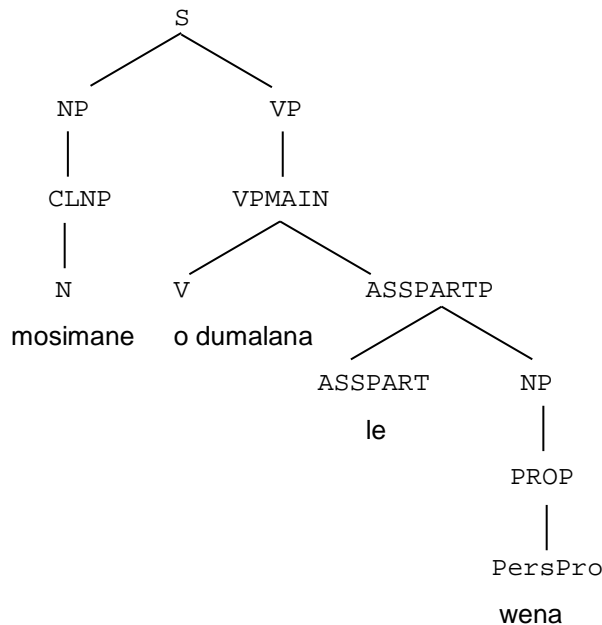
le wena

NPre1-boy AgrSubj1-agree-RecSuf-VEnd

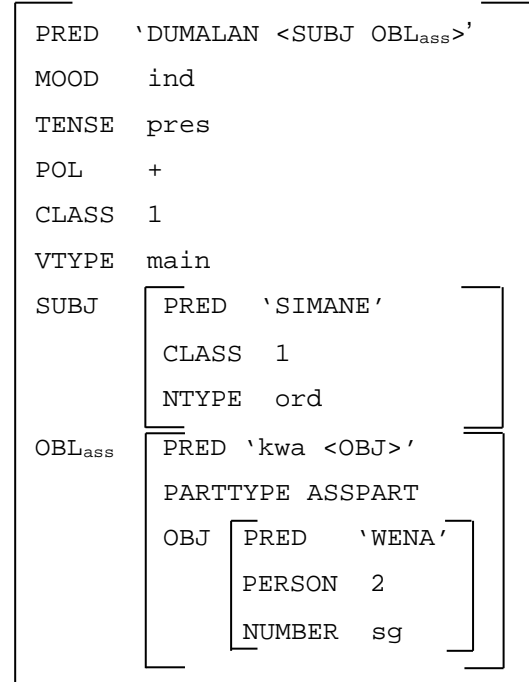
AssPart PersProP2sg

The boy agrees with you.

(6-8) **c-structure**



f-structure



VPMAIN: a verb (V) that subcategorises for an OBJ_θ and OBJ

Two NPs that perform the OBJ_θ and OBJ GFs, respectively, can follow the main verb in the VPMAIN. This valency for OBJ_θ and OBJ results from the meaning of the verb, derived from the basic verb root or the extended verb root, formed by suffixing one or more extensions, such as the causative (cf. §4.6.1.5 (a)), applicative (cf. §4.6.1.5 (b)) or both (cf. §4.6.1.5 (e) (i)). For the extended verb root **rekel-** (buy for) in (6-9), the noun **bana** (children) performs the OBJ_θ GF function and **dijo** (food), the OBJ GF (6-10).

(6-9) Basadi ba rekela bana dijo.

women they buy for children food

ba-sadi ba-rek-el-a

ba-ana

di-j-o

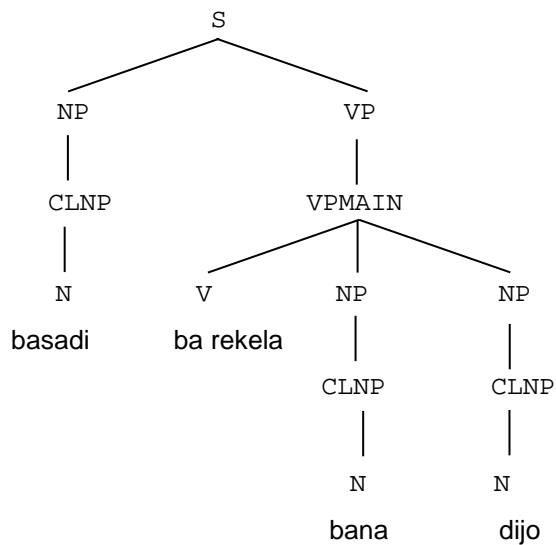
NPre2-woman AgrSubj2-buy-ApplSuf-VEnd

NPre2-child

NPre10-eat-DevSuf

The women buy the children food.

(6-10) **c-structure**



f-structure

PRED	'REKEL <SUBJ OBJ _θ OBJ>'						
MOOD	ind						
TENSE	pres						
POL	+						
CLASS	1						
VTYP	main						
SUBJ	<table border="1"><tr><td>PRED</td><td>'SADI'</td></tr><tr><td>CLASS</td><td>1</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'SADI'	CLASS	1	NTYPE	ord
PRED	'SADI'						
CLASS	1						
NTYPE	ord						
OBJ _θ	<table border="1"><tr><td>PRED</td><td>'ANA'</td></tr><tr><td>CLASS</td><td>2</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'ANA'	CLASS	2	NTYPE	ord
PRED	'ANA'						
CLASS	2						
NTYPE	ord						
OBJ	<table border="1"><tr><td>PRED</td><td>'J'</td></tr><tr><td>CLASS</td><td>8</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'J'	CLASS	8	NTYPE	ord
PRED	'J'						
CLASS	8						
NTYPE	ord						

Certain possessive constructions also provide valency for both these GFs (Pretorius, R., *et al.*, 2012:208, 210). In possessive constructions in Setswana, the possession usually occurs in the initial position while the possessor follows the possessive particle. In (6-11), **lenao** (leg) is the possession and the possessor is **mosimane** (boy). However, inalienable possession may also occur where the first noun represents a possessor affected by the action of the verb. In this possessive construction, the possessor is followed by the possession and no possessive particle is used. In (6-12), the possessor **mosimane** (boy) is followed by the possession **lenao** (leg). The noun **mosimane** (boy) performs the OBJ_θ GF and **lenao** (leg) the OBJ GF.

(6-11) Kotsi e golafaditse lenao la mosimane.

accident it injured let leg of boy

(ne)-kotsi

e-golofal-is-il-e

le-nao

NPre9-accident

AgrSubj9-injure-CausSuf-PerfSuf-VEnd

NPre5-leg

la

mo-simane

PossPart5

NPre1-boy

They accident got the boy's leg injured.

(6-12) Kotsi e golafaditse mosimane lenao.

accident it injured boy leg

(ne)-kotsi

e-golofal-is-il-e

mo-simane

NPre9-accident

AgrSubj9-injure-CausSuf-PerfSuf-VEnd

NPre1-boy

le-nao

NPre5-leg

The accident got the boy's leg injured.

VPMAIN: a verb (V) that subcategorises for an OBJ and OBL_{loc}

An NP and a phrase indicating locality (a LOCCLNP, LocPartP_{go}, LocPartP_{ga}, LocPartP_{ka}, LocPartP_{fa}, LocPartP_{mo} or LocPartP_{kwa}) that perform the OBJ and OBL_{loc} GFs, respectively, can follow the verb in the VPMAIN. This valency for OBJ and OBL_{loc} results from the meaning of the verb, derived from the basic verb root or the extended verb root, formed by suffixing one or more extensions, such as the causative (cf. §.4.6.1.5(a)), the causative and applicative (cf. §.4.6.1.5(e) (i)) or the causative, applicative and reciprocal extensions (cf. §.4.6.1.5(e) (vii)). For the extended verb root **tseny-** (put) in (6-13), the CLNP **buka** (book) and the LOC_{mo} **mo kgetsing** (in the bag) perform the OBJ and OBL_{loc} GFs respectively (6-14).

(6-13) Mosadi o tsenya buka mo kgetsing.

woman she enter let book in bag

mo-sadi

o-tsen-y-a

(ne)-buka

NPre1-woman

AgrSubj1-enter-CausSuf-VEnd

NPre9-book

mo

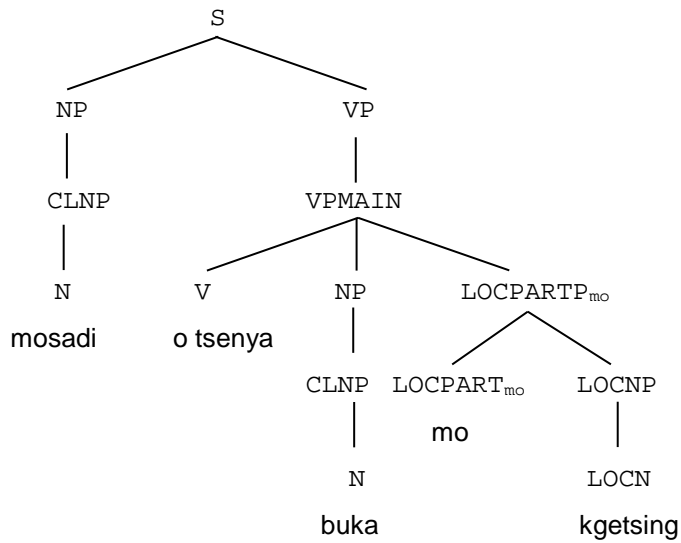
(ne)-kgetsing

LocPart_{mo}

NPre9-bag-LocSuf

The woman puts the book in the bag.

(6-14) **c-structure**



f-structure

PRED	'TSENY <SUBJ OBJ _θ OBJ>'										
MOOD	ind										
TENSE	pres										
POL	+										
CLASS	1										
VTYPE	main										
SUBJ	<table border="1"> <tr> <td>PRED</td> <td>'SADI'</td> </tr> <tr> <td>CLASS</td> <td>1</td> </tr> <tr> <td>NTYPE</td> <td>ord</td> </tr> </table>	PRED	'SADI'	CLASS	1	NTYPE	ord				
PRED	'SADI'										
CLASS	1										
NTYPE	ord										
OBJ	<table border="1"> <tr> <td>PRED</td> <td>'BUKA'</td> </tr> <tr> <td>CLASS</td> <td>9</td> </tr> <tr> <td>NTYPE</td> <td>ord</td> </tr> </table>	PRED	'BUKA'	CLASS	9	NTYPE	ord				
PRED	'BUKA'										
CLASS	9										
NTYPE	ord										
OBL _{loc}	<table border="1"> <tr> <td>PRED</td> <td>'MO <OBJ>'</td> </tr> <tr> <td>OBJ</td> <td> <table border="1"> <tr> <td>PRED</td> <td>'KGETSI'</td> </tr> <tr> <td>CLASS</td> <td>9</td> </tr> <tr> <td>NTYPE</td> <td>ord</td> </tr> </table> </td> </tr> </table>	PRED	'MO <OBJ>'	OBJ	<table border="1"> <tr> <td>PRED</td> <td>'KGETSI'</td> </tr> <tr> <td>CLASS</td> <td>9</td> </tr> <tr> <td>NTYPE</td> <td>ord</td> </tr> </table>	PRED	'KGETSI'	CLASS	9	NTYPE	ord
PRED	'MO <OBJ>'										
OBJ	<table border="1"> <tr> <td>PRED</td> <td>'KGETSI'</td> </tr> <tr> <td>CLASS</td> <td>9</td> </tr> <tr> <td>NTYPE</td> <td>ord</td> </tr> </table>	PRED	'KGETSI'	CLASS	9	NTYPE	ord				
PRED	'KGETSI'										
CLASS	9										
NTYPE	ord										

VPMAIN: a verb (V) that subcategorises for an OBJ and OBL_{ass}

An NP and an ASSPARTP that perform the OBJ and OBL_{recip} GFs, respectively, can follow the verb in the VPMAIN. For the basic verb root **nw-** (drink) in (6-15), the noun **kofi** (coffee) and the ASSPARTP **le wena** (with you) perform the OBJ and OBL_{ass} grammatical function respectively (6-16).

(6-15) Monna o tla nwa kofi le wena.

man he will drink coffee with you

mo-nna

o-nw-a

(ne)-kofi

le

NPre1-man AgrSubj1-drink-VEnd

NPre9-coffee

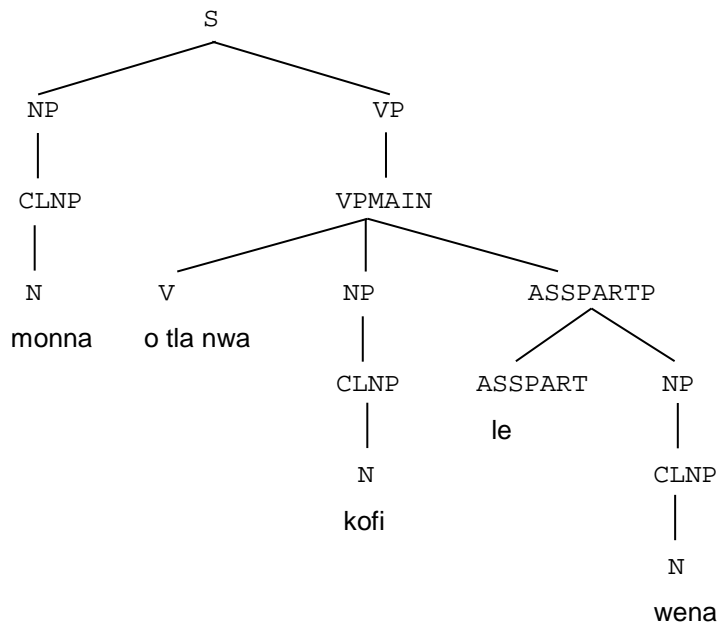
AssPart

wena

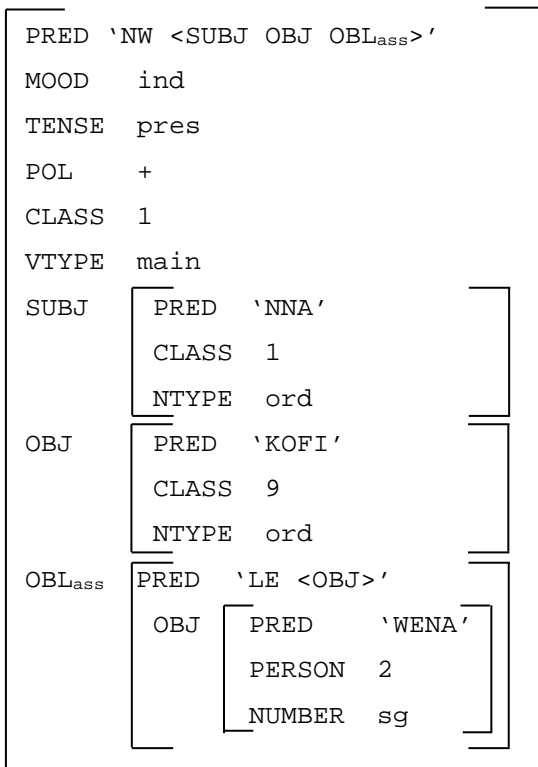
PersProP2sg

The man will drink coffee with you.

(6-16) **c-structure**



f-structure



VPMAIN: a verb (V) that subcategorises for an OBJ_θ, OBJ and OBL_{loc}

Two NPs that perform the OBJ_θ and OBJ GFs, respectively, and a phrase indicating locality (OBL_{loc}) can follow the main verb in the VPMAIN. This valency for OBJ_θ, OBJ and OBL_{loc} results from the meaning of the verb that is derived from the basic verb root or the extended verb root, formed by suffixing the causative and the applicative extensions (cf. §.4.6.1.5(e) (i)). For the

extended verb root **tsenyets-** (deposit) in (6-17), the noun **mme** (mother) performs the OBJ_θ GF, **madi** (money) the OBJ GF, and **mo bankeng** (in the bank) the OBL_{Loc} GF (6-18).

(6-17) Rre o tsenyetsa mme madi mo bankeng

father he enter let for mother money in bank

(-)-rre o-tsen-is-el-a

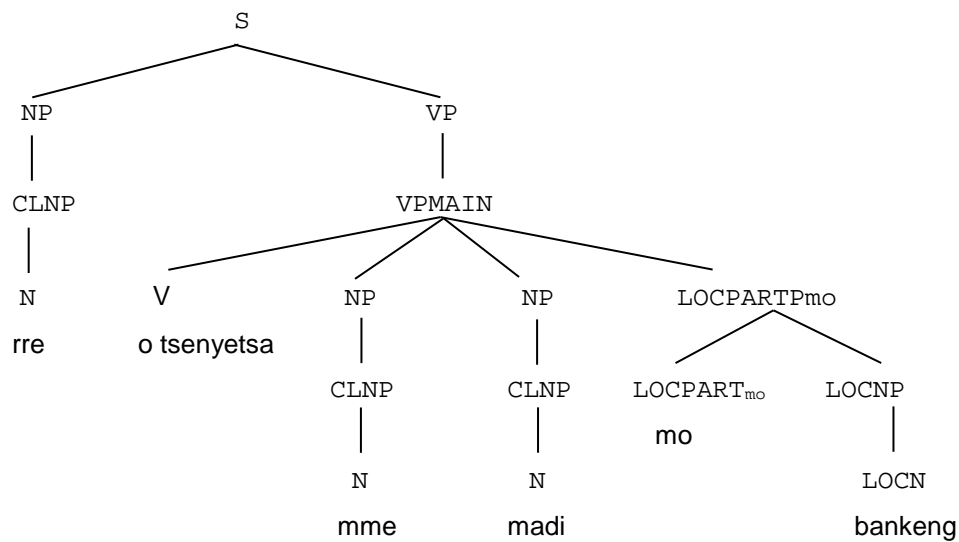
NPre1a-father AgrSubj1a-enter-CausSuf-ApplSuf-VEnd

(-)-mme ma-di mo (ne)-banka-ing

NPre1a-mother NPre6-money LocPart_{mo} NPre10-bank-LocSuf

Father deposits money in the bank for mother.

(6-18) **c-structure**



f-structure

PRED	'TSENYETS <SUBJ OBJ _θ OBJ OBL _{loc} >'										
MOOD	ind										
TENSE	pres										
POL	+										
CLASS	1										
VTYPE	main										
SUBJ	<table border="1"><tr><td>PRED</td><td>'RRE'</td></tr><tr><td>CLASS</td><td>1a</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'RRE'	CLASS	1a	NTYPE	ord				
PRED	'RRE'										
CLASS	1a										
NTYPE	ord										
OBJ _θ	<table border="1"><tr><td>PRED</td><td>'MME'</td></tr><tr><td>CLASS</td><td>1a</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'MME'	CLASS	1a	NTYPE	ord				
PRED	'MME'										
CLASS	1a										
NTYPE	ord										
OBJ	<table border="1"><tr><td>PRED</td><td>'DI'</td></tr><tr><td>CLASS</td><td>6</td></tr><tr><td>NTYPE</td><td>ord</td></tr></table>	PRED	'DI'	CLASS	6	NTYPE	ord				
PRED	'DI'										
CLASS	6										
NTYPE	ord										
OBL _{loc}	<table border="1"><tr><td>PRED</td><td>'MO <OBJ>'</td></tr><tr><td>OBJ</td><td><table border="1"><tr><td>PRED</td><td>'BENKELE'</td></tr><tr><td>CLASS</td><td>5</td></tr><tr><td>NTYPE</td><td>locn</td></tr></table></td></tr></table>	PRED	'MO <OBJ>'	OBJ	<table border="1"><tr><td>PRED</td><td>'BENKELE'</td></tr><tr><td>CLASS</td><td>5</td></tr><tr><td>NTYPE</td><td>locn</td></tr></table>	PRED	'BENKELE'	CLASS	5	NTYPE	locn
PRED	'MO <OBJ>'										
OBJ	<table border="1"><tr><td>PRED</td><td>'BENKELE'</td></tr><tr><td>CLASS</td><td>5</td></tr><tr><td>NTYPE</td><td>locn</td></tr></table>	PRED	'BENKELE'	CLASS	5	NTYPE	locn				
PRED	'BENKELE'										
CLASS	5										
NTYPE	locn										

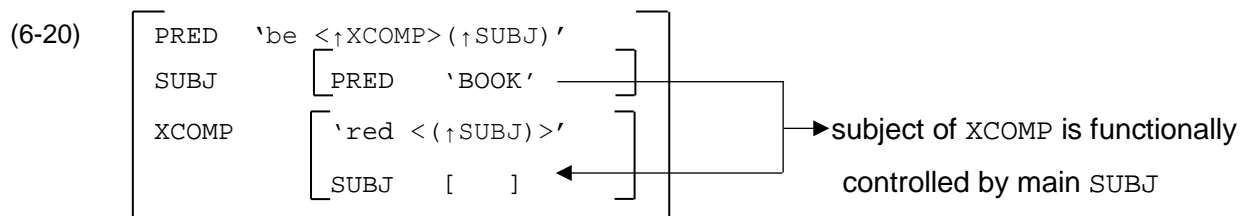
6.2.2.2 Identifying, describing and associative copulative verb phrases

In Setswana, three subcategories of copulative verbs, viz. identifying, describing and associative copulative verbs, are distinguished (cf. §4.6.3). In each case, the copulative verb is the head of the copulative verb phrase, VPIDCOP, VPDESCOP or VPASSCOP, respectively. Each one of these heads take specific complements, and these phrases constitute minimal sentences (cf. §5.8.2 to §5.8.4). The important question that now arises is what the GFs of these complements are. Different approaches have been considered by, for example, Butt *et al.* (1999:69–70), Dalrymple (2004), Attia (2008:141–143) and Sulger (2009:26–40) to analyse copulative constructions.

In the so-called *single-tier* analysis (for example, Dalrymple, 2004), the copulative verb is considered optional. The complement is treated as the head in the sentence and it subcategorises for a SUBJ (6-19), as is the case for the Japanese sentence meaning *The book is red*, taken from Dalrymple *et al.* (2004:191–192), who states that "Japanese adjectives do not require the copula ... [T]he adjective provides the main PRED for the clause".

$$(6-19) \left[\begin{array}{l} \text{PRED} \quad \text{'RED} < \uparrow \text{SUBJ} > \text{' } \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'BOOK' } \end{array} \right] \end{array} \right]$$

In the so-called *double-tier* analysis (Butt *et al.*, 1999:69–70), both the SUBJ and complement of a copulative verb function as arguments in the sentence. The complement is either open (XCOMP) or closed (PREDLINK). In the open complement analysis, the SUBJ of XCOMP is unified with the SUBJ that precedes the copulative verb, and control equations are defined between the SUBJ of the sentence and the SUBJ of the XCOMP. The SUBJ of the XCOMP does not include a value but it receives a value by functional control through the copulative verb (Attia, 2008:143; Sulger, 2009:30–31). The open complement analysis of the sentence "The book is red" is presented in (6-20).



The closed complement analysis posits a GF PREDLINK, modelling:

the fact that a particular property is predicated of the subject in a syntactically reasonable way. ... As PREDLINK is a closed category, there is no control equation between the SUBJ and the PREDLINK (Butt *et al.*, 1999:70).

Hence there is no need for the complements of the copulative verbs to have SUBJ arguments (Attia, 2008:143; Sulger, 2009:33–32). The closed complement analysis of the sentence "The book is red" is presented in (6-21).

$$(6-21) \left[\begin{array}{l} \text{PRED} \quad \text{'be} < \text{SUBJ}, \text{PREDLINK} > \text{' } \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'BOOK' } \end{array} \right] \\ \text{PREDLINK} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'RED' } \end{array} \right] \end{array} \right]$$

The single-tier analysis is the preferred analysis for languages that do not have a copula, while languages that use overt copulas can choose between the two variants of the double-tier analysis (Attia, 2008:142; Sulger, 2009:29). We apply the closed complement (PREDLINK) analysis for Setswana, as the complement of the copulative verb does not subcategorise for a SUBJ. For example, in (6-22) the adjective **montle** (beautiful) is predicated of the SUBJ **mosadi** (woman). The f-structure of (6-22) that includes a describing copulative verb **o** is presented in (6-23).

(6-22) Mosadi o montle.
 woman is beautiful
 mo-sadi o mo-ntle
 NPre2-woman DesCopV2 NPre2-adjective
 The woman is beautiful.

(6-23)
$$\left[\begin{array}{l} \text{PRED} \quad \text{'O <SUBJ PREDLINK>'} \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'SADI'} \\ \text{CLASS} \quad 1 \\ \text{NTYPE} \quad \text{ord} \end{array} \right] \\ \text{PREDLINK} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'NTLE'} \\ \text{CLASS} \quad 1 \\ \text{NTYPE} \quad \text{adjective} \end{array} \right] \end{array} \right]$$

6.2.2.3 Auxiliary verb phrase

The minimal structure of a VPAUX consists of the auxiliary verb and an obligatory complement, viz. the VPMAIN, VPIDCOP, VPDESCOP, VPASSCOP or another VPAUX (cf. §5.8.5). In terms of subcategorisation frames, there are two main approaches to analyse auxiliaries in LFG (Butt *et al.*, 1999:60–63). The auxiliary verb is treated as a special type of raising verb that takes a SUBJ and an XCOMP⁷⁵ argument, or is considered a feature-carrying element (Butt *et al.*, 1999:61–63). Setswana auxiliary verbs are considered to be feature-carrying elements (cf. §4.6.2). The auxiliary verb does not have a subcategorisation frame and a flat f-structure analysis is followed because the complement following the auxiliary verb is considered the main predicate of the sentence. The f-structure of the auxiliary verb **re ne** in (6-24) is presented in (6-25).

(6-24) Rona re ne re bua Seesimane.
 we we were we speak English
 rona re-ne re-bu-a Se-esimane
 PersProPlpl AgrSubjPlpl-Aux AgrSubjPlpl-speak-VEnd NPre7-english
 We were speaking English.

(6-25)
$$\left[\begin{array}{l} \text{AUX} \quad + \\ \text{TENSE} \quad \text{RelPast} \end{array} \right]$$

⁷⁵ The XCOMP function is as an open clausal function that does not contain an internal subject phrase (Dalrymple, 2001:24).

6.2.2.4 Inclusion of adjuncts in the verbal phrases

The VPMAIN, VPIDCOP, VPDESCOP, VPASSCOP and VPAUX can also include phrases that refer to time, locality, manner, instrument and association, an INTERJP or an IDEOP. These phrases are modifiers in the VPs and therefor provide additional, though not essential, information and are not argument functions. They perform an ADJUNCT function. In (6-26), the INSTRPARTP **ka karata** (with a card) performs the ADJUNCT function. The c- and f-structure are presented in (6-27).

(6-26) Mosadi o duela ditlhako ka karata.

woman she pay shoes with card

mo-sadi

o-duel-a

di-tlhako

NPre1-woman AgrSubj1-pay-VEnd

NPre10-shoe

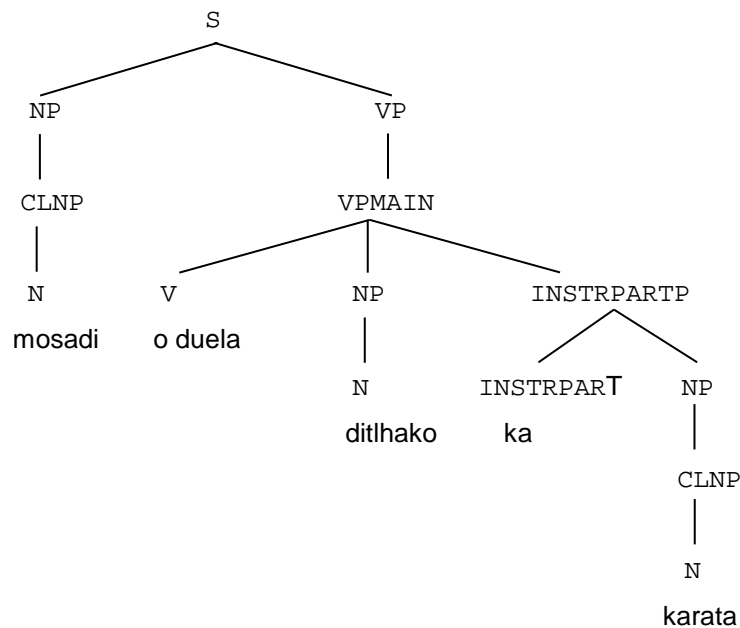
ka

(ne)-karata

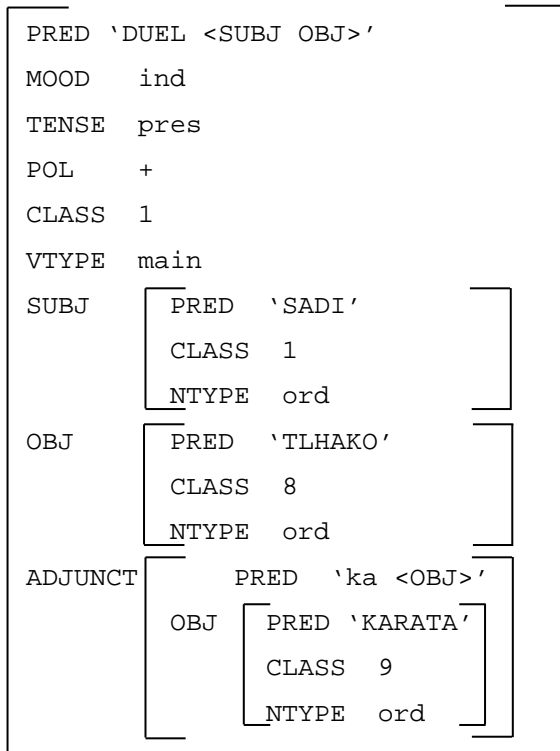
InstrPart NPre9-card

The woman pays the shoes with a card.

(6-27) **c-structure**



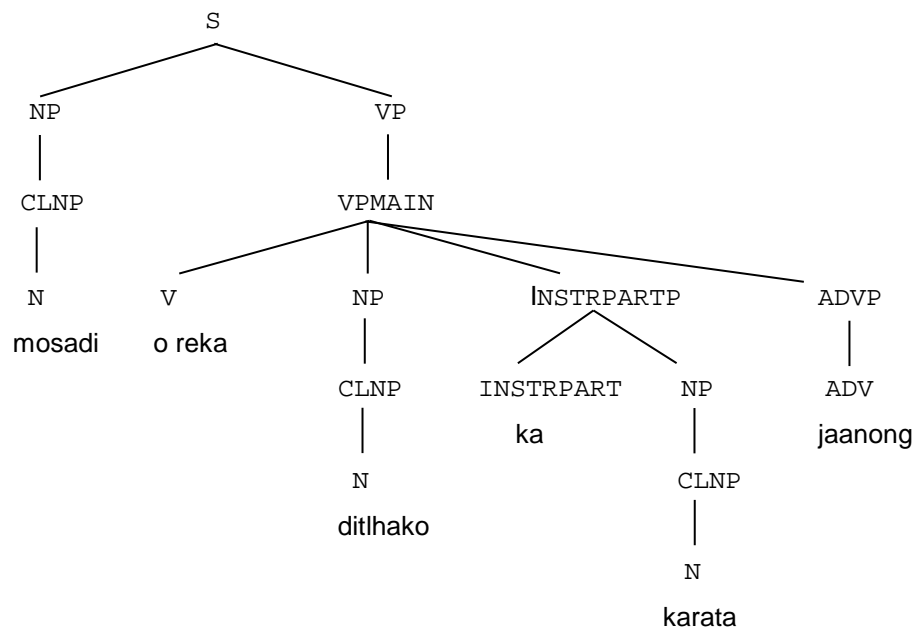
f-structure



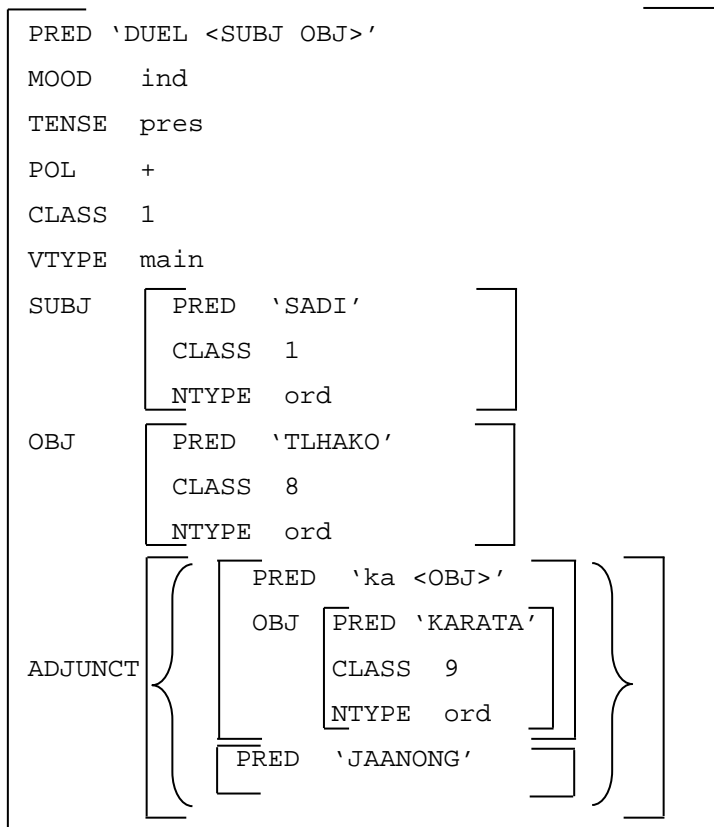
More than one ADJUNCT can be included in a VP (cf. §6.2.3) and it then forms part of a set (cf. §3.4.1). For example, in (6-28) the ADJUNCTS are **ka karata** (with a card) and **jaanong** (now). The c- and f-structure are presented in (6-29).

- (6-28) Mosadi o duela ditlhako ka karata jaanong.
 woman she pay shoes with card now
mo-sadi o-rek-a di-tlhako
 NPre1-woman AgrSubj1-pay-VEnd NPre10-shoe
ka (ne)-karata jaanong
 InstrPart NPre9-card Adv
 The woman now pays the shoes with/by a card.

(6-29) **c-structure**



f-structure



The phrases that can function as ADJUNCTS in the VPMAIN, VPIDCOP, VPDESCOP, VPASSCOP and VPAUX are given below.

Phrases indicating time

Phrases such as some CLNPs, some ADVPS, a TEMPPARTP or a LOCPARTP_{mo} indicate the time as well as the duration and frequency of an action (Le Roux, 2007:93). In (6-30), a LOCPARTP_{mo} **mo mosong** (in the morning) indicates when the action of the verb **ba taboga** (they run) is executed.

(6-30) Bana ba taboga mo mosong.

children they run in morning

ba-ana *ba-tabog-a* *mo* *mo-so-ng*

NPre2-child AgrSubj2-run-VEnd LocPart_{mo} NPre2-morning-LocSuf

The children run in the morning.

Phrases indicating locality

A LOCCLNP, ADVP, POSSPARTP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{fa}, LOCPARTP_{mo}, or a LOCPARTP_{kwa} expresses the place, position, direction or distance where an action is performed. Only the pronouns of classes 16-20 indicate locality. Therefore, only the absolute, demonstrative, inclusive quantitative, exclusive quantitative, separative quantitative and communal possessive pronouns of noun classes 16-20 are included as phrases indicating locality. In (6-31), the LOCPARTP_{mo} **mo bankeng** (in the bank) indicates the locality of the action.

(6-31) Ba dira mo bankeng.

they work in bank

ba-dir-a *mo* *(ne)-banka-ing*

AgrsSubj2-work-VEnd LocPart_{mo} NPre9-bank-LocSuf

They work in the bank.

Phrases indicating manner

Some phrases, usually an ADVP, an INSTRPARTP or a COMPPARTP⁷⁶, indicate the manner in which an action or process is carried out (Louwrens, 1994a: 101–102; Le Roux, 2007:95). Example (6-32) includes the ADVP **thata** (hard) that indicates how the action of the verb is executed.

⁷⁶ Cole (1955:369) refers to examples where the comparative particle and complement are used in sentences as “manner-comparative adverbs”.

ADJUNCT⁷⁹. In this section, we present a summary of the arguments that can be included in a simple sentence ($S \rightarrow NP VP$) and in what order they occur. An NP that maps to the SUBJ usually occupies the initial position in the sentence. In the discussion in §4.6.1.5 and §6.2.1 we observed the following orders in the VPMAIN:

- main verb, OBJ
- main verb, OBJ, OBL_{loc}
- main verb, OBJ, OBL_{ass}
- main verb, OBJ_θ, OBJ
- main verb, OBJ_θ, OBJ, OBL_{loc}
- main verb, OBL_{loc}
- main verb, OBL_{ass}

In the VPIDCOP, VPDESCOP and VPASSCOP, a phrase that functions as a PREDLINK follows the copulative verb (cf. §5.8.2, §5.8.3, §584 and §6.2.2.2). In the VPAUX, the auxiliary verb is followed by a VPMAIN, one of the copulative verb phrases or another VPAUX (cf. §5.8.5 and §6.2.2.3).

The basic position of phrases (ADJUNCTS) indicating locality, time, manner, instrument and association, and the IDEOP is usually at the end of a sentence. Although these ADJUNCTS are not arguments, they form part of the structure of the different verb phrases. The order in which ADJUNCTS occur is mostly free although Cole (1955:452) mentions that a phrase that indicates manner usually precedes phrases that indicate locality and time, and phrases that indicate locality and manner can be used interchangeably. In this study, free ADJUNCT order is assumed. The INTERJP that also functions as an ADJUNCT, usually occurs at the beginning of a sentence.

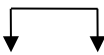
6.4 SUBJECT-VERB AGREEMENT

Agreement is a central typological feature of verbs in Bantu language syntax, including Setswana, in which verbs exhibit agreement with their preceding SUBJS.


In the VPMAIN, the verb agrees with the SUBJ with respect to class or person, and number. This agreement is established by the subject agreement morpheme (Appendix A: Table 4-4 to Table 4-7, p.191-192) which is included in the morphological structure of the main verb (6-35).

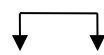
at the beginning of the sentence to emphasise the specific place, time, manner, instrument or association. We do not include the altered orders in this study as our focus is not on information structure.

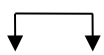
⁷⁹ Krüger (1967:153) also use the term “descriptive determiner” or “descriptive” to refer to the ADJUNCT function.

- (6-35)  **Lekau le** reka dibuka.
 young man he buys books
 le-kau le-rek-a di-buka
 NPre5-young man AgrSubj5-buy-VEnd NPre10-book
 The young man buys books.

In the *VPIDCOP*, the identifying copulative verb in the present tense of the indicative mood is **ke** (is/are) in the positive and **ga se** (is not/are not) in the negative for all the noun classes ((6-36) and (6-37)). However, the identifying copulative verb for the first and second persons (Appendix A: Table 4-11, p.196) exhibits agreement with the specific personal pronoun. In (6-38), the identifying copulative verb **re** (are) shows agreement concerning the second person plural **rona** (we).

- (6-36)  **Mosadi ke** morutabana.
 woman is teacher
 mo-sadi ke mo-rutabana
 NPre1-woman IdCopV NPre1-teacher
 The woman is a teacher.

- (6-37)  **Basadi ke** barutabana.
 women are teachers
 ba-sadi ke ba-rutabana
 NPre2-woman IdCopV NPre2-teacher
 The women are teachers.

-  (6-38) **Rona re** bareki.
 we are buyers
 rona re bareki
 PersProPlpl IdCopVP1pl NPre2-buy-DevSuf
 We are buyers.

In the *VPDESCOP*, the describing copulative verb agrees with the *SUBJ* with respect to class, person and number (6-39) (Appendix A: Table 4-12 and Table 4-13, p.196). If the complement is a changeable adjective, the prefix of this adjective also agrees with the class of the *SUBJ* (6-40).

(6-39) **Metsi a** bollo.
 water is warm
 ma-itsi a bo-llō
 NPre6-water DesCopV6 NPre14-adjective
 The water is warm.

(6-40) **Setlhare se** segolo.
 tree is big
 se-tlhare se se-golo
 NPre7-tree DesCopV7 NPre7-adjective
 The tree is big.

In the *VPASSCOP*, the associative copulative verb agrees with the *SUBJ* with respect to class, person and number. This agreement is established by the subject agreement morpheme (6-41).

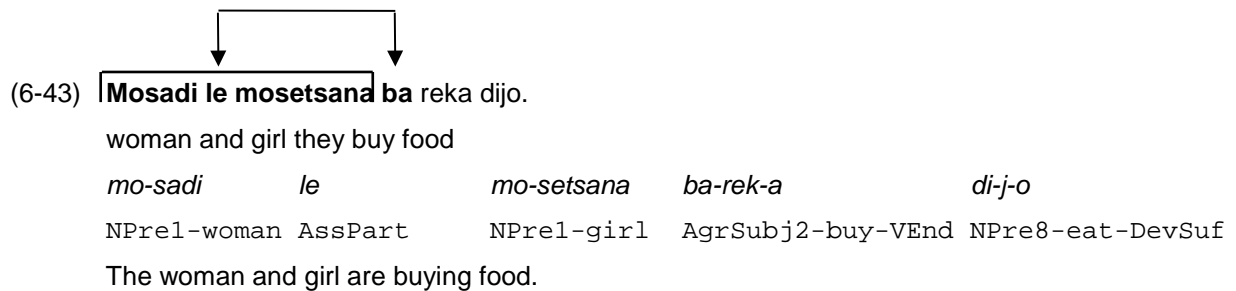
(6-41) **Bana ba** na le dibaesekele.
 children are with bicycles
 ba-ana ba-na le di-baesekele
 NPre2-child AgrSubj2-AssCopV AssPart NPre10-bicycle
 The children have bicycles.

In the *VP AUX*, the auxiliary verb and its complements agree with the *SUBJ* with respect to class, person and number. In (6-42), the obligatory subject agreement morpheme of the auxiliary verb and that of the main verb agree with the *SUBJ basadi* (woman), a noun in class 2.

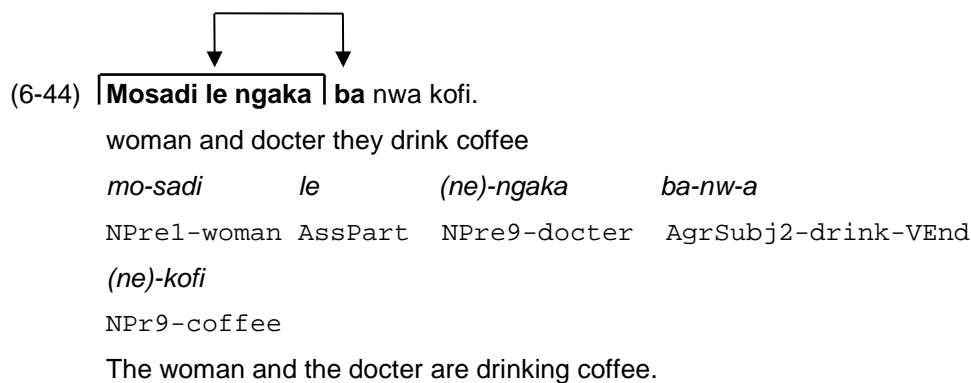
(6-42) **Basadi ba** ne **ba** ruta bana dipalo.
 women they were they teach children maths
 ba-sadi ba-ne ba-rut-a di-bal-o
 NPre2-woman AgrSubj2-Aux AgrSubj2-teach-VEnd NPre10-count-DEVSuf
 The women were teaching the children maths.

Coordination (cf. §5.9) remains a complex issue in Bantu languages due to, amongst others, the strategies followed in class gender agreement when the conjuncts belong to different noun

classes (Cole, 1955:428–430; Poulos & Louwrens, 1994; Katamba, 2003; De Vos & Mitchley, 2012; Mitchley, 2015). If both nouns are from the same singular noun class, the subject agreement morpheme of the corresponding plural noun class is used. For example, in (6-43) both nouns in the coordinate phrase are from class 1 and therefore the subject agreement morpheme is from class 2.



If the nouns in the coordinate phrase are from different classes but indicate humans, the subject agreement morpheme of class 2 is applied (6-44).

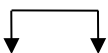


If the nouns in the coordinate phrase are from different classes but indicate non-personal and non-animal nouns, the subject agreement morpheme of class 10 is applied. If the coordinate phrase includes both people or persons and animals or inanimate subjects, one of the nouns is placed before the verb, while the others are placed after the verb as a complement of the particle **le** (together with) in an ASSPARTP. For example, in (6-45), **mosimane** indicates a human in class 1 and **ntšwa** an animal in class 10. Therefore, **le ntšwa** is included as an ADJUNCT in the sentence.

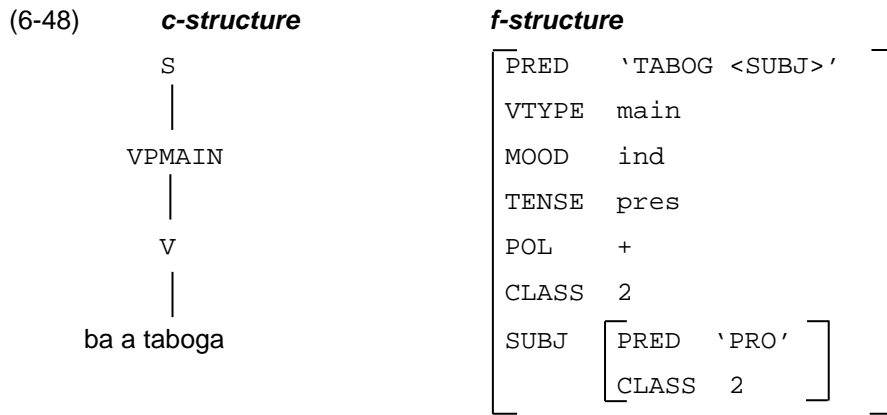
- (6-45) Mosimane o tshameka le ntšwa.
 boy he plays together with you
mo-simane o-tshamek-a le ntšwa
 NPre1-boy AgrSubj1-play-VEnd AsPartP (ne)-dog
 The boy and dog are playing.

6.5 STATUS OF THE SUBJECT AND OBJECT AGREEMENT MORPHEMES

The subject agreement morpheme not only exhibits the agreement between a verb and its *SUBJ*, but also performs an anaphoric function when it refers to an omitted *SUBJ* (Krüger, 2006:32, 53). A subject agreement morpheme in Setswana thus behaves either as an agreement marker or as an incorporated pronoun. It behaves as an agreement marker if an overt *SUBJ* precedes the verb, as illustrated in (6-46). The subject agreement morpheme acts as an incorporated pronoun if the *SUBJ* is omitted from the sentence structure, as is evident in the sentence in (6-47). In this instance, the incorporated pronoun functions as the *SUBJ* in the sentence and the special *PRED* value 'PRO' is attributed to it in the f-structure (Butt *et al.* 1999: 13). The f-structure of the sentence in (6-47) is presented in (6-48).

- (6-46)  **Batho ba** a taboga.
 people they run
ba-tho ba-a-tabog-a
 NPre2-person AgrSubj2-PresPre-run-VEnd
 The people are running.

- (6-47) **Ba** a taboga.
 they run
ba-a-tabog-a
 AgrSubj2-PresPre-run-VEnd
 They people are running.



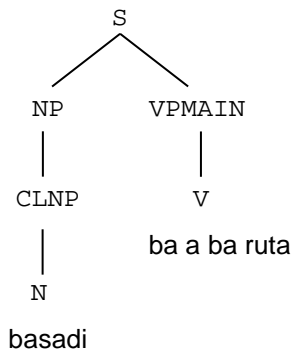
Setswana is therefore a so-called pro-drop language and all verbs therefore subcategorise for a SUBJ. It is not obligatory for an overt SUBJ to be included in a sentence, but if the SUBJ is omitted, the subject agreement morpheme provides a pronominal interpretation of the missing argument. The use of the subject agreement morpheme thus enables the reconstruction of the missing SUBJ.

An object agreement morpheme (Appendix A: Table 4-8 and Table 4-9, p.193) as a verbal prefix, shows correspondence to an omitted OBJ (Cole, 1955:229). The verb does not show agreement with an overt OBJ as the object agreement morpheme and OBJ do not typically co-occur in a sentence. The OBJ **bana** (children) in (6-49) is omitted in (6-50) and the object agreement morpheme **ba** (them) is used to refer to the omitted OBJ. In this instance, an anaphoric relationship exists between the object agreement morpheme and the omitted noun (Louwrens, 1994a:11). The object agreement morpheme behaves as an incorporated pronoun and the incorporated pronoun functions as the OBJ in the sentence. The special PRED value 'PRO' is included in the f-structure (Butt *et al.*, 1999: 13; Mchombo, 2004:20–22). The c-structure of the sentence in (6-50) is presented in (6-51) and it shows that the VPMAIN consists of only a verb. However, the f-structure shows that the verb in this VPMAIN subcategorises for an OBJ, as an object agreement morpheme is treated as an incorporated pronoun.

(6-49) Basadi ba ruta bana.
 women they teach children
ba-sadi ba-rut-a ba-ana
 NPre2-woman AgrSubj2-teach-VEnd NPre2-child
 The women teach the children.

(6-50) Basadi ba a ba ruta.
 women they them teach
ba-sadi ba-a-ba-rut-a
 NPre2-woman AgrSubj2-PresPre-AgrObj2-teach-VEnd
 The women teach them.

(6-51) **c-structure**



f-structure

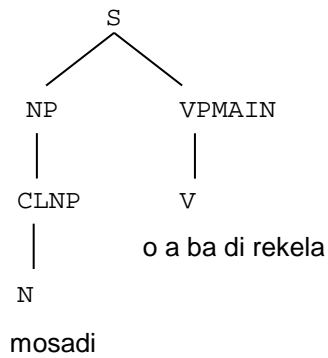
PRED	'RUT <SUBJ OBJ>'
MOOD	ind
TENSE	pres
POL	+
CLASS	2
SUBJ	[PRED 'SADI' CLASS 2]
OBJ	[PRED 'PRO' CLASS 2]

Sentences in Setswana may include verbs that subcategorise for double objects, i.e. the OBJ_θ and OBJ (6-52). Both these objects may be replaced by their respective object agreement morphemes (6-53). The object agreement morpheme of the OBJ_θ is followed by the object agreement morpheme of the OBJ (Berg *et al.*, 2013:119–125). These object agreement morphemes are treated as incorporated pronouns and the verb therefore subcategorises for an OBJ_θ and an OBJ (6-54).

(6-52) Mosadi o rekela bana dijo.
 woman she buy for children food
mo-sadi o-rek-el-a ba-ana di-j-o
 NPre1-woman AgrSubj1-buy-ApplSuf-VEnd NPre2-child NPre10-eat-DevSuf
 The woman buys food for the children.

(6-53) Mosadi o a ba di rekela.
 woman she them it buy for
mo-sadi o-a-ba-di-rek-el-a
 NPre1-woman AgrSubj1-PresPre-AgrObj2-AgrObj8-buy-ApplSuf-VEnd
 The woman buys it for them.

(6-54) **c-structure**



f-structure

PRED	'REKEL <SUBJ OBJ _θ OBJ>'
MOOD	ind
TENSE	pres
POL	+
CLASS	1
SUBJ	[PRED 'SADI' CLASS 1]
OBJ _θ	[PRED 'PRO' CLASS 2]
OBJ	[PRED 'PRO' CLASS 8]

The reflexive morpheme **i(n)-** (self) reflects the action of a transitive verb onto the SUBJ (Krüger, 2006:179). The morpheme has a referential value because it refers to an OBJ that is identical to the SUBJ and therefore a co-reference relation exists between the subject and the object. This morpheme is not class or person bound as it has one form for all classes and persons (Cole, 1955:232). We treat the reflexive morpheme as an incorporated pronoun and, as is the case with the object agreement morpheme, the PRED value 'PRO' is included in the f-structure of a verb with a reflexive morpheme. The c- and f-structure of the sentence in (6-55) is presented in (6-56).

(6-55) Bana **ba a ithuta**.

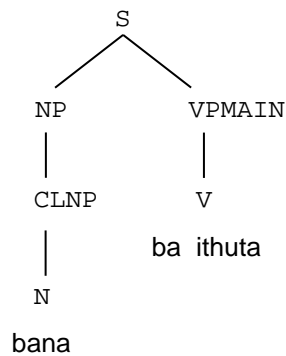
children they self learn

ba-ana ba-i-rut-a

NPre2-child AgrSubj2-ReflPre-learn-VEnd

The children learn.

(6-56) **c-structure**



f-structure

PRED	'RUT <SUBJ OBJ>'
MOOD	ind
TENSE	pres
POL	+
CLASS	2
SUBJ	[PRED 'ANA' CLASS 2]
OBJ	[PRED 'PRO']

← co-reference
relation between
SUBJ and OBJ

6.6 SUMMARY

The aim of this chapter was to describe the c-and f-structure of the simple sentence, which in Setswana consists of an NP (SUBJ) followed by a specific VP (predicate). A concise summary of this syntactic structure is presented in Table 6-1 (p.153). This description forms the basis for the XLE implementation presented in Chapter 7.

ADJUNCT	SUBJECT	PREDICATE				
		VERB	COMPLEMENT	OBJECT (INDIRECT & DIRECT)	OBLIQUE	ADJUNCTS
<u>Other</u> INTERJJP	CLNP PROP POSSPARTP QUALPARTP	Main verb	X	CLNP PROP POSSPARTP QUALPARTP	<u>OBL_{loc}</u> LOCCLNP LOCPARTPgo LOCPARTPga LOCPARTPka LOCPARTPfa LOCPARTPmo LOCPARTPkwa <u>OBL_{recip}</u> ASSPARTP	<u>Time</u> CLNP ADVP TEMPPARTP LOCPARTPmo <u>Place</u> LOCCLNP ADVP Absolute pronoun (16 - 20) Demonstrative pronoun (16 - 20) Inclusive quantitative pronoun (16 - 20) Exclusive quantitative pronoun (16 - 20) Separative quantitative pronoun (16 - 20) Communal possessive pronoun (16 - 20) POSSPARTP LOCPARTPgo LOCPARTPga LOCPARTPka LOCPARTPfa LOCPARTPmo LOCPARTPkwa
		Auxiliary verb	Main verb VPAUX	CLNP PROP POSSPARTP QUALPARTP	<u>OBL_{loc}</u> LOCCLNP LOCPARTPgo LOCPARTPga LOCPARTPka LOCPARTPfa LOCPARTPmo LOCPARTPkwa <u>OBL_{ass}</u> ASSPARTP	Manner ADVP INSTRPARTP COMPPARTP <u>Instrument</u> INSTRPARTP <u>Association</u> ASSPARTP COMPPARTP AGPARTP <u>Other</u> INTERJJP IDEOP
			VPIDCOP VPDESCOP VPASSCOP	X	X	
		Identifying copulative verb	CLNP PROP POSSPARTP QUAPARTP	(PREDLINK)	X	X
		Describing copulative verb	Adjective LOCPARTPgo, LOCPARTPga, LOCPARTPka, LOCPARTPfa LOCPARTPmo, LOCPARTPkwa LOCCLNP Absolute pronoun (16 - 20) Demonstrative pronoun (16 - 20) Quantitative pronoun (16 - 20) Adverbs of place Interrogatives kae?, jang?, go mang? POSSPARTP (16 - 20)	(PREDLINK)	X	X
		Associative copulative verb	Positive: ASSPARTP Negative: CLNP PROP ADVP POSSPARTP QUALPARTP	(PREDLINK)	X	X

Table 6-1: The syntactic structure of the Setswana simple sentence

CHAPTER 7

XLE IMPLEMENTATION OF THE SYNTACTIC STRUCTURE OF THE SETSWANA SIMPLE SENTENCE

7.1 INTRODUCTION

In this chapter, we describe the XLE implementation of the syntactic structure of the Setswana simple sentence in the LFG framework, as covered in Chapters 4 to 6. Together with the framing of Setswana syntax in LFG, it constitutes one of the main contributions of this study.

XLE is a computational environment for developing LFG grammars and provides, amongst others, functionality for building parsers and interfacing with finite state tokenisers and morphological analysers. XLE applies specifications of grammar rules, lexical entries and morphology to input sentences and other strings and then outputs the following:

- c-structure information, including the validity of the f-structures;
- an f-structure chart containing all the complete or incomplete bracketing of the input string that the grammar allows;
- morphological information, including all the possible morphological analyses of each lexical item;
- f-structure information, including the validity or not of the f-structure, the validity of which is determined by the presence of any inconsistency, incompleteness and incoherency in the grammar.

A complete description of the architecture and extensive capabilities of XLE⁸⁰ falls outside the scope of this study. Instead, the focus is on the *user interface* (§7.2) and those core aspects of XLE that are user defined and enable the development, testing and evaluation of an LFG grammar, specified in what we refer to as the *grammar file*. In §7.3 we give the macro structure of this file, while §7.4 to §7.7 is devoted to specific sections of the grammar file. The contents of this chapter is largely based on Crouch *et al.* (2015).

7.2 USER INTERFACE

XLE provides a powerful *user interface* for defining and manipulating linguistic rules and representations, as well as other grammar development functionality such as analysing

⁸⁰ The complete architecture of the XLE system and its user interface are explained in Butt *et al.* (1999:157-173).

performance, processing test suites and outputting analyses in a number of formats. We return to the latter development functionality in Chapter 8.

By way of explanation, we provide a Setswana example to illustrate typical XLE output. Figure 7-1 (p.158) shows the representation of the c- and f-structure of a parsed Setswana sentence (7-1). The four windows represent the following information:

- The upper left window presents the c-structure tree. The tree is displayed with the root at the top and the terminal nodes (leaves) that consist of the lexical items, at the bottom. The number of trees as well as the validity of a tree are presented in the title bar. The `Commands` menu includes menu items such as `resize` or `copy` the window, `print` the tree in different formats (`print postscript`, `print SExp`), `show` the input, `show` a morphology window, `chart` window or a `bracket` window and `check disjunctions`. The `Views` menu controls how the tree is displayed and includes a `node numbers` and a `partials` menu item. The `node numbers` menu item causes the tree to be drawn with or without visible node-numbers. However, the default display includes the node-numbers. "Each c-structure node that the parser discovers is assigned a unique node-number" and they "serve to correlate the units of the f-structure with the c-structure nodes they correspond to" (Walkthrough – Crouch *et al.*, 2015). The `partials` menu item "determines whether or not the partial constituents used internally by XLE to deal with multiple daughters are displayed" (XLE User Documentation – Crouch *et al.*, 2015).
- The lower left window shows the f-structure. This structure is displayed as an AVM structure in the standard LFG format. The number of solutions and the overall number of f-structures and their validity are displayed in the title bar. The `Commands` menu includes menu items to `resize` or `copy` the window and to `print` the f-structure in different formats (`print postscript`, `print LFG`, `print Prolog`). The `Views` menu controls how the f-structure is displayed and includes an `abbreviate attributes`, a `constraints`, a `node numbers`, a `subc constraints` and a `restrictions` menu item. "The abbreviated menu item suppresses all of the attributes except those that appear in the `abbrevAttributes Tcl` variable. The `constraints` menu item determines whether negated and sub-c constraints are included in the display. The `node numbers` menu item determines whether node numbers for each f-structure are displayed in a column along the left side of the f-structure. The `sub-c constraints` menu item determines whether or not sub-c constraints are included in the display" (XLE User Documentation – Crouch *et al.*, 2015).
- The upper right window shows the f-structure chart. This chart indexes the packed solutions by their constraints. Each constraint appears once in an f-structure that is annotated by all of the choices where that constraint holds. The `Views` menu controls how the f-structure is

displayed and includes an `abbreviate` attributes, a constraints, and a linear menu item. "The `abbreviate` menu item suppresses all of the attributes except those that appear in the `abbrevAttributes` Tcl variable. The constraints menu item determines whether negated and sub-c constraints are included in the display. Finally, the linear menu item changes the display into a line of tokens with corresponding f-structures" (XLE User Documentation – Crouch *et al.*, 2015).

- The lower right window shows the f-structure chart choices window. This window indexes the packed solutions by the alternative choices. Choices are labelled as `a:1`, `a:2...b:1`, `b:2` etc. (Figure 7-2, p.157). (XLE User Documentation – Crouch *et al.*, 2015). If there is only one solution applicable to a sentence, the chart does not show any information (Figure 7-1, p.158).

Figure 7-1 (p.158) shows that example (7-1) has a single valid tree and therefore one valid f-structure. As there is only one solution, the lower right window does not present any information. However, example (7-2) has two valid trees, and therefore the lower right window in Figure 7-2 (p.157) shows that there are two solutions (`a:1` and `a:2`).

(7-1) Mosadi o reka mosese.
 woman she buys dress
mo-sadi *o-rek-a* *mo-sese*
 NPre1-woman AgrSubj1-buy-VEnd NPre3-dress
 The woman buys a dress.

(7-2) O reka mosese.
 she buy dress
o-rek-a *mo-sese*
 AgrSubj1-buy-VEnd NPre3-dress
 She buys a dress.⁸¹

⁸¹ The subject agreement morpheme in this example can also be that of the second person singular:

(7-2) O reka mosese.
 you buy dress
o-rek-a *mo-sese*
 AgrSubjP2sg-buy-VEnd NPre3-dress
 You buy a dress.

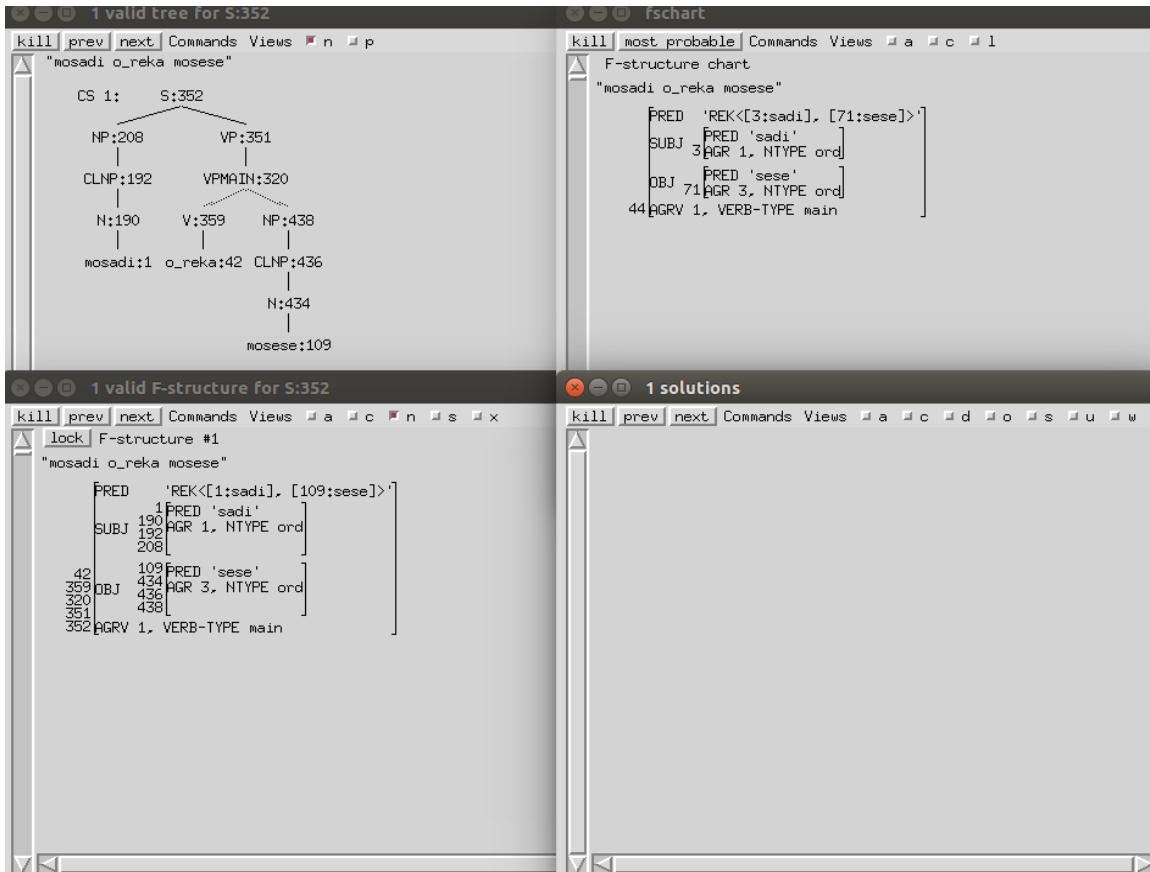


Figure 7-1: One solution for a sentence

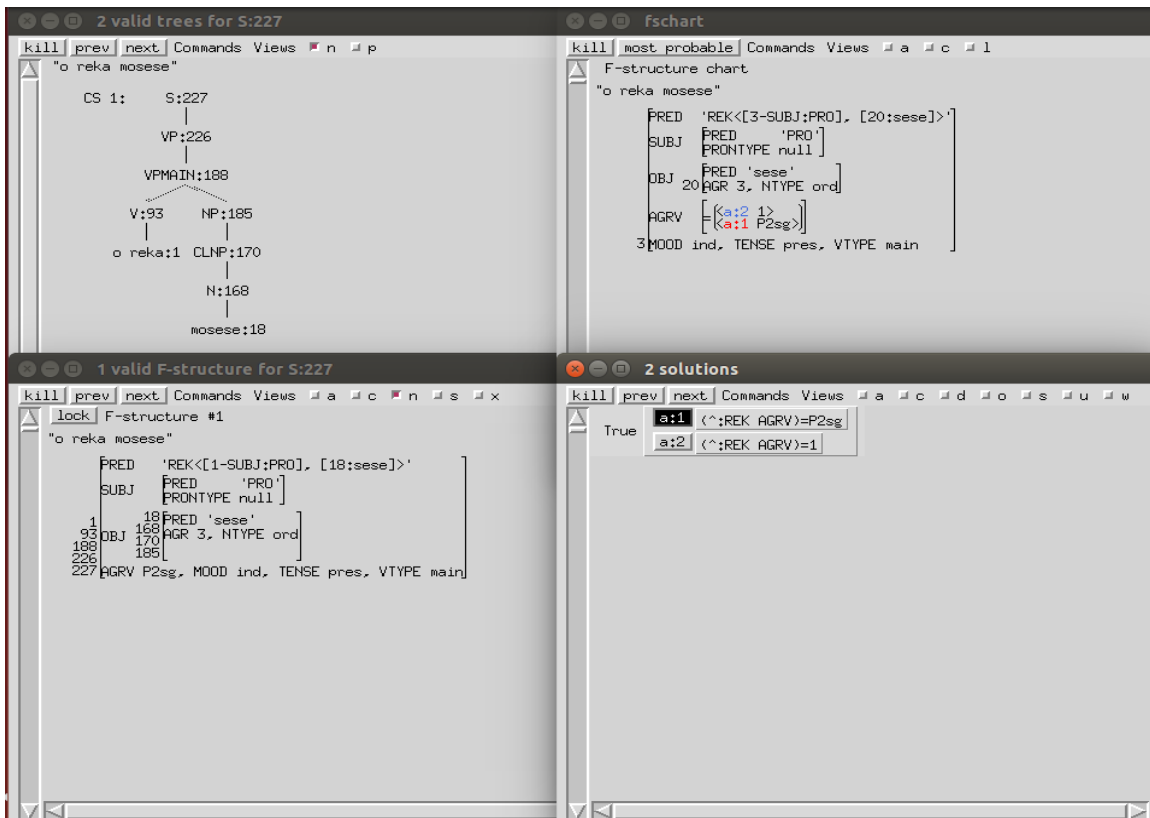


Figure 7-2: Two solutions for a sentence

7.3 GRAMMAR FILE

A grammar is implemented in a grammar (.lfg) file, divided into different sections (Diagram 7-1). Each section begins with a heading that includes the grammar version (SIMPLE_SENTENCE), the language (SETSWANA), the type of the section as well as the version (1.0). The different sections are CONFIG (§7.4), FEATURES (§7.4), RULES (c-structure rules with corresponding f-structure annotations and sublexical rules) (§7.8), TEMPLATES (§7.7), LEXICON (lexical entries) (§7.6) and MORPHOLOGY (tokenisation and morphological analysis) (§7.5).

```
SIMPLE_SENTENCE SETSWANA CONFIG (1.0)
ROOTCAT S.
FILES .
LEXENTRIES (SIMPLE_SENTENCE SETSWANA).
RULES (SIMPLE_SENTENCE SETSWANA).
TEMPLATES (SIMPLE_SENTENCE SETSWANA).
MORPHOLOGY (SIMPLE_SENTENCE SETSWANA).
GOVERNABLERELATIONS SUBJ OBJ OBJ-TH OBLloc OBLass PREDLINK.
SEMANTICFUNCTIONS ADJUNCT.
NONDISTRIBUTIVES NUM PERS.
EPSILON e.
OPTIMALITYORDER .
----
SIMPLE_SENTENCE SETSWANA FEATURES (1.0)
----
SIMPLE_SENTENCE SETSWANA RULES (1.0)
----
SIMPLE_SENTENCE SETSWANA TEMPLATES (1.0)
----
SIMPLE_SENTENCE SETSWANA LEXICON (1.0)
----
SIMPLE_SENTENCE SETSWANA MORPHOLOGY (1.0)
----
```

Diagram 7-1: Macro structure of the grammar file

7.4 CONFIGURATION AND FEATURES

The Configuration section (CONFIG) includes the following information (Walkthrough – Crouch et al., 2015):

- **ROOTCAT:** The root category shows the default category when a string is parsed. The value of the root category for the Setswana grammar is the sentence (S).
- **FILES:** A list of files to be included. In the Setswana grammar, no such files are used.
- **LEXENTRIES, RULES, TEMPLATES, MORPHOLOGY:** Here (`SIMPLE_SENTENCE SETSWANA`) indicates that the information in the relevant `SIMPLE_SENTENCE SETSWANA` sections of the grammar file is used.
- **GOVERNABLERELATIONS:** A list of grammatical relations that must be subcategorised for. In the Setswana grammar `SUBJ`, `OBJ`, `OBJ-TH`, `OBLloc`, `OBLass` and `PREDLINK` are included.
- **SEMANTICFUNCTIONS:** A list of attributes whose values must contain a `PRED`. For the Setswana grammar, `ADJUNCT` is included.
- **NONDISTRIBUTIVES:** A list of attributes that do not distribute when coordinated. For the Setswana grammar, `NUM` and `PERS` are included.
- **EPSILON:** The epsilon `e` indicates an empty string in the c-structure of the grammar and "will allow you to hang equations in rules where there is not a convenient constituent on which to do so" (Walkthrough – Crouch *et al.*, 2015).
- **OPTIMALITYORDER:** The ranking of optimality constraints. In the Setswana grammar, such information is not used⁸².

The `CONFIGURATION` section is followed by a feature declaration that specifies attributes and their possible values. The features that are applicable to the Setswana word categories (cf. §4.3 to §4.11) are listed in the `FEATURES` section (Diagram 7-2, p.160).

⁸² The ambiguity rate in a grammar can increase significantly when a grammar gets bigger. Ambiguity in an XLE grammar can be controlled by using optimality (OT) marks. The OT marks in a grammar are associated with particular rules, parts of rules or lexical entries and are used with a grammar to approve or disapprove a construction (Kaplan *et al.*, 2002:33–34). XLE uses a form of Optimality Theory that allows the grammar writer to indicate that certain constructions or lexical items are preferred, undesirable or totally unacceptable (Butt *et al.*, 1999:196–202; Falk, 2001:195–199).

```

-----
SIMPLE_SENTENCE  SETSWANA  FEATURES  (1.0)

CLASS: -> $ {1 2 1a 2a 3 4 5 6 7 8 9 10 11 14 15 16 17 18 19 20}.
PERS: --> $ {p1 p2}.
NUM: --> $ {sg pl}.
TENSE: -> $ {past pres fut relpast relfut}.
ASPECT: -> $ {prog pers perf}.
MOOD: -> $ {ind cons imp hab}.
POL: -> $ {pos neg}.
VERB-TYPE: -> {main rel aux idcopv descopv asscopv}.
PRESPRE: -> $ {+ -}.
INFLOBJ-TH: -> $ {+ -}.
INFLOBJ: -> $ {+ -}.
NTYPE: -> {ord adjective locn loccln}.
PRONTYPE: -> $ {perspro abspro dempro incqpro excqpro selqpro sepqpro intqpro
posspro null}.
PART-TYPE: -> $ {qualpart posspart asspart instrpart altpart agenpart
locpartgo locpartga locpartka locpartfa locpartmo locpartkwa tempart
compart}.
APPL: -> $ {+ -}.
RECIP: -> $ {+ -}.
CAUS: -> $ {+ -}.
PERF: -> $ {+ -}.
ADV: -> $ {+ -}.
AUX: -> {+ -}.
REFL: -> {+ -}.
INTERJ: -> {+ -}.
PASS: -> {+ -}.
INT: -> {+ -}.

-----

```

Diagram 7-2: Setswana grammar features and their values

7.5 MORPHOLOGY

The body of the MORPHOLOGY section in the Setswana grammar file consists of an obligatory tokenisation subsection (TOKENIZE) (Grammatical Notations – Crouch *et al.*, 2015) and a morphological analysis subsection (ANALYZE) for Setswana. In XLE, the tokenisation and morphological analysis constitute the pre-processing steps (Kaplan, *et al.*, 2004) and the availability of a finite-state tokeniser and a finite-state morphological analyser is assumed. In §1.1.4, we already confirmed the existence of such computational tools for Setswana.

Tokenisation is the segmentation of running text into *tokens* such as words, numbers, punctuation marks, parentheses and similar entities. In §1.1.1, we differentiated between an orthographic and a linguistic word. We reiterate that this is an important difference for Setswana tokenisation since the verb usually consists of multiple orthographic words but only one linguistic word (cf. (7-6)). Therefore, we consider tokens to be linguistic words, as explicated in detail in Pretorius (2014). Our example (7-3) represents a single token.

(7-3) o a o reka
 she it buy
o-a-o-rek-a
 AgrSubj1-PresPre-AgrObj3-buy-VEnd
 she buys it

Morphological analysis is the segmentation of a token (linguistic word) into its constituent morphemes (e.g. root, stem, prefixes and suffixes) as minimal meaningful units. The deep morphological structure of the various Setswana word categories is discussed in §4.4 to §4.11.

In example (7-4), the identified tokens (7-5) are separated by the @ symbol as the token boundary⁸³. The morphological analyses of these tokens are presented in (7-6). The token **yo** is ambiguous and therefore has two valid analyses, i.e. a demonstrative pronoun and a qualificative particle. Disambiguation occurs at the level of the relevant c-structure rules where the context comes into play.

(7-4) Mosadi yo o reka mosese.
 woman this she buys dress
mo-sadi yo o-rek-a mo-sese
 NPre1-woman DemPro1D1 AgrSubj1-buy-VEnd NPre3-dress
 This woman is buying a dress.

(7-5) mosadi@yo@o reka@mosese@

(7-6) mosadi
 sadi+NPre1

 yo
 yo+DemPro1D1
 yo+QualPart1

 o reka
 rek+Pres+AgrSubj1+VEnd

 mosese
 sese+NPre3

⁸³ The convention of using "@" as the token boundary follows, for example, Attia (2008), Çetinoğlu (2009) and Sulger (2015).

An important aspect is the adherence of our grammar to the *Lexical Integrity Principle* (cf. §3.2.2). In other words, how we ensure that tokens (typically verbs) that include disjunctively written morphemes occupy terminal node positions in the c-structure and do not allow these morphemes to be terminal nodes. The example (7-3) shows the token **o a o reka** (she buys it) as the only terminal node in Figure 7-3. Indeed, the verb prefixes **o**, **a** and **o** do not occupy terminal node positions in the c-structure.

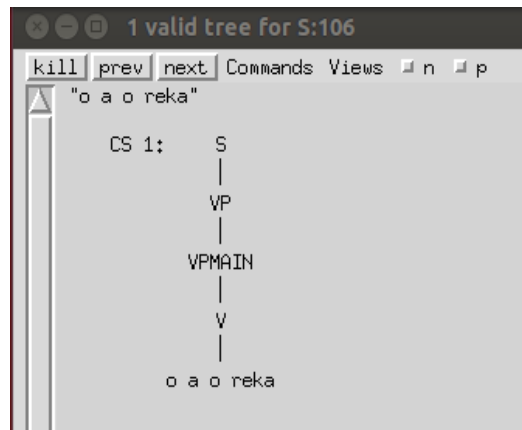


Figure 7-3: Terminal node of sentence **o a o reka** (she buys it)

7.6 LEXICON

Lexical entries reside in the LEXICON section of the grammar file. The basic structure of a lexical entry is as follows (Walkthrough – Crouch *et al.*, 2015):

word Category Morphcode Schemata.

The *word* is the base form of the word, *Category* is the c-structure or sublexical category, *Morphcode* is as below and *Schemata* are f-structure constraints as a list of attributes and values.

The Setswana lexical entries include stems and roots (basic and extended roots for verbs) (cf. §7.6.1 and §7.6.2), as well as entries that represent the morphological tagset (§7.6.3). A stem or a root entry consists of the following (Butt *et al.*, 1999:164-165; Grammatical Notations – Crouch *et al.*, 2015):

- **Base form:** In Setswana this base form is a stem for a word in a closed word category, a basic root for a noun and a basic or an extended root for a verb.
- **C-structure or sublexical category:** The c-structure categories include a noun, pronoun, verb, particle, conjunction, adverb, interjection and ideophone, as well as their

subcategories, where applicable (cf. §4.3 to §4.11). The sublexical categories have the form $x-SFX$, where x is a c-structure category.

- **Morphcode:** It is a marker that specifies the origin of the morphological information of the word (Butt *et al.*, 1999:169). XLE shows that the morphological analyser supplies the morphological information while an asterisk $*$ indicates that the information does not come from the morphological analyser and that the lexical item must be parsed as is. All the Setswana entries are marked with the morph code XLE as the morphological information comes from the morphological analyser.
- **F-structure constraints as a list of attributes and values:** An obligatory attribute is $(\wedge PRED)$ and its value is a so-called semantic form, an expression enclosed in single-quotes, used to make visible in the syntax only those aspects of semantic representation that interact in some way with syntactic properties. The semantic form can take one of three forms of which the following two are used in the Setswana lexical entries: '*function*' and '*function*<*a1..ak*>' where *function* is a designator for a semantic function and <*a1..ak*> are optional designators for one or more thematic argument functions *a1... ak*. In the Setswana grammar, the designator for the semantic function resembles the base form (for convenience). The optional designators are typically function-application expressions such as $(\wedge SUBJ)$, $(\wedge OBJ)$, etc. The attribute names in these expressions denote governable GFs such as $SUBJ$, OBJ , $OBJ-TH$, $OBLloc$, $OBLass$ and $PREDLINK$ ⁸⁴. The remainder of a lexical entry may also include any other relevant f-structure information in the form of attribute value pairs (Butt *et al.*, 1999:165).

7.6.1 -UNKNOWN LEXICAL ENTRY

The $-unknown$ lexical entry is a special form that matches any base form in the morphology that does not have an overt lexical entry (Walkthrough – Crouch *et al.*, 2015). Two types of $-unknown$ are distinguished (Butt *et al.*, 1999:166–167), i.e. unknown words that are not encoded in the lexicon, but that are known to the morphological analyser, and unknown words that are not encoded in the lexicon and that are furthermore not known to the morphological analyser.

Different *word categories* in the $-unknown$ entry are separated by the $;$ (semi-colon) symbol that is used to indicate where the annotations for a particular category end. In the Setswana grammar, the $-unknown$ entry provides for the following categories: the noun (N), locative noun ($LOCN$), locative class noun ($LOCCLN$), all the subcategories of the pronoun (viz. $(PERSPRO)$, $(ABSPRO)$,

⁸⁴ A detailed description of semantic forms may be found in the section *The functional description language* of Grammatical Notations in Crouch *et al.* (2015).

(DEMPRO), (POSSPRO), (INCQPRO), (EXCQPRO), (SELQPRO), (SEPQPRO), (INTQPRO), all the subcategories of the particles (viz. (POSSPART), (QUALPART), (INSTRPART), (TEMPPART), (LOCPART_{GO}), (LOCPART_{GA}), (LOCPART_{KA}), (LOCPART_{FA}), (LOCPART_{MO}), (LOCPART_{KWA}), (AGPART), (ASSPART), (COMPPART), (HORTPART), (INTPART), all the categories of the copulative verbs (viz. (IDCOPV), (DESCOPV), (ASSCOPV), the adverb (ADV), interjection (INTERJ), ideophone (IDEO) and the conjunctive (CONJ). An abbreviated version of the -unknown entry is given in (7-7).

The *morphcode* is XLE since our -unknown represents base forms, for which the morphological analyser provides the information. The (generic) value of the semantic form attribute (^PRED) is '%stem' and optional thematic argument function designators SUBJ, OBJ and PREDLINK. '%stem' is a special variable that is instantiated to -unknown and, ultimately, to the base form provided by the morphological analyser (Sulger, 2015:66). Other relevant f-structure information is also included, e.g. (^NTYPE)= ord.

```
(7-7)  -unknown N XLE (^PRED)='%stem' (^NTYPE)= ord;
...
PERSPRO XLE (^PRED)= '%stem' (^PRONTYPE)= perspro;
...
POSSPART XLE (^PRED)='%stem<(^OBJ)>' (^PARTTYPE)= posspart;
...
IDCOPV XLE (^PRED)='%stem<(^SUBJ) (^PREDLINK)>' (^VTYPE)= idcopv;
...
INT XLE (^PRED)= '%stem' (^INT)= +;
ADV XLE (^PRED)= '%stem' (^ADVERB)= +;
INTERJ XLE (^PRED) = '%stem' (^INTERJ)= +;
IDEO XLE (^PRED) = '%stem' (^IDEO)= +;
CONJ XLE (^PRED) = '%stem' (^CONJ)= +.
```

In §5.5, we stated that a particle phrase consists of a specific particle as head word, followed by a complement. In the Setswana grammar, this is modelled by allowing the particle to subcategorise for an OBJ, as shown in (7-7) for the POSSPART. This is in accordance with Butt *et al.* (1999:45) who state that non-verbal subcategorisation is allowed in LFG as predicates other than verbs can subcategorise for arguments too.

In §6.2.2.2, we argued that copulative verbs subcategorise for a SUBJ and a PREDLINK function. This is shown for IDCOPV in the -unknown entry in (7-7).

We conclude the description of the `-unknown` entry by making the following observations:

- This generic entry allows us to contain the size of the XLE lexicon by, for example, not repeating base forms that are already in the morphological analyser.
- Entries that have no or predictable subcategorisation frames are included in the `-unknown` entry (7-7).
- Closed categories are included for convenience. The Setswana noun class system of 20 classes requires 20 different words for most closed categories that are noun class specific.
- Main verbs are not included because their subcategorisation frames are semantically determined (cf. §4.6.1.5 and §6.2.2.1), and therefore are not predictable.
- Auxiliary verbs are not included since each auxiliary verb is used in a unique way to represent TAMP or a specific semantic value (cf. §4.6.2).

7.6.2 LEXICAL ENTRIES FOR MAIN AND AUXILIARY VERBS

In §4.6.1.5 and §6.2.2.1, we explained how the morphological complexity of the main verb, its basic and extended roots, verb extensions and the resulting argument structure lead to its associated subcategorisation frames. In the XLE grammar, the lexical entry for the main verb captures this information, as shown in (7-8) to (7-10).

(7-8) `batl V XLE { (^PRED)='BATL<(^SUBJ)>'`
`| (^PRED)='BATL<(^SUBJ) (^OBJ)>' } (^VTYPE)= main.`

(7-9) `batlis V XLE (^PRED)='BATLIS<(^SUBJ) (^OBJ-TH) (^OBJ)>'`
`(^VTYPE)= main.`

(7-10) `batlisan V XLE (^PRED)='BATLISAN<(^SUBJ) (^OBJ)>'`
`(^VTYPE)= main.`

In (7-8), the basic root **batl-** (seek) subcategorises for a `SUBJ` or a `SUBJ` and an `OBJ`. The extended root **batlis-** (let seek) in (7-9) consists of the basic root **batl-** (seek) and a causative extension **-is-** (let). This extended root subcategorises for a `SUBJ`, an `OBJ-TH` and an `OBJ`. However, the extended root **batlisan-** (let seek) in (7-10) consists of the basic root **batl-** (seek) and two verbal extensions, i.e. the causative **-is-** (let) and the reciprocal **-an-**. This extended root subcategorises for a `SUBJ` and an `OBJ`.

On the other hand, the semantic value of the basic verb root **lel-** in (7-11) and the fact that it is an intransitive verb allows only for a `SUBJ` as a `GF`.

(7-11) le1 V XLE (^PRED)='LEL<(^SUBJ)>' (^VTYPE)= main.

Figure 7-4 displays the c- and f-structure for sentence (7-12) with the verb root **batlis-** (helps seek).

(7-12) Rre o batlisa mme dinotlolo.

father he seek let mother keys

(-)-rre o-batl-is-a (-)-mme

NPre1a-father AgrSubj1a-seek-CausSuf-VEnd NPre1a-mother

di-notlolo

NPre8-key

Father helps mother seek the keys.

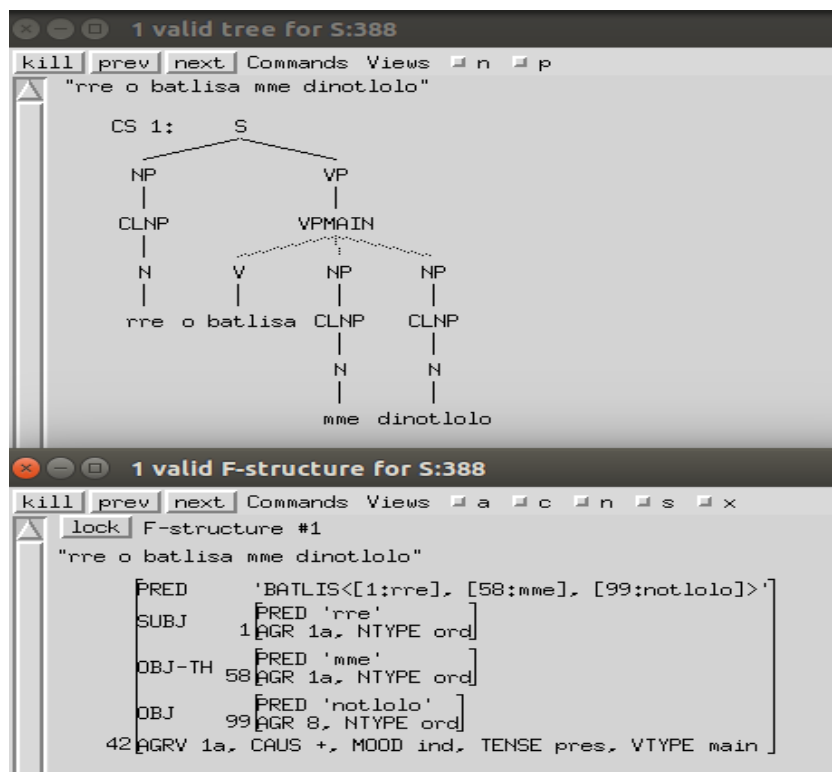


Figure 7-4: Verb subcategorises for a subject, an indirect object and a direct object

Auxiliary verbs were discussed in §4.6.2 and §5.8.5. Occurring together with a complementary verb, they enrich the meaning of this complement by adding TAMP or semantic information. In this sense, they are feature-carrying elements that do not have their own argument structure and allow a flat f-structure (Butt *et al.*, 1999:63). For example, in (7-13) and (7-14), the lexical entries for the auxiliary verbs **ne** and **tlabo** include f-descriptions that specify the relevant respective tenses.

(7-13) ne AUX XLE (^AUX)= + (^TENSE)= relpast.

(7-14) tlabo AUX XLE (^AUX)= + (^TENSE)= relfut.

7.6.3 LEXICAL ENTRIES FOR MORPHOLOGICAL TAGS

The purpose of lexical entries for morphological tags is to provide sublexical information for use in c- and f-structure constraints, thus forming an interface between the morphological analysis and the grammar. This is exemplified in (7-15) where the head is +PresPre, the sublexical category is V_SFX, the morph code is XLE and the functional information specifies the present tense, the present tense prefix and the indicative mood.

(7-15) +PresPre V_SFX XLE (^TENSE)=pres (^PRESPRE)=+ (^MOOD)=ind.

7.7 TEMPLATES

A template captures a linguistic generalisation and represents a shorthand definition of a set of information. Templates in XLE are defined as (arbitrarily complex) functions – they can take arguments that are in turn substituted for the variables in the template definition (Butt *et al.*, 1999:188). Templates reside in the TEMPLATES section and are mainly called from the lexicon, but can also be called by rules⁸⁵. For example, in (7-16), a call @(AGRV-C 1) to template AGRV-C with argument 1 is included as f-description information in the lexical entry for the tag +AgrSubj1. Similarly, in (7-18), a call @(NOMAGR-C 1) to template NOMAGR-C with argument 1 is included as f-description in the lexical entry for the tag +NPre1. In (7-17) and (7-19) the templates AGRV-C and NOMAGR-C are defined as functions of one variable _C. The generalisation is simply that when (^AGRV) = _C or (^AGR) = _C is executed with _C equal to 1, the agreement is checked with respect to noun class 1. Similarly, these templates can also be used to check subject-verb agreement for other noun classes, person and number.

(7-16) +AgrSubj1 V_SFX XLE @(AGRV-C 1).

(7-17) VAGR-C(_C) = (^AGRV) = _C.

(7-18) +NPre1 N_SFX XLE @(NOMAGR-C 1).

(7-19) NOMAGR-C(_C) = (^AGR) = _C.

⁸⁵ Templates are explained in some detail in Butt *et al.* (1999:188-192).

7.8 RULES

The c-structure rules and their f-structure annotations (cf. Chapters 3 to 6), as well as the sublexical rules (cf. §7.8.4) are implemented in the `RULES` section. A description of the complete XLE syntax for writing these rules, as well as the XLE notational conventions, may be found in Crouch *et al.* (2015). In the remainder of this section we describe various salient features of the implementation of the `SIMPLE_SENTENCE SETSWANA` grammar (cf. Diagram 7-1. p.158 and Diagram 7-2, p.160).

According to Crouch *et al.* (2015), the number of arcs is a good indication of the size of the grammar. Moreover, the number of disjuncts gives an indication of how many different rules the grammar would have if disjunctive constraints had not been allowed. The Setswana grammar has 67 rules, 210 states, 395 arcs and 398 disjuncts. We return to the scope and coverage of the grammar and the size of the test lexicon in Chapter 8.

Although the detailed description of XLE rule syntax falls outside the scope of this study, we briefly note the general form of these rules (Walkthrough - Crouch *et al.*, 2015):

```
category --> category1: schemata1;
           category2: schemata2;
           ... .
```

Categories represent the constituents in c-structure rules, for example `S`, `NP` and `VP` in (7-20). Schemata represent the f-structure annotations of these categories. For example, in (7-20) `NP` has two schemata specifying its f-structure, i.e. that it is the `SUBJ` of the sentence and agrees with the verb in the required way.

7.8.1 SIMPLE SENTENCE

The immediate constituents of the simple sentence, their GFs (cf. §6.2.1 to §6.2.2.3), the word order followed (cf. §6.3) and subject-verb agreement (cf. §6.4) are included in the top-level rule (7-20) of the simple sentence grammar. In this rule, it is specified that the first constituent is either an `NP` or⁸⁶ the empty string `e` and the second constituent is a `VP`.

The `NP` functions as the `SUBJ` of the sentence ($(\wedge \text{SUBJ}) = !$)⁸⁷. The agreement of the verb with the `SUBJ` is modelled by means of the constraining equality $(\wedge \text{SUBJ} \text{ AGR}) =_c (\wedge \text{AGR} \text{ V})$. The agreement information is obtained from the morphological analyser via morphological tags (as

⁸⁶ A disjunction includes choices placed between curly brackets `{ }` and separated by the pipe symbol `|`.

⁸⁷ See §3.5 for LFG notation relating c- and f-structure as well as (Grammatical Notations – Crouch *et al.*, 2015).

sublexical entries) (cf. 7-16 and 7-18), and transferred through these entries to the grammar via calls to the `VAGR-C` and `NOMAGR-C` templates ((7-17) and (7-19)) with the appropriate arguments.

In §6.5, we discussed Setswana as pro-drop language in which the subject agreement marker of the verb performs an anaphoric (pronominal) function if a sentence has no overt NP as subject. In XLE, we model this with the empty string `e` where the subject has `PRED` value `'PRO'`.

Finally, a set of optional interjections (`INTERJP*`), functioning as a single `ADJUNCT`, can be included at the beginning of a simple sentence (cf. §5.7 and §6.3).

```
(7-20) S --> [ INTERJP*: ! $ (^ADJUNCT) ]
          { NP: (^SUBJ)=! (^SUBJ AGR) =c (^AGRV)
            | e: (^SUBJ PRED)='PRO' (^SUBJ PRONTYPE)=null; }
          VP: ^=! .
```

The remainder of this section is devoted to the grammar rules that further refine this NP and VP. The next level rules in (7-21) and (7-22) are discussed in subsequent sections:

```
(7-21) NP --> { CLNP
                | PROP
                | QUALPARTP
                | POSSPARTP } .
```

```
(7-22) VP --> { VPMAIN
                | VPAUX
                | VPIDCOP
                | VPDESCOP
                | VPASSCOP } .
```

7.8.2 NOUN PHRASES

In this section, we consider the four different types of noun phrases (NPs) that are included in the rule in (7-21).

7.8.2.1 CLNP

The rule for `CLNP` (7-23) allows for `CLNP` to be a noun (N) or a noun (N) followed by one or more modifiers such as `PROP`, `POSSPARTP` etc. (cf. §5.2) that function as `ADJUNCTS`. The noun-modifier agreement feature is modelled with the constraining equality `(^AGR) =c (!AGR)`, which ensures

that the class of the modifier agrees with the class of the noun. Finally, $\sim(\text{^PRONTYPE}) = \text{PersPro}$) prevents a personal pronoun to function as a modifier. PersPro is a value of the attribute PRONTYPE , as specified in §7.4.

(7-23) $\text{CLNP} \rightarrow \text{N} (\{ \text{PROP: ! } \$ (\text{^ADJUNCT}) (\text{^AGR}) =_c (!\text{AGR}) \sim(\text{^PRONTYPE}) = \text{PersPro}$
 $\quad | \text{POSSPARTP: ! } \$ (\text{^ADJUNCT}) (\text{^AGR}) =_c (!\text{AGR})$
 $\quad \dots \})$.

7.8.2.2 PROP

The rule for PROP (7-24) implements all the different pronoun subcategories, as discussed in §5.4, as well as other pronouns or phrases that modify these pronouns. These modifiers function as ADJUNCTS . As in §7.8.2.1, the constraining equality $(\text{^AGR}) =_c (!\text{AGR})$ ensures appropriate agreement.

(7-24) $\text{PROP} \rightarrow \{ \text{PERSPRO} (\text{INCQPRO: ! } \$ (\text{^ADJUNCT}) (\text{^AGR}) =_c (!\text{AGR}))$
 $\quad | \text{ABSPRO} (\text{DEMPRO: ! } \$ (\text{^ADJUNCT}) (\text{^AGR}) =_c (!\text{AGR}))$
 $\quad \dots \}$.

7.8.2.3 POSSPARTP

The POSSPARTP rule (7-25) ensures that a possessive particle is followed by an obligatory complement (cf. §5.5.1) that functions as an OBJ . In addition, the absolute and personal pronouns are prevented to function as complements (cf. §4.5 and §5.5.1). The defining equality $(\text{^OBJ AGR}) =_{16}$ implements the class restrictions that are discussed in Table 5-7 (p.111) with respect to successive possessive particles, as in example (5-9).

(7-25) $\text{POSSPARTP} \rightarrow \text{POSSPART} \{ \text{PROP: } (\text{^OBJ}) = ! \sim(\text{^OBJ PRONTYPE}) = \text{AbsPro}$
 $\quad \quad \quad \sim(\text{^OBJ PRONTYPE}) = \text{PersPro}$
 $\quad | \text{POSSPRO: } (\text{^OBJ}) = !$
 $\quad | \text{ADV P: } (\text{^OBJ}) = !$
 $\quad | \text{CLNP: } (\text{^OBJ}) = !$
 $\quad | \text{QUALPARTP: } (\text{^OBJ}) = ! \}$
 $\quad | \text{POSSPARTP: } (\text{^OBJ}) = ! (\text{^OBJ AGR}) =_{16} \}$.

7.8.2.4 QUALPARTP

The QUALPARTP rule (7-26) ensures that a qualificative particle is followed by an obligatory complement that functions as an OBJ (cf. §5.5.2). The agreement feature between the

qualificative particle and some complements is modelled with the constraining equality (\wedge AGR) =_c (!AGR) as before.

(7-26) QUALPARTP --> QUALPART { ADJECTIVE: ! \$ (\wedge OBJ) (\wedge AGR) =_c (!AGR)
 | RELV: ! \$ (\wedge OBJ) (\wedge AGR) =_c (!AGR)
 | ADVP: (\wedge OBJ)=!
 | SEPPRO: ! \$ (\wedge OBJ) (\wedge AGR) =_c (!AGR)}.

7.8.3 VERB PHRASES

In this section we consider the five different types of verb phrases (VPS) (cf. §5.8) that occur in the rule in (7-22).

7.8.3.1 VPMAIN

The VPMAIN rule (7-27) implements a main verb and optional phrases that function as OBJ-TH, OBJ, OBL or ADJUNCT. It allows for all the subcategorisation frames discussed in Chapters 4 and 6. We highlight the following essential aspects encoded in the schemata in this rule:

- The constraining equality $\sim(\wedge$ PRESPRE)= + ensures that the present tense morpheme **a** is not included in any verb that is followed by an OBJ-TH, OBJ, OBL or ADJUNCT (cf. §4.6.1.2). Figure 7-5 (p.172) shows the incorrect use of the present tense morpheme **a**, and how this error manifests in the f-structure.
- The schemata (\wedge INFOBJ PRED)='PRO' (\wedge INFOBJ PRONTYPE)=null ensure that an object agreement morpheme is modelled as an *inflectional object*, which shows correspondence to the class, person and number of the omitted object, and has a pronominal value (cf. §6.5), as exemplified in Figure 7-6 (p.173). Similarly, the schemata (\wedge INFOBJ-TH PRED)='PRO' (\wedge INFOBJ-TH PRONTYPE)=null and (\wedge INFOBJ PRED)='PRO' (\wedge INFOBJ PRONTYPE)=null model the occurrence of two object agreement morphemes (cf. §6.5) and is exemplified in Figure 7-7 (p.173).
- ADJUNCT phrases are modelled as the grouping [ADVP*: ! \$ (\wedge ADJUNCT), TEMPPARTP*: ! \$ (\wedge ADJUNCT), ...]* where * is the Kleene star and the , the shuffle operator (Grammatical Notations – Crouch *et al.*, 2015) so that an arbitrary number of ADJUNCTs can occur in free order. Figure 7-8 and Figure 7-9 (p.174) show the inclusion of three ADJUNCTs in different orders. Figure 7-8 (p.174) shows the use of the three ADJUNCTs, i.e. **kwa lebenkeleng** (in the shop), **ka karata** (with a card) and **jaanong** (now) and Figure 7-9 (p.174) shows the same ADJUNCTs in another order, namely **ka karata** (with a card), **kwa lebenkeleng** (in the shop) and **jaanong** (now).

```

(7-27) VPMAIN --> V {({NP: (^OBJ)=! ~(^PRESPRE)= +
| NP: (^OBJ-TH)=! ~(^PRESPRE)= + ;
| NP: (^OBJ)=! ~(^PRESPRE)= +
| NP: (^INFOBJ PRED)='PRO' (^INFOBJ PRONTYPE)=null
| (^PRESPRE)= +
| NP: (^INFOBJ-TH PRED)='PRO'
| (^INFOBJ-TH PRONTYPE)=null (^PRESPRE)= + ;
| NP: (^INFOBJ PRED)='PRO' (^INFOBJ PRONTYPE)=null
...})
| {LOCPARTPka: (^OBLloc)=! ~(^PRESPRE)= +
| LOCPARTPfa: (^OBLloc)=! ~(^PRESPRE)= +
...})}
[ADVP*: ! $ (^ADJUNCT),
TEMPPARTP*: ! $ (^ADJUNCT),
... ]*.

```

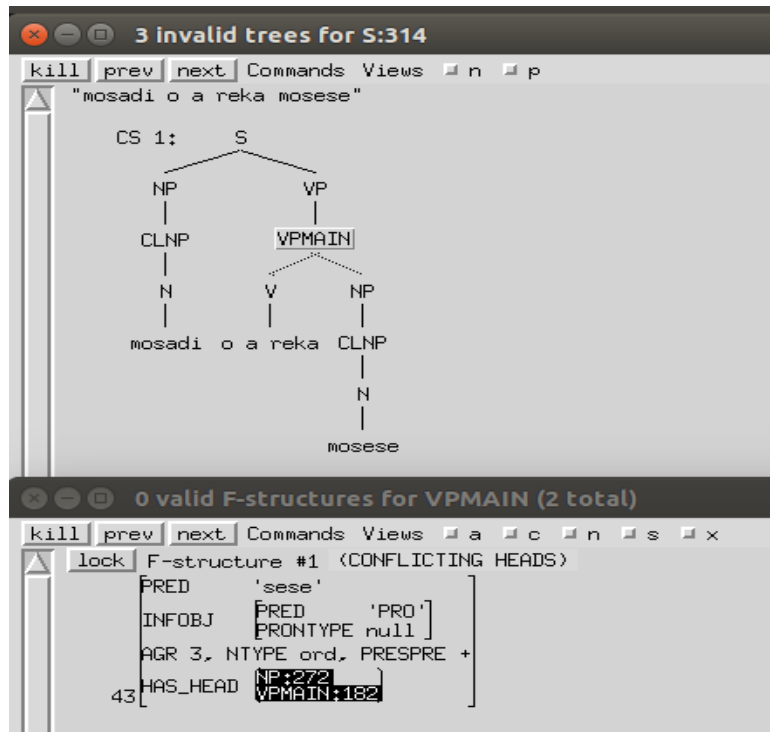


Figure 7-5: The incorrect use of the present tense morpheme a

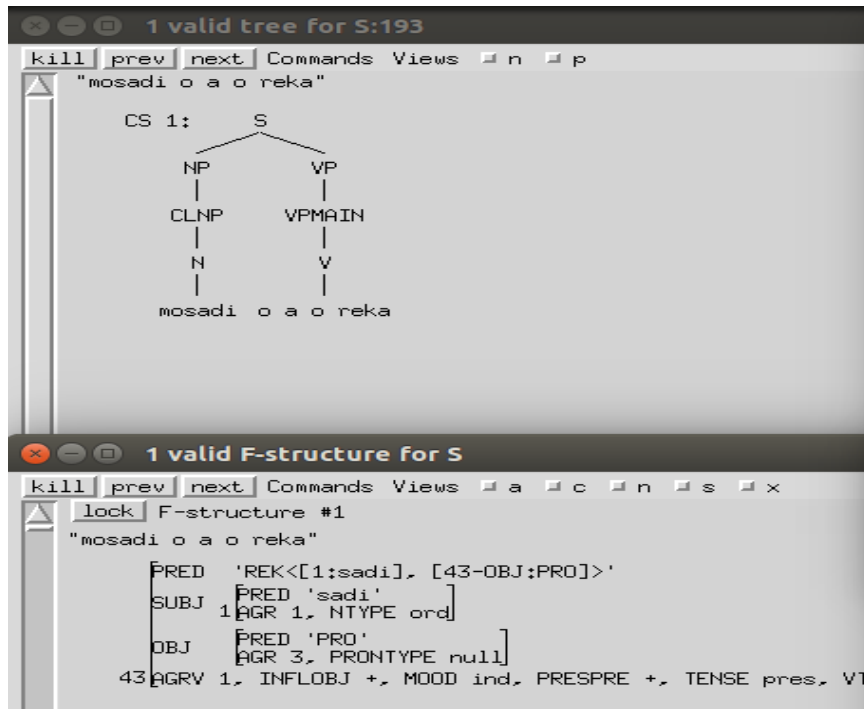


Figure 7-6: One object agreement morpheme

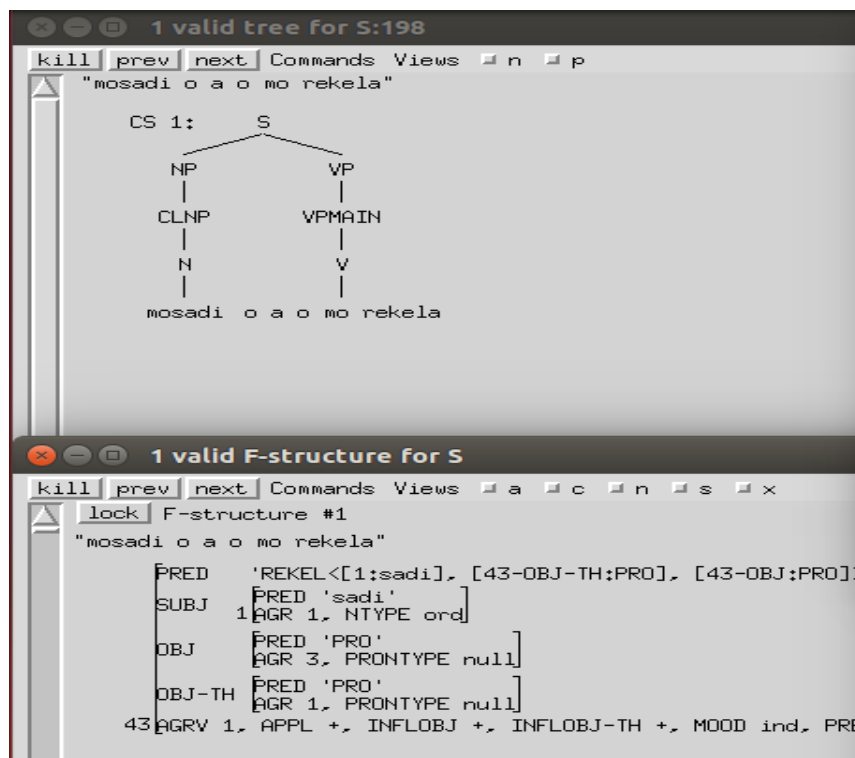


Figure 7-7: Two object agreement morphemes

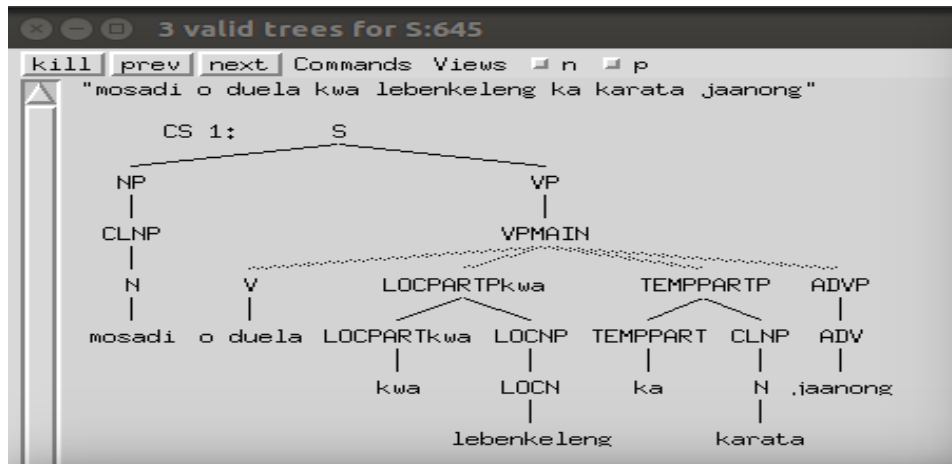


Figure 7-8: Sentence with three adjuncts

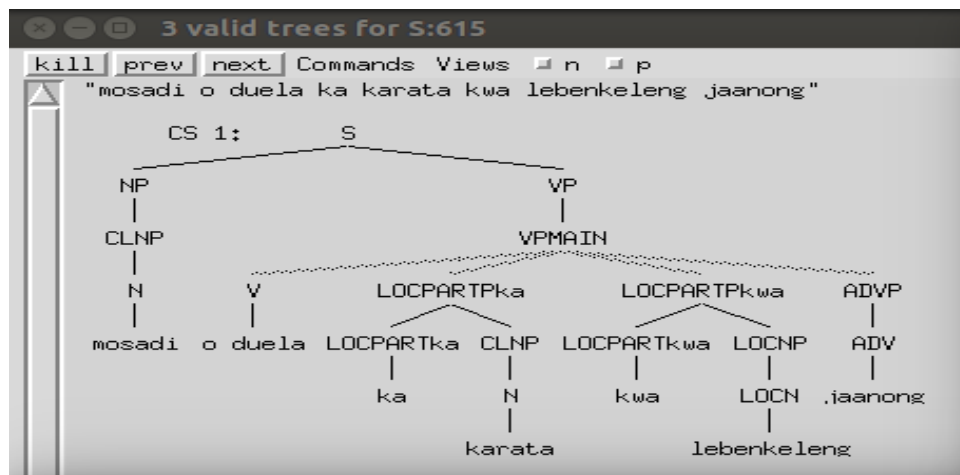


Figure 7-9: Inclusion of three adjuncts with a different order

7.8.3.2 VPAUX

The VPAUX (7-28) rule ensures that an auxiliary verb AUX (cf. §7.6.2) is followed by an obligatory VP as complement (cf. §5.8.5).

(7-28) VPAUX --> AUX VP.

7.8.3.3 VPIDCOP, VPDESCOP and VPASSCOP

As discussed in (cf. §6.2.2.2), the closed complement analysis (([^]PREDLINK)=!) is used to implement the VPIDCOP, VPDESCOP and VPASSCOP rules ((7-29) to (7-31)).

The VPIDCOP rule (7-29) requires the identifying copulative verb IDCOPV (cf. §7.6.1) to be followed by an obligatory NP as complement (cf. §5.8.2).

(7-29) VPIDCOP --> IDCOPV NP: (^PREDLINK)=!
 [ADV* : ! \$ (^ADJUNCT),
 TEMPPART* : ! \$ (^ADJUNCT),
 ...]*.

The VPDESCOP rule (7-30) requires the describing copulative verb DESCOPV (cf. §7.6.1) to be followed by an obligatory complement, i.e. ADJECTIVE and LOCCLNP, as listed in §5.8.3.

(7-30) VPDESCOP --> DESCOPV {ADJECTIVE: (^PREDLINK)=!
 | LOCCLNP: (^PREDLINK)=!
 ...}
 [ADV* : ! \$ (^ADJUNCT),
 TEMPPART* : ! \$ (^ADJUNCT),
 ...]*.

The VPASSCOP rule (7-31) requires an associative copulative verb ASSCOPV (cf. §7.6.1) to be followed by an obligatory complement, i.e. an ASSPART in the positive and an NP or ADVP in the negative (§5.5.7).

(7-31) VPASSCOP --> ASSCOPV {{ASSPART: (^PREDLINK)=! (^POL)=pos
 | {NP: (^PREDLINK)=! (^POL)=neg
 | ADVP: (^PREDLINK)=! (^POL)=neg}}}
 [ADV* : ! \$ (^ADJUNCT),
 TEMPPART* : ! \$ (^ADJUNCT),
 ...]*.

As in the case of VPMAIN, an arbitrary number of ADJUNCTS in free order can follow the respective complements of the copulative verbs.

7.8.3.4 Phrases functioning as obliques and adjuncts

We reiterate that a main verb may subcategorise for an OBL, as shown in (7-27). In this section, the focus is on those phrases that perform this function, viz. LOCNP, LOCCLNP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{fa}, LOCPARTP_{mo}, LOCPARTP_{kwa}, and ASSPART. Moreover, the VPMAIN (7-27), VPIDCOP (7-29), VPDESCOP (7-30) and VPASSCOP (7-31) rules also allow for phrases that function as ADJUNCTS, viz. LOCNP, LOCCLNP, LOCPARTP_{go}, LOCPARTP_{ga}, LOCPARTP_{ka}, LOCPARTP_{fa}, LOCPARTP_{mo}, LOCPARTP_{kwa}, ASSPART, ADVP, TEMPPARTP, INSTRPARTP, COMPPARTP, INTERJP, IDEOP, AGPARTP, PROP (§7.8.2.2) and

POSSPARTP (§7.8.2.3). In Chapter 5, the structures of these phrases were discussed in detail. In this section, we use XLE rule fragments to illustrate salient aspects of these phrases, i.e. agreement and non-verbal subcategorisation.

Phrases that require agreement include:

```
LOCNP --> LOCN ( {PROP: ! $ (^ ADJUNCT) (^AGR) =c (!AGR) ~(^PRONTYPE)= PersPro
                  | ...} ).
LOCCLNP --> LOCCLN (POSSPARTP: ! $ (^ ADJUNCT) (^AGR) =c (!AGR)).
```

Phrases with particles that have non-verbal subcategorisation frames (cf. §7.6.1) involving complements that function as OBJs include:

```
INSTRPARTP --> INSTRPART {ADVP: (^OBJ)=! | ...}.
TEMPPARTP --> TEMPPART CLNP: (^OBJ)=!.
LOCPARTPgo --> LOCPARTgo {PROP: (^OBJ)=! | CLNP: (^OBJ)=!}.
LOCPARTPga --> LOCPARTga {PROP: (^OBJ)=! | CLNP: (^OBJ)=!}.
LOCPARTPka --> LOCPARTka {CLNP: (^OBJ)=! | LOCPARTPfa: (^OBJ)=! | ...}.
LOCPARTPmo --> LOCPARTmo {LOCNP: (^OBJ)=! | LOCCLNP: (^OBJ)=! | ...}.
LOCPARTPfa --> LOCPARTfa {LOCNP: (^OBJ)=! | LOCCLNP: (^OBJ)=! | ...}.
LOCPARTPkwa --> LOCPARTkwa {LOCNP: (^OBJ)=! | LOCCLNP: (^OBJ)=! | ...}.
ASSPARTP --> ASSPART {NP: (^OBJ)=! | ADVP: (^OBJ)=! | ...}.
COMPPARTP --> COMPPART {NP: (^OBJ)=! | ADVP: (^OBJ)=! | ...}.
AGPARTP --> AGPART {NP: (^OBJ)=! | ADVP: (^OBJ)=! | ...}.
```

Phrases that consist of lexical entries only include:

```
ADVP --> {ADV | INT}.
INTERJP --> INTERJ.
IDEOP --> IDEO.
```

7.8.4 SUBLEXICAL RULES

Sublexical rules provide the interface between the finite state morphological analyser and the XLE grammar by parsing the output of the morphological analyser into a base form and as many tags as the morphology provides. In the Setswana grammar, the base form is either a root or a stem depending on the specific word category to which the surface form belongs (cf. §7.6) and the tags that represent the prefixes and suffixes that occur in the output. The order provided by the morphology is followed (Butt *et al.*, 1999:164–165). Sublexical rules reside in the RULES section of the grammar file.

For example, the output of the morphological analysis of the main verb **o reka** (she buys) (7-32) includes the verb root **rek** and the morphological tags **+Pres**, **+AgrSubj1** and **+VEnd**. The sublexical rule (7-33) parses such a Setswana main verb into a root (**V_BASE**) and morphological tags (**V_SFX_BASE+**).

(7-32) o reka
 rek+Pres+AgrSubj1+VEnd

(7-33) V --> V_BASE
 V_SFX_BASE+.

The sublexical information is not automatically displayed in the default c-structure representation. However, this information is available through an expanded display mode in the c-structure output of XLE (Figure 7-10).

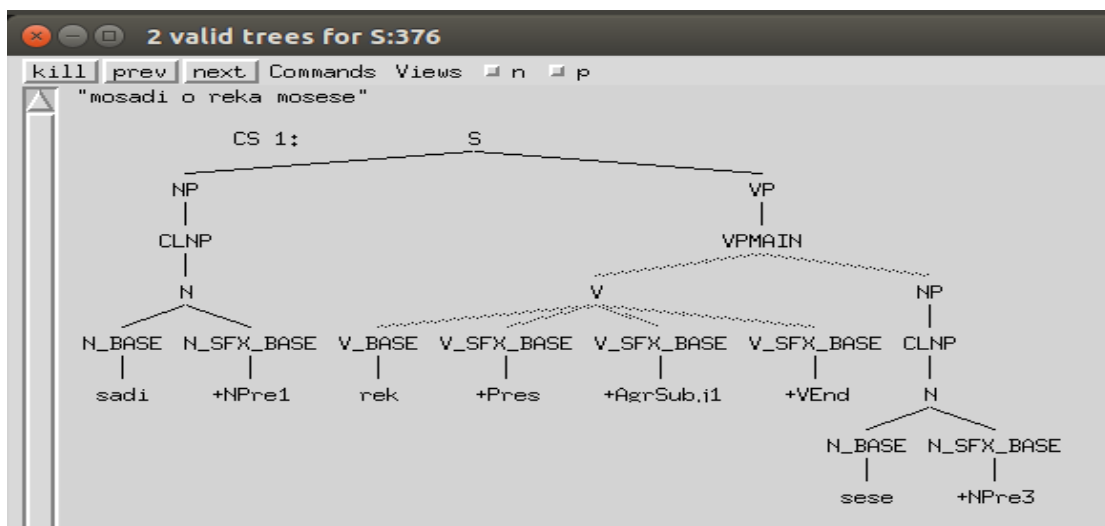


Figure 7-10: Expanded display mode of tree showing sublexical information of example (7-1)

7.9 SUMMARY

In this chapter, we firstly discussed the XLE grammar development platform with specific reference to the user interface available to the grammar developer. Secondly, we used the structure of the Setswana grammar file to describe our implementation of the syntactic structure of the Setswana simple sentence in XLE. In Chapter 8, we test and evaluate the grammar and report on a novel treebank for Setswana, which is another contribution of this study.

CHAPTER 8

TESTING THE GRAMMAR

8.1 INTRODUCTION

In this chapter, we describe the testing of the Setswana grammar. We explicate why and how we handcrafted a test suite that covers all the salient syntactic forms modelled and implemented in the grammar. The test suite contains both grammatical and ungrammatical forms for confirming the accuracy of the grammar. We also discuss the results obtained when applying the parser (i.e. our XLE implemented computational grammar) to the test suite. Finally, we show how the grammatical phrases and sentences form the basis for a novel treebank for Setswana.

8.2 TESTING A COMPUTATIONAL GRAMMAR

As mentioned in §1.1, a grammar is a representation of the rules that govern the construction of syntactic units, including phrases and sentences, in a particular framework such as LFG. In turn, a computational grammar (also referred to as a rule-based parser, cf. §1.1.2) is an implementation of this grammar by means of a grammar development platform such as XLE.

Broadly speaking, two main methodologies are used for testing a computational grammar:

- Firstly, a corpus-based test procedure assumes the existence of a gold standard, i.e. a text corpus appropriately annotated with syntactic information, often referred to as a treebank (cf., for example, Abiellé, 2003; Dickinson *et al.*, 2015). The grammar is then applied to the gold standard corpus (treebank) and measures such as accuracy, recall, precision and F-score, are computed (Manning & Schütze, 2000:267–270). A grammar with high scores is considered better than one with low scores.
- Secondly, test suites are often used (cf. for example, Lehmann *et al.*, 1996; Lloberes *et al.*, 2014). In this case, a grammar is considered tested if specially constructed linguistic examples, both grammatical and ungrammatical, are included in the test suite for each of the major types of constructions that form part of the grammar (Butt *et al.*, 1999:205; Marciniak *et al.*, 2003:130; Lloberes *et al.*, 2014:87). We follow Butt and King (2003:163) by distinguishing between three basic sources for test suites, i.e. hand-crafted data by grammar writers themselves, standard test suites if available, and selected real world data often from specialised domains.

As discussed in §1.1.2, Setswana is an under-resourced language for which, as far as we know, no computational grammar and no treebanks or test suites for deep syntactic analysis exist.

Therefore, our methodology *ipso facto* entails the development of a comprehensive test suite that covers the constructions discussed in Chapters 4 to 6, with the exception of coordination, which forms part of our future work.

8.3 GRAMMAR SCOPE AND COVERAGE

The scope of the study is the Setswana simple sentence and its syntactic structure. We reiterate that within this scope, the computational grammar covers all the relevant syntactic constructions (cf. §8.4), except coordination.

As discussed in §1.5, the Setswana lexicon in XLE is restricted. However, it is, nevertheless carefully crafted to include all the salient features of a comprehensive Setswana lexicon needed to implement and test the parser. Included are:

- 43 verb roots, basic and extended, specifically selected to cover the complexity of the verb in terms of possible subcategorisation frames;
- six auxiliary verbs that are used to indicate TAMP in the simple sentence, as discussed in §6.2.2.3 and §7.6.1;
- all other word categories and subcategories that do not have overt lexical entries, as discussed in § 7.6.1, and are catered for by the `-unknown` lexical entry;
- all the morphological tags that occur in the morphological analysis, provide sublexical information and act as an interface between the XLE grammar and the morphological analyser, as discussed in §7.6.3.

As a pre-processing step, tokenisation isolates valid words in a sentence or phrase. XLE inputs these tokens to the morphological analyser (cf. §8.3.4) to obtain sublexical information. While a finite-state tokeniser for Setswana is available and will in due course be included in a broad-coverage Setswana grammar, it does not have an accuracy of 100%. Therefore, for the purposes of testing the parser it is undesirable to allow errors of tokenisation to impact on such testing. For this purpose, a tokeniser was fine-tuned to provide *perfect tokenisation for all the test items in the test suite*.

Morphological analysis is a second pre-processing step in XLE. As is the case for the tokeniser, the available finite-state morphological analyser for Setswana is also fine-tuned to provide the *correct analyses for all the tokens in the test suite*. This is done to prevent errors of morphological analysis to influence the accuracy of the grammar. More specifically, we determine whether it has the intended coverage (parses all specified valid units) and whether it overgenerates (parses any invalid units) (Butt & King, 2003:171).

The next step in our testing methodology is to *construct a test suite* of grammatical and ungrammatical phrases and sentences that will cover the implemented grammar and ensure its correctness, i.e. whether the implemented grammar has the intended coverage (parses all specified valid units) and whether it overgenerates (parses any invalid units) (Butt & King, 2003:171).

8.4 TEST SUITE

In §8.2 we argued for a comprehensive test suite as the preferred testing methodology for our grammar. Our test strategy is based on two general principles (Butt *et al.*, 1999:205; Butt & King, 2003:163; Marciniak *et al.*, 2003:130): Firstly, test simple structures and phrases, and, secondly, test complex structures by systematically combining already tested (simple) structures.

In Table 8-1 (p.181), we summarise important *linguistic characteristics* of Setswana syntax (cf. §1.1.1) that should be covered by the test structures in our comprehensive test suite. We reiterate that the focus is on the noun class system, concordial agreement, and phrase structure coverage of the Setswana simple sentence.

In Part 1 of Table 8-1 (p.181), we focus on subject-verb agreement. The large number of grammatical test items results from noun class information, the inclusion of TAMP features as well as combinations of these. For testing purposes, we limit the noun classes to 1, 1a, 2 and 2a, and also test for person and number (first and second person, singular and plural) without loss of generality in order to keep the test suite as small as possible. In the ungrammatical test items, we violate agreement by simply reversing class 1 and class 2, class 1a and class 2a, the first person singular and plural, and the second person singular and plural. We consider one such ungrammatical item in each case as sufficient to test subject-verb agreement. In terms of subcategorisation, we randomly selected a small number of test items and removed one GF from each frame.

Part 2 of the table concerns phrase internal agreement in the various types of noun phrases, as well as in the locative noun phrase. The large number of grammatical test items results from agreement constraints imposed on modifiers and complements in the various noun phrases and locative noun phrases. The number of ungrammatical test items is somewhat smaller since ungrammaticality can only arise in multiword phrases.

Part 3 of the table shows that we only check for phrasal structure coverage using grammatical test items. The number of test items for the adverb, agentive particle, interjection and ideophone

phrases are small since extending these is a matter of extending the lexicon. A large number of ADJUNCTS is included to make provision for their free order and arbitrary number. We tested the inclusion of one to five ADJUNCTS in sentences that have a main or copulative verb. Each pro-drop item depends on a subject agreement morpheme and a verb. Finally, test items should allow for one or two object agreement morphemes in a main verb phrase or an auxiliary verb phrase. Ungrammatical test items are mostly irrelevant in Part 3. However, multiword locative class noun phrases include ungrammatical test items since these phrases are subject to agreement constraints.

Linguistic characteristics	Grammatical test items	Ungrammatical test items
1. Subject-verb agreement (noun classes and person and number)		
mood (indicative, consecutive, imperative and habitual)	154	154
tense (present, past, future, relative past and relative future)	68	68
aspect (perfective, progressive and persistive)	41	41
polarity (positive and negative structures)	163	163
main verbs and their subcategorisation frames	82	15
copulative verbs	47	47
auxiliary verbs	15	15
2. Phrase internal agreement		
noun phrases	115	64
locative noun phrase	24	22
3. Phrasal rule (c-structure) coverage		
particle phrases	97	-
locative class noun phrase	9	5
adverb phrase	6	-
agentative particle phrase	4	-
interjection phrase	1	-
ideophone phrase	1	-
ADJUNCTS in the sentence	70	-
pro-drop	50	-
function of object agreement morphemes	33	-

Table 8-1: Number of test items for various linguistic characteristics in the test suite

In terms of the *lexicon*, we note that a hand-crafted test suite does not test for lexical coverage (Balkan *et al.*, 1994:5), but rather focusses on capturing the *salient features* of the Setswana lexicon by selecting a small number of entries. The number of lexical entries that occur in the test suite are given in Table 8-2 (p.182).

Lexical entries	Number
Nouns in classes 1 to 15	35
Locative nouns in classes 4, 5, 9 and 10	5
Locative class nouns in classes 16–20	5
Adjectives	6
Absolute, demonstrative and quantitative pronouns in classes 1 to 15	34
Personal pronouns of the first and second person (singular and plural)	4
Possessive pronouns	5
Adverbs	8
Interjection	1
Ideophone	1
Particles	41
Main verbs (basic and extended verb roots)	43
Auxiliary verbs	6
Copulative verbs	7

Table 8-2: Number of lexical entries in the test suite

The test suite contains 828 grammatical⁸⁸ and 365 ungrammatical test items. A summary of the number of grammatical test items per word length⁸⁹ in the test suite is given in Table 8-3.

Number of test items	Word length
38	1
98	2
145	3
175	4
173	5
133	6
43	7
8	8
1	9
4	10
5	11
5	12

Table 8-3: Number of test items per word length in test suite

⁸⁸ This test suite is available at <http://repo.sadilar.org/handle/20.500.12185/478>

⁸⁹ Word length refers to the the number of words in a sentence

The noun stems and verb roots that occur in the test suite are given in Appendix B (Table 8-4, p.199 and Table 8-5, p.200).

8.5 RESULTS

To summarise, in terms of our testing methodology, we hand-crafted a comprehensive test suite to demonstrate the scope, coverage and accuracy of our XLE grammar for the Setswana simple sentence. Our expectation is that, by construction, each grammatical item will obtain at least one valid parse. In this section, we briefly discuss the outcome of applying the XLE grammar to the test suite.

On executing the “parse-testfile” command in XLE, we obtain three files. One file includes the test items and statistics, one file includes just the statistics and one file includes the differences between the current test and the previous test. The statistics file shows performance information about each one of the test items (XLE User documentation – Crouch *et al.*, 2015). For example, the performance information of the parsed CLNP **mosadi yo** (this woman), as presented in (8-1), indicates that there is one solution (i.e. valid parse) and that it took 0.003 seconds to process 11 subtrees. The performance information of the parsed sentence **basadi ga ba ketla ba reka mosese** (the women will not buy a dress) indicates two solutions and it took 0.002 seconds to parse 40 subtrees (8-2).

(8-1) CLNP: mosadi yo (1 0.003 11)

(8-2) basadi ga ba ketla ba reka mosese (2 0.002 40)

Table 8-6 (p.184) provides a summary of the number of items in the test suite with the number of valid parses per item. All grammatical items in the test suite were parsed, of which 518 had only one grammatical parse. Moreover, 220 had two grammatical parses, mainly due to Setswana mood ambiguity. Similarly, the items with 3 to 8 grammatical parses are also due mainly to mood ambiguity as well as multiple ADJUNCTS including ADJUNCTS with **ka** as either an instrumental, temporal or locative particle (cf. §4.7), all of which take a CLNP as complement. In addition to mood ambiguity, the items with 12 to 22 grammatical parses have multiple ADJUNCTS including ADJUNCTS with **ga** as either a locative particle or possessive particle of classes 15 to 20 (cf. §4.7), all of which also take a CLNP as complement.

Test items	Valid parses
518	1
220	2
30	3
32	4
15	6
4	8
4	12
3	14
1	20
1	22

Table 8-6: Test items and number of valid parses

The following three examples ((8-3) to (8-5)) were obtained from the statistics file that is generated by XLE when the test file containing the grammatical items, was parsed:

(8-3) #229 (the women buy a dress)
 basadi ba reka mosese (2 0.002 29)

(8-4) #647 (she buys in the shop)
 o reka ka lebenkele (6 0.004 41)

(8-5) #646 (the woman buys the dress at mother's place)
 mosadi o reka mosese ga mme (12 0.003 57)

We conclude the discussion of the results obtained by confirming that all the ungrammatical items in the test suite showed zero parses and that our grammar has the required accuracy.

8.6 TREEBANK

In §8.2, we presented a treebank as a resource for testing a computational grammar, but noted that there are no treebanks for Setswana. Now, having developed and tested a computational grammar for Setswana, we use our XLE grammar to construct the first treebank for Setswana, annotated with deep syntactic information. For this purpose, we parse the test file of grammatical items and then utilise the functionality provided by the XLE user interface to create the treebank (cf. Test Files and Tree Banks in XLE User documentation – Crouch *et al.*, 2011).

For every grammatical item in our test suite, the interface allows us to inspect, amongst others, its parse tree(s) (the upper left window) and f-structure(s) (lower left window). Some items only have one preferred tree and f-structure while others can have more than one. We then select and store the preferred analysis/analyses in a user-defined folder. We use three of the available formats, viz. `.SExp` for trees, `.lfg` for f-structures and `.pl` for c- and f-structures in prolog (see Appendix B for examples of formats for stored analyses). The collection of files in the user-defined folder then constitutes the treebank, which can serve as a gold standard for future grammar testing and evaluation. This novel treebank consists of 2903 trees and f-structures for 828 phrases and sentences⁹⁰.

8.7 SUMMARY

In this chapter, we summarised the scope and coverage of our XLE grammar and we proposed a testing methodology based on a carefully hand-crafted test suite of grammatical and ungrammatical items. We proceeded to explicate the alignment between the linguistic structure of the Setswana simple sentence and the test suite to demonstrate the correctness of our grammar. Finally, we used the grammatical items of the test suite to construct a novel treebank, annotated with deep syntactic information.

⁹⁰ This treebank is available at <http://repo.sadilar.org/handle/20.500.12185/478>

CHAPTER 9

CONCLUSION

9.1 INTRODUCTION

In this final chapter, we reflect on the research reported on in this thesis by showing that the research problem, stated in Chapter 1, has been adequately addressed. Furthermore, we critically assess the contributions of this study. We also propose directions for future work.

9.2 ADDRESSING THE RESEARCH PROBLEM

We reiterate that the aim of our study was a rule-based computational syntactic analysis of Setswana. The research problem was fourfold: the accurate formulation of rules representing the structure of the Setswana simple sentence in the LFG framework; the implementation of these rules to develop a computational grammar; the testing of this grammar and the development of a treebank, annotated with deep syntactic information.

In Chapters 4, 5 and 6 we convincingly showed that the Setswana simple sentence is by no means syntactically “simple” and that many of the salient characteristics of Setswana syntax already present themselves here. More specifically, the systems of nominal classification and concordial agreement that govern Setswana syntax, the phrasal structure, as well as the verb as the morphologically most complex word category in Setswana, were covered in detail within the framework of LFG. We anticipate that this work would largely carry over to compound and complex sentences. Of specific importance is our analysis of verb extensions and their significance for argument structure and subcategorisation frames of extended verb roots, a common phenomenon in Setswana.

While the previous chapters focussed on syntactic structure, we devoted Chapter 7 to an explanation of how we developed a computational grammar based on the previously discussed syntactic structure. A discussion of the rich XLE user interface available to the grammar developer, for amongst others, defining and manipulating linguistic rules and representations, is followed by an overview of the grammar development process. The structure of the so-called “grammar file” provided the outline for our discussion since it contains all the essential components of an XLE grammar. The purpose of each component was briefly described and illustrated with code snippets from the Setswana grammar.

When building a computational artefact, such as a computational grammar, which models linguistic behaviour, it is obligatory to assess to what extent the artefact is correct (does what it was designed to do), is accurate (does it well) and is useful (is good and easy to use in other applications). In Chapter 8, an appropriate testing methodology was described and applied to our computational grammar. By systematically hand-crafting a test suite containing grammatical as well as ungrammatical test items, we were able to establish the correctness and the accuracy of the XLE grammar by demonstrating that all the grammatical items were correctly parsed and that all the ungrammatical ones did not produce valid parses. Its evaluation, i.e. its usefulness, was deferred to future work. Finally, we employed functionality of the XLE platform to create a treebank out of the grammatical test items. As far as we know, it is the first Setswana treebank, annotated with deep syntactic information.

Although our scope was limited to the Setswana simple sentence, the work reported on in this thesis constitutes a first step on our journey towards the future development of a broad-coverage LFG/XLE parser for Setswana.

9.3 CRITICAL REFLECTIONS

We briefly reflect on what the broader impact of our contribution might be and mention three aspects. Firstly, for an under-resourced language such as Setswana a first LFG/XLE grammar, test suite and treebank are important, however basic, contributions, which may form the basis for much work that needs to be done in future. Secondly, our work furthers the technological development of the South African Bantu languages, which forms a strategic part of building a truly multilingual society in South Africa in the 21st century. Our contribution to Setswana will also enable similar projects for other South African Bantu languages that share their syntactic structure with Setswana. Thirdly, we hope to include, in due course, the framework of LFG as well as the XLE grammar development platform in the teaching and learning of Setswana, both at under- and postgraduate level.

9.4 RESEARCH CONTRIBUTIONS

A summary of the research contributions of this study is as follows:

- The syntax of the Setswana simple sentence couched in the LFG framework, with specific reference to the exposition of Setswana phrasal structure and the argument structure, and subcategorisation frames of main verbs with both basic and extended verb roots;
- A first computational grammar for Setswana, by implementing the rules that represent the syntactic structure of Setswana simple sentence in XLE;

- A hand-crafted test suite for checking the correctness and accuracy of our Setswana grammar;
- A novel Setswana treebank that is annotated with deep syntactic information; and
- Two research publications (Berg, *et al.*, 2012 and Berg, *et al.*, 2013) emanated from research for this study.

9.5 FUTURE WORK

The research reported on in this thesis can be extended in numerous ways of which we mention five:

- (1) Broadening the scope and coverage of the Setswana XLE grammar by, for example, including the syntactic analysis of coordination, compound and complex sentences, and by ensuring that the lexicon is sufficiently comprehensive;
- (2) A thorough testing of the XLE grammar in (1) with real-world data and an extensive evaluation thereof by using it in HLT applications;
- (3) Participating in the ParGram⁹¹ Project, the aim of which is to produce broad-coverage grammars for a variety of languages;
- (4) Bootstrapping of the Setswana grammar to other structurally similar Bantu languages, for example Sesotho sa Leboa and Sesotho; and
- (5) Extending the study of Setswana in the LFG framework.

⁹¹ Cf. <https://pagram.w.uib.no/> - In the ParGram project broad-coverage grammars for a variety of languages are written collaboratively within the linguistic framework of LFG and with a commonly-agreed-upon set of grammatical features.

APPENDIX A: TABLES FROM CHAPTER 1 AND CHAPTER 4

Class	Class prefix	Semantic content
1	mo-	Exclusively personal nouns
2	ba-	
1a	(-)-	Proper names, kinship terms, and personified names
2a	bo-	
3	mo-	Miscellaneous: impersonal, parts of the body, tools, instruments, animals, natural phenomena, most names of trees and plants
4	me-	
5	le-	Miscellaneous: some animals, many plants, natural phenomena, collective nouns, strangers, regions
6	ma-	
7	se-	Miscellaneous: parts of the body, names of languages, customs, habits, languages, some plants, tools, instruments, some personal nouns
8	di-	
9	(ne)-	Miscellaneous: mostly animal names, some plants, parts of the body, natural phenomena, personal nouns indicating people with a special rank
10	di-	
11	lo-	Miscellaneous: few nouns belong to this class (This class is disappearing.)
10	di-	
14	bo-	Mainly abstract concepts, also collective nouns
6	ma-	
15	go-	Infinitive
16	fa-	Location or place
17	go-	
18	mo-	
19	ga-	
20	(-)-	

Table 1-1: Setswana noun classes

Class	Absolute pronoun ⁹²	Demonstrative pronouns			Quantitative pronouns					Possessive pronoun
		Distance 1 ⁹³ (distal)	Distance 2 (proximal)	Distance 3 (post-distal)	Inclusive (all)	Exclusive (no)	Selective (only)	Separative (unknown)	Interrogative (which?)	
1	ene	yo	yoo	yole	-	ope	esi	osele	ofe?	gagwe
2	bone	ba	bao	bale	botlhe	bape	bosi	basele	bafe?	bone
3	one	o	oo	ole	otlhe	ope	osi	osele	ofe?	one
4	yone	e	eo	ele	yotlhe	epe	yosi	esele	efe?	yone
5	lone	le	leo	lele	lotlhe	lepe	losi	lesele	lefe?	lone
6	one	a	ao	ale	otlhe	ape	osi	asele	afe?	one
7	sone	se	seo	sele	sotlhe	sepe	sosi	sesele	sefe?	sone
8	tsone	tse	tseo	tsele	tsotlhe	dipe	tsosi	disele	dife?	tsone
9	yone	e	eo	ele	yotlhe	epe	yosi	esele	efe?	yone
10	tsone	tse	tseo	tsele	tsotlhe	dipe	tsosi	disele	dife?	tsone
11	lone	lo	loo	lole	lotlhe	lope	losi	losele	lofe?	lone
10	tsone	tse	tseo	tsele	tsotlhe	dipe	tsosi	disele	dife?	tsone
14	bone	bo	boo	bole	botlhe	bope	bosi	bosele	bofe?	bone
6	one	a	ao	ale	otlhe	ape	osi	asele	afe?	one
15	gone	go	goo	gole	gotlhe	gope	gosi	gosele	gofe?	gone
16	gone	fa	fao	fale	gotlhe	gope	gosi	gosele	gofe?	gone
17	gone	go	goo	gole	gotlhe	gope	gosi	gosele	gofe?	gone
18	gone	mo	moo	mole	gotlhe	gope	gosi	gosele	gofe?	gone
19	gone	-	kwa	kwale	gotlhe	gope	gosi	gosele	gofe?	gone
20	gone	-	kwa	kwale	gotlhe	gope	gosi	gosele	gofe?	gone

Table 4-1: Pronouns for noun classes

⁹² The absolute pronouns have a variant form where the ending is **-a** and not **-e**.

⁹³ Distance 1 also has a variant form (cf. Krüger, 2006:132).

Person	Personal pronouns	Possessive pronouns	
		Ordinary	Communal
First person, singular	nna	ka/me	-
First person, plural	rona	rona	etsho/getsho
Second person, singular	wena	gago	-
Second person, plural	lona	bona	eno/gaeno
Third person, singular	ene	gagwe	-
Third person, plural	bone	bona/bone	gabo

Table 4-2: Pronouns for persons

Class	Subject agreement morpheme
1	o
2	ba
3	o
4	e
5	le
6	a
7	se
8	di
9	e
10	di
11	lo
10	di
14	bo
6	a
15	go
16	fa
17	go
18	go
19	go
20	go

Table 4-4: Subject agreement morphemes of noun classes

Person	Subject agreement morpheme
First person, singular	ke
First person, plural	re
Second person, singular	o
Second person, plural	lo

Table 4-5: Subject agreement morphemes of personal pronouns

Class	Consecutive subject agreement morphemes
1	a
2	ba
3	wa
4	ya
5	la
6	a
7	sa
8	tsa
9	ya
10	tsa
11	la/lwa
10	tsa
14	ba
6	a
15	ga
16	ga
17	ga
18	ga
19	ga
20	ga

Table 4-6: Consecutive subject agreement morphemes of noun classes

Person	Consecutive subject agreement morpheme
First person, singular	ka
First person, plural	ra
Second person, singular	wa
Second person, plural	lwa

Table 4-7 Consecutive subject agreement morphemes of personal pronouns

Class	Object agreement morpheme
1	mo
2	ba
3	o
4	e
5	le
6	a
7	se
8	di
9	e
10	di
11	lo
10	di
14	bo
6	a
15	go
16	fa
17	go
18	go
19	go
20	go

Table 4-8: Object agreement morphemes of noun classes

Person	Object agreement morpheme
First person, singular	ni
First person, plural	re
Second person, singular	go
Second person, plural	lo

Table 4-9: Object agreement morphemes of personal pronouns

Indicative mood				
Tense & Aspect	Polarity	Identifying copulative	Describing copulative	Associative copulative
Present tense	Positive	Monna ke kgosi. (The man is a king.)	Monna o mogolo. (The man is big.)	Monna o na le koloi. (The man has a car.)
	Negative	Monna ga se kgosi. (The man is not a king.)	Monna ga a mogolo. (The man is not big.)	Monna ga a na koloi. (The man does not have a car.)
Progressive	Positive	Monna e sa le kgosi. (The man is still a king.)	Monna o sa le mogolo. (The man is still big.)	Monna o sa na le koloi. (The man still has a car.)
		Monna e santse e le kgosi. (The man is still a king.)	Monna o santse a le mogolo. (The man is still big.)	Monna o santse a na le koloi. (The man still has a car.)
	Negative	Monna ga e tlhole e le kgosi. (The man is not still a king.)	Monna ga a tlhole a le mogolo. (The man is not still big.)	Monna ga a tlhole a na le koloi. (The man still does not have a car.)
Potential	Positive	Monna e ka nna kgosi. (The man can be a king.)	Monna a ka nna mogolo. (The man can be big.)	Monna a ka nna le koloi. (The man can have a car.)
	Negative	Monna e ka se nne kgosi. (The man cannot be a king.)	Monna a ka se nne mogolo. (The man cannot be big.)	Monna a ka se nne le koloi. (The man cannot have a car.)
Past tense	Positive	Monna e nnile kgosi. (The man was a king.)	Monna o nnile mogolo. (The man was big.)	Monna o nnile le koloi. (The man had a car.)
	Negative	Monna ga e a nna kgosi. (The man was not a king.)	Monna ga a nna mogolo. (The man was not big.)	Monna ga a nna le koloi. (The man did not have a car.)
Future tense	Positive	Monna e tla nna kgosi. (The man will be a king.)	Monna o tla nna mogolo. (The man will be big.)	Monna o tla nna le koloi. (The man will have a car.)
	Negative	Monna ga e ketla e nna kgosi. (The man will not be a king.)	Monna ga a ketla a nna mogolo. (The man will not be big.)	Monna ga a ketla a nna le koloi. (The man will not have a car.)
		Monna ga e nke e nna kgosi. (The man will not be a king.)	Monna ga a nke a nna mogolo. (The man will not be big.)	Monna ga a nke a nna le koloi. (The man will not have a car.)
Relative past tense	Positive	Monna e ne e le kgosi. (The man was a king.)	Monna o ne a le mogolo. (The man was big.)	Monna o ne a na le koloi. (The man had a car.)
	Negative	Monna e ne e se kgosi. (The man was not a king.)	Monna o ne a se mogolo. (The man was not big.)	Monna o ne a se na koloi. (The man did not have a car.)
Relative future tense	Positive	Monna e tlabo e le kgosi. (The man will be a king.)	Monna o tlabo a le mogolo. (The man will be big.)	Monna o tlabo a na le koloi. (The man will have a car.)
	Negative	Monna e tlabo e se kgosi. (The man will not be a king.)	Monna o tlabo a se mogolo. (The man will not be big.)	Monna o tlabo a se na koloi. (The man will not have a car.)

(continue on next page)

Consecutive mood				
Tense & Aspect	Polarity	Identifying copulative	Describing copulative	Associative copulative
	Positive	Monna ya nna kgosi. (The man is then a king.)	Monna a nna mogolo. (The man is then big.)	Monna a nna le koloi. (The man then has a car.)
	Negative	Monna ya se nne kgosi. (The man is then not a king.)	Monna a se nne mogolo. (The man is then not big.)	Monna a se nne le koloi. (The man then does not have a car.)
Imperative mood				
Tense & Aspect	Polarity	Identifying copulative	Describing copulative	Associative copulative
	Positive	Monna, нна kgosi! (Man, be a king!)	Monna, нна mogolo. (Man, be big!)	Monna, нна le koloi. (Man, has a car!)
	Negative	Monna, se nne kgosi! (Man do not be a king!)	Monna, se nne mogolo. (Man do not be big!)	Monna, se nne le koloi. (Man, do not have a car!)
Habitual mood				
Tense & Aspect	Polarity	Identifying copulative	Describing copulative	Associative copulative
	Positive	Monna e nne kgosi. (The man is usually a king.)	Monna a nne mogolo. (The man is usually big.)	Monna a nne le koloi. (The man usually has a car.)
	Negative	Monna e se nne kgosi. (The man is not usually a king.)	Monna a se nne mogolo. (The man is not usually big.)	Monna a se nne le koloi. (The man usually do not have a car.)

Table 4-10: Examples of copulative verbs in sentences

Person	Personal pronoun	Identifying copulative verb
First person, singular	nna	ke
First person, plural	rona	re
Second person, singular	wena	o
Second person, plural	lona	lo

Table 4-11: Identifying copulative verbs of personal pronouns

Class	Describing copulative verb
1	o
2	ba
3	o
4	e
5	le
6	a
7	se
8	di
9	e
10	di
11	lo
10	di
14	bo
6	a
15	go
16	fa
17	go
18	go
19	go
20	go

Table 4-12: Describing copulative verbs of noun classes

Person	Personal pronoun	Describing copulative verb
First person, singular	nna	ke
First person, plural	rona	re
Second person, singular	wena	o
Second person, plural	lona	lo

Table 4-13: Describing copulative verbs of personal pronouns

Class	Possessive particle
1	wa
2	ba
3	wa
4	ya
5	la
6	a
7	sa
8	tsa
9	ya
10	tsa
11	la
10	tsa
14	ba/jwa
6	a
15	ga
16	ga
17	ga
18	ga
19	ga
20	ga

Table 4-14: Possessive particles

Class	Qualificative particle
1	yo
2	ba
3	o
4	e
5	le
6	a
7	se
8	tse
9	e
10	tse
11	lo
10	tse
14	bo
6	a
15	go
16	go
17	go
18	go
19	go
20	go

Table 4-15: Qualificative particles

APPENDIX B: TABLES AND TREEBANK FORMATS FROM CHAPTER 8

NOUNS			
Singular		Plural	
mang	(who-singular)	bomang	(who-plural)
mme	(mother)	bomme	(mothers)
monna	(man)	banna	(men)
mosadi	(woman)	basadi	(women)
mosetsana	(girl)	basetsana	(girls)
motho	(human, person)	batho	(humans, persons)
malome	(uncle)	bomalome	(uncles)
rre	(father. sir)	borre	(fathers. sirs)
John	(John)	boJohn	(John and company)
mosese	(dress)	mesese	(dresses)
moso	(tomorrow)		
lebenkele	(shop)	mabenkele	(shops)
letsatsi	(day)		
madi	(money)		
setlhare	(tree)		
setlhopha	(team)		
Setswana	(Setswana)		
		dinotlolo	(keys)
buka	(book)	dibuka	(books)
dijo	(food)		
karata	(card)	dikarata	(cards)
kgosi	(king)	dikgosi	(kings)
toropo	(town)		
kgetsi	(bag)	dikgetsi	(bags)
LOCATIVE CLASS NOUNS			
fatshe	(down, on the ground)		
godimo	(above, on top)		
morago	(behind)		
gare	(in the middle, between)		
teng	(here, inside)		

Table 8-4: Nouns and locative class nouns

VERBS				
batl- (seek)	batlis- (let seek)	batlisan- (help each other seek)		
bo- (return)	boil- (returned)			
bon- (see)	bonan- (see each other)			
bu- (speak)				
didimal- (keep quiet)				
duel- (pay)				
dumalan- (agree)				
j- (eat)	jesan- (feed each other)			
kwal- (write)	kwadis- (let write)	kwalel- (write for)	kwalelan- (write for each other)	kwalelanw- (being written to each other)
lel- (cry)				
rat- (love, like)	ratan- (love/like each other)			
rek- (buy)	rekel- (buy for)	rekw- (being bought)	rekelw- (being bought for)	
rom- (send)	romiw- (being sent)			
tabog- (run)	tabogis- (let run)	tabogisiw- (being trained, "being let run for")		
tl- (come)	tlis- (let come)	tlisets- (bring for, "let come for")		
tsamay- (walk)	tsamael- (walk towards)			
tsey- (put in)	tseyets- (put in for)	tseyetsan- (put in for each other)	tseyetsanw- (being put in for each other)	
tshamek- (play)	tshamekel- (play for)			
y- (go)				

Table 8-5: Verbs (basic and extended verb roots)

EXAMPLES OF FORMATS FOR STORED ANALYSES

```
(CLNP:132
(N:88 mosadi:1)
(PROP:124
(DEMPRO:122 yo:43)))
```

Tree in .SExp format

```
FEATURE STRUCTURE LEXICON (1.0)
```

```
FS1 FS *
  (^ ADJUNCT) = %1
  (^ AGR) = 1
  (^ NTYPE) = ord
  (^ PRED) = 'sadi'
%2 $ %1
  (%2 AGR) = 1
  (%2 PRED) = 'yo'
  (%2 PRONTYPE) = DemPro.
```

```
----
```

f-structure in .lfg format

```
% -*- coding: iso-8859-1 -*-
```

```
fstructure('mosadi yo',
  % Properties:
  [
    'sentence_id'('S5'),
    'markup_free_sentence'('mosadi yo'),
    'xle_version'('XLE release of May 07, 2013 13:40.'),
    'grammar'('/home/ansu/xle/toytswanaAnsu30Oktober2017/toy-tsw.lfg'),
    'grammar_date'('Oct 30, 2017 19:13'),
    'word_count'('2'),
    'statistics'('1 solutions, 0.001 CPU seconds, 0.150MB max mem, 11 subtrees unified'),
    'rootcategory'('CLNP'),
    'hostname'('ansu-HP-ProBook-450-G2')
  ],
  % Choices:
  [
  ],
  % Equivalences:
  [
  ],
  % Constraints:
  [
    cf(1,eq(attr(var(0),'PRED'),semform('sadi',0,[],[]))),
    cf(1,eq(attr(var(0),'ADJUNCT'),var(1))),
    cf(1,eq(attr(var(0),'AGR'),'1')),
    cf(1,eq(attr(var(0),'NTYPE'),'ord')),
    cf(1,in_set(var(2),var(1))),
    cf(1,eq(attr(var(2),'PRED'),semform('yo',1,[],[]))),
    cf(1,eq(attr(var(2),'AGR'),'1')),
    cf(1,eq(attr(var(2),'PRONTYPE'),'DemPro'))
  ],
  % C-Structure:
  [
    cf(1,subtree(132,'CLNP',91,124)),
    cf(1,phi(132,var(0))),
    cf(1,subtree(124,'PROP',-,122)),
    cf(1,phi(124,var(2))),
    cf(1,subtree(122,'DEMPRO',121,44)),
    cf(1,phi(122,var(2))),
  ]
)
```

(continue on next page)

```

cf(1,subtree(121,'DEMPRO',-,77)), cf(1,phi(121,var(2))),
cf(1,subtree(91,'CLNP',-,88)), cf(1,phi(91,var(0))),
cf(1,subtree(88,'N',87,5)),
cf(1,phi(88,var(0))),
cf(1,subtree(87,'N',-,42)),
cf(1,phi(87,var(0))),
cf(1,subtree(77,'DEMPRO_BASE',-,47)),
cf(1,phi(77,var(2))),
cf(1,terminal(47,'yo',[43])),
cf(1,phi(47,var(2))),
cf(1,terminal(45,'+DemPro1D1',[43])),
cf(1,phi(45,var(2))),
cf(1,subtree(44,'DEMPRO_SFX_BASE',-,45)),
cf(1,phi(44,var(2))),
cf(1,subtree(42,'N_BASE',-,7)),
cf(1,phi(42,var(0))),
cf(1,terminal(7,'sadi',[1])),
cf(1,phi(7,var(0))),
cf(1,subtree(5,'N_SFX_BASE',-,3)),
cf(1,phi(5,var(0))),
cf(1,terminal(3,'+NPre1',[1])),
cf(1,phi(3,var(0))),
cf(1,semform_data(0,42,1,7)),
cf(1,semform_data(1,77,8,10)),
cf(1,fspan(var(2),8,10)),
cf(1,fspan(var(0),1,10)),
cf(1,surfaceform(1,'mosadi',1,7)),
cf(1,surfaceform(43,'yo',8,10))
]).

```

c- and f-structure in .p1 format (prolog)

APPENDIX C: MORPHOLOGICAL TAGS

Tag	Explanation
AbsPro	absolute pronoun
adjective	adjective
Adv	adverb
AgPart	agentative particle
AgrObj	object agreement morpheme
AgrSubj	subject agreement morpheme
ApplSuf	applicative suffix
AssCopV	associative copulative verb
AssPart	associative particle
AugSuf	augmentative suffix
Aux	auxiliary
CausSuf	causative suffix
ComplSuf	completive suffix
CompPart	comparative particle
Conj	conjunction
ConsAgrSubj	consecutive subject agreement morpheme
D1, D2, D3	distance 1, distance 2, distance 3
DemPro	demonstrative pronoun
DesCopV	describing copulative verb
DevSuf	deverbatives suffix
DimSuf	diminutive suffix
FemSuf	feminine suffix
FutPre	future prefix
HortPart	hortative particle
IdCopV	identifying copulative verb
Ideo	ideophone
IncQPro	inclusive quantitative pronoun
Instrpart	instrumental particle
IntensSuf	intensive suffix
Interj	interjection
IntPart	interrogative particle

(continue on next page)

LocPart _{fa}	locative particle fa
LocPart _{ga}	locative particle ga
LocPart _{go}	locative particle go
LocPart _{ka}	locative particle ka
LocPart _{kwa}	locative particle kwa
LocPart _{mo}	locative particle mo
LocSuf	locative suffix
NegPrea	negative prefix a
NegPrega	negative prefix ga
NegPresa	negative prefix sa
NPre	noun prefix
P1	first person
P2	second person
PassSuf	passive suffix
PerfSuf	perfect suffix
pl	plural
PossPart	possessive particle
PossPro	possessive particle
PresPre	present tense prefix a
ProgPre	progressive prefix
QualPart	qualificative particle
RecSuf	reciprocal suffix
SepQPro	separative quantitative particle
sg	singular
TempPart	temporal particle
VEnd	verbal ending
1 2 3 20	noun classes 1 to 20

Table: Morphological tags

REFERENCE LIST

- Ackerman, F. & Moore, J. 2013. Proto-Properties in a comprehensive theory of argument realization. (*In King, T.H. & De Paiva, V., eds. From quirky case to representing space: papers in honor of Annie Zaenen. Stanford: CSLI. p.9–20).*
- Ackermann, F. 1992. Complex predicates and morpholexical relatedness: locative alternation in Hungarian. (*In Sag, I. & Szabolcsi, A., eds. Lexical matters. Stanford: CSLI. p.55–83).*
- Alsina, A. & Mchombo, S.A. 1988. Lexical mapping in the Chichewa applicative construction. Berkeley: Departments of Linguistics, Stanford University and University of California. Revised and expanded from a paper presented at the 19th Annual African Linguistics Conference, Boston University, April 1417.
- Alsina, A. & Mchombo, S.A. 1990a. Object asymmetries in Chichewa. Berkeley: Stanford University.
- Alsina, A. & Mchombo, S.A. 1990b. The syntax of applicatives in Chichewa: problems for a theta theoretic asymmetry. *Natural Language and Linguistic Theory*, 8(4):493–506.
- Alsina, A. & Mchombo, S.A. 1993. Object asymmetries and the Chichewa applicative construction. (*In Mchombo, S.A., ed. Theoretical aspects of Bantu grammar, CSLI Lecture Notes, number 38. Stanford: CSLI. p.1745).*
- Alsina, A. 1996. The role of argument structure in grammar: evidence from Romance. Stanford: CSLI.
- Anon. 2011. META-NORD, Summary. <http://www.meta-net.eu/projects/meta-nord/summary>
Date of access: 18 June 2016.
- Asudeh, A. 2006. Direct compositionality and the architecture of LFG. (*In Butt, M., Dalrymple, M. and King, T.H., eds. Intelligent linguistic architectures: variations on themes by Ronald M. Kaplan. Stanford: CSLI. p.363–387).*
- Asudeh, A. & Toivonen, I. 2015. Lexical functional grammar. (*In Heine, B. & Narrog, H., eds. The Oxford handbook of linguistic analysis. Oxford: Oxford University. p.373–406).*

Attia, M.A. 2008. Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation. Manchester: University of Manchester. (Thesis – PHD).

Austin, P. 2001. Lexical Functional Grammar. (*In* Smelser, N.J. & Bates, P., eds. International encyclopaedia of the social and behavioural sciences. Oxford: Elsevier. p.8748–8754).

Babarczy, A., Gàbor, B., Gàbor, H., Karpati, A., Rung, A. & Szakadat, I. 2007. Hunpars: a rule-based sentence parser for Hungarian. (*In* Proceedings of the 45th Annual Meeting of the ACL Conference on Computational Linguistics. Prague. p.91–96).

Balkan, L., Meijer, S., Arnold, D., Dauphin, E., Estival, D., Falkedal, K., Lehmann, S. & Regnier-Post, S. 1994. Test suite design guidelines and methodology. Report to LRE 62-089 (D-WP2.1) Test Suites for Natural Language Processing (TSNLP), University of Essex, July 1994.

Beesley, K.R. & Karttunen, L. 2003. Finite-state morphology. Stanford: CSLI.

Berg, A., Pretorius, R. & Pretorius, L. 2012. Exploring the treatment of selected typological characteristics of Tswana in LFG. (*In* Butt, M. & King, T.H., eds. Proceedings of the LFG12 Conference. Stanford: CSLI. p.85–94).

Berg, A., Pretorius, R. & Pretorius, L. 2013. The representation of Setswana double objects in LFG. (*In* Butt, M. & King, T.H., eds. Proceedings of the 18th International Lexical Functional Grammar Conference. Stanford: CSLI. p.111–130).

Borin, L., Forsberg, M. & Kokkinakis, D. 2010. Diabase: Towards a diachronic BLARK in support of historical studies. (*In* Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. & Tapias, D., eds. Proceedings of the Seventh International Conference on Language Resources and Evaluation. Valletta. European Language Resources Association (ELRA). p.35–42. http://lrec-conf.org/proceedings/lrec2010/pdf/156_Paper.pdf Date of access: 18 June 2016).

Borin, L., Forsberg, M. & Lönngrén, L. 2008. The hunting of the BLARK – SALDO, a freely available lexical database for Swedish language technology. (In Nivre, J., Dahllöf, M. & Megyesi, B. eds. Resourceful language technology. Festschrift in honour of Anna Sågvald Hein. Uppsala: Uppsala Universitet. p.21-31. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.148.5870&rep=rep1&type=pdf#page=21> Date of access: 18 June 2015.

Botswana Central Statistics Office. 2009. Botswana Demographic Survey 2006. Government Printer: Gaborone.

Bresnan, J. 1982. The mental representation of grammatical relations. Cambridge: The MIT.

Bresnan, J. 1987. On locative inversion in Chichewa. *CSLI Monthly*, 2(8).

Bresnan, J. 1997. The emergence of the unmarked pronoun: Chichewa pronominals in optimality theory. Presented at the BLS 23 Special Session on Syntax and Semantics in Africa. <http://www-csli.stanford.edu/~bresnan/jb-bls-roa.ps> Date of access: 12 June 2016.

Bresnan, J. 2001. *Lexical-Functional Grammar*. Oxford: Blackwell.

Bresnan, J., Asudeh, A., Toivonen, I. & Wechsler, S. 2016. *Lexical-Functional Syntax*. 2nd edition. Oxford: Wiley-Blackwell.

Bresnan, J. & Kanerva, J. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry*, 20(1):1-50.

Bresnan, J. & Mchombo, S.A. 1985. On topic, pronoun and agreement in Chichewa. (In Choi, S., Devitt, D., Janis, W., McCoy, T. & Zhang, S., eds. *Proceedings of the Second Eastern States Conference on Linguistics*. Buffalo: The State University of New York. p.276–312).

Bresnan, J. & Mchombo, S.A. 1986. Grammatical and anaphoric agreement. *Papers from the Parasession on Pragmatics and Grammatical Theory. Chicago Linguistic Society*, 22(2): 278–297.

Bresnan, J. & Mchombo, S.A. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63(4):741–782.

Bresnan, J. & Mchombo, S.A. 1995. The lexical integrity principle: evidence from Bantu. *Natural Language and Linguistic Theory*, 13(2):181-254.

Bresnan, J. & Moshi, L. 1990. Object asymmetries in comparative Bantu syntax. *NLinguistic Inquiry*, 21(2):147-185.

Brits, J.H. 2006. Outomatiese Setswana Lemma-identifisering. Potchefstroom: North-West University. (Dissertation – MA).

Brits, J.H., Pretorius, R.S. & Van Huyssteen, G.B. 2005. Automatic lemmatization in Setswana: towards a prototype. *South African Journal of African Languages*, 25(11):37-47.

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

Butt, M. & King, T.H. 2003. Grammar writing, testing and evaluation. (In Farghaly, A., ed. Handbook for language engineers. California: CSLI. p.129-180.)

Butt, M., King, T.H., Nino, M. & Segond, F. 1999. A grammar writers cookbook. Stanford: CSLI.

Census Namibia. 2011. Population and Housing Census, 2011. <http://www.nsa.org.na> Date of access: 24 April 2014.

Çetinoğlu, Ö. 2009. Building a lexical functional grammar for Turkish. Istanbul: Sabancı University. (PhD – Thesis).

Chavula, J.J. 2016. Verbal derivation and valency in Citimbuka. Leiden: University of Leiden. (Thesis – PhD).

Chomsky, N. 1957. Syntactic structures. The Hague: Mouton.

Chomsky, N. 1995. The minimalist program. Cambridge: MIT.

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

Cole, D.T. 1961. Doke's classification of Bantu languages. (*In* Doke, C.M. & Cole, D.T., eds. Contributions to the history of Bantu linguistics. Johannesburg: Witwatersrand University. p.80-108).

Creissels, D. 2006. Tswana verb morphology and the lexical integrity principle. *Lingue e Linguaggio*, (1):1–18.

Crouch, D, Dalrymple, M, Kaplan, R.M., King, T.H., Maxwell, J.T. & Newman, P. 2015. XLE Documentation. California: Palo Alto Research Centre. <http://www2.parc.com/isl/groups/nlft/xle/doc/xle.toc.html> Date of access: 30 September 2015.

Daelemans, W., Binnenpoorte, D., De Vriend, F., Sturm, J., Strik, H. & Cucchiari, C. 2003. Establishing priorities in the development of HLT resources: The Dutch-Flemish experience. (*In* Multilingualism and electronic language management: Proceedings of the 4th International MIDP Colloquium, Bloemfontein, South Africa. Pretoria: Van Schaik. p.9_23. <http://www.cnts.ua.ac.be/papers/2005/dbd05.pdf> Date of access: 18 June 2016.)

Dalrymple, M. 2001. Lexical Functional Grammar. Syntax and Semantics, Volume 34. California: Academic.

Dalrymple, M. 2004. Copular complements: closed or open? (*In* Butt, M. & King, T.H., eds. Proceedings of the LFG04 Conference. p.188-198 Stanford: CSLI. <http://csli-publications.stanford.edu/> Date of access: 9 May 2013).

Dalrymple, M. 2006. Lexical functional grammar. (*In* Brown, K., ed. Encyclopaedia of language and linguistics, 2nd edition. Oxford: Elsevier. p.82-94).

Dalrymple, M., Kaplan, R.M., Maxwell, J.T. & Zaenen, A. 1995. Formal architecture. (*In* Dalrymple, M., Kaplan, R.M., Maxwell, J.T. & Zaenen, A., eds. Formal issues in lexical-functional grammar. p.1-5).

De Vos, M. & Mitchley, H. 2012. Subject marking and preverbal coordination in Sesotho: a perspective from optimality theory. *Southern African Linguistics and Applied Language Studies*, 30(2):155-170.

Department of African Languages and Literature, Botswana University. 2000. The structure of Setswana sentences. Gaborone: Lightbooks.

Dickinson, M., Hajič, J., Kübler, S. and Przepiórkowski, A., eds. 2015. Treebanks and linguistic theories. (*In Proceedings of the 14th international workshop on treebanks and linguistic theories*. Bloomington, USA. p.1–133).

Doke, C.M. 1927. Textbook of Zulu grammar. Johannesburg: University of the Witwatersrand.

Dowty, D. 1993. Thematic proto-roles and argument selection. *Language*, 67(3): 547–619.

Du Plessis, J.A. & Visser, M. 1992a. Sesotho syntax. Pretoria: Via Afrika.

Du Plessis, J.A. & Visser, M. 1992b. Xhosa syntax. Pretoria: Via Afrika.

Du Plessis, J.A., Musehane, N.M. & Visser, M. 1992. Venda syntax. Pretoria: Via Afrika.

Du Plessis, J.A., Nxumalo, N.E. & Visser, M. 1995. Tsonga syntax. Pretoria: Via Afrika.

Eiselen, E.R. & Puttkammer, M.J. 2014. Developing text resources for ten South African languages. (*In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J. & Piperidis, S., eds. Proceedings of the 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland. European Language Resources Association. p.3698–3703 <http://www.lrec-conf.org/proceedings/lrec2014/index.html> Date of access: 9 September 2015*).

Eiselen, E.R. 2016. South African Language Resources: Phrase Chunking. (*In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28. p.689–693*).

Faaß, G. & Prinsloo, D.J. 2011. A computational implementation of the Northern Sotho Infinitive. *South African Journal of African Languages*, 31(2):281–301.

Faaß, G. 2010. A morphosyntactic description of Northern Sotho as a basis for an automated translation from Northern Sotho into English. Pretoria: UP. (Thesis – PHD)

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

Farghaly, A. 2003. Introduction. (*In Farghaly, A., ed. Handbook for language engineers. California: CSLI. p.1–20).*

Fernando, M. 2008. An analysis of verbal affixes in Kikongo with special reference to form and function. Pretoria: University of South Africa. (Dissertation – MA).

Forst, M. 2011. Computational Aspects of Lexical Functional Grammar. *Language and Linguistics Compass*, 5(1):1–18.

Guthrie, M. 1971. Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages, Volume 2. Farnborough: Gregg International.

Haegeman, L. 1994. Introduction to government and binding theory. Oxford: Wiley-Blackwell.

Her, O. 2013. Lexical Mapping Theory revisited. (*In King, T.H. and De Paiva, V., eds. From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen. Stanford: CSLI. p.47–60).*

Hulden, M. 2009. Foma: a finite-state compiler and library. (*In Proceedings of the EACL 2009 Demonstrations Session. p.29–32).*

Hurst, P. 2006. The syntax of the Malagasy reciprocal construction: an LFGg account. (*In Butt, M. & King, T.H., eds. Proceedings of LFG06. Stanford: CSLI. p.256-274. <http://csli-publications.stanford.edu/> Date of access: 14 August 2016).*

Hurst, P. 2010. The syntax of lexical reciprocal constructions. (*In Butt, M. & King, T.H., eds. Proceedings of LFG10. Stanford: CSLI. p.311–331 <http://csli-publications.stanford.edu/> Date of access: 16 September 2016).*

Hyman, L. 2002. Suffix ordering in Bantu: A morphocentric approach. (*In Booij, G. & Van Marle, J., eds. Yearbook of morphology. Netherlands: Springer. p.245–281).*

Joshi, A.K. & Schabes, Y. 1997. Tree-adjointing grammars. (*In Rozenberg, G. & Salomaa, A., eds. Handbook of formal languages, Volume 3: beyond words. Berlin: Springer-Verlag. p.69–124).*

Jurafsky, D. & Martin, J.H. 2009. Speech and Language Processing. New Jersey: Pearson.

- Kaplan, R.M. 1995. The formal architecture of lexical-functional grammar. (*In* Dalrymple, M., Kaplan, R.M., Maxwell, J.T. & Zaenen, A., eds. Formal issues in lexical-functional grammar. Stanford: Stanford University. p.7–27).
- Kaplan, R.M. & Bresnan, J. 1982. Lexical functional grammar: a formal system for grammatical representation. (*In* Bresnan, J. ed., The mental representation of grammatical relations. Cambridge MIT. p.173–281).
- Kaplan, R.M. & Bresnan, J. 1995. A formal system for grammatical representation. (*In* Dalrymple, M., Kaplan, R.M., Maxwell, J.T. & Zaenen, A., eds. Formal issues in lexical-functional grammar. Stanford: Stanford University. p.29130).
- Kaplan, R.M., King, T.H. & Maxwell, J.T. 2002. Adapting existing grammars: the XLE experience. (*In* Oostijk, N., & Sutcliff, R., eds. Proceedings of the Workshop on Grammar Engineering and Evaluation. p.29–35).
- Kaplan, R.M., Maxwell, J.T., King, T.H., & Crouch, R. 2004. Integrating finite-state technology with deep LFG grammars. (*In* Proceedings of the ESSLLI 2004 workshop on Combining Shallow and Deep Processing for NLP. ESSLLI. p.168–173).
- Karlsson, F., Voutilainen, A., Heikkilä, J. & Antilla, A. 1995. Constraint grammar. Berlin: Mouton de Gruyter.
- Katamba, F. 2003. Bantu nominal morphology. (*In* Nurse, D. & Philippson, G., eds. The Bantu languages. London: Routledge. p.103–120).
- Khoali, M.H.E. 1994. Interrogative structures in Setswana – a functional approach. Potchefstroom: Potchefstroom University for Higher Education. (Dissertation – MA).
- Khumalo, L. 2007. An analysis of the Ndebele passive construction. Oslo: University of Oslo. (Thesis – PhD).
- Khumalo, L. 2014. On the reciprocal in Ndebele. *Nordic Journal of African Studies*, 23(3): 140–161.
- Kosch, I.M. 1991. A survey of Northern Sotho grammatical descriptions since 1876. Pretoria: University of South Africa. (Thesis – PhD).

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

Krauwer, S. 2003. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. Utrecht: Utrecht Institute of Linguistics/ELSNET.
<http://www.elsnet.org/dox/krauwer-specom2003.pdf>. Date of access: 2 February 2010.

Kroeger, P.R. 2004. Analyzing syntax: a lexical-functional approach. Cambridge: Cambridge University.

Krüger, C.J.H. 1961. Die sintaktiese struktuur van die enkelvoudige sin in Tswana: Potchefstroom: Potchefstroom University for Higher Education. (Dissertation – MA).

Krüger, C.J.H. 1967. Die struktuur van die woordgroep in Tswana. Pretoria: UP. (Thesis – PhD).

Krüger, C.J.H. 1994. Notes on morphology with special reference to Tswana. *South African Journal of African Languages*, 14(1):15–23.

Krüger, C.J.H. 2006. Introduction to the morphology of Setswana. München: Lincom Europe.

Krüger, C.J.H. 2013a. Setswana syntax: a survey of word group structures: Volume 1. München: Lincom Europe.

Krüger, C.J.H. 2013b. Setswana syntax: a survey of word group structures: Volume 2. München: Lincom Europe.

Kübler, S. 2004. Memory-based parsing. Amsterdam: John Benjamins.

Lam, O.S. 2007. The applicative affix and morpheme ordering in Chichewa. (In Butt, M. & King, T.H., eds. The Proceedings of the LFG07 Conference. p.364–82. <http://csli-publications.stanford.edu/LFG/6/lfg01.html> Date of access: 12 June 2016).

Le Roux, J.C. 2007. A grammatical analysis of the Tswana adverbial. Pretoria: University of South Africa. (Thesis – PhD).

Le Roux, J.C. 2011. The border dispute between adverbials and conjunctions in Tswana grammar. *South African Journal of African Languages*, 31(1): 54–65.

Lehmann, S., Oepen, S., Regnier-Prost, S., Netter, K., Lux, V., Klein, J., Falkedal, K., Fouvry, F., Estival, D., Dauphin, E., Compagnion, H., Baur, J., Balkan, L. and Arnold, D. 1996. TSNLP: test suites for natural language processing. (*In The 16th International Conference on Computational Linguistics*. Copenhagen, Denmark. p.711-716).

Letsholo, R. & Matlhaku, K. 2014. The syntax of the Setswana noun phrase. *Marang: Journal of Language and Literature*, 24:22–41.

Lindén, K., Silfverberg, M. & Pirinen, T. 2009. HFST tools for morphology: an efficient open-source package for construction of morphological analyzers. (*In Mahlow, C. & Piotrowski, M., eds. State of the art in computational morphology*. Berlin: Springer. p.28–47).

Lipps, J. 2011. A lexical-functional analysis of Swahili relative clauses. Oxford: Oxford University. (Dissertation – MPhil).

Lloberes, M., Castellón, I. & Padró, L. 2014. Suitability of ParTes test Suite for parsing evaluation. (*In Proceedings of the 14th International Conference on Parsing Technologies*, Association for Computational Linguistics. Bilbao, Spain. p.61–65).

Lombard, D., Van Wyk, E.B. & Mokgokong, P.C. 1985. Introduction to the grammar of Northern Sotho. Pretoria: Van Schaik.

Louwrens, L.J. 1991. Aspects of Northern Sotho grammar. Pretoria: Via Afrika.

Louwrens, L.J. 1994a. Dictionary of Northern Sotho grammatical terms. Pretoria: Via Afrika.

Louwrens, L.J. 1994b. Aspect as an essential sub-category of the verb in Northern Sotho. *Journal for African Languages*, 14(3):116–127.

Louwrens, L.J., Kosch, I.M. & Kotzé, A.E. 1995. Northern Sotho. München: Lincom Europe.

Machobane, M.E. 2010. Thuto-polelo ea Sesotho. Cape Town:Printing Solutions.

Maegaard, B., Krauwer, S., Choukri, K. & Damsgaard Jørgensen, L. 2006. The BLARK concept and BLARK for Arabic. (*In Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genova: European language resources distribution agency. p.733–778).

https://www.researchgate.net/profile/Bente_Maegaard/publication/228379950_The_BLARK_concept_and_BLARK_for_Arabic/links/02e7e517b7f20f11f3000000.pdf Date of access: 18 June 2016.

Manning, C.D. & Schütze, H. 2000. Foundations of statistical natural language processing. Cambridge: MIT.

Manova, S. & Aronoff, M. 2010. Modeling affix order. *Morphology*, 20:109–131.

Marciniak, M., Mykowiecka, A., Prezepiorkowski, A. & Kupsc, M. 2003. An HSPG-annotated test suite for Polish. (*In Abeillé, A., ed. Treebanks: building and using parsed corpora*. Boston: Kluwer. p.129–146).

Mchombo, S.A. 1991. Reciprocalization in Chichewa: a lexical account. *Linguistic Analysis*, 21:3–22.

Mchombo, S.A. 2001. Effects of head-marking on constituent order in Chichewa. (*In Butt, M. & King, T.H., eds. The Proceedings of the LFG '01 Conference*. University of Hong Kong. p.221–237. <http://csli-publications.stanford.edu/LFG/6/lfg01.html> Date of access: 12 June 2016).

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

Mchombo, S.A. 2007. Argument binding and morphology in Chichewa. (*In Hoyt, F., Seifert, N., Teodorescu, A. & White, J., eds. Texas linguistics society 9: morphosyntax of underrepresented languages*. CSLI. p.203–221.)

Mitchley, H. 2015. Agreement and coordination in Xitsonga, Sesotho and isiXhosa: an optimality theoretic perspective. Rhodes: Rhodes University. (Dissertation – MA).

Moyane, L.H. 1995. Grammatical description in Setswana: an overview. Johannesburg: Rand Afrikaans University. (Dissertation- MA).

- Nordlinger, R. & Bresnan, J. 2011. Lexical-functional grammar: interactions between morphology and syntax. (In Borsley, R & Börsjars, K., eds. *Non-transformational syntax: a guide to current models*. Oxford: Blackwells. p.112–140).
- Nurse, D. 2008. *Tense and aspect in Bantu*. Oxford: Oxford University.
- Olejarnik, M. 2009. *Complex predicates in Swahili: an LFG approach*. Warsaw: University of Warsaw. (Thesis – PhD).
- Pollard, C. & Sag, I.A. 1994. *Head-driven phrase structure grammar*. Chicago: CSLI.
- Posthumus, L.C. 1990. Time reference in Zulu. *South African Journal for African Languages*, 10(1):22–28.
- Posthumus, L.C. 1993. The hierarchy of the essential verb categories in Zulu. *South African Journal for African Languages*, 13(3):95–102.
- Posthumus, L.C. 1994. Word-based versus root-based morphology in African Languages. *South African Journal for African Languages*, 14(1):28–36.
- Posthumus, L.C. 2000. The so-called adjective in Zulu. *South African Journal for African Languages*, 20(2):148–158.
- Poulos, G. 1990. *A linguistic analysis of Venda*. Pretoria: Via Afrika.
- Poulos, G. & Louwrens, L.J. 1994. *A linguistic analysis of Northern Sotho*. Pretoria: Via Afrika.
- Poulos, G. & Msimang, C.T. 1998. *A linguistic analysis of Zulu*. Pretoria: Via Afrika.
- Pretorius, L., Marais, L. & Berg, A.S. 2017. A GF miniature resource grammar for Tswana: modelling the proper verb. *Language Resources and Evaluation*, 51(1):159–189.
- Pretorius, L., Viljoen, B., Pretorius, R. & Berg, A. 2008. Towards a computational morphological analysis of Setswana compounds. *Literator*, 29(1):1–20.

Pretorius, L., Viljoen, B., Pretorius, R. & Berg, A. 2010. A finite-state approach to Setswana verb morphology. (*In Finite-State Methods and Natural Language Processing Revised Selected Papers, Lecture Notes in Computer Science. Volume 6062/2010. Berlin: Springer. p.131–138*).

Pretorius, R., Berg, A. & Pretorius, L. 2012. Multiple objects in Setswana: a computational approach. *Southern African Linguistics and Applied Language Studies*, 30(2):201–218.

Pretorius, R., Berg, A., Pretorius, L. & Viljoen, B. 2009. Setswana tokenisation and computational verb morphology: facing the challenge of a disjunctive orthography. (*In AfLaT2009 - Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages. p.66–73*).

Pretorius, R., Viljoen, B. & Pretorius, R. 2005. A finite state morphological analysis of Setswana nouns. *South African Journal of African Languages*, 25(1):37–47.

Pretorius, R.S. 1997. Auxiliary verbs as a subcategory of the verb in Tswana. Potchefstroom: Potchefstroom University for Christian Higher Education. (Thesis – PhD).

Pretorius, R.S. 2003. Absolute tense forms in Tswana. *Journal for language teaching*, 37(1):13–25.

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

Pretorius, W.J. 2000. Die identifisering en beskrywing van die begrippe stam en wortel in die Afrikatale, met besondere verwysing na die Sothotale. *Journal for Language Teaching*, 34(1):51–61.

Pretorius, W.J. 2004. 'n Dimensionele grammatiese beskrywing van hulpwerkwoorde in Noord-Sotho. Johannesburg: Rand Afrikaans University. (Thesis – PhD).

Prys, D. 2006. The BLARK matrix and its relation to the language resources situation for the Celtic languages. (*In Strategies for developing machine translation for minority languages - 5th SALTMIL Workshop on Minority Languages. Fifth International Conference on Language Resources and Evaluation, Italy, Genoa. p.31–32. <http://mt-archive.info/LREC-2006-SALTMIL-WS.pdf#page=37>. Date of access: 18 June 2016.*

- Ranta, A. 2011. Grammatical framework: programming with multilingual grammars. Stanford: CSLI.
- Ras, W.A.P. 1991. Die ideofoon in Tswana. Potchefstroom: Potchefstroom University for Higher Education. (Dissertation – MA).
- Rice, K. 2011. Principles of affix ordering: an overview. *Word Structure*, 4(2):169–200.
- Rose, S., Beaudoin-Lietz, C. & Nurse, D. 2002. A glossary of terms for Bantu verbal categories with special emphasis on tense and aspect. Munich: Lincom Europa.
- Samuelsson, C. & Voutilainen, A. 1997. Comparing a linguistic and a stochastic Tagger. (*In Proceedings of the 8th Conference on European Chapter of the ACL, Spain, Madrid.* p.246–253).
- Schachter, P. 1992. Comments on Bresnan and Kanerva’s “Locative inversion in Chichewa: a case study of factorization in grammar”. (*In Stowell, T. & Wehrli, R., eds. Syntax and semantics No. 26: syntax and the lexicon.* New York: Academic. p.103–110).
- Setshedi, J.E. 1974. The auxiliary verbs and the deficient verbs in Tswana. Pietersburg: University of Natal. (Dissertation – MA).
- Statistics South Africa. 2011. Census 2011. <http://www.statssa.gov.za.html>. Date of access: 14 September 2012.
- Streiter, O., Stuflessen, M. & Lan Weng, Q. 2006. Models of cooperation for the development of NLP resources: a comparison of BLARK and XNLRDF. (*In Strategies for developing machine translation for minority languages - 5th SALT MIL Workshop on Minority Languages. Fifth International Conference on Language Resources and Evaluation, Italy, Genoa.* p.109–112). <http://mt-archive.info/LREC-2006-SALTMIL-WS.pdf#page=37> Date of access: 18 June 2016.
- Strik, H., Daelemans, W., Binnenpoorte, D., Sturm, J., De Vriend, F. & Cucchiarini, C. 2002. Dutch HLT resources: From BLARK to priority lists. (*In Proceedings of ICSLP-2002, Denver, USA, p.1549-1552*). <http://hstrik.ruhosting.nl/wordpress/wp-content/uploads/2013/04/a92.pdf> Date of access: 18 June 2016).

Sulger, S. 2009. Irish clefting: the LFG perspective. Konstanz: University of Konstanz. (MA –Dissertation).

Sulger, S. 2015. Modeling nominal predications in Hindi/Urdu. Konstanz: University of Konstanz. (Phd – Thesis).

Sulger, S., King, T.H., Dione, C.B., Dyvik, H., Rosen, V., De Smedt, K., Patejuk, A., Çetinoğlu, O., Arka, Mistica, M. 2013. ParGramBank: The ParGram Parallel Treebank. (*In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria. Association for Computational Linguistics.* p.550–560.
http://www.aclweb.org/website/old_anthology/P/P13/P13-1054.pdf Date of access: 10 July 2016).

Taljaard, P.C. & Bosch, S.E. 1988. Handbook of isiZulu. Pretoria: Van Schaik.

Taljard, E. 2001a. 'n Diakroniese perspektief op die identifiserende kopulatief in Bantoe. *South African Journal of African Languages*, 19(1-2):83–95.

Taljard, E. 2001b. Diachronic perspective on variable copulative prefixes with special reference to Sepedi. *South African Journal of African Languages*, 21(3):298–309.

Taljard, E. 2003. 'To be (with)' or 'to have'? a case for grammaticalization in Northern Sotho. *South African Linguistics and Applied Language Studies*, 21(3):169–181.

University of Leipzig. 2015. Leipzig glossing rules. University of Leipzig: Department of linguistics and Max Planck Institute for Evolutionary Anthropology: Department of Linguistics.
<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf> Date of access: 25 July 2015.

Van Wyk, E.B. 1958. Woordverdeling in in Noord-Sotho en Zoeloe. 'n bydrae tot die vraagstuk van woordidentifikasie in die Bantoetaale. Pretoria: University of Preoria. (Thesis – PhD).

Van Wyk, E.B. 1962. Die Bantoetaalkunde as beskrywende taalwetenskap. Pretoria: UP.

Van Wyk, E.B. 1964. Sinsleer. (*In Van der Merwe, H.J.J.M., ed. Studierigtings in die Taalkunde.* Pretoria: Van Schaik).

Van Wyk, E.B. 1966. The word classes of Northern Sotho. *Lingua*, 17(2): 230–261.

Van Wyk, E.B. 1967. Northern Sotho. *Lingua*, 17(1-2):1–32.

Vermeulen, G.J.G. 1984. Voegwoorde in Tswana. Potchefstroom: Potchefstroom University for Christian Higher Education. (Thesis – PhD).

Watters, J. 2000. Syntax. (*In* Heine, B and Nurse, D., eds. African languages: an introduction. Cambridge: Cambridge University. p.194–230).

Zerbian, S. 2006. Expression of information structure in the Bantu language Northern Sotho. Berlin: Humboldt University. (Thesis – PhD).