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
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
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

<table>
<thead>
<tr>
<th>Diagram</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagram 3-1</td>
<td>Mother and daughter nodes</td>
<td>27</td>
</tr>
<tr>
<td>Diagram 3-2</td>
<td>Correspondences between features of argument functions</td>
<td>34</td>
</tr>
<tr>
<td>Diagram 4-1</td>
<td>Absolute tenses</td>
<td>61</td>
</tr>
<tr>
<td>Diagram 4-2</td>
<td>Relative tense</td>
<td>63</td>
</tr>
<tr>
<td>Diagram 7-1</td>
<td>Macro structure of the grammar file</td>
<td>158</td>
</tr>
<tr>
<td>Diagram 7-2</td>
<td>Setswana grammar features and their values</td>
<td>160</td>
</tr>
</tbody>
</table>
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

Setswana\(^1\) (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

\[ ba-tho \quad ba-bo-il-e \quad maabane \]
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

\[ ba-tho \quad ba \]
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,

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).
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 \textit{o reka} (he buys) is included in the structure.

\begin{verbatim}
(1-3)
\begin{center}
\textbf{Independent clause}
\end{center}
Monna o reka khomputara.
man he buy computer
\textit{mo-nna} \textit{o-rek-a} \textit{(ne)-khomputara}
NPrel-man AgrSubj-buy-VEnd NPre9-computer
The man buys a computer.
\end{verbatim}

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.

\begin{verbatim}
(1-4)
\begin{center}
\textbf{Independent clause}
\end{center}
Monna o reka khomputara.
man he buy computer
\textit{mo-nna} \textit{o-rek-a} \textit{(ne)-khomputara}
NPrel-man AgrSubj-buy-VEnd NPre9-computer
The man buys a computer.
\end{verbatim}

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 \textit{mme} (and).
The wind comes in when we open the door.

They thank the woman and they go home.

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

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), Categorial 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 & Voutilannen, 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 $\phi$ (phi) function (Falk, 2001:64). The $\phi$ function is defined as the composition of the $\mu$, $\alpha$ and $\lambda$ functions. The $\mu$ function specifies the mapping from the c-structure tree to the m-structure, the $\alpha$ function specifies the mapping from the m-structure to the a-structure while the $\lambda$ 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)](image-url)
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 verbs\textsuperscript{14}. 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

XLE\textsuperscript{15} 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

\textsuperscript{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.

\textsuperscript{15} The XLE documentation is available at: http://www2.parc.com/isl/groups/nltt/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).

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.
1.2 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.

A comprehensive account of LFG is presented in the books of Dalrymple (2001) and Bresnan (2001)\textsuperscript{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.

\textsuperscript{16} A second edition of this book is published in 2016 (Bresnan \textit{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 LFG\textsuperscript{17} scholarly community. The Essex LFG website\textsuperscript{18}, the Google+ ILFGA website\textsuperscript{19}, and the LFG Facebook website\textsuperscript{20} 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 framework\textsuperscript{21}. The proceedings of the LFG conferences are published from 1996 online by CSLI Publications\textsuperscript{22}. 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

\textsuperscript{17} The official website for ILFGA is https://sites.google.com/site/ilfgalfg/home/
\textsuperscript{18} http://www.essex.ac.uk/linguistics/external/LFG/index.html
\textsuperscript{19} Google+ ILFGA page: https://plus.google.com/109464318749972104499
\textsuperscript{20} The LFG Facebook page: https://www.facebook.com/lfgpage
\textsuperscript{21} The LFG bibliography is available at http://www.essex.ac.uk/linguistics/external/LFG/Bibliography/bibliography.html
\textsuperscript{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) categories\(^{23}\) (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 group\(^{24}\) 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. Kruger (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).

Kruger (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.


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.

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](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/](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/](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 RGL\textsuperscript{29} (Pretorius, L., \textit{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.

\textsuperscript{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(functional)-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.

\[
(3-1) \quad S \rightarrow NP \ VP
\]

<table>
<thead>
<tr>
<th>Mother node</th>
<th>Symbol</th>
<th>Daughter nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S )</td>
<td>( \rightarrow )</td>
<td>( NP \ VP )</td>
</tr>
</tbody>
</table>

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 → {N | N 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 → (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 (ADVP) in a Setswana VP.
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 specified. Either the absolute pronoun can be followed by a demonstrative pronoun (3-8) or it can be preceded by the demonstrative pronoun (3-9).

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.

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

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).
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.

\[(3-11) \quad \text{VP} \rightarrow \text{V XP}\]

\[(3-12) \quad \text{XP} \equiv \{\text{NP} \mid \text{ADVP} \mid \text{INSTRPARTP} \mid \text{TEMPPARTP}\}\]

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).

\[(3-13) \quad S \leftarrow \text{root node (nonterminal)}\]

\[\text{NP} \leftarrow \text{branches}\]

\[\text{VP} \leftarrow \text{nodes (nonterminal)}\]

\[\text{N} \leftarrow \text{branches}\]

\[\text{V} \leftarrow \text{nodes (nonterminal)}\]

\[\text{mosadi o a reka} \leftarrow \text{branches}\]

\[\text{terminal nodes (leaves)}\]

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 &
Toivonen, 2015:377). The terminal nodes of a phrase structure tree are thus the individual words filling a single c-structure node (3-14).

(3-14)

```
S
 / 
NP VP
 / 
N V NP
 / 
mosadi o reka mosese
```

The exposition in (3-15) shows the establishment of phrase structure rules and their phrase structure tree representations for the sentence in (3-2).

(3-15) **Rule**

```
S → NP VP
```

**Tree representation**

```
S
 / 
NP VP
 / 
V NP
 / 
N
```

### 3.3 ARGUMENT STRUCTURE

Argument structure deals with predicate-argument information such as the number and type of arguments of a predicator. In terms of semantics, a-structure represents the core participants in actions designated by the predicator. In terms of syntax, it represents the nucleus of information required to characterise the syntactic dependents of an argument-taking head (Bresnan, 2001:304). In LFG, these two representations are linked through the so-called Lexical Mapping Theory (cf. §3.3.2). Before explaining this theory, we introduce the central notion of grammatical functions (GFs) in LFG.
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 (\textit{SUBJ}), object (\textit{OBJ}), secondary object (\textit{OBJ}_θ), oblique (\textit{OBL}_θ), complement (\textit{COMP}), open complement (\textit{XCOMP}), adjunct (\textit{ADJUNCT}) and open adjunct (\textit{XADJ}). The subscript θ of a particular argument refers to the semantic role associated with the argument, for example, in \textit{OBJ}_{\text{theme}} θ refers to the theme semantic role, whereas θ refers to the locative semantic role in the argument \textit{OBL}_{\text{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: \textit{SUBJ}, \textit{OBJ}, \textit{OBJ}_θ, \textit{OBL}_θ, \textit{COMP} and \textit{XCOMP}. A GF that is not governable is often referred to as a modifier (nonargument function): \textit{ADJUNCT} and \textit{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 \textit{SUBJ}, \textit{OBJ} and \textit{OBJ}_θ while noncore GFs (nonterms) are \textit{OBL}_θ, \textit{COMP} and \textit{XCOMP}.
- A semantically restricted GF can only be associated with a limited set of semantic roles (cf. §3.3.2): \textit{OBJ}_θ and \textit{OBL}_θ. A semantically unrestricted GF may be connected to any semantic role or even more than one semantic role: \textit{SUBJ} and \textit{OBJ}.
- An open GF does not contain an internal subject phrase and the \textit{SUBJ} must be specified externally to its phrase: \textit{XCOMP} and \textit{XADJ}. A closed GF contains an internal subject phrase: \textit{SUBJ}, \textit{OBJ}, \textit{OBJ}_θ, \textit{OBL}_θ, \textit{COMP} and \textit{ADJUNCT}.
- The objective GFs are \textit{OBJ} and \textit{OBJ}_θ while the nonobjective GFs are \textit{SUBJ} and \textit{OBL}_θ.

LFG also allows for grammaticalised discourse functions: topic (\textit{TOP}), focus (\textit{FOC}) and \textit{SUBJ}. The \textit{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 \textit{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 \textit{XCOMP} function. A more recent development in LFG has been the introduction of the so-called \textit{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

\footnotesize{32 We refer to the secondary object or the indirect object as \textit{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. \textsc{predlink} is a closed category and "there is no control equation between the \textsc{subj} and the \textsc{predlink} and hence no need for NPs, APs, and PPs to have subject arguments" (Butt \textit{et al.} 1999:70).

### 3.3.2 LEXICAL MAPPING THEORY


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

\begin{equation}
\text{rekel- (buy for) < agent, patient, theme >}
\end{equation}

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

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

<table>
<thead>
<tr>
<th></th>
<th>SUBJ</th>
<th>OBJ₀</th>
<th>SUBJ/OBJ</th>
</tr>
</thead>
</table>

(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.\textsuperscript{34}) on the relation between thematic roles and GFs:

- **Subject condition**
  Every verb must have a \texttt{SUBJ}. The most prominent thematic role ([−o]) is the agent, which is realised as the \texttt{SUBJ}. If an agent is not available, the thematic role that follows ([−r]) is realised as the \texttt{SUBJ}. This means that the patient/theme thematic role also maps to a \texttt{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.

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

\[
(3-20) \quad f:M \text{ where } M = \begin{bmatrix}
\text{ATTRIBUTE}_1 & \text{VALUE}_1 \\
\text{ATTRIBUTE}_2 & \text{VALUE}_2
\end{bmatrix}
\]

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 \( (f \text{ attribute}_2) = \text{value}_2 \).

---

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) \[
\text{NUM } \begin{array}{l} \text{sg} \end{array}
\]

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\textsuperscript{37} in (3-22).

(3-22) \[
\text{PRED } \text{‘SADI’}
\]

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) \[
\text{PRED } \text{‘REK <SUBJ OBJ>’}
\]

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.

\textsuperscript{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).
(3-24) Re tla tsamaya gone jaanong.
we shall walk right now

re-tla-tsamay-a gone jaanong
AgrSubjPlp1-FutPre-walk-VEnd Adv Adv

We shall walk/leave right now.

(3-25)

The value of an attribute may also be an f-structure (Bresnan, 2001:47; Dalrymple, 2001:31). For example, in (3-26) the value of the SUBJ attribute is represented with an AVM. The f-structure in this instance consists of the attributes PRED and CLASS. The value of PRED is the noun root -sadi (woman) and it is a class 1 noun.

(3-26) SUBJ [PRED ‘SADI’]
       [CLASS 1]

We conclude this section with the f-structure (3-27) of (3-2), which we use in subsequent examples.

(3-27) [PRED ‘REK <SUBJ OBJ>’]
       [CLASS 1]
       [TENSE pres]

3.4.2 CONDITIONS OF WELL-FORMEDNESS

Requirements on f-structures in LFG are enforced by a set of conditions referred to as conditions of well-formedness. To ensure that an f-structure representation is valid, it must conform to three conditions, i.e. the conditions of completeness, coherence and consistency (uniqueness). These conditions are used to confirm grammatically correct structures or to reject ungrammatical structures.
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.

The woman buys for the girl.

(3-29)

\[
\begin{array}{c}
\text{PRED} \quad \text{REK <SUBJ OBJ_θ OBJ>}
\end{array}
\]

\[
\begin{array}{c}
\text{CLASS} \quad 1
\end{array}
\]

\[
\begin{array}{c}
\text{TENSE} \quad \text{pres}
\end{array}
\]

\[
\begin{array}{c}
\text{SUBJ} \quad \text{PRED} \quad \text{SADI}
\end{array}
\]

\[
\begin{array}{c}
\text{CLASS} \quad 1
\end{array}
\]

\[
\begin{array}{c}
\text{OBJ_θ} \quad \text{PRED} \quad \text{SETSANA}
\end{array}
\]

\[
\begin{array}{c}
\text{CLASS} \quad 1
\end{array}
\]

\[
\text{OBJ}
\]

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.
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).
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).

\[(\text{f CLASS}) = c \ 1\]

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.

\[(\text{f TENSE})\]

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.

\[\neg(\text{f TENSE})\]

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).

\[\{(\text{f PRONTYPE}) = \text{AbsPro} \mid (\text{f PRONTYPE}) = \text{PossPro}\}\]

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.

\[(3-38)\] \(
((\text{SUBJ} \ \text{PRED}) = \text{’PRO’})
\)

(3-39) Batho ba a taboga.
people they run

\(\text{ba-tho} \quad \text{ba-a-tabog-a}\)

NPre2-person AgrSubj2-PresPre-run-VEnd

The people are running.

(3-40) Ba a taboga.
they run

\(\text{ba-a-tabog-a}\)

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**.

\[(3-41)\] \(
(f \ \text{TENSE}) \neq \text{PRES}
\)

or

\(\neg (f \ \text{TENSE}) = \text{PRES}\)

### 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 \(\phi\) function\(^{38}\).

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 \(\phi\) 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 \(\phi\) function is also referred to as the correspondence function, the mapping function or the projection function.
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 $\phi$ 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 $\phi(V)=f_5=f_1$.

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 $\uparrow\text{SUBJ}=1$ and $\downarrow=1$ 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 ↑ and ↓. The ↓ refers to the f-structure associated with the daughter node under which the schema containing the ↓ appears (see (3-44)). The ↑ refers to the f-structure associated with the mother node of the daughter node under which the schema containing the ↑ appears.

(3-44) \[ S \rightarrow NP \quad VP \\
(↑SUBJ)=↓ \quad ↑=↓ \]

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

- The annotation \((↑SUBJ)=↓\) 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 \(↑=↓\) 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).

(3-45)

A lexical entry in LFG includes a "categorial 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 ↑ (Dalrymple, 2001:121). Examples (3-46) and (3-47) show the lexical entries for -sadi, the noun root for mosadi (woman) and -rek-, 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 (↑ PRED) = ‘SADI’
    (↑ NTYPE) = ord
    (↑ CLASS) = 1

(3-47) rek V (↑ PRED) = ‘REK’<SUBJ>
    (↑ VTYPE) = main
    (↑ CLASS) = 1
    (↑ 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 representation\(^{39}\). 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 $\phi$, 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’s\textsuperscript{40} 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).

\textsuperscript{40} Cole was Doke’s student.
Although numerous studies 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 ideophones. 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 \textit{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 \textit{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 \textit{eng}? (what?), \textit{kae}? (where?), \textit{leng}? (when?) and \textit{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
mo-lao me-lao
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 *ditseleng* (in the roads) is formed from the noun *ditsela* (roads).

(4-4)  
ditsela  >  mo ditseleng\(^{43}\)  
roads  in roads locality  
di-tsela  mo  di-tsela-ing  
NPre10-road  LocPartmo  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 *ditselana* (small roads) is formed from the noun *ditsela* (roads).

(4-5)  
ditsela  >  ditselana  
roads  roads small  
di-tsela  di-tsela-ana  
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, chieftainess) in (4-6) is formed from the noun *kgosi* (chief).

(4-6)  
kgosi  >  kgosigadi  
king  chief feminine  
(ne)-kgosi  (ne)-kgosi-gadi  
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 *podi* (goat).

---

\(^{43}\) Owing to morphophonological alternation, the morpheme sequence *di-tsela-ing* results in the surface form *ditseleng*. This phenomenon is common in our examples.
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.
the milk is cold

\[
\text{ma-šwi a tsididi}
\]

NPre6-milk DesCopV6 adjective

(4-9) kofi e tsididi
the coffee is cold

\[
\text{(ne)-kofi e tsididi}
\]

NPre9-coffee DesCopV9 adjective

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
a big person

\[
\text{mo-tho yo mo-golo}
\]

NPre1-person QualPart1 NPre1-adjective

(4-11) batho ba bagolo
people that are big

\[
\begin{array}{ccc}
ba-tho & ba & ba-golo \\
NPre2-person & QualPart2 & NPre2-adjective
\end{array}
\]
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 \textit{go} as well as a verb root and an ending \textit{-a}. The verb root in (4-12) is \textit{tshamek-} (play). The infinitive noun has valence for locative and/or diminutive suffixes (Krüger 2006:93).

(4-12) go tshameka
to play

\[
\begin{array}{c}
go-tshamek-a \\
NPre15-play-VEnd
\end{array}
\]
the playing

Nouns in classes 16 to 20 constitute the \textit{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

\[
\begin{array}{c}
mo-rago \\
NPre18-behind
\end{array}
\]
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
        \[ PRED \quad \text{`SADI'} \]
        \[ NTYPE \quad \text{ord} \]
        \[ CLASS \quad 1^{45} \]

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)."
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).

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.

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

\begin{align*}
\text{(4-22) rona} & \quad \text{PRED} \quad \text{\textquote{RONA}} \\
& \quad \text{PRONTYPE} \quad \text{PersPro} \\
& \quad \text{PERS} \quad \text{P1} \\
& \quad \text{NUM} \quad \text{pl}
\end{align*}

\begin{align*}
\text{(4-23) yo} & \quad \text{PRED} \quad \text{\textquote{YO}} \\
& \quad \text{PRONTYPE} \quad \text{DemPro} \\
& \quad \text{CLASS} \quad \text{1}
\end{align*}

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 suffixes\textsuperscript{47} 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 \textsc{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

\textsuperscript{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 \textit{tabogisa} (let run) and consists of the verb root \textit{tabog-}, the causative extension \textit{-is-} and the verbal ending \textit{-a}.

(4-24) \hspace{1em} ba a taboga
\hspace{1em} they run
\hspace{1em} \textit{ba-tabog-a}
\hspace{1em} AgrSubj2-run-VEnd
\hspace{1em} they run

(4-25) \hspace{1em} ba tabogisa
\hspace{1em} they run let
\hspace{1em} \textit{ba-tabog-is-a}
\hspace{1em} AgrSubj2-run-CausSuf-VEnd
\hspace{1em} they let run

On the other hand we adhere to the \textit{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.

\textbf{4.6.1.1 Mood}

We distinguish eight verbal moods in Setswana, viz. the indicative, consecutive, imperative, habitual, participial, subjunctive, infinitive and relative moods\textsuperscript{48}. 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.

\textsuperscript{48} There are different opinions on the modal classification. See also, for example, Louwrens, \textit{et al.} (1995).
<table>
<thead>
<tr>
<th>Prefixes</th>
<th>Root</th>
<th>Suffixes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>NegPre</th>
<th>AgrSubj</th>
<th>NegPre</th>
<th>PresPre</th>
<th>ProgPre</th>
<th>PotPre</th>
<th>FutPre</th>
<th>AgrObj</th>
<th>ReflPre</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>ga</td>
<td>Person</td>
<td>se</td>
<td>a</td>
<td>sa</td>
<td>a</td>
<td>tla</td>
<td>Person</td>
<td>Class</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>sa</td>
<td>a</td>
<td>sa</td>
<td>ka</td>
<td></td>
<td>Class</td>
<td>i- (reflexive)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CausSuf</th>
<th>ApplSuf</th>
<th>RecSuf</th>
<th>PerfSuf</th>
<th>PassSuf</th>
<th>Mood</th>
<th>Tense</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>-is-</td>
<td>-el-</td>
<td>-an-</td>
<td>-il-</td>
<td>-iw-</td>
<td>-a</td>
<td>-e</td>
<td>(e or ê)</td>
</tr>
<tr>
<td>-y-</td>
<td></td>
<td></td>
<td></td>
<td>-w-</td>
<td>-ng (relative)</td>
<td>-ng (imperative)</td>
<td></td>
</tr>
</tbody>
</table>

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

A. Linear arrangement
B. Morphological items
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

\[
\begin{align*}
\text{ba-sadi} & \quad \text{ba-rut-a} & \quad \text{ba-ana} & \quad \text{di-bal-o} \\
\text{NPre2-woman} & \quad \text{AgrSubj2-teach-VEnd} & \quad \text{NPre2-child} & \quad \text{NPre10-count-DevSuf}
\end{align*}
\]

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

\[
\begin{align*}
\text{mo-nna} & \quad \text{a-kop-a} & \quad \text{(ne)-thuso} \\
\text{NPre1-man} & \quad \text{ConsAgrSubj1-seek-VEnd} & \quad \text{NPre9-help-DevSuf}
\end{align*}
\]

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.
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  
LocPartkwa  
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.
4.6.1.2 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).

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<sub>e</sub>, OBJ, OBL<sub>e</sub>, 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<sub>e</sub> 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.
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). 

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

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) 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.
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

gar-e-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).

The relative past tense indicates an action that is incomplete and continuing at a certain moment in the past (Krüger, 2013b:396). 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 indefinite tense (Pretorius, 1997:249; Krüger, 2013b:396).
The negative of the relative past tense is formed by means of the negative morpheme sa and the verbal ending -e (4-40).

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).

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

---

\(^{52}\) 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.

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.

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.

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.

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.

The children play tennis.

\(^{55}\) It should be noted that the root r- of the stem re and the root its- of the stem itsa also occur with the verbal ending -e in the positive. Such stems are labelled as defective stems (Pretorius, 1997:26).
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*
   NPrel-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*
   NPrel-woman AgrSubj1-run-CausSuf-VEnd NPrel-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 OBJθ 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

\[
\begin{array}{ccc}
  \text{ba-ana} & \text{ba-kwal-a} & (\text{ne})-\text{teko} \\
  \text{NPre2-child} & \text{AgrSubj2-write-VEnd} & \text{NPre9-test} \\
\end{array}
\]

The children are writing a test.

(4-54) Mosadi o kwadisa bana teko.
woman she write let children test

\[
\begin{array}{ccc}
  \text{mo-sadi} & \text{o-kwal-is-a} & \text{ba-ana} & (\text{ne})-\text{teko} \\
  \text{NPre1-woman} & \text{AgrSubj1-write-CausSuf-VEnd} & \text{NPre2-child} & \text{NPre9-test} \\
\end{array}
\]

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) kwal- (write) \[< \text{agent theme} >\]

\[
\begin{array}{ccc}
  \text{IRC:} & [-o] & [-r] \\
  \text{DRC:} & [-r] \\
  \text{bi-uniqueness:} & \text{SUBJ} & \text{OBJ} \\
\end{array}
\]

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, the secondary (indirect) object with \(\theta\) has the value patient. In terms of word order, the secondary object appears nearest to the causative verb.
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 OBLθ, 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)-kgetsi-ing
LocPartmo NPre9-bag-LocSuf

The woman puts the book in the bag.

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

NPrela-Katlego  AgrSubj1a-PresPre-play-VEnd

Katlego is playing.

(4-60) Katlego o tshamekela setlhopa 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

NPrela-Katlego  AgrSubj1a-play-ApplSuf-VEnd  NPre7-choose-DevSuf

sa  ntlha  (sa  se-kolo  sa  rona)

PossPart7  Adv  (PossPart7  NPre7-school  PossPart7  PossProP1pl)

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).
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 OBJθ in (4-66).
An applicative verb can also introduce a locative thematic role as it signifies a *movement or transition with respect to place, state or condition* (Krüger, 2006:244; Chavula, 2016:147–149). In (4-67) the verb *ba a tsamaya* (they walk) is intransitive. The inclusion of the applicative extension -el- changes this verb to a transitive verb *ba tsamaela* (they walk towards) in (4-68).

(4-67) Batho ba a tsamaya.

people they walk

*ba-tho*  *ba-a-tsamay-a*

NPre2-person AgrSubj2-PresPre-walk-VEnd

The people are walking.

(4-68) Batho ba tsamaela kwa nokeng.

people they walk towards to river

*ba-tho*  *ba-tsamay-el-a*  *kwa*

NPre2-person AgrSubj2-walk-ApplSuf-VEnd LocPart{kwa}

*(ne)-noka-ing*

NPre9-river-LocSuf

The people walk to the river.

In (4-67) *batho* (people) is the agent and it maps to *SUBJ* in (4-69).

(4-69) *tsamay-*(walk) < agent >

<table>
<thead>
<tr>
<th>IRC</th>
<th>DRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-o]</td>
<td>[-r]</td>
</tr>
</tbody>
</table>

In (4-68), the phrase *kwa nokeng* (to the river) has a locative thematic role. By applying the IRC and DRC, the respective semantic features are [-o] and [+r] and it is then mapped to *OBL* in example (4-70) where *θ* is *locative* in this instance.
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-`.

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

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.
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)\textsuperscript{56}.

(4-75) Bana ba a ratana.

\begin{verbatim}
  ba-ana   ba-a-rat-an-a
NPre2-children AgrSubj2-love-RecSuf-VEnd
\end{verbatim}

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.

\textsuperscript{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).
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.

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-young 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).

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), 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** Ø |

In the second valency reducing approach, both the agent and patient (or beneficiary) 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 Ø 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.

\[
\begin{align*}
(4-81) \quad \text{ratan- (love each other)} & < \text{agent patient} > \\
\text{Argument unification:} & \quad \\downarrow \\
\text{IRC:} & \quad [-o] \\
\text{DRC:} & \quad [-r] \\
\text{Subject condition:} & \quad \text{SUBJ}
\end{align*}
\]

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 (PROrecip) (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.

\[
\begin{align*}
(4-82) \quad \text{ratan- (love each other)} & < \text{agent patient} > \\
\text{(PROrecip)} & \quad \text{[[-o] [-r]} \\
\text{IRC:} & \quad \text{[-r]} \\
\text{DRC:} & \quad \text{SUBJ/SUBJ/OBJ} \\
\text{bi-uniqueness:} & \quad \text{SUBJ OBJ}
\end{align*}
\]

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 OBLass (4-83).

(4-83) **dumalan** (agree with each other)  < [ agent] >  [ ]
IRC:  [-o]  [-o]
DRC:  [+r]  [+r]

(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.
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 ADJUCT 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).

Subject condition:

Subject condition: SUBJ
seen as an extended root followed by an extension, again resulting in an extended root. For example, tli\textit{s}etsa\textsuperscript{60} (brings for) has a basic root and two extensions, but we may also think of this verb as an extended root \textit{tlis-} (let come) followed by the extension \textit{-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)\textsuperscript{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.

\textbf{(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 \textit{cause to do} to the basic verb root. When an applicative extension is suffixed to the causative verb, a meaning of \textit{doing something for someone is added to the verb} so that a derivation of \textit{cause to do for} is possible (Chavula, 2016:204).

In (4-88), the verb o tlisa (she brings) includes the causative extension \textit{-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).

\textsuperscript{60} Owing to morphophonological alternation, the morpheme sequence \textit{tl-is-el-a} results in the surface form tlisetsa.

\textsuperscript{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).
The woman brings the food.

In (4-90), an applicative extension is suffixed to the causative verb **o tlisa** (she brings) resulting in the verb **o tlisetsa** (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 **OBJₐ** in (4-91).

In the following example, (4-92), the verb **o tsenya** (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 **OBLloc** (4-93).
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 OBJθ (4-95).

(4-95) tsenyets- (put for) < agent beneficiary theme locative >
IRC:        [-o]    [+o]    [-r]    [-o]
DRC:        [-r]    [+r]    [+r]    [+r]
            SUBJ   OBJθ   SUBJ/OBJ   OBLloc
bi-uniqueness: SUBJ   OBJθ   OBJ   OBLloc

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 OBJθ.
(ii) Causative + Reciprocal

The causative verb may also take a reciprocal extension which ads 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 koloi (car keys) the theme. The agent, patient and theme thematic roles are mapped as before (4-97).

(4-96) Rre o batlisa\(^{62}\) mme dinotlolo tsa koloi.
father he seek let mother keys of car
\[-\text{rre} \quad \text{\textit{o-batl-is-a}} \quad \text{\textit{(-)-mme}}\]
\(\text{NPre1a-father} \quad \text{AgrSubj1a-seek-CausSuf-VEnd} \quad \text{NPre1a-mother} \)
di-notlolo tsa (ne)-koloi
\(\text{NPre8-key PossPart8 NPre10-car} \)
Father helps mother to seek the car keys.

(4-97) batlisa- \(^{63}\) let seek < agent patient theme >
IRC: [-o] [+o] [-r]
DRC: [-r] [+r]
\(\text{SUBJ OBJ } \text{SUBJ/OBJ} \)
bi-uniqueness: \(\text{SUBJ OBJ OBJ} \)

In (4-98), the causative extension allows the inclusion of a theme dinotlolo tsa koloi (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 batlisana\(^{63}\) dinotlolo tsa koloi.
Father and mother they seek let each other keys of car
\[-\text{rre} \quad \text{\textit{le}} \quad \text{\textit{(-)-mme}}\]
\(\text{NPre1a-father Conj NPre1a-mother} \)
ba-batl-is-an-a di-notlolo tsa (ne)-koloi
\(\text{AgrSubj2-want-CausSuf-RecSuf-VEnd NPre9-key PossPart9 NPre9-car} \)
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

The children feed each other.

(4-101) jesan- (let eat each other) < [ agent] >

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

The coach trains the athletes today.
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
ba-tabog-i ba-tabog-is-iw-a
NPre2-run-DevSuf AgrSubj2-run-CausSuf-PassSuf-VEnd
ke mo-katis-i gompieno
AgPart NPre1-train-DevSuf Adv
The athletes are trained by the coach today.

(4-105) tabogisiw- (trained) < agent patient >

(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, and OBJ respectively (4-107).
Ditsala di kwalela batho makwalo.

friends write for people letters

di-tsala    di-kwal-el-a
NPre10-friend AgrSubj10-write-ApplSuf-VEnd
ba-tho       ma-kwal-o
NPre2-person NPre6-write-DevSuf

The friends write letters to the people.

kwalela- (write for)  < agent  beneficiary  theme >

IRC:                         [-o]       [+o]       [-r]
DRC:                         [-r]       [+r]       [+] SUBJ  OBJ  SUBJ/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 (PROrecip) 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).

Ditsala di kwalelana makwalo.

friends write for each other letters

di-tsala    di-kwal-el-an-a
NPre10-friend AgrSubj10-write-ApplSuf-RecSuf-VEnd
ma-kwal-o
NPre6-write-DevSuf

The friends write letters to each other.

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

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

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
The sentence is mosadi (woman) and the theme is dijo (food). The agent maps to SUBJ, the beneficiary to OBJ, and the patient to OBJ (4-111).

(4-110) Mosadi o rekela bana dijo.

woman she buy for children food

mo-sadi o-rekel-a ba-ana

NPrel-woman AgrSubj1-buy-ApplSuf-VEnd NPre2-child
di-j-o

NPrel-eat-DevSuf

The woman buys food for the children.

(4-111) rekel- (buy for) < agent beneficiary theme >

IRC: [-o] [+o] [-r]
DRC: [-r] [+r] |
| SUBJ OBJ OBJ

bi-uniqueness: SUBJ OBJ OBJ

The verb di rekelwa (it is bought for) in the passive sentence in (4-112) includes an applicative extension and a passive extension. The beneficiary bana (children) maps to SUBJ (4-113) while the agent thematic role mosadi (woman) is suppressed. The agent is the complement in the agentative particle phrase and is assigned an ADJUNCT function. The applicative extension still requires the inclusion of the theme thematic role, i.e. dijo (food) and it maps to OBJ.

(4-112) Bana ba rekelwa dijo ke mosadi.

Children they buy for food by woman

ba-ana ba-rekel-w-a di-j-o

NPrel2-children AgrSubj2-buy-ApplSuf-PassSuf-VEnd NPrel0-eat-DevSuf
dijo

ke mo-sadi

AgPart NPre1-woman

The children are bought food by the woman.

---

64 The following structure is also grammatical. The theme maps to SUBJ and the beneficiary to OBJ.

Dijo di rekelwa bana ke mosadi.

food it buy for is children by woman

di-j-o di-rekel-w-a ba-ana

NPrel0-eat-DevSuf AgrSubj10-buy-ApplSuf-PassSuf-VEnd NPre2-child

de-j-o

AgPart NPre1-woman

The food is bought for the children by the woman.
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 \textit{SUBJ} while the theme thematic role maps to \textit{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 \textit{makwalo} (letters) then maps to the \textit{SUBJ} while the agent \textit{ditsala} (friends) is suppressed (4-115).

(4-114) \textit{Makwalo a kwalelanwa ke ditsala}.
- letters it write for each are by friends
  \textit{ma-kwal-o} \textit{a-kwal-el-an-w-a}
  NPre6-write-DevSuf AgrSubj6-write-ApplSuf-RecSuf-PassSuf-VEnd
  \textit{ke ditsala}
  AgPart NPre9-friend
  Letters are written to each other by the friends.

(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) \textit{rre} (father) is the agent, \textit{mme} (mother) is the beneficiary, \textit{madi} (money) is the theme, and \textit{mo bankeng} (in the bank), the locative. These thematic roles map to \textit{SUBJ}, \textit{OBJ}, \textit{OBJ} and \textit{OBL}_{loc} (cf. (4-95)).
example (4-116), a reciprocal extension is added to the extended root. The reciprocal extension necessitates a plural denoting **SUBJ** and therefore an **OBJ** 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 **OBLloc** (4-117).

(4-116) **Batho ba tsenyetsana madi mo bankeng.**
people they enter let for each other money in bank

\[ ba-tho \quad ba-tsen-is-el-an-a \]

NPre2-person AgrSubj2-enter-CausSuf-ApplSuf-RecSuf-VEnd

\[ ma-di \quad mo \quad (ne)-banka-ing \]

NPre6-money LocPartmo NPre9-bank-LocSuf

The people put money in the bank for each other.

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

IRC: 

\[ [-o] \quad [-r] \quad [-o] \]

DRC: 

\[ [-r] \quad [+r] \]

bi-uniqueness:

\[ SUBJ \quad SUBJ/OBJ \quad OBLloc \]

\( (viii) \text{ Causative + Applicative + Reciprocal + Passive} \)

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 agentative 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 **OBLloc**, 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

\[ ma-di \quad a-tsen-is-el-an-w-a \]

NPre6-money AgrSubj6-enter-CausSuf-ApplSuf-RecSuf-PassSuf-VEnd

\[ ke \quad ba-tho \quad mo \quad (ne)-banka-ing \]

AgPart NPre2-person LocPartmo NPre9-bank-LocSuf

Money is being put in the bank for each other by the people.
(4-119) tsenyetsanw- (is being put for each other) < agent theme locative >

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

Ø SUBJ/OBJ OBLloc

Subject condition:

[viii] Extension sequences not adhering to CARP

Extension sequences that do not conform to the CARP template are rare. Pretorius (2014) identifies a small number 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
o-batl-a o-batl-isis-a
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.
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

---

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.

\(\text{o-ne} \ \text{a-rut-a} \ \text{ba-ana}\)

She was teaching the children mathematics.

\(\text{AgrSubj1-Aux} \ \text{AgrSubj1-teach-VEnd} \ \text{NPre2-child}\)

\(\text{di-bal-o}\)

\(\text{NPre10-count-DevSuf}\)

4.6.3 \textbf{COPULATIVE VERBS}

We distinguish between three copulative verbs, viz. the identifying, describing, and associative copulative verb\(^{68}\). These copulative verbs can be used in the same tense forms as those distinguished for main verbs (Krüger, 2013b:422–486).

\textit{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.

\(\text{ba-tho} \ \text{ba} \ \text{ke} \ \text{ba-dir-i}\)

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).

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

(4-124) Rona re badiri.

we are workers

rona re ba-dir-i
PersProP1pl 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

mo-simane o mo-golo
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

ba-setsana ba-na le di-buka
NPre2-girl AgrSubj2-AssCopV AssPart NPre10-book

The girls have books.

94
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 (\textsc{vtype}) feature that is one of \{main, idcopv, descopv, asscopv, aux\}.
- The agreement feature \textsc{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 \textsc{mood} assumes one of the following values: \{ind, cons, imp, hab\}.
- The tense feature \textsc{tense} assumes one of the following values: \{pres, past, fut, relpast, relfut\}.
- The aspect feature \textsc{asp} has one of the following values: \{perf, prog, pers\}.
- The polarity feature \textsc{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 \textit{o tabogisa} (he let run) in sentence (4-50) is presented in (4-127).

\begin{equation}
\text{(4-127) o tabogisa} \quad \begin{array}{l}
PRED \quad \text{TABOG <SUBJ OBJ>’69} \\
\text{VTYPE} \quad \text{main} \\
\text{CLASS} \quad 1 \\
\text{MOOD} \quad \text{ind} \\
\text{TENSE} \quad \text{pres} \\
\text{POL} \quad \text{pos} \\
\text{CAUS} \quad + \\
\end{array}
\end{equation}

\section*{4.7 PARTICLES}

In Setswana no prepositions are distinguished, instead the use of so-called particles expresses the various meanings that are related to prepositions\textsuperscript{70}. 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. §4.6.1).

\textsuperscript{69} The subcategorisation frames of main verbs are elaborated on in Chapter 6.

\textsuperscript{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 qualitative 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 qualitative 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-nts1
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 koloi

with car

ka  (ne)-koloi
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
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   NPre-la-father
to father

(4-134) ga ntate
at father
   ga    (-)-ntate
   LocPart_ga   NPre-la-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$_f$a  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$_m$o  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$_k$wa  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

<table>
<thead>
<tr>
<th>ke-dir-a</th>
<th>le</th>
<th>(-)-mme</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgrSubj1-work-VEnd</td>
<td>AssPart</td>
<td>NPrela-mother</td>
</tr>
</tbody>
</table>

I work with mother.

(4-141) Mosadi o na le koloi.

woman she is with car

<table>
<thead>
<tr>
<th>mo-sadi</th>
<th>o-na</th>
<th>le</th>
<th>(ne)-koloi</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPre1-woman</td>
<td>AgrSubj1-AssCopV</td>
<td>AssPart</td>
<td>NPre9-car</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>o-bu-a</th>
<th>fela</th>
<th>jaaka</th>
<th>wena</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgrSubj1-talk-VEnd</td>
<td>Adverb</td>
<td>CompPart</td>
<td>PersProP2sg</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>a</th>
<th>nke</th>
<th>ke-bon-e</th>
</tr>
</thead>
<tbody>
<tr>
<td>HortPart</td>
<td>Aux</td>
<td>AgrSubjPlsg-see-VEnd</td>
</tr>
</tbody>
</table>

Let me see!

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

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


- The CLASS feature of possessive and qualifying 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`  
<table>
<thead>
<tr>
<th>PRED</th>
<th><code>LA</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTTYPE</td>
<td>PossPart</td>
</tr>
<tr>
<td>CLASS</td>
<td>5</td>
</tr>
</tbody>
</table>

### 4.8 CONJUNCTION

A conjunction\(^{71}\) 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`  
father and mother

---

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

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).

4.9 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 quiet 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).
### 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).

(5-1) CLNP

\[
\begin{array}{c}
\text{dikoloi tsotlhe} \\
cars all \\
di-koloi \quad tsotlhe \\
NPre10\text{-car IncQPro10} \\
all the cars
\end{array}
\]

(5-2) LOCNP

\[
\begin{array}{c}
\text{mo ofising e} \\
in office "location" this \\
mo \quad (ne)-ofisi-ing \quad e \\
LocPartmo \quad NPre9\text{-office-LocSuf DemPro9} \\
in this office
\end{array}
\]

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.

(5-3) CLNP

\[
\begin{array}{c}
\text{POSSPARTP} \\
dibuka tsa bana \\
books of children \\
di-buka \quad tsa \quad ba-ana \\
NPre10\text{-book PossPart10 NPre2-child} \\
the children's books
\end{array}
\]
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).

(5-4)

![Diagram of CLNP structure](image)

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.

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

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) LOCCLNP

\[
\text{LocPart}_{fa} \text{ godimo } ga \text{ tafole}
\]

at top/above of table

\[
fa \text{ go-dimo } ga \text{ (ne)-tafole}
\]

LocPart_{fa} NPre17-above PossPart17 NPre9-table

on top of the table/on the table

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.

<table>
<thead>
<tr>
<th>Head</th>
<th>Modifier</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolute pronoun</td>
<td>demonstrative pronoun</td>
<td>The class of the modifying pronoun agrees with the class of the absolute pronoun.</td>
</tr>
<tr>
<td></td>
<td>inclusive quantitative pronoun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>selective quantitative pronoun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>separative quantitative pronoun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>POSSPARTP</td>
<td>The class of the possessive particle agrees with the class of the absolute pronoun.</td>
</tr>
<tr>
<td></td>
<td>QUALPARTP</td>
<td>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.</td>
</tr>
</tbody>
</table>

Table 5-3: PROP structure and agreement with absolute pronoun as head
<table>
<thead>
<tr>
<th>Head</th>
<th>Modifier</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative pronoun</td>
<td>inclusive quantitative pronoun</td>
<td>The class of the modifying pronouns agree with the class of the demonstrative pronoun.</td>
</tr>
<tr>
<td></td>
<td>selective quantitative pronoun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>separative quantitative pronoun</td>
<td></td>
</tr>
<tr>
<td>POSSPARTP</td>
<td></td>
<td>The class of the possessive particle agrees with the class of the demonstrative pronoun.</td>
</tr>
<tr>
<td>QUALPARTP</td>
<td></td>
<td>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.</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Head</th>
<th>Modifier</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>inclusive quantitative pronoun</td>
<td>POSSPARTP</td>
<td>The class of the possessive particle agrees with the class of the inclusive quantitative pronoun.</td>
</tr>
<tr>
<td></td>
<td>POSSPARTP</td>
<td>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.</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Head</th>
<th>Modifier</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>exclusive quantitative pronoun</td>
<td>POSSPARTP</td>
<td>The class of the possessive particle agrees with the class of the exclusive quantitative pronoun.</td>
</tr>
<tr>
<td></td>
<td>QUALPARTP</td>
<td>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.</td>
</tr>
</tbody>
</table>

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 tsoné (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


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.
There is no agreement between the possessive particle and its complement.

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.

(5-8) POSSPARTP

\[\text{madi\ a\ batho}\]

money of people

\[\text{ma-di\ a\ ba-tho}\]

NPre6-money PossPart6 NPre2-person

the people's money

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.

<table>
<thead>
<tr>
<th>Head</th>
<th>Complement</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualitative particle</td>
<td>adjective</td>
<td>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.</td>
</tr>
<tr>
<td>relative verb</td>
<td></td>
<td>The class of the complement agrees with the class of the qualificative particle.</td>
</tr>
<tr>
<td>identifying copulative relative verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>describing copulative relative verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>associative copulative relative verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADVP</td>
<td></td>
<td>There is no agreement between the complement and the qualificative particle.</td>
</tr>
</tbody>
</table>

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.

(5-10) QUALPARTP

setulo se sentle
chair that is nice/beautiful

se-dul-o se sentle
NPre7-sit-DevSuf QualPart7 NPre7-adjective
a nice chair
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.

<table>
<thead>
<tr>
<th>Head</th>
<th>Complement</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>instrumental particle</td>
<td>Interrogative eng?</td>
<td>There is no agreement between the complement and the instrumental particle.</td>
</tr>
<tr>
<td></td>
<td>ADVF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CLNP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PROF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>POSSPARTP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>QUALPARTP</td>
<td></td>
</tr>
</tbody>
</table>

Table 5-9: INSTRPARTP structure and agreement

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

(5-11) INSTRPARTP

```
  ka koloi
  with 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 LOCPARTPgo consists of the locative particle go (with someone personally, to) as head, followed by a PROF 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.
LOCPARTP\textsubscript{go} consists of the locative particle \textit{ga} (at someone's place, at) as head, followed by a \textsc{prop} or a \textsc{clnp} as complement belonging to class 1a or 2a. In (5-13), the locative particle \textit{ga} (at someone's place) is followed by a noun in class 1a.

\begin{verbatim}
(5-12) LOCPARTP\textsubscript{go} \hline
go mme to mother
\end{verbatim}

\begin{verbatim}
(5-13) LOCPARTP\textsubscript{ga} \hline
\textit{ga} malome to uncle
\end{verbatim}

The \textit{LOCPARTP\textsubscript{ka}} consists of the locative particle \textit{ka} (goes through, over, out of, into) as head, followed by a \textsc{clnp}, a \textit{LOCPARTfa}, \textit{LOCPARTPmo} or \textit{LOCPARTPkwa} as complement. In (5-14) the locative particle \textit{ka} (into) is followed by a noun.

\begin{verbatim}
(5-14) LOCPARTP\textsubscript{ka} \hline
\textit{ka} motse into village
\end{verbatim}

The head in a \textit{LOCPARTPmo}, is \textit{mo}, in the \textit{LOCPARTPfa} \textit{fa} and in the \textit{LOCPARTPkwa} \textit{kwa}, and these heads can be followed by a \textsc{locnp}, a \textsc{locclnp}, a \textit{LOCPARTPgo} or a \textit{LOCPARTP\textsubscript{ga}} as complement. For example, in (5-15), the locative particle \textit{mo} (in) is followed by \textit{lebenkeleng} (shop "location"), a locative noun.

\begin{verbatim}
(5-15) LOCPARTP\textsubscript{mo} \hline
\textit{mo} lebenkeleng into the shop
\end{verbatim}
5.5.5 TEMPORAL PARTICLE PHRASE

The TEMPPARTP consists of the temporal particle ka (on, by) followed by a CLNP as the complement (Cole, 1955:358). There is no class agreement between the complement and the temporal particle. For example, in (5-16), the CLNP letsatsi leo (that day) is the complement of the temporal particle ka (on) as head.

(5-16) TEMPPARTP

\[
\begin{array}{c}
\text{ka letsatsi leo} \\
\text{on that day} \\
\text{ka le-tsatsi leo} \\
\text{TempPart NPre3-day DemPro3D2} \\
\text{on that day}
\end{array}
\]

5.5.6 AGENTATIVE PARTICLE PHRASE

The AGPARTP consists of the agentative particle ke (by) as head, followed by a CLNP, PROP, ADVP or QUALPARTP as the complement. There is no class agreement between the complement and the preceding agentative particle. The AGPARTP is typically included in passive sentences and in (5-17), it consists of the agentative particle ke, followed by a noun.

(5-17) AGPARTP

\[
\begin{array}{c}
\text{Mosimane o romilwe ke mosadi} \\
\text{boy he was sent by woman} \\
\text{mo-simane o-rom-il-iw-e ke mo-sadi} \\
\text{NPre1-boy AgrSubj1-send-PerfSuf-PassSuf-VEnd AgPart NPre1-woman} \\
\text{The boy was sent by the woman.}
\end{array}
\]
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_{ko}, LOCPARTP_{la} or LOCPARTP_{ka} 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).

(5-18) ASSPARTP

\[ Ke \, ya \, le \, mme \]
\[ SubjAgrPlsg-go \, AssPart \, NPre1a-mother \]
I go with mother

\[ ke-ya \quad le \quad mme \]
I am going with mother

5.6 ADVERB PHRASE

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

(5-19) ADVP

\[ Mosimane \, o \, dira \, theta \]
\[ NPrel-boy \, AgrSubj1-work-VEnd \, Adv \]
The boy works hard.

5.7 INTERJECTION AND IDEOPHONE PHRASES

Interjection phrases (INTERJP) (5-20) and the ideophone phrases (IDEOP) in a simple sentences include only the interjection or ideophone (5-21) (cf. §4.10 and §4.11).
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, VP IDCOP for the identifying copulative verb, VP DESCOP for the describing copulative verb, VP ASSCOP 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 (VPMAIN) consists of a main verb of which the (complex) morphological structure is presented in §4.6.1. The VPMAIN 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 VPMAIN 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).
5.8.2 VERB PHRASE WITH AN IDENTIFYING COPULATIVE VERB

The minimal identifying copulative verb phrase (VPIDCOP) 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)  

\[ \text{VPIDCOP} \]

\[ \text{Ke baatletiki.} \]

is athletes

\[ ke \quad \text{ba-atletiki} \]

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
- 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 LOCCLNP mo setsidifatsing (in the refrigerator).

(5-24)  

\[ \text{VPDESCOP} \]

\[ \text{a mo setsidifatsing} \]

They are talking.
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

\[
\begin{array}{c}
\text{O na le koloi.} \\
\text{he is with car}
\end{array}
\]

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

\[
\begin{array}{c}
\text{Ga a na koloi.} \\
\text{not he has car}
\end{array}
\]

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, VIDCOP, VPDESCOP, VPASSCOP or
another VPAUX. The auxiliary verb **one** (she was) in (5-27) is followed by the main verb **areka** (she buys).

(5-27)  

```
                     VPAUX
                      ↑
  O ne  a reka.
   she was she buy
    o-ne    a-rek-a
AgrSubj1-Aux    AgrSubj1-buy-VEnd
  She was buying.
```

### 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**.

(5-28)  

```
                     CLNP   CLNP
                      ↑   ↑
  monna  le  mosadi
   man and woman
  mo-nna  le  mo-sadi
NPre1-man   Conj   NPre1-woman
  the man and the woman
```

Example (5-29) is a coordinate phrase that is structured with two **PROPs**, i.e. the personal pronouns **nna** (I) and **wena** (you).
In (5-30), the coordinate phrase is structured with the POSSPARTP *ya me* (of me) and the POSSPARTP *ya gagwe* (of him).

\[
\begin{array}{c}
\text{(5-30) POSSPARTP POSSPARTP} \\
\uparrow \quad \uparrow \\
\text{ya me} \quad \text{le} \quad \text{ya gagwe}
\end{array}
\]

of me and of his

\[
\begin{array}{c}
\text{ya me} \quad \text{le} \quad \text{ya gagwe}
\end{array}
\]

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).

\[
\begin{array}{c}
\text{(5-31) QUALPARTP QUALPARTP} \\
\uparrow \quad \uparrow \\
\text{yo monnye} \quad \text{le} \quad \text{yo mogolo}
\end{array}
\]

that is small and that is big

\[
\begin{array}{c}
\text{yo monnye} \quad \text{le} \quad \text{yo mogolo}
\end{array}
\]

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.
5.10 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):
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 theta (that is strong) modify the head.

\[(5-37)\]

\[
\text{Monna yo molele yo o leng theta}
\]

\[
\text{mo-nna yo mo-leele yo o-le-ng theta}
\]

\[
\text{the tall strong man}
\]

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 \(\text{PROP} > \text{POSSPARTP} > \text{QUALPARTP}\) also applies here. The CLNP\(_1\) in (5-38) consists of the head dibuka (books) and its modifiers QUALPARTP, PROP and POSTPARTP. In turn, the POSSPARTP of CLNP\(_1\) consists of the possessive particle tsa (of) as head and the CLNP\(_2\) 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).
5.11 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.
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

\[ ba-tho \quad ba-a-bu-a \]
NPre2-person AgrSubj2-PresPre-talk-VEnd

The people are talking.

\[ (6-2) \]
\[ \text{c-structure} \quad f-structure \]

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) \textit{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 OBJθ, OBJ, OBLθ 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 OBLloc (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
\[
\text{ba-tho} \quad \text{ba-bu-a} \quad \text{se-esimane}
\]
NPre2-person AgrSubj2-speak-VEnd NPre7-English
The people speak English.

(6-4) c-structure
\[
\begin{array}{c}
\text{S} \\
\quad \text{NP} \quad \text{VP} \\
\quad \quad \text{CLNP} \quad \text{VPMAIN} \\
\quad \quad \quad \text{N} \quad \text{V} \quad \text{NP} \\
\text{batho} \quad \text{ba} \quad \text{bua} \\
\end{array}
\]

f-structure
\[
\begin{array}{c}
PRED \quad \text{‘BU <SUBJ OBJ>’} \\
MOOD \quad \text{ind} \\
TENSE \quad \text{pres} \\
POL \quad + \\
CLASS \quad 2 \\
VTYPE \quad \text{main} \\
\text{SUBJ} \quad \text{PRED \ ‘THO’} \\
\quad \text{CLASS} \quad 2 \\
\quad \text{NTYPE} \quad \text{ord} \\
OBJ \quad \text{PRED \ ‘ESIMANE’} \\
\quad \text{CLASS} \quad 7 \\
\quad \text{NTYPE} \quad \text{ord}
\end{array}
\]
**VPMAIN: a verb (V) that subcategorises for an OBLloc**

A LOCPARTPkwa, LOCPARTPgo, LOCPARTPa, LOCPARTPka, LOCPARTPfa, LOCPARTPme or LOCPARTPkwa can follow the main verb in the VPMAIN. As in the previous case, this valency for an OBLloc GF\(^ {74} \) 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 OBLloc GF (not an OBJ), i.e. the LOCPARTPkwa kwa toropong (to town) in (6-6).

(6-5) Batho ba ya kwa toropong.
people they go to town locality

\[
\text{PRED} \quad \text{SUBJ} \quad \text{OBLloc} \\
\text{MOOD} \quad \text{ind} \\
\text{TENSE} \quad \text{pres} \\
\text{POL} \quad + \\
\text{CLASS} \quad 2 \\
\text{VTYPE} \quad \text{main} \\
\text{SUBJ} \quad \text{PRED} \quad \text{THO} \\
\text{CLASS} \quad 2 \\
\text{NTYPE} \quad \text{ord} \\
\text{OBLloc} \quad \text{PRED} \quad \text{KWA} \quad \text{OBJ} \\
\text{PARTTYPE} \quad \text{LOCPARTPkwa} \\
\text{OBJ} \quad \text{PRED} \quad \text{TOROPO} \\
\text{CLASS} \quad 9 \\
\text{NTYPE} \quad \text{locn} \\
\]

(6-6) Batho ba ya kwa toropong.
people they go to town locality

\[
\text{NPre2-person} \quad \text{AgrSubj2-go-VEnd} \quad \text{LocPartkwa} \quad \text{NPre9-town-LocSuf}
\]

The people go to town.

**VPMAIN: a verb (V) that subcategorises for an OBLass**

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 OBLass. For example, in (6-7), the ASSPARTP le wena (with you) is the OBLass (6-8).

---

\(^ {74} \) The OBL\( \_\theta \) 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)
Mosimane o dumalana le wena.

The boy agrees with you.

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

\[
\begin{align*}
&\text{S} \\
&\text{NP} \\
&\text{VP} \\
&\text{CLNP} \\
&\text{VPMAIN} \\
&\text{N} \\
&\text{V} \\
&\text{ASSPARTP} \\
&\text{mosimane} \\
&\text{o dumalana} \\
&\text{ASSPART} \\
&\text{le} \\
&\text{PROP} \\
&\text{PersPro} \\
&\text{wen}a
\end{align*}
\]

(6-9) **f-structure**

\[
\begin{align*}
&\text{PRED} '\text{DUMALAN <SUBJ OBLass>}' \\
&\text{MOOD} \ ind \\
&\text{TENSE} \ pres \\
&\text{POL} \ + \\
&\text{CLASS} \ 1 \\
&\text{VTYPE} \ main \\
&\text{SUBJ} \ [\text{PRED} '\text{SIMANE}' \\
&\text{CLASS} \ 1 \\
&\text{NTYPE} \ ord] \\
&\text{OBLass} \ [\text{PRED} '\text{kwa <OBJ>}' \\
&\text{PARTTYPE} \ ASSPART \\
&\text{OBJ} \ [\text{PRED} '\text{WENA}' \\
&\text{PERSON} \ 2 \\
&\text{NUMBER} \ sg]
\end{align*}
\]

**VPMAIN: a verb (V) that subcategorises for an OBJ\textsubscript{θ} and OBJ**

Two NPs that perform the OBJ\textsubscript{θ} and OBJ GFs, respectively, can follow the main verb in the VPMAIN. This valency for OBJ\textsubscript{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\textsubscript{θ} GF function and dijo (food), the OBJ GF (6-10).

(6-9) Basadi ba rekela bana dijo.

The women buy the children food.

The women buy the children food.
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 OBJθ GF and lenao (leg) the OBJ GF.

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

They accident got the boy’s leg injured.
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_ga, 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 *tsenya-* (put) in (6-13), the CLNP *buka* (book) and the LOCCLNP *mo kgetsin* (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)-kgetsi-ing

LocPart_mo NPre9-bag-LocSuf

The woman puts the book in the bag.
(6-14) **c-structure**

```
S
  | NP
  | VP
  | CLNP
  | VPMAIN
  | N  V  NP  LOCPART\_no
  | mosadi  o tsenya
  | CLNP  LOCPART\_no  LOCNP
  | mo    LOCN
  | buka  kgetsing
```

(6-15) **f-structure**

```
PRED  ‘TSENY <SUBJ OBJ, OBJ>’
MOOD   ind
TENSE  pres
POL    +
CLASS  1
VTYPE  main

SUBJ | FRED  ‘SADI’
CLASS 1
  | NTYPE  ord
OBJ  | FRED  ‘BUKA’
CLASS 9
  | NTYPE  ord
OBLloc | FRED  ‘MO <OBJ>’
OBJ  | FRED  ‘KGETSI’
CLASS 9
  | NTYPE  ord
```

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

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  AgrSubjl-drink-VEnd  NPre9-coffee  AssPart
wena
PersProP2sg
```

The man will drink coffee with you.
(6-16) *c*-structure

```
  S
   NP   VP
      CLNP   VPMAIN
          N   V   NF   ASSPART
                monna   o tla nwa
                N   V   NP   ASSPART
                      CLNP   le
                      N   kofi
                      N   wena
```

*f*-structure

```
PRED 'NW <SUBJ OBJ OBLass>'
MOOD   ind
TENSE  pres
POL    +
CLASS  1
VTYPE  main
SUBJ   [PRED 'NNA'
          CLASS 1
          NTYPE ord]
OBJ    [PRED 'KOFI'
          CLASS 9
          NTYPE ord]
OBLass [PRED 'LE <OBJ>'
          OBJ [PRED 'WENA'
              PERSON 2
              NUMBER sg]]
```

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

Two NPs that perform the OBJ, OBJ GFs, respectively, and a phrase indicating locality (OBL_{loc}) can follow the main verb in the VPMAIN. This valency for OBJ, 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$_6$ 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
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, $VP_{IDCOP}$, $VP_{DESCOP}$ or $VP_{ASSCOP}$, 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".

---

**f-structure**

```
PRED 'TSENYETS <SUBJ OBJ OBJ OBLloc>'
MOOD ind
TENSE pres
POL +
CLASS 1
VTYPE main
SUBJ
PRED 'RRE'
CLASS la
NTYPE ord
OBJ
PRED 'MME'
CLASS la
NTYPE ord
OBJ
PRED 'DI'
CLASS 6
NTYPE ord
OBLloc
PRED 'MO <OBJ>'
OBJ
PRED 'BENKELE'
CLASS 5
NTYPE locn
```
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 copulatives 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).

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)  \[
\begin{array}{l}
\text{PRED} \ 'O <\text{SUBJ PREDLINK}>' \\
\text{SUBJ} \\
\quad \text{PRED} \ 'SADI' \\
\quad \text{CLASS} 1 \\
\quad \text{NTYPE} \text{ ord} \\
\quad \text{PREDLINK} \\
\quad \text{PRED} \ 'NTLE' \\
\quad \text{CLASS} 1 \\
\quad \text{NTYPE} \text{ adjective}
\end{array}
\]

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)  \[
\begin{array}{l}
\text{AUX} + \\
\text{TENSE} \text{ RelPast}
\end{array}
\]

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 therefore 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 dithako 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

```
S
  |       |
  NP   VP
   |
CLNP  VPMAIN
   |       |
  N   V   NP
   |
mosadi  o duela
   |
N
   |
dithako  ka
   |
CLNP
   |
kara
```
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.
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 **TEMPARTP** or a **LOCPARTPmo** indicate the time as well as the duration and frequency of an action (Le Roux, 2007:93). In (6-30), a **LOCPARTPmo** *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.

* The children run in morning.

**Phrases indicating locality**

A **LOCCLNP**, **ADVP**, **POSSPARTP**, **LOCPARTPg0**, **LOCPARTPgA**, **LOCPARTPka**, **LOCPARTPfa**, **LOCPARTPmo**, or a **LOCPARTPkwa** 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 **LOCPARTPmo** *mo bankeng* (in the bank) indicates the locality of the action.

(6-31) Ba dira mo bankeng.

* They work in the bank.

**Phrases indicating manner**

Some phrases, usually an **ADVP**, an **INSTRPARTP** or a **COMPPARTP**\(^76\), 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".
(6-32) Ba dira thata.
they work hard
\[
\begin{align*}
\text{ba-dir-a} & \quad \text{thata} \\
\text{AgrSubj2-work-VEnd} & \quad \text{Adv}
\end{align*}
\]
They work hard.

Phrase indicating instrument

An **INSTRPARTP** (cf. §5.5.3) expresses the instrument by which an action can be accomplished (Le Roux, 2007:96). For example, in (6-33) the **INSTRPARTP** is **ka karata** (with the card).

(6-33) Ene\textsuperscript{77} o duela ka karata.
she she pay with card
\[
\begin{align*}
en & \quad o-duel-a & \quad ka & \quad (ne)-karata \\
\text{AbsPro1 AgrSubj1-pay-VEnd InstrPart NPre9-card}
\end{align*}
\]
She pays with a card.

Phrases indicating association

In a sentence, specific phrases that indicate *with whom*, *by whom* or *similar to (like)* can be included. The **ASSPARTP** indicates *with whom* the action of the verb is executed, the **COMPPARTP** indicates a similarity and the **AGPARTP** is included in passive sentences to indicate *by whom*. We refer to these types of phrases as the phrases that indicate a certain type of association. In (6-34), the **ASSPARTP** le nna (with me) indicates *with whom* the action is carried out (Le Roux, 2007:97, 122). The associative particle le in this phrase establishes an associative link between two or more individuals or things (Louwrens, 1994a:16).

(6-34) Ba tla tsamaya le nna.
they will walk with me
\[
\begin{align*}
ba-tla-tsamay-a & \quad le & \quad nna \\
\text{AgrSubj2-FutPre-walk-VEnd AssPart PersProPlsg}
\end{align*}
\]
They will go with me.

6.3 WORD ORDER IN THE SIMPLE SENTENCE

The SVO and SVOX word orders are characteristic of most Bantu languages (Watters, 2000:196–201). Setswana is typically classified as an SVO language (Cole, 1955: 426, Krüger, 2006:11–12) or if an **ADJUNCT** is included in a sentence, the order is SVOX\textsuperscript{78} where X indicates the

\textsuperscript{77} This absolute pronoun is translated as either he or she as pronouns and agreement morphemes do not indicate gender.

\textsuperscript{78} For purposes of emphasis, the order can be altered. Zerbian (2006:361) discusses inversion when the **SUBJ** appears post verbally in Sesotho sa Leboa. The **OBJ** may also be placed at the beginning of a sentence (Cole, 1955:428). **ADJUNCTs** can also be placed
In this section, we present a summary of the arguments that can be included in a simple sentence \((S \rightarrow \text{NP} \text{ VP})\) and in what order they occur. An NP that maps to the \text{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 \text{VPMAIN}:

- main verb, OBJ
- main verb, OBJ, OBL\text{loc}
- main verb, OBJ, OBL\text{ass}
- main verb, OBJ\text{a}, OBJ
- main verb, OBJ\text{a}, OBJ, OBL\text{loc}
- main verb, OBL\text{loc}
- main verb, OBL\text{ass}

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

The basic position of phrases (\text{ADJUNCTs}) indicating locality, time, manner, instrument and association, and the \text{IDEOP} is usually at the end of a sentence. Although these \text{ADJUNCTs} are not arguments, they form part of the structure of the different verb phrases. The order in which \text{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 \text{ADJUNCT} order is assumed. The \text{INTERJP} that also functions as an \text{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 \text{SUBJ}s.

In the \text{VPMAIN}, the verb agrees with the \text{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).

---

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

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).
The water is warm.

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).

The children have bicycles.

In the **VPAUX**, 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.

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.

(6-43)  **Mosadi le mosetsana ba reka dijo.**
woman and girl they buy food

- **mo-sadi**  NPre1-woman
- **le**  AssPart
- **mo-setsana**  NPre1-girl
- **ba-rek-a**  AgrSubj2-buy-VEnd
- **di-j-o**  NPre9-eat-DevSuf

The woman and girl are buying food.

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).

(6-44)  **Mosadi le ngaka ba nwa kofi.**
woman and docter they drink coffee

- **mo-sadi**  NPre1-woman
- **le**  AssPart
- **(ne)-ngaka**  NPre9-docter
- **ba-nw-a**  AgrSubj2-drink-VEnd

**>(ne)-kofi**  NPr9-coffee

The woman and the docter are drinking coffee.

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.
6.5 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 ‘PR0’ 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.
(6-50) 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)  
\[
\text{c-structure} \quad \text{f-structure}
\]

\[
\begin{array}{c}
S \\
\text{NP} \quad \text{VPMAIN} \\
\text{CLNP} \quad \text{V} \\
\text{N} \\
\text{basadi}
\end{array}
\begin{array}{c}
\text{FRED} \quad '\text{RUT} <\text{SUBJ OBJ}>' \\
\text{MOOD} \quad \text{ind} \\
\text{TENSE} \quad \text{pres} \\
\text{POL} \quad + \\
\text{CLASS} \quad 2 \\
\text{SUBJ} \quad \text{FRED} \quad '\text{SADI}' \\
\text{CLASS} \quad 2 \\
\text{OBJ} \quad \text{FRED} \quad '\text{PRO}' \\
\text{CLASS} \quad 2
\end{array}
\]

Sentences in Setswana may include verbs that subcategorise for double objects, i.e. the OBJ\(_0\) 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\(_0\) 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\(_0\) 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
   NPrel1-woman AgrSubj1-buy-ApplSuf-VEnd       NPrel2-child NPrel0-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
   NPrel1-woman AgrSubj1-PresPre-AgrObj2-AgrObj8-buy-ApplSuf-VEnd
   The woman buys it for them.
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.
<table>
<thead>
<tr>
<th>ADJUNCT</th>
<th>SUBJECT</th>
<th>VERB</th>
<th>COMPLEMENT</th>
<th>OBJECT (INDIRECT &amp; DIRECT)</th>
<th>OBBIQUE</th>
<th>ADJUNCTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>INTERJP</td>
<td>Main verb</td>
<td>X</td>
<td>CLNP PROP POSSPARTP QUALPARTP</td>
<td>OBL_{v_0}</td>
<td>LOCCLNP, LOCPARTPga, LOCPARTPka, LOCPARTPfa, LOCPARTPmo, LOCPARTPkwa, OBL_{v_0} ASSPARTP</td>
</tr>
<tr>
<td></td>
<td>INTERJP</td>
<td>Auxiliary verb</td>
<td>VPAUX</td>
<td>CLNP PROP POSSPARTP QUALPARTP</td>
<td>OBL_{v_0}</td>
<td>LOCCLNP, LOCPARTPga, LOCPARTPka, LOCPARTPfa, LOCPARTPmo, LOCPARTPkwa, OBL_{v_0} ASSPARTP</td>
</tr>
<tr>
<td></td>
<td>INTERJP</td>
<td>Identifying copulative verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INTERJP</td>
<td>Describing copulative verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INTERJP</td>
<td>Associative copulative verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 XLE\(^\text{80}\) 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 (SETSWA.NA), 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 ε 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 used.

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).
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.
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 boundary. 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

NPrel-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)](image)

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 (^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 (^SUBJ), (^OBJ), etc. The attribute names in these expressions denote governable GFs such as SUBJ, OBJ, OBJ-TH, OBLloc, OBLass and PREDLINK\(^\text{84}\). 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, INTQPRO, all the subcategories of the particles (viz. POSSPART, QUALPART, INSTRPART, TEMPPART, LOCPARTga, LOCPARTka, LOCPARTfa, LOCPARTka, LOCPARTmo, 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)= +;
   IDEOGRAPH 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) \[ \text{batl} V \text{XLE} \quad \{(\text{^PRED})='\text{BATL}(\text{^SUBJ})>' \]
\[ \quad \mid (\text{^PRED})='\text{BATL}(\text{^SUBJ} \ (\text{^OBJ})>') \} \quad (\text{^VTYPE})= \text{main}. \]

(7-9) \[ \text{batlis} V \text{XLE} \quad (\text{^PRED})='\text{BATLIS}(\text{^SUBJ} \ (\text{^OBJ-TH}) \ (\text{^OBJ})>') \]
\[ \quad (\text{^VTYPE})= \text{main}. \]

(7-10) \[ \text{batlisan} V \text{XLE} \quad (\text{^PRED})='\text{BATLISAN}(\text{^SUBJ} \ (\text{^OBJ})>') \]
\[ \quad (\text{^VTYPE})= \text{main}. \]

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) \lel V XLE\ (^{PRED}=\lel<^{SUBJ}>)\ (^{VTYPE}=\text{main})

Figure 7-4 displays the c- and f-structure for sentence (7-12) with the verb root batlisa- (helps seek).

(7-12) \text{Rre o batlisa mme dinotlolo.}
father he seek let mother keys
\text{(-)-rre o-batl-is-a (-)-mme}
\text{NPre1a-father AgrSubj1a-seek-CausSuf-VEnd NPre1a-mother}
di-notlolo
\text{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 \textit{et al.}, 1999:63). For example, in (7-13) and (7-14), the lexical entries for the auxiliary verbs \textit{ne} and \textit{tlabo} include f-descriptions that specify the relevant respective tenses.
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 rules. 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 or the empty string e and the second constituent is a VP.

The NP functions as the SUBJ of the sentence \((^\text{SUBJ}) = !\)\(^7\). The agreement of the verb with the SUBJ is modelled by means of the constraining equality \((^\text{SUBJ AGR}) =c (^\text{AGR})\). 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 \(\mid\).

\(^{87}\) 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) \[ S \rightarrow \{ \text{INTERJP}*: !$ \} (\text{^ADJUNCT}) \]
\[ \text{NP}: (\text{^SUBJ})=! (\text{^SUBJ AGR}) =c (\text{^AGRV}) \]
\[ | \ e: (\text{^SUBJ PRED})='PRO' (\text{^SUBJ PRONTYPE})=null; \}
\[ \text{VP}: ^=! . \]

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) \[ \text{NP} \rightarrow \{ \text{CLNP} \]
\[ | \text{PROP} \]
\[ | \text{QUALPARTP} \]
\[ | \text{POSSPARTP} \} . \]

(7-22) \[ \text{VP} \rightarrow \{ \text{VPMAIN} \]
\[ | \text{VPAUX} \]
\[ | \text{VFIDCOP} \]
\[ | \text{VPDESCOP} \]
\[ | \text{VPASSCOP} \} . \]

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, \( ~(\text{PRONTYPE}) = \text{PersPro} \) prevents a personal pronoun to function as a modifier. PersPro is a value of the attribute PRONTYPE, as specified in §7.4.

(7-23) \( \text{CLNP} \rightarrow \text{N} \{ \text{PROP: } ! \$ (\text{^ADJUNCT}) (\text{^AGR}) =c (\text{^AGR}) \sim (\text{PRONTYPE}) = \text{PersPro} \) \\
| \( \text{POSSPARTP: } ! \$ (\text{^ADJUNCT}) (\text{^AGR}) =c (\text{^AGR}) \\
| \ldots \} \).

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) \( \text{PROP} \rightarrow \{ \text{PERSPRO (INCQPRO: } ! \$ (\text{^ADJUNCT}) (\text{^AGR}) =c (\text{^AGR}) \) \\
| \text{ABSPRO (DEMPRO: } ! \$ (\text{^ADJUNCT}) (\text{^AGR}) =c (\text{^AGR}) \\
| \ldots \} \).

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) \( \text{POSSPARTP} \rightarrow \text{POSSPART} \{ \text{PROP: } (\text{^OBJ})=! -(\text{^OBJ PRONTYPE}) = \text{AbsPro} \\
| -(\text{^OBJ PRONTYPE}) = \text{PersPro} \\
| \text{POSSPRO: } (\text{^OBJ})=! \\
| \text{ADVP: } (\text{^OBJ})=! \\
| \text{CLNP: } (\text{^OBJ})=! \\
| \text{QUALPARTP: } (\text{^OBJ})=! \\
| \text{POSSPARTP: } (\text{^OBJ})=! (\text{^OBJ AGR}) =16 \} \).

7.8.2.4 QUALPARTP

The QUALPARTP rule (7-26) ensures that a qualifying 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 \((^\text{AGR}) =c (!\text{AGR})\) as before.

\[(7-26)\quad \text{QUALPARTP} \rightarrow \text{QUALPART} \mid \text{ADJECTIVE: } !$ (^OBJ) (^AGR) =c (!\text{AGR}) \mid \text{RELV: } !$ (^OBJ) (^AGR) =c (!\text{AGR}) \mid \text{ADV: } (^\text{OBJ})=! \mid \text{SEPQPRO: } !$ (^OBJ) (^AGR) =c (!\text{AGR})].\]

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 \(~(^\text{PRESPRE})= +\) 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 \((^\text{INFOBJ PRED})='\text{PRO}' (^\text{INFOBJ PRONTYPE})=\text{null}\) 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 \((^\text{INFOBJ-TH PRED})='\text{PRO}' (^\text{INFOBJ-TH PRONTYPE})=\text{null}\) and \((^\text{INFOBJ PRED})='\text{PRO}' (^\text{INFOBJ PRONTYPE})=\text{null}\) 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 \([\text{ADVP*: } !$ (^\text{ADJUNCT}), \text{TEMPPARTP*: } !$ ('^\text{ADJUNCT}), \ldots ]^*\) 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).
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
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) \[ \text{VPIDCOP} \rightarrow \text{IDCOPV NP: (}^\text{PREDLINK})=! \]
\[ \text{[ADVP*: ! $ (}^\text{ADJUNCT}), \]
\[ \text{TEMPPARTP*: ! $ (}^\text{ADJUNCT}), \]
\[ \ldots \}^* . \]

The \text{VPDESCOP} rule (7-30) requires the describing copulative verb \text{DESCOPV} (cf. §7.6.1) to be followed by an obligatory complement, i.e. \text{ADJECTIVE} and \text{LOCCLNP}, as listed in §5.8.3.

(7-30) \[ \text{VPDESCOP} \rightarrow \text{DESCOPV} \{ \text{ADJECTIVE: (}^\text{PREDLINK})=! \]
\[ \mid \text{LOCCLNP: (}^\text{PREDLINK})=! \]
\[ \ldots \}^* \]
\[ \text{[ADVP*: ! $ (}^\text{ADJUNCT}), \]
\[ \text{TEMPPARTP*: ! $ (}^\text{ADJUNCT}), \]
\[ \ldots \}^* . \]

The \text{VPASSCOP} rule (7-31) requires an associative copulative verb \text{ASSCOPV} (cf. §7.6.1) to be followed by an obligatory complement, i.e. an \text{ASSPARTP} in the positive and an \text{NP} or \text{ADVP} in the negative (§5.5.7).

(7-31) \[ \text{VPASSCOP} \rightarrow \text{ASSCOPV} \{ \{ \text{ASSPARTP: (}^\text{PREDLINK})=! (}^\text{POL})=pos \]
\[ \mid \{ \text{NP: (}^\text{PREDLINK})=! (}^\text{POL})=neg \]
\[ \mid \text{ADVP: (}^\text{PREDLINK})=! (}^\text{POL})=neg \} \}
\[ \text{[ADVP*: ! $ (}^\text{ADJUNCT}), \]
\[ \text{TEMPPARTP*: ! $ (}^\text{ADJUNCT}), \]
\[ \ldots \}^* . \]

As in the case of \text{VPMAIN}, an arbitrary number of \text{ADJUNCT}s in free order can follow the respective complements of the copulative verbs.

\textbf{7.8.3.4 Phrases functioning as obliques and adjuncts}

We reiterate that a main verb may subcategorise for an \text{OBL}, as shown in (7-27). In this section, the focus is on those phrases that perform this function, viz. \text{LOCNP}, \text{LOCCLNP}, \text{LOCPARTP}_g, \text{LOCPARTP}_ga, \text{LOCPARTP}_ka, \text{LOCPARTP}_fa, \text{LOCPARTP}_no, \text{LOCPARTP}_{kwa}, and \text{ASSPARTP}. Moreover, the \text{VPMAIN (7-27), VPIDCOP (7-29), VPDESCOP (7-30) and VPASSCOP (7-31)} rules also allow for phrases that function as \text{ADJUNCT}s, viz. \text{LOCNP}, \text{LOCCLNP}, \text{LOCPARTP}_g, \text{LOCPARTP}_ga, \text{LOCPARTP}_ka, \text{LOCPARTP}_fa, \text{LOCPARTP}_no, \text{LOCPARTP}_{kwa}, \text{ASSPARTP}, \text{ADVP}, \text{TEMPPARTP}, \text{INSTRPARTP}, \text{COMPPARTP}, \text{INTERJP}, \text{IDEOP}, \text{AGPARTP}, \text{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.
IDEO --> 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)](image)

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, agentative 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.

<table>
<thead>
<tr>
<th>Linguistic characteristics</th>
<th>Grammatical test items</th>
<th>Ungrammatical test items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Subject-verb agreement (noun classes and person and number)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mood (indicative, consecutive, imperative and habitual)</td>
<td>154</td>
<td>154</td>
</tr>
<tr>
<td>tense (present, past, future, relative past and relative future)</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>aspect (perfective, progressive and persistive)</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>polarity (positive and negative structures)</td>
<td>163</td>
<td>163</td>
</tr>
<tr>
<td>main verbs and their subcategorisation frames</td>
<td>82</td>
<td>15</td>
</tr>
<tr>
<td>copulative verbs</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>auxiliary verbs</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td><strong>2. Phrase internal agreement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>noun phrases</td>
<td>115</td>
<td>64</td>
</tr>
<tr>
<td>locative noun phrase</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td><strong>3. Phrasal rule (c-structure) coverage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>particle phrases</td>
<td>97</td>
<td>-</td>
</tr>
<tr>
<td>locative class noun phrase</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>adverb phrase</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>agentative particle phrase</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>interjection phrase</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>ideophone phrase</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>ADJUNCTS in the sentence</td>
<td>70</td>
<td>-</td>
</tr>
<tr>
<td>pro-drop</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>function of object agreement morphemes</td>
<td>33</td>
<td>-</td>
</tr>
</tbody>
</table>

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).
<table>
<thead>
<tr>
<th>Lexical entries</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns in classes 1 to 15</td>
<td>35</td>
</tr>
<tr>
<td>Locative nouns in classes 4, 5, 9 and 10</td>
<td>5</td>
</tr>
<tr>
<td>Locative class nouns in classes 16–20</td>
<td>5</td>
</tr>
<tr>
<td>Adjectives</td>
<td>6</td>
</tr>
<tr>
<td>Absolute, demonstrative and quantitative pronouns in classes 1 to 15</td>
<td>34</td>
</tr>
<tr>
<td>Personal pronouns of the first and second person (singular and plural)</td>
<td>4</td>
</tr>
<tr>
<td>Possessive pronouns</td>
<td>5</td>
</tr>
<tr>
<td>Adverbs</td>
<td>8</td>
</tr>
<tr>
<td>Interjection</td>
<td>1</td>
</tr>
<tr>
<td>Ideophone</td>
<td>1</td>
</tr>
<tr>
<td>Particles</td>
<td>41</td>
</tr>
<tr>
<td>Main verbs (basic and extended verb roots)</td>
<td>43</td>
</tr>
<tr>
<td>Auxiliary verbs</td>
<td>6</td>
</tr>
<tr>
<td>Copulative verbs</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 8-2: Number of lexical entries in the test suite

The test suite contains 828 grammatical[^88^] and 365 ungrammatical test items. A summary of the number of grammatical test items per word length[^89^] in the test suite is given in Table 8-3.

<table>
<thead>
<tr>
<th>Number of test items</th>
<th>Word length</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>145</td>
<td>3</td>
</tr>
<tr>
<td>175</td>
<td>4</td>
</tr>
<tr>
<td>173</td>
<td>5</td>
</tr>
<tr>
<td>133</td>
<td>6</td>
</tr>
<tr>
<td>43</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
</tr>
</tbody>
</table>

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](http://repo.sadilar.org/handle/20.500.12185/478)

[^89^]: Word length refers to 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 moseše (the women will not buy a dress) indicates two solutions and it took 0.002 seconds to parse 40 subtrees (8-2).

\[(8-1) \quad \text{CLNP: mosadi yo (1 0.003 11)} \]
\[(8-2) \quad \text{basadi ga ba ketla ba reka moseše (2 0.002 40)} \]

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.
<table>
<thead>
<tr>
<th>Test items</th>
<th>Valid parses</th>
</tr>
</thead>
<tbody>
<tr>
<td>518</td>
<td>1</td>
</tr>
<tr>
<td>220</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>

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 sentences\textsuperscript{90}.

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.

\textsuperscript{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 ParGram\textsuperscript{91} 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.

\textsuperscript{91} Cf. https://pargram.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

<table>
<thead>
<tr>
<th>Class</th>
<th>Class prefix</th>
<th>Semantic content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mo-</td>
<td>Exclusively personal nouns</td>
</tr>
<tr>
<td>2</td>
<td>ba-</td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>(-)-</td>
<td>Proper names, kinship terms, and personified names</td>
</tr>
<tr>
<td>2a</td>
<td>bo-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>mo-</td>
<td>Miscellaneous: impersonal, parts of the body, tools, instruments, animals, natural phenomena, most names of trees and plants</td>
</tr>
<tr>
<td>4</td>
<td>me-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>le-</td>
<td>Miscellaneous: some animals, many plants, natural phenomena, collective nouns, strangers, regions</td>
</tr>
<tr>
<td>6</td>
<td>ma-</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>se-</td>
<td>Miscellaneous: parts of the body, names of languages, customs, habits, languages, some plants, tools, instruments, some personal nouns</td>
</tr>
<tr>
<td>8</td>
<td>di-</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(ne)-</td>
<td>Miscellaneous: mostly animal names, some plants, parts of the body, natural phenomena, personal nouns indicating people with a special rank</td>
</tr>
<tr>
<td>10</td>
<td>di-</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>lo-</td>
<td>Miscellaneous: few nouns belong to this class (This class is disappearing.)</td>
</tr>
<tr>
<td>12</td>
<td>di-</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>bo-</td>
<td>Mainly abstract concepts, also collective nouns</td>
</tr>
<tr>
<td>6</td>
<td>ma-</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>go-</td>
<td>Infinitive</td>
</tr>
<tr>
<td>16</td>
<td>fa-</td>
<td>Location or place</td>
</tr>
<tr>
<td>17</td>
<td>go-</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>mo-</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ga-</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(-)-</td>
<td></td>
</tr>
</tbody>
</table>

Table 1-1: Setswana noun classes
### Table 4-1: Pronouns for noun classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Absolute pronoun&lt;sup&gt;92&lt;/sup&gt;</th>
<th>Demonstrative pronouns</th>
<th>Quantitative pronouns</th>
<th>Possessive pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distance 1&lt;sup&gt;93&lt;/sup&gt;</td>
<td>Distance 2 (proximal)</td>
<td>Inclusive (all)</td>
<td>Selective (only)</td>
</tr>
<tr>
<td></td>
<td>(distal)</td>
<td>Distance 3 (post-distal)</td>
<td>Exclusive (no)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Selective (only)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Separative (unknown)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Interrogative (which?)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ene</td>
<td>yo</td>
<td>-</td>
<td>ope</td>
</tr>
<tr>
<td>2</td>
<td>bone</td>
<td>ba</td>
<td>bao</td>
<td>bale</td>
</tr>
<tr>
<td>3</td>
<td>one</td>
<td>o</td>
<td>oo</td>
<td>ole</td>
</tr>
<tr>
<td>4</td>
<td>yone</td>
<td>e</td>
<td>eo</td>
<td>ele</td>
</tr>
<tr>
<td>5</td>
<td>lone</td>
<td>le</td>
<td>leo</td>
<td>lele</td>
</tr>
<tr>
<td>6</td>
<td>one</td>
<td>a</td>
<td>ao</td>
<td>ale</td>
</tr>
<tr>
<td>7</td>
<td>sone</td>
<td>se</td>
<td>seo</td>
<td>sele</td>
</tr>
<tr>
<td>8</td>
<td>tsone</td>
<td>tse</td>
<td>tseo</td>
<td>tsele</td>
</tr>
<tr>
<td>9</td>
<td>yone</td>
<td>e</td>
<td>eo</td>
<td>ele</td>
</tr>
<tr>
<td>10</td>
<td>tsone</td>
<td>tse</td>
<td>tseo</td>
<td>tsele</td>
</tr>
<tr>
<td>11</td>
<td>lone</td>
<td>lo</td>
<td>loo</td>
<td>lole</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>bone</td>
<td>bo</td>
<td>boo</td>
<td>bole</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>gone</td>
<td>go</td>
<td>goo</td>
<td>gole</td>
</tr>
<tr>
<td>16</td>
<td>gone</td>
<td>fa</td>
<td>fao</td>
<td>fale</td>
</tr>
<tr>
<td>17</td>
<td>gone</td>
<td>go</td>
<td>goo</td>
<td>gole</td>
</tr>
<tr>
<td>18</td>
<td>gone</td>
<td>mo</td>
<td>moo</td>
<td>mole</td>
</tr>
<tr>
<td>19</td>
<td>gone</td>
<td>-</td>
<td>kwa</td>
<td>kwale</td>
</tr>
<tr>
<td>20</td>
<td>gone</td>
<td>-</td>
<td>kwa</td>
<td>kwale</td>
</tr>
</tbody>
</table>

---

<sup>92</sup> The absolute pronouns have a variant form where the ending is -a and not -e.

<sup>93</sup> Distance 1 also has a variant form (cf. Krüger, 2006:132).
<table>
<thead>
<tr>
<th>Person</th>
<th>Personal pronouns</th>
<th>Possessive pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ordinary</td>
</tr>
<tr>
<td>First person, singular</td>
<td>nna</td>
<td>ka/me</td>
</tr>
<tr>
<td>First person, plural</td>
<td>rona</td>
<td>rona</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>wena</td>
<td>gago</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lona</td>
<td>bona</td>
</tr>
<tr>
<td>Third person, singular</td>
<td>ene</td>
<td>gagwe</td>
</tr>
<tr>
<td>Third person, plural</td>
<td>bone</td>
<td>bona/bone</td>
</tr>
</tbody>
</table>

Table 4-2: Pronouns for persons

<table>
<thead>
<tr>
<th>Class</th>
<th>Subject agreement morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o</td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
</tr>
<tr>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>e</td>
</tr>
<tr>
<td>5</td>
<td>le</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>se</td>
</tr>
<tr>
<td>8</td>
<td>di</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
</tr>
<tr>
<td>10</td>
<td>di</td>
</tr>
<tr>
<td>11</td>
<td>lo</td>
</tr>
<tr>
<td>10</td>
<td>di</td>
</tr>
<tr>
<td>14</td>
<td>bo</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>15</td>
<td>go</td>
</tr>
<tr>
<td>16</td>
<td>fa</td>
</tr>
<tr>
<td>17</td>
<td>go</td>
</tr>
<tr>
<td>18</td>
<td>go</td>
</tr>
<tr>
<td>19</td>
<td>go</td>
</tr>
<tr>
<td>20</td>
<td>go</td>
</tr>
</tbody>
</table>

Table 4-4: Subject agreement morphemes of noun classes
<table>
<thead>
<tr>
<th>Person</th>
<th>Subject agreement morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>First person, singular</td>
<td>ke</td>
</tr>
<tr>
<td>First person, plural</td>
<td>re</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>o</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lo</td>
</tr>
</tbody>
</table>

Table 4-5: Subject agreement morphemes of personal pronouns

<table>
<thead>
<tr>
<th>Class</th>
<th>Consecutive subject agreement morphemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
</tr>
<tr>
<td>3</td>
<td>wa</td>
</tr>
<tr>
<td>4</td>
<td>ya</td>
</tr>
<tr>
<td>5</td>
<td>la</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>sa</td>
</tr>
<tr>
<td>8</td>
<td>tsa</td>
</tr>
<tr>
<td>9</td>
<td>ya</td>
</tr>
<tr>
<td>10</td>
<td>tsa</td>
</tr>
<tr>
<td>11</td>
<td>la/lwa</td>
</tr>
<tr>
<td>12</td>
<td>tsa</td>
</tr>
<tr>
<td>13</td>
<td>ba</td>
</tr>
<tr>
<td>14</td>
<td>a</td>
</tr>
<tr>
<td>15</td>
<td>ga</td>
</tr>
<tr>
<td>16</td>
<td>ga</td>
</tr>
<tr>
<td>17</td>
<td>ga</td>
</tr>
<tr>
<td>18</td>
<td>ga</td>
</tr>
<tr>
<td>19</td>
<td>ga</td>
</tr>
<tr>
<td>20</td>
<td>ga</td>
</tr>
</tbody>
</table>

Table 4-6: Consecutive subject agreement morphemes of noun classes

<table>
<thead>
<tr>
<th>Person</th>
<th>Consecutive subject agreement morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>First person, singular</td>
<td>ka</td>
</tr>
<tr>
<td>First person, plural</td>
<td>ra</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>wa</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lwa</td>
</tr>
</tbody>
</table>

Table 4-7 Consecutive subject agreement morphemes of personal pronouns
<table>
<thead>
<tr>
<th>Class</th>
<th>Object agreement morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mo</td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
</tr>
<tr>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>e</td>
</tr>
<tr>
<td>5</td>
<td>le</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>se</td>
</tr>
<tr>
<td>8</td>
<td>di</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
</tr>
<tr>
<td>10</td>
<td>di</td>
</tr>
<tr>
<td>11</td>
<td>lo</td>
</tr>
<tr>
<td>12</td>
<td>di</td>
</tr>
<tr>
<td>14</td>
<td>bo</td>
</tr>
<tr>
<td>15</td>
<td>go</td>
</tr>
<tr>
<td>16</td>
<td>fa</td>
</tr>
<tr>
<td>17</td>
<td>go</td>
</tr>
<tr>
<td>18</td>
<td>go</td>
</tr>
<tr>
<td>19</td>
<td>go</td>
</tr>
<tr>
<td>20</td>
<td>go</td>
</tr>
</tbody>
</table>

Table 4-8: Object agreement morphemes of noun classes

<table>
<thead>
<tr>
<th>Person</th>
<th>Object agreement morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>First person, singular</td>
<td>ni</td>
</tr>
<tr>
<td>First person, plural</td>
<td>re</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>go</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lo</td>
</tr>
</tbody>
</table>

Table 4-9: Object agreement morphemes of personal pronouns
<table>
<thead>
<tr>
<th>Tense &amp; Aspect</th>
<th>Polarity</th>
<th>Identifying copulative</th>
<th>Describing copulative</th>
<th>Associative copulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present tense</td>
<td>Positive</td>
<td>Monna ke kgosi.</td>
<td>Monna o mogolo.</td>
<td>Monna o na le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man is a king.)</td>
<td>(The man is big.)</td>
<td>(The man has a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna ga se kgosi.</td>
<td>Monna ga a mogolo.</td>
<td>Monna ga a na koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man is not a king.)</td>
<td>(The man is not big.)</td>
<td>(The man does not have a car.)</td>
</tr>
<tr>
<td>Progressive</td>
<td>Positive</td>
<td>Monna e sa le kgosi.</td>
<td>Monna o sa le mogolo.</td>
<td>Monna o sa na le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man is still a king.)</td>
<td>(The man is still big.)</td>
<td>(The man still has a car.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monna e santse e le kgosi.</td>
<td>Monna o santse a le mogolo.</td>
<td>(The man still has a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna ga e thole e le kgosi.</td>
<td>Monna ga a thole a le mogolo.</td>
<td>(The man still does not have a car.)</td>
</tr>
<tr>
<td>Potential</td>
<td>Positive</td>
<td>Monna e ka nna kgosi.</td>
<td>Monna a ka nna mogolo.</td>
<td>Monna a ka nna le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man can be a king.)</td>
<td>(The man can be big.)</td>
<td>(The man can have a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna e ka se nne kgosi.</td>
<td>Monna a ka se nne mogolo.</td>
<td>Monna a ka se nne le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man cannot be a king.)</td>
<td>(The man cannot be big.)</td>
<td>(The man cannot have a car.)</td>
</tr>
<tr>
<td>Past tense</td>
<td>Positive</td>
<td>Monna e nnile kgosi.</td>
<td>Monna o nnile mogolo.</td>
<td>Monna o nnile le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man was a king.)</td>
<td>(The man was big.)</td>
<td>(The man had a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna ga e a nna kgosi.</td>
<td>Monna ga a nna mogolo.</td>
<td>Monna ga a nna le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man was not a king.)</td>
<td>(The man was not big.)</td>
<td>(The man did not have a car.)</td>
</tr>
<tr>
<td>Future tense</td>
<td>Positive</td>
<td>Monna e tla nna kgosi.</td>
<td>Monna o tla nna mogolo.</td>
<td>Monna o tla nna le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man will be a king.)</td>
<td>(The man will be big.)</td>
<td>(The man will have a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna ga e ketla e nna kgosi.</td>
<td>Monna a ketla a nna mogolo.</td>
<td>Monna a ketla a nna le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man will not be a king.)</td>
<td>(The man will not be big.)</td>
<td>(The man will not have a car.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monna ga e nke e nna kgosi.</td>
<td>Monna a nke a nna mogolo.</td>
<td>Monna a nke a nna le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man will not be a king.)</td>
<td>(The man will not be big.)</td>
<td>(The man will not have a car.)</td>
</tr>
<tr>
<td>Relative past tense</td>
<td>Positive</td>
<td>Monna e ne e le kgosi.</td>
<td>Monna o ne a le mogolo.</td>
<td>Monna o ne a na le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man was a king.)</td>
<td>(The man was big.)</td>
<td>(The man had a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna e ne e se kgosi.</td>
<td>Monna o ne a se mogolo.</td>
<td>Monna o ne a se na koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man was not a king.)</td>
<td>(The man was not big.)</td>
<td>(The man did not have a car.)</td>
</tr>
<tr>
<td>Relative future tense</td>
<td>Positive</td>
<td>Monna e tlabo e le kgosi.</td>
<td>Monna o tlabo a le mogolo.</td>
<td>Monna o tlabo a na le koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man will be a king.)</td>
<td>(The man will be big.)</td>
<td>(The man will have a car.)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Monna e tlabo e se kgosi.</td>
<td>Monna o tlabo a se mogolo.</td>
<td>Monna o tlabo a se na koloi.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(The man will not be a king.)</td>
<td>(The man will not be big.)</td>
<td>(The man will not have a car.)</td>
</tr>
</tbody>
</table>
### Consecutive mood

<table>
<thead>
<tr>
<th>Tense &amp; Aspect</th>
<th>Polarity</th>
<th>Identifying copulative</th>
<th>Describing copulative</th>
<th>Associative copulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Monna ya nna kgosi. (The man is then a king.)</td>
<td>Monna a nna mogolo. (The man is then big.)</td>
<td>Monna a nna le koloi. (The man then has a car.)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Monna ya se nne kgosi. (The man is then not a king.)</td>
<td>Monna a se nne mogolo. (The man is then not big.)</td>
<td>Monna a se nne le koloi. (The man then does not have a car.)</td>
<td></td>
</tr>
</tbody>
</table>

### Imperative mood

<table>
<thead>
<tr>
<th>Tense &amp; Aspect</th>
<th>Polarity</th>
<th>Identifying copulative</th>
<th>Describing copulative</th>
<th>Associative copulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Monna, nna kgosi! (Man, be a king!)</td>
<td>Monna, nna mogolo. (Man, be big!)</td>
<td>Monna, nna le koloi. (Man, has a car!)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Monna, se nne kgosi! (Man do not be a king!)</td>
<td>Monna, se nne mogolo. (Man do not be big!)</td>
<td>Monna, se nne le koloi. (Man, do not have a car!)</td>
<td></td>
</tr>
</tbody>
</table>

### Habitual mood

<table>
<thead>
<tr>
<th>Tense &amp; Aspect</th>
<th>Polarity</th>
<th>Identifying copulative</th>
<th>Describing copulative</th>
<th>Associative copulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Monna e nne kgosi. (The man is usually a king.)</td>
<td>Monna a nne mogolo. (The man is usually big.)</td>
<td>Monna a nne le koloi. (The man usually has a car.)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Monna e se nne kgosi. (The man is not usually a king.)</td>
<td>Monna a se nne mogolo. (The man is not usually big.)</td>
<td>Monna a se nne le koloi. (The man usually do not have a car.)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-10: Examples of copulative verbs in sentences
### Table 4-11: Identifying copulative verbs of personal pronouns

<table>
<thead>
<tr>
<th>Person</th>
<th>Personal pronoun</th>
<th>Identifying copulative verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>First person, singular</td>
<td>nna</td>
<td>ke</td>
</tr>
<tr>
<td>First person, plural</td>
<td>rona</td>
<td>re</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>wena</td>
<td>o</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lona</td>
<td>lo</td>
</tr>
</tbody>
</table>

### Table 4-12: Describing copulative verbs of noun classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Describing copulative verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o</td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
</tr>
<tr>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>e</td>
</tr>
<tr>
<td>5</td>
<td>le</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>se</td>
</tr>
<tr>
<td>8</td>
<td>di</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
</tr>
<tr>
<td>10</td>
<td>di</td>
</tr>
<tr>
<td>11</td>
<td>lo</td>
</tr>
<tr>
<td>12</td>
<td>di</td>
</tr>
<tr>
<td>14</td>
<td>bo</td>
</tr>
<tr>
<td>15</td>
<td>go</td>
</tr>
<tr>
<td>16</td>
<td>fa</td>
</tr>
<tr>
<td>17</td>
<td>go</td>
</tr>
<tr>
<td>18</td>
<td>go</td>
</tr>
<tr>
<td>19</td>
<td>go</td>
</tr>
<tr>
<td>20</td>
<td>go</td>
</tr>
</tbody>
</table>

### Table 4-13: Describing copulative verbs of personal pronouns

<table>
<thead>
<tr>
<th>Person</th>
<th>Personal pronoun</th>
<th>Describing copulative verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>First person, singular</td>
<td>nna</td>
<td>ke</td>
</tr>
<tr>
<td>First person, plural</td>
<td>rona</td>
<td>re</td>
</tr>
<tr>
<td>Second person, singular</td>
<td>wena</td>
<td>o</td>
</tr>
<tr>
<td>Second person, plural</td>
<td>lona</td>
<td>lo</td>
</tr>
<tr>
<td>Class</td>
<td>Possessive particle</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>wa</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>wa</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ya</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>la</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>sa</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>tsa</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ya</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>tsa</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>la</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>tsa</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ba/jwa</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>ga</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-14: Possessive particles
<table>
<thead>
<tr>
<th>Class</th>
<th>Qualificative particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yo</td>
</tr>
<tr>
<td>2</td>
<td>ba</td>
</tr>
<tr>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>e</td>
</tr>
<tr>
<td>5</td>
<td>le</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>se</td>
</tr>
<tr>
<td>8</td>
<td>tse</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
</tr>
<tr>
<td>10</td>
<td>tse</td>
</tr>
<tr>
<td>11</td>
<td>lo</td>
</tr>
<tr>
<td>14</td>
<td>bo</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
</tr>
<tr>
<td>15</td>
<td>go</td>
</tr>
<tr>
<td>16</td>
<td>go</td>
</tr>
<tr>
<td>17</td>
<td>go</td>
</tr>
<tr>
<td>18</td>
<td>go</td>
</tr>
<tr>
<td>19</td>
<td>go</td>
</tr>
<tr>
<td>20</td>
<td>go</td>
</tr>
</tbody>
</table>

Table 4-15: Qualificative particles
### Nouns

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>mang (who-singular)</td>
<td>bomang (who-plural)</td>
</tr>
<tr>
<td>mme (mother)</td>
<td>bomme (mothers)</td>
</tr>
<tr>
<td>monna (man)</td>
<td>banna (men)</td>
</tr>
<tr>
<td>mosadi (woman)</td>
<td>basadi (women)</td>
</tr>
<tr>
<td>mosetsana (girl)</td>
<td>basetsana (girls)</td>
</tr>
<tr>
<td>motho (human, person)</td>
<td>batho (humans, persons)</td>
</tr>
<tr>
<td>malome (uncle)</td>
<td>bomalome (uncles)</td>
</tr>
<tr>
<td>rre (father, sir)</td>
<td>borre (fathers, sirs)</td>
</tr>
<tr>
<td>John (John)</td>
<td>boJohn (John and company)</td>
</tr>
<tr>
<td>mosese (dress)</td>
<td>mesese (dresses)</td>
</tr>
<tr>
<td>moso (tomorrow)</td>
<td>mosenkele (shops)</td>
</tr>
<tr>
<td>lebenkele (shop)</td>
<td>mabenkele (shops)</td>
</tr>
<tr>
<td>letsatsi (day)</td>
<td></td>
</tr>
<tr>
<td>madi (money)</td>
<td></td>
</tr>
<tr>
<td>selihare (tree)</td>
<td></td>
</tr>
<tr>
<td>setlhopho (team)</td>
<td></td>
</tr>
<tr>
<td>Setswana (Setswana)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dinotlolo (keys)</td>
</tr>
<tr>
<td>buka (book)</td>
<td>dibuka (books)</td>
</tr>
<tr>
<td>dijo (food)</td>
<td></td>
</tr>
<tr>
<td>karata (card)</td>
<td>dikarata (cards)</td>
</tr>
<tr>
<td>kgosi (king)</td>
<td>dikgosi (kings)</td>
</tr>
<tr>
<td>toropo (town)</td>
<td></td>
</tr>
<tr>
<td>kgetsi (bag)</td>
<td>dikgetsi (bags)</td>
</tr>
</tbody>
</table>

### Locative Class Nouns

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>fatshe</td>
<td>(down, on the ground)</td>
</tr>
<tr>
<td>godimo</td>
<td>(above, on top)</td>
</tr>
<tr>
<td>morago</td>
<td>(behind)</td>
</tr>
<tr>
<td>gare</td>
<td>(in the middle, between)</td>
</tr>
<tr>
<td>teng</td>
<td>(here, inside)</td>
</tr>
</tbody>
</table>

Table 8-4: Nouns and locative class nouns
<table>
<thead>
<tr>
<th>Root</th>
<th>Basic Verb</th>
<th>Extended Verb</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>batl-</td>
<td>batlis-</td>
<td>batlisan-</td>
<td>(seek) (let seek) (help each other seek)</td>
</tr>
<tr>
<td>bo-</td>
<td>boil-</td>
<td>(return) (returned)</td>
<td></td>
</tr>
<tr>
<td>bon-</td>
<td>bonan-</td>
<td>(see) (see each other)</td>
<td></td>
</tr>
<tr>
<td>bu-</td>
<td>(speak)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>didimal-</td>
<td>(keep quiet)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>duel-</td>
<td>(pay)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dumalan-</td>
<td>(agree)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j-</td>
<td>jesan-</td>
<td>(eat) (feed each other)</td>
<td></td>
</tr>
<tr>
<td>kwal-</td>
<td>kwadis-</td>
<td>kwalel-</td>
<td>kwalelan-</td>
</tr>
<tr>
<td>lel-</td>
<td>(cry)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rat-</td>
<td>ratan-</td>
<td>(love, like) (love/like each other)</td>
<td></td>
</tr>
<tr>
<td>rek-</td>
<td>rekel-</td>
<td>rekw-</td>
<td>rekelw-</td>
</tr>
<tr>
<td>rom-</td>
<td>romiw-</td>
<td>(send) (being sent)</td>
<td></td>
</tr>
<tr>
<td>tabog-</td>
<td>tabogis-</td>
<td>tabogisiw-</td>
<td>(run) (let run) (being trained, “being let run for”)</td>
</tr>
<tr>
<td>tl-</td>
<td>tlis-</td>
<td>tlisets-</td>
<td>(come) (let come) (bring for, “let come for”)</td>
</tr>
<tr>
<td>tsamay-</td>
<td>tsamael-</td>
<td>(walk) (walk towards)</td>
<td></td>
</tr>
<tr>
<td>tseny-</td>
<td>tsenyets-</td>
<td>tsenyetsan-</td>
<td>tsenyetsanw-</td>
</tr>
<tr>
<td>tshamek-</td>
<td>tshamekel-</td>
<td>(play) (play for)</td>
<td></td>
</tr>
</tbody>
</table>

Table 8-5: Verbs (basic and extended verb roots)
EXAMPLES OF FORMATS FOR STORED ANALYSES

Tree in .SExp format

FEATURE STRUCTURE LEXICON (1.0)

FS1 FS *
  (^ ADJUNCT) = %1
  (^ AGR) = 1
  (^ NTYPE) = ord
  (^ FRED) = 'sadi'
%2 $ %1
  (%2 AGR) = 1
  (%2 PRED) = 'yo'
  (%2 PRONTYPE) = DemPro.

f-structure in .lfg format
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, subtree(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)**

203
### APPENDIX C: MORPHOLOGICAL TAGS

<table>
<thead>
<tr>
<th>Tag</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsPro</td>
<td>absolute pronoun</td>
</tr>
<tr>
<td>adjective</td>
<td>adjective</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
</tr>
<tr>
<td>AgPart</td>
<td>agentative particle</td>
</tr>
<tr>
<td>AgrObj</td>
<td>object agreement morpheme</td>
</tr>
<tr>
<td>AgrSubj</td>
<td>subject agreement morpheme</td>
</tr>
<tr>
<td>ApplSuf</td>
<td>applicative suffix</td>
</tr>
<tr>
<td>AssCopV</td>
<td>associative copulative verb</td>
</tr>
<tr>
<td>AssPart</td>
<td>associative particle</td>
</tr>
<tr>
<td>AugSuf</td>
<td>augmentative suffix</td>
</tr>
<tr>
<td>Aux</td>
<td>auxiliary</td>
</tr>
<tr>
<td>CausSuf</td>
<td>causative suffix</td>
</tr>
<tr>
<td>ComplSuf</td>
<td>completive suffix</td>
</tr>
<tr>
<td>CompPart</td>
<td>comparative particle</td>
</tr>
<tr>
<td>Conj</td>
<td>conjunction</td>
</tr>
<tr>
<td>ConsAgrSubj</td>
<td>consecutive subject agreement morpheme</td>
</tr>
<tr>
<td>D1, D2, D3</td>
<td>distance 1, distance 2, distance 3</td>
</tr>
<tr>
<td>DemPro</td>
<td>demonstrative pronoun</td>
</tr>
<tr>
<td>DesCopV</td>
<td>describing copulative verb</td>
</tr>
<tr>
<td>DevSuf</td>
<td>deverbatives suffix</td>
</tr>
<tr>
<td>DimSuf</td>
<td>diminutive suffix</td>
</tr>
<tr>
<td>FemSuf</td>
<td>feminine suffix</td>
</tr>
<tr>
<td>FutPre</td>
<td>future prefix</td>
</tr>
<tr>
<td>HortPart</td>
<td>hortative particle</td>
</tr>
<tr>
<td>IdCopV</td>
<td>identifying copulative verb</td>
</tr>
<tr>
<td>Ideo</td>
<td>ideophone</td>
</tr>
<tr>
<td>IncQPro</td>
<td>inclusive quantitative pronoun</td>
</tr>
<tr>
<td>Instrpart</td>
<td>instrumental particle</td>
</tr>
<tr>
<td>IntensSuf</td>
<td>intensive suffix</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection</td>
</tr>
<tr>
<td>IntPart</td>
<td>interrogative particle</td>
</tr>
<tr>
<td>Tag</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>LocPartfa</td>
<td>locative particle fa</td>
</tr>
<tr>
<td>LocPartga</td>
<td>locative particle ga</td>
</tr>
<tr>
<td>LocPartgo</td>
<td>locative particle go</td>
</tr>
<tr>
<td>LocPartka</td>
<td>locative particle ka</td>
</tr>
<tr>
<td>LocPartkwa</td>
<td>locative particle kwa</td>
</tr>
<tr>
<td>LocPartmo</td>
<td>locative particle mo</td>
</tr>
<tr>
<td>LocSuf</td>
<td>locative suffix</td>
</tr>
<tr>
<td>NegPrea</td>
<td>negative prefix a</td>
</tr>
<tr>
<td>NegPrega</td>
<td>negative prefix ga</td>
</tr>
<tr>
<td>NegPresa</td>
<td>negative prefix sa</td>
</tr>
<tr>
<td>NPre</td>
<td>noun prefix</td>
</tr>
<tr>
<td>P1</td>
<td>first person</td>
</tr>
<tr>
<td>P2</td>
<td>second person</td>
</tr>
<tr>
<td>PassSuf</td>
<td>passive suffix</td>
</tr>
<tr>
<td>PerfSuf</td>
<td>perfect suffix</td>
</tr>
<tr>
<td>pl</td>
<td>plural</td>
</tr>
<tr>
<td>PossPart</td>
<td>possessive particle</td>
</tr>
<tr>
<td>PossPro</td>
<td>possessive particle</td>
</tr>
<tr>
<td>PresPre</td>
<td>present tense prefix a</td>
</tr>
<tr>
<td>ProgPre</td>
<td>progressive prefix</td>
</tr>
<tr>
<td>QualPart</td>
<td>qualificative particle</td>
</tr>
<tr>
<td>RecSuf</td>
<td>reciprocal suffix</td>
</tr>
<tr>
<td>SepQPro</td>
<td>separative quantitative particle</td>
</tr>
<tr>
<td>sg</td>
<td>singular</td>
</tr>
<tr>
<td>TempPart</td>
<td>temporal particle</td>
</tr>
<tr>
<td>VEnd</td>
<td>verbal ending</td>
</tr>
<tr>
<td>1 2 3 ...... 20</td>
<td>noun classes 1 to 20</td>
</tr>
</tbody>
</table>

Table: Morphological tags
REFERENCE LIST


Bresnan, J. 1982. The mental representation of grammatical relations. Cambridge: The MIT.


208


210


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.


