



Developing Sepedi Phrase Structure Rules: lexical and phrasal categories

Napjadi Letsoalo



Published online: 21/08/2022

To cite this OER: Letsoalo, Napjadi. Developing Sepedi Phrase Structure Rules: lexical and phrasal categories. *Unisa Open Teaching and Learning Materials*, 21 Aug. 2022, CC BY-NC-SA 2.5 ZA.

Developing Sepedi Phrase Structure Rules: lexical and phrasal categories

Napjadi Letsoalo

Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa

Correspondance: letsone@unisa.ac.za | letsoalone@gmail.com

1. Introduction

The grammatical patterns that characterise different kinds of phrases can be expressed in terms of **phrase-structure rules** that represent speakers' knowledge about well-formed structures or expressions in a language. These rules describe the way phrases can be successfully constructed. The infinite set of sentences that speakers can produce and understand are made possible by the fact that there are only a limited set of phrase-structure rules that specify the order and structure of the various constituents in our language. The purpose of this paper is to generate rewrite rules or phrase structure rules that can be used to describe Sepedi's syntax.

2. Basic Categories

The most basic categories that are important to the composition of labelling notation for Sepedi are noun (*leina*), verb (*lediri*), adjective (*lehlaodi*), adverb (*lehlathi*) and concord (*lekgokedi*). Consider the sentence below,

Banna ba batelele ba rema mohlare bonolo.

Ba-nna ba ba-telele ba rema mo-hlare bonolo.

2-Man Dem2 SC2-tall SC cut 3-tree easy

'Tall men cut the tree easily.'

Considering the sentence *Banna ba batelele ba rema mohlare bonolo*, a breakdown is provided to indicate the lexical items that represent Sepedi lexical categories; these are shown in square brackets and are highlighted:

[Banna] ba batelele ba rema **[mohlare]** bonolo. (*leina*)

Banna **[ba batelele]** ba rema mohlare bonolo. (*lehlaodi*)

Banna ba batelele **[ba]** rema mohlare bonolo. (*lekgokedi*)

Banna ba batelele ba **[rema]** mohlare bonolo. (*lediri*)

Banna ba batelele ba rema mohlare **[bonolo]**. (*lehlathi*)

3. Category Labeling

In order to have a clear system of syntactic representation, each category has a label. For the purpose of understanding, this paper adopts initialism, a type of word formation by taking the initial letters that is uttered letter by letter (Adha and Dania, 2020). Notice that the lexical categories presented above start with the noun class 5 prefix *le-*. Initialism is only productive if the prefix is omitted to remain with *-ina* from *leina*, *-hlaodi* from *lehlaodi*, *-kgokedi* from *lekgokedi*, *-diri* from *lediri* and *-hlathi* from *lehlathi*. These labels are an easy way to indicate the lexical categories of a sentence. As mentioned before, syntactic analysis of a sentence requires a system of notations to represent syntactic analyses of sentences. According to Letsoalo (2019), the categories are represented as follows:

I	<i>leina</i>	noun
D	<i>lediri</i>	verb
Hd	<i>lehlaodi</i>	adjective
Ht	<i>lehlathi</i>	adverb
Kg	<i>lekgokedi</i>	concord

Note that the first three letters of *-hlaodi* and *-hlathi* are the same; in order to distinguish the two, the initial letters of syllables are used. The category labels of these words are a result of the shortening of a word made by omitting certain letters and bringing together the initial letters of syllables:

<i>hlaodi</i>	(Hd)
<i>hlathi</i>	(Ht)

These labels are an easy way to indicate the lexical categories of a sentence. Syntactic analysis of a sentence requires a system of notations to represent syntactic analyses of sentences. This system of analysis reveals how a sentence conforms to the following: (1) words follow each other in a specific linear order; (2) words belong to a category or class; and (3) certain strings of words form larger groupings.

4. Linearity Principle

Words in a sentence occur in a specific linear order. Sepedi has an acceptable linear order of words in a sentence, and a sentence can be deemed ungrammatical if the permissible word order of sentences is violated. Words in each sentence should follow each other in a certain sequence, and if the sequence is changed, the sentence either becomes ungrammatical or changes meaning. Look at the following example:

(a) *Ba-nna* *ba* *ba-telele* *ba* *rema* *mo-hlare* *bonolo*.
 2-Man Dem2 SC2-tall SC cut 3-tree easy
 Men *who (are) tall they cut (the) tree easily*.

(b) **Banna* *ba* *mo-hlare* *rema* *bonolo* *ba* *ba-telele*.
 2-Man Dem2 3-tree cut easy SC SC2-tall
 Men *who tree cut easy they tall*.

The second sentence is ungrammatical because the permissible word order of Sepedi sentences has been violated. The classification of the various constituents of a sentence, words and phrases, will be indicated by means of category labels such as I (for *leina*), D (for *lediri*), Kg (for *lekgokedi*), Hd (for *lehlaodi*), Ht (for *lehlathi*), etc. This labeling system is called **labeled bracketing**, a way of representing the structure of an expression by writing square brackets ('[' and ']') to the left- and right-hand side of its component parts, i.e. words or constituents (Glottopedia, 2009). For example,

(c) [I *Banna*] [Hd *ba batelele*] [Kg *ba*] [D *rema*] [I *mohlare*] [Ht *bonolo*].

Syntactic analysis also involves labeling of larger groupings of certain strings of words. The phrases (*sehlophantšu* – singular form), in the same manner as lexical items are labelled, are indicated by means of phrasal labels such as SI (for *sehlophantšu ina*) a noun phrase, SD (for *sehlophantšu diri*) a verb phrase, SHd (for *sehlophantšu hlaodi*) an adjectival phrase, SHt (for *sehlophantšu hlathi*) etc. The larger groupings of the constituents of a sentence are indicated by square brackets below:

(d) [SI *Banna ba batelele*] [SD *ba rema mohlare bonolo*]

The further syntactic analysis of the sentence above can be presented as:

(e) [SI [I *Banna*] [Hd *ba batelele*]] [SD [Kg *ba*] [D *rema*] [SI [I *mohlare*]] [SHt [Ht *bonolo*]]].

The syntactic label of a sentence is (L for *lefoko*). Therefore, the complete syntactic analysis of the sentence above can be presented as:

(f) [L [SI [I *Banna*] [Hd *ba batelele*]] [SD [Kg *ba*] [D *rema*] [SI [I *mohlare*]] [SHt [Ht *bonolo*]]]]

5. Phrase Structure Rules

From the analysis of the sentence *Banna ba batelele ba rema mohlare bonolo* provide above, the following **basic Sepedi phrasal rules** are deduced:

i. $L \rightarrow SI SD$

This rule says that a sentence (L for *lefoko*) is composed of a noun phrase (SI for *sehlophantšu ina*) and a verb phrase (SD for *sehlophantšu diri*).

ii. $SI \rightarrow I Hd$

This rule indicates that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*) and an adjective (Hd for *lehlaodi*).

iii. $SD \rightarrow Kg D SI SHt$

This rule says that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a verb (D for *lediri*), a noun phrase (SI for *sehlophantšu ina*) and an adverbial phrase (SHt for *sehlophantšu hlathi*).

iv. SI → I

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*).

v. SHt → Ht

This rule says that an adverbial phrase (SHt for *sehlophantšu hlathi*) is composed of an adverb (Ht for *lehlathi*).

6. Additional Phrase Structure Rules

The rules above, however, are specific to the sentence above and those syntactically similar to it. There are phrases in Sepedi that cannot be represented by the rules above. Therefore, there should be rules that account for those phrasal structures. There are other noun phrases that need separate rules from the ones above.

(a) *Matome le Lesedi*
 1a-Matome CP 1a-Lesedi
 Matome and Lesedi

(b) *Matome le Lesedi yo mo-golo*
 1a-Matome CP 1a-Lesedi Dem1 SC1-big
 Matome and Lesedi the elder

(c) *Matome yo mo-golo le Lesedi*
 1a-Matome Dem1 SC1-big CP 1a-Lesedi
 Matome the elder and Lesedi

(d) *Matome yo mo-golo le Lesedi yo mo-telele*
 1a-Matome Dem1 SC1-big CP 1a-Lesedi Dem1 SC1-tall
 Matome the elder and Lesedi the tall

6.1 *Sehlophantšu Ina* (Noun Phrases)

The phrases above are compound noun phrases; the nouns are brought together by a connective particle (*lekopanyi*) *le*. Using the initialism strategy referred to above, the category label for *lekopanyi* is K. The phrasal rules that account for the structures above are the following, respectively:

vi. SI → I K I

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*), a connective particle (K for *lekopanyi*) and a noun (I for *leina*).

vii. SI → I K I Hd

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*), a connective particle (K for *lekopanyi*), a noun (I for *leina*) and an adjective (Hd for *lehlaodi*).

viii. SI → I Hd K I

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*), an adjective (Hd for *lehlaodi*), a connective particle (K for *lekopanyi*) and a noun (I for *leina*).

ix. SI → I Hd K I Hd

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*), an adjective (Hd for *lehlaodi*), a connective particle or conjunction (K for *lekopanyi*), a noun (I for *leina*) and an adjective (Hd for *lehlaodi*).

6.2 *Sehlophantšu Dir* (SD) - Verb Phrase (VP)

There are verb phrases which are different to that in the sentence (*Banna ba batelele ba rema mohlare bonolo*) above. Consider the examples below:

(a) *o a sega*
 SC3SG TM laugh
 She is laughing.

(b) *ba sega kudu*
 SC3PL laugh very much
 They laugh very much.

(c) *le di bone*
 SC2PL OC8 saw
 You saw them.

(d) *re di bone gabotse*
 SC1PL OC8 saw clear
 We saw them clearly.

(e) *o phela a sega*
 SC3SG always SC3SG laugh
 She is always laughing.

(f) *ba a dula ba sega*
 SC3PL TM sit SC3PL laugh
 They are sitting, they laugh.

Some of the phrases above have a tense marker (*lebaka*) *a*. Using the initialism strategy that was applied above, the label for *lebaka* is B. The phrasal rules that account for the structures above are the following, respectively:

x. $SD \rightarrow Kg B D$

This rule says that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a tense marker (B for *lebaka*) and a verb (D for *lediri*).

xi. $SD \rightarrow Kg D SHt$

This rule stipulates that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a verb (D for *lediri*) and an adverbial phrase (SHt for *sehlophantšu hlathi*).

xii. $SD \rightarrow Kg Kg D$

This espouses that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a concord (Kg for *lekgokedi*) and a verb (D for *lediri*).

xiii. $SD \rightarrow Kg Kg D SHt$

This rule says that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a concord (Kg for *lekgokedi*), a verb (D for *lediri*) and an adverbial phrase (SHt for *sehlophantšu hlathi*).

xiv. $SD \rightarrow Kg D Kg D$

This rule stipulates that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a verb (D for *lediri*), a concord (Kg for *lekgokedi*), and a verb (D for *lediri*).

xv. $SD \rightarrow Kg B D SD (Kg D)$

This rule says that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a tense marker (B for *lebaka*), a verb (D for *lediri*), and another verb phrase (SD for *sehlophantšu diri*) which is composed of a concord (Kg for *lekgokedi*) and a verb (D for *lediri*).

7. Three Basic Phrase Structure Rules

From the rules above, the following rules can be said to be the basic phrase structure rules:

$L \rightarrow SI SD$

This rule states that a sentence (L for *lefoko*) is composed of a noun phrase (SI for *sehlophantšu ina*) and a verb phrase (SD for *sehlophantšu diri*).

$SI \rightarrow I (Hd)$

This rule says that a noun phrase (SI for *sehlophantšu ina*) is composed of a noun (I for *leina*), and the noun can optionally be modified by an adjective (Hd for *lehlaodi*).

$SD \rightarrow Kg B D (SI) (SHt)$

This rule says that a verb phrase (SD for *sehlophantšu diri*) is composed of a concord (Kg for *lekgokedi*), a tense marker (B for *lebaka*) and a verb (D for *lediri*), and can

optionally be composed of additional noun phrase (SI for *sehlophantšu ina*) and (SHt for *sehlophantšu hlathi*).

These phrase structure rules are important to guide the syntactic analysis of phrases and sentences.

8. Summary

Phrase structure rules are important in the analysis of syntactic structures. The following constituent nodes were produced: L for *lefoko* (sentence), SI for *sehlophantšu ina* (noun phrase), SD for *sehlophantšu diri* (verb phrase), I for *leina* (noun), D for *lediri* (verb), Kg for *lekgokedi* (connective particle), K for *lekopanyi* (conjunction), Hd for *lehlaodi* (adjective), Ht for *lehlathi* (adverb), and SHt for *sehlophantšu hlathi* (adverbial phrase).

References

Letsoalo, N. 2019. Interrogative construction in Sepedi. PhD Thesis, Turfloop: University of Limpopo. <http://hdl.handle.net/10386/3181>

Adha, A.D. and Dania, R. 2020. Morphological Analysis of Word Formation Found in VOA News Articles. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 8(2), pp.458-470.

Glottopedia contributors, "Labeled bracketing," Glottopedia, 16 Feb, 2009, http://glottopedia.org/index.php?title=Labeled_bracketing&oldid=7879 (accessed August 21, 2022).