Chapter 4

Features of verbal phrases

4.1 Introduction

A significant sub-task in the preparation of parsing is that of gathering information which helps resolve ambiguities. It has been argued in paragraph 2.1 (on page 20) that distinguishing word classes according to their distribution supports the development of morphosyntactic rules leading to a lower number of resulting analyses than - for example - distinguishing them according to their morphological characteristics. Moreover, it was shown in paragraph 3.2.1.6 that labelling the transitivity of verbs - for example by defining labels of previously underspecified verb stem properties in greater detail - supports the identification of unmarked sentence borders. If lexical items are however intrinsically ambiguous , i.e. if they appear in a number of word classes, it will be necessary to find their distributional patterns on the basis of their context to support disambiguation on a lexical level. For example, Northern Sotho makes use of quite a number of highly ambiguous function words, of which e.g. *a* may either represent a subject concord of noun class 1, set 2 and 3 (2CS01, 3CS01) or of noun class 6 (all three sets: 1CS06, 2CS06, 3CS06), or a demonstrative concord (CDEM06), or a possessive concord (CPOSS06), both of noun class 6, or a hortative particle (PART_hort) etc. Even without taking the different subject concord sets into consideration, *a* already has a ninefold ambiguity, as Table 4.1 (taken from Faaß et al. (2009)) demonstrates. According to the definitions of this study, this token could be interpretated in 12 ways. In order to support identification of the correct POS of *a*, contextual partners must therefore be considered.

Any work on defining patterns in the co-occurrence of parts of speech could also assist in developing generalisations on the path to a more general linguistic modelling of Northern Sotho verbs than is undertaken in this study. This chapter could therefore form the basis

	Description	Example
1	subject concord of nominal cl. 1	*ge monna* ***a*** *fihla* conjunctive + noun cl. 1 + **subject concord cl. 1** + verb stem if/when + man + **subj-cl1** + arrive 'when the man arrives'
2	subject concord of nominal cl. 6	*masogana* ***a*** *thuša basadi* noun cl. 6 + **subject concord cl. 6** + verb stem + noun cl.2 young men + **subj-cl6** + help women 'the young men help the women'
3	object concord of	*moruti o* ***a*** *biditše* noun cl. 1 + subject concord cl. 1 + **object concord cl. 6** + verb stem teacher + subj-cl1 + **obj-cl6** + called 'the teacher called them'
4	possessive concord of nominal cl. 6	*maoto* ***a*** *gagwe* noun cl. 6 + **possessive concord cl. 6** + possessive pronoun cl. 1 feet + **of** + his 'his feet'
5	demonstrative concord of nominal cl. 6	*ba nyaka masogana* ***a*** subject concord cl. 2 + verb stem + noun cl. 6 + **demonstrative concord** they + look for + young men + **these** 'they are looking for these young men'
6	present tense morpheme	*morutiši o* ***a*** *bitša* noun cl. 1 + subject concord cl.1 + **present tense marker** + verb stem teacher + subj-cl1 + **pres** + call 'the teacher is calling'
7	past tense morpheme	*morutiši ga o* ***a*** *bitša masogana* noun cl. 1 + negation morpheme + subject concord cl.1 + **past tense marker** + verb stem + noun cl. 6 teacher + neg + subj-cl1 + **past** + call + young men 'the teacher did not call the young men'
8	hortative particle	***a*** *ba tsene* **hortative particle** + subject concord cl. 2 + verb stem let + subj-cl2 + come in 'let them come in'
9	interrogative particle	***a*** *o tseba Sepedi* **interrogative particle** + subject concord 2nd pers sg. + verb stem + noun cl. 7 **ques** + subj-2nd-pers-sg + know + Sepedi 'do you know Sepedi'

Table 4.1: The polysemy of *a* (taken from Faaß et al. (2009))

for future work, where other parsing strategies will come into use, concentrating on clusters of parts of speech found in text rather than using pre-defined VP-rules. In this chapter, Northern Sotho verbal constellations are summarised according to some of their features detectable from a computational perspective and we will attempt to provide a first set of generalisations for them. However first it is necessary to look at how a parser works and how the knowledge outlined above can be utilized.

4.2 Parsers: approaches to describe natural languages

Parsers, i.e. operational grammars, are capable of comparing a given text (usually a sentence) with pre-defined grammatical and lexical knowledge of a system, i.e. about the units and how they combine. "Rule-based" parsers contain a lexicon and a set of rules (e.g. context free grammars (CFG) as described in paragraph 1.4.4). Other parsers do not contain a separate rules-section, as the units of language in those systems (lexical elements and the constituents they form, i.e. phrases) are described using the same schemata. Such parsers are called "lexicon-based" (Hellwig, 1989, p. 422).

In principle, parsers match specific sets of properties allocated to pre-defined units of a language with flowing text provided as their input. During the first processing step, this flowing text is usually separated into units (tokens). As described in paragraph 1.4.2.1, such separation is no trivial task, however solvable with tokenization tools. The units identified in the sentence and the order in which they appear are compared to the system's knowledge on legal constellations. If there is a match, the parser will accept the sentence. Nowadays, most parsers will not only inform the user whether the sentence was accepted or not but also assign one or more 'well-formed' - i.e. correct - analyses (parses). Some parsers work shallowly, their analysis resulting in flat, non-recursive 'chunks' rather than phrases or sentences, like, e.g. noun phrases, these are often called shallow parsers or 'chunkers'. In a rule-based parser, grammatical knowledge is represented as constraints, and as long the constituent in question - chunk or sentence - conforms to these constraints, it is considered to be well-formed (Sag et al., 2003, p. 83).

The system's knowledge utilised by the parser often provides these constraints as sets of properties containing attribute-value pairs. Analyses result from certain principles, including the principle of uniqueness (cf. paragraph 5.1.2.2). For example, the English verb form 'sleeps' is often stored in parser lexicons with two attribute-value pairs related to the

expected subject, 'person'='third' and 'number'='singular'. Any noun that could possibly be the verb's subject (i.e. a noun that appears in a position where the grammar rule expects the subject to appear) will only be accepted as such by the parser if it is not stored with information contradicting these constraints. A phrase like *'pupils sleeps' will not be authorised because of such contradicting constraints of the units, as the noun 'pupils' is usually stored with the attribute-value pair 'number'='plural' and the uniqueness principle only allows for one value to appear with each attribute. In other words, the uniqueness principle proihibits units contributing to a structure (e.g. a phrase) from providing contradicting values for one attribute. Therefore, some constraint-based grammars are also called unification-based grammars.

If words or a sentence may be interpreted in several ways, they are ambiguous, lexically and/or structurally. Humans can disambiguate a number of such ambiguities using their world knowledge, so while humans consider the noun phrase 'mothers and children under 13' (cf. example (77)) to only have one reading, a parser will usually assign three analyses, as the mothers, the children or the mothers and the children may – from a syntactic perspective – all be under 13 ("pp-attachment", cf. e.g. Schütze (1995)).

(77)

bomme$_{\text{N02b}}$	*le*$_{\text{PART\_con}}$	*bana*$_{\text{N02}}$	*ba*$_{\text{CDEM02}}$	*ba*$_{\text{CS02}}$	*lego*$_{\text{VCOP}}$	*fase*$_{\text{NLOC}}$
mothers	`con`	children	`dem-3rd-cl2`	`subj-3rd-cl2`	who are	under
ga$_{\text{CPOSSLOC}}$	*mengwaga*$_{\text{N04}}$	*ye*$_{\text{CDEM04}}$	*13*$_{\text{NUM}}$			
of	years	`dem-3rd-cl4`	13			

'mothers and children under 13'

However, even where world knowledge, as in (77) is not necessary for the disambiguation of a sentence, tools analysing human language are often still confronted with lexical and structural ambiguities on all representation levels. Applying generalisations on features of constellations, i.e. signalling elements or signalling element clusters may help to reduce such ambiguities. Considering Northern Sotho, a typical example is the situative $^{\text{SIT}}$VP (cf. paragraph 3.2.6) which in a number of cases is identical on the surface with an indicative $^{\text{IND}}$VP (cf. paragraph 3.2.5). Such phrases are therefore ambiguous, however, as soon as the context is taken into account, their ambiguity can often be resolved; in the case of the situative, it is often preceded by the conjunction *ge* -'when' - which never occurs with an indicative. A feature catalogue of the Northern Sotho phrases, i.e. a **data category inventory** should support an effective reduction of ambiguities; some basics will be described in section 4.4.

For an overview of the many possible parsing strategies and algorithms, see e.g. (Hellwig, 1989). This section will solely describe the parsing algorithm (a right-corner parser) utilized in the example analysis shown in section 4.3, paragraphs 4.2.1 to 4.2.1.2 briefly explain the necessary terms.

4.2.1 Vertical parsing directions

There are top-down and bottom-up parsers. Following a top-down strategy, a set of all sentences that are possible according to the knowledge of the system (i.e. lexicon and rules) is generated first. The sentence in question, i.e. the sentence to be processed is then examined in order to find out if it is a member of the given set. A bottom-up strategy on the other hand begins with the surface sentence and combines the units of the sentence by utilizing the knowledge step by step until the highest possible node, the S-node is found (cf. paragraph 1.4.4 on page 13). If a system contains few rules and a small lexicon, a top-down strategy can be sufficient. However, in the case of Northern Sotho, where quite a few morphosyntactic rules have been defined (cf. chapter 3), we have chosen a bottom-up strategy to avoid over-generation[1].

4.2.1.1 Horizontal parsing directions

Depending on the language in question, some parsers process sentences from left to right, others from right to left. Here, it is not relevant how a human would read the sentence, rather how the number of (intermediate) illicit analyses can be kept low while processing the sentence. This issue would constitute a research question by itself for parsing Northern Sotho, exceeding the scope of this study. We have opted rather arbitrary to try a right-to-left analysis in section 4.3.

4.2.1.2 Right-corner parsing

Applications of parsing algorithms often combine both processing strategies - bottom-up and top-down - like, e.g. the left-corner parser which begins with the leftmost element of the sentence and, while constantly identifying lexicon entries fitting to the surface forms it meets, it “builds sentence structure in a left-to-right, bottom-up fashion, piecing together

[1]Over-generation describes a situation where illicit analyses are generated by a system. These have to be filtered out, either manually or by additional algorithms.

the left corner of a structural description first" (Petrick, 1989, p. 690). Our sample analysis proceeds similar to this parser, hence we call it 'right-corner parser'.

4.3 A sample analysis

For the sake of convenience, we repeat example (47) of page 105 (cf. Lombard (1985, p. 147)) as (78) here in order to illustrate the reduction of lexical and structural ambiguities step by step.

(78)

ge	*ba*	*bona*	*noga*	*ba*	*a*	*e*	*bolaya*
when	`subj-3rd-cl2`	see	$\text{snake}_{\text{N09}}$	`subj-3rd-cl2`	`pres`	`obj-3rd-cl9`	kill

'when they see a snake, they kill it'

The parser is assumed to know about the rules defined in chapter 3. The assumed system's lexicon contains the possible parts of speech for all of the contained tokens. Note that ':' in the following list indicates POS-ambiguity, e.g. *bona* may be an emphatic pronoun of class 2 (PROEMP02) or a possessive pronoun of the same class (PROPOSS02), or a transitive verb (V_tr, meaning '[to] see').

- *ge*
  CONJ
- *ba*
  1CS02:2CS02:3CS02:CDEM02:CO02:CPOSS02:V_AUX:VCOP
- *bona*
  PROEMP02:PROPOSS02:V_tr
- *noga*
  N09
- *ba*
  1CS02:2CS02:3CS02:CDEM02:CO02:CPOSS02:V_AUX:VCOP
- *a*
  1CS06:2CS06:3CS06:2CS01:3CS01:CDEM06:CO06:CPOSS06:
  MORPH_past:MORPH_pres:PART_hort:PART_que
- *e*
  1CS04:2CS04:1CS09:2CS09:CO04:CO09:CSNEUT
- *bolaya*
  V_tr

Our example analysis[2] begins with the rightmost element, which is $bolaya_{V\_tr}$ 'kill'. The parser first attempts to identify rules ending in this POS as hypothetical analyses. As none are found (no phrase consists of a transitive verb alone), the second element, *e*, is considered. In this study (cf. paragraph 3.2.1.1), a basic verbal phrase, VBP was defined as the smallest element of any verbal phrase containing a verbal stem and possibly its object(s). The token *e* which is highly ambiguous (1CS04:2CS04:1CS09:2CS09:CO04:CO09:CSNEUT) will therefore be assumed to be an object concord of either class 4 or class 9 (CO04:CO09), i.e the first (lexical) ambiguity is partially resolved, as shown in Figure 4.1.

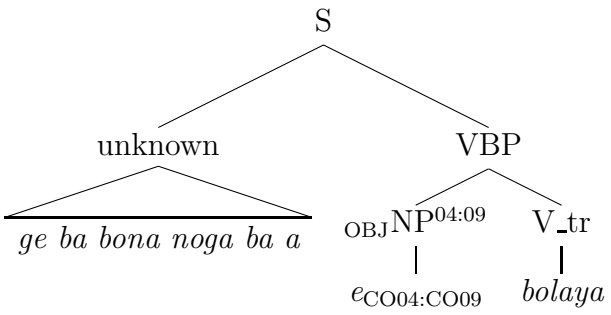

Figure 4.1: A partial analysis of (*ge ba bona noga ba a*) *e bolaya* '`obj-3rd-cl4:9` kill'

A number of possible analyses remain, as Table 4.2 shows. All verbal phrases that contain VBPs defined in chapter 3 containing a verb stem ending in *a* are considered.

The next element, *a*, is highly ambiguous; the set of possible parts of speech contains twelve elements. However, the particles may both be excluded from the set as PART_hort always precedes a subject concord, and we have already identified *e* as an object concord and PART_que ordinarily only appears as either the first or the last element of the sentence (the parser is informed about the total length, i.e. the number of tokens contained in the sentence). Due to the grammar rules describing NPs, the parser is aware that the possessives CPOSS01:CPOSS06 only occur in noun phrases where they precede nouns or possessive pronouns. If we look at the knowledge gained so far, a demonstrative concord CDEM06 could only appear here preceding a subject concord of class 6, which is not the case. This reading can therefore also be excluded from the set . The element *e* has already been identified as an object concord, therefore *a* cannot be CO06 (Northern Sotho only allows for one object concord to occur in the verbal phrase). Only seven possibilities

[2]Note that the following explanations are fairly abstract and thus simplified.

Table 4.2: The remaining set of possible analyses when considering the VBP ending in *a*

	$^{\text{IND}}$VP			
	VIE		**VBP**	
descr.	**zero-2**	**zero-1**	**zero**	**zero +1**
pres.pos.long	$1\text{CS}_{\text{categ}}$	MORPH_pres	VBP^{p}	\$.
perf.neg. 1	$ga_{\text{MORPH\_neg}}$ $se_{\text{MORPH\_neg}}$ $3\text{CS}_{\text{categ}}$		VBP	
perf.neg. 3	$ga_{\text{MORPH\_neg}}$ $3\text{CS}_{\text{categ}}$		VBP	
perf.neg. 4	$ga_{\text{MORPH\_neg}}$ $1\text{CS}_{\text{categ}}$ a		VBP	
fut.pos	$1\text{CS}_{\text{categ}}$	tlo/tla MORPH_fut	VBP	
	$^{\text{SIT}}$VP			
pres.pos.	$2\text{CS}_{\text{categ}}$		VBP	
perf.neg.1	$2\text{CS}_{\text{categ}}$ $se_{\text{MORPH\_neg}}$ $3\text{CS}_{\text{categ}}$		VBP	
perf.neg.2	$2\text{CS}_{\text{categ}}$ $se_{\text{MORPH\_neg}}$ $1\text{CS}_{\text{categ}}$		VBP	
perf.neg.3	$2\text{CS}_{\text{categ}}$ $sa_{\text{MORPH\_neg}}$		VBP	
fut.pos.	$2\text{CS}_{\text{categ}}$	tlo/tla MORPH_fut	VBP	
	$^{\text{REL}}$VP			
perf.neg.1	$2\text{CS}_{\text{categ}}$ $sego/seng$ MORPH_neg $3\text{CS}_{\text{categ}}$		VBP	
fut.pos.1	$2\text{CS}_{\text{categ}}$	$tlago/tlogo$ MORPH_fut	VBP	
fut.neg.1	$2\text{CS}_{\text{categ}}$ $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	$tlago/tlogo$ MORPH_fut	VBP	

Table 4.3: The remaining set of possible analyses when considering *a*

	$^{\text{PRES}}$VP			
	VIE		**VBP**	
descr.	**zero-2**	**zero-1**	**zero**	**zero +1**
pres.pos.long	$1CS_{categ}$	MORPH_pres	VBP^{p}	$.
perf.neg. 1	ga_{MORPH_neg} se_{MORPH_neg} $3CS_{categ}$		VBP	
perf.neg. 3	ga_{MORPH_neg} $3CS_{categ}$		VBP	
perf.neg. 4	ga_{MORPH_neg} $1CS_{categ}$ MORPH_past		VBP	
	$^{\text{SIT}}$VP			
pres.pos.	$2CS_{categ}$		VBP	
perf.neg.1	$2CS_{categ}$ se_{MORPH_neg} $3CS_{categ}$		VBP	
perf.neg.2	$2CS_{categ}$ se_{MORPH_neg} $1CS_{categ}$		VBP	
	$^{\text{REL}}$VP			
perf.neg.1	$2CS_{categ}$ *sego/seng* $MORPH\_neg$ $3CS_{categ}$		VBP	

therefore remain (2/3CS01:1/2/3CS06:MORPH_past/pres) as illustrated in Figure 4.2. By accepting that the constellation either contains a subject concord or a tense morpheme, all respective constellations can be excluded from the set of hypothetical analyses. The remaining set of possible analyses however still contains all predicative VPs (except for copulative constellations which can be excluded, because a main verb is contained), cf. Table 4.3 and Figure 4.2.

When examining the next preceding token, *ba*, some of its possible parts of speech may again be excluded on the basis of the information collected so far. In the given sentence, we can exclude the object concord reading, as *e* has already been successfully identified as an object concord which only appears alone in Northern Sotho verbs, secondly, these concords must appear directly in front of the verb stem. If *a* should however represent a tense morpheme, *ba* may be a subject concord of class 2 (1CS02:2CS02:3CS02) and as such it may function as the subject of the clause. The ambiguity remaining for the following element, *a*, can now be mostly resolved as, if preceded by a subject concord it can only

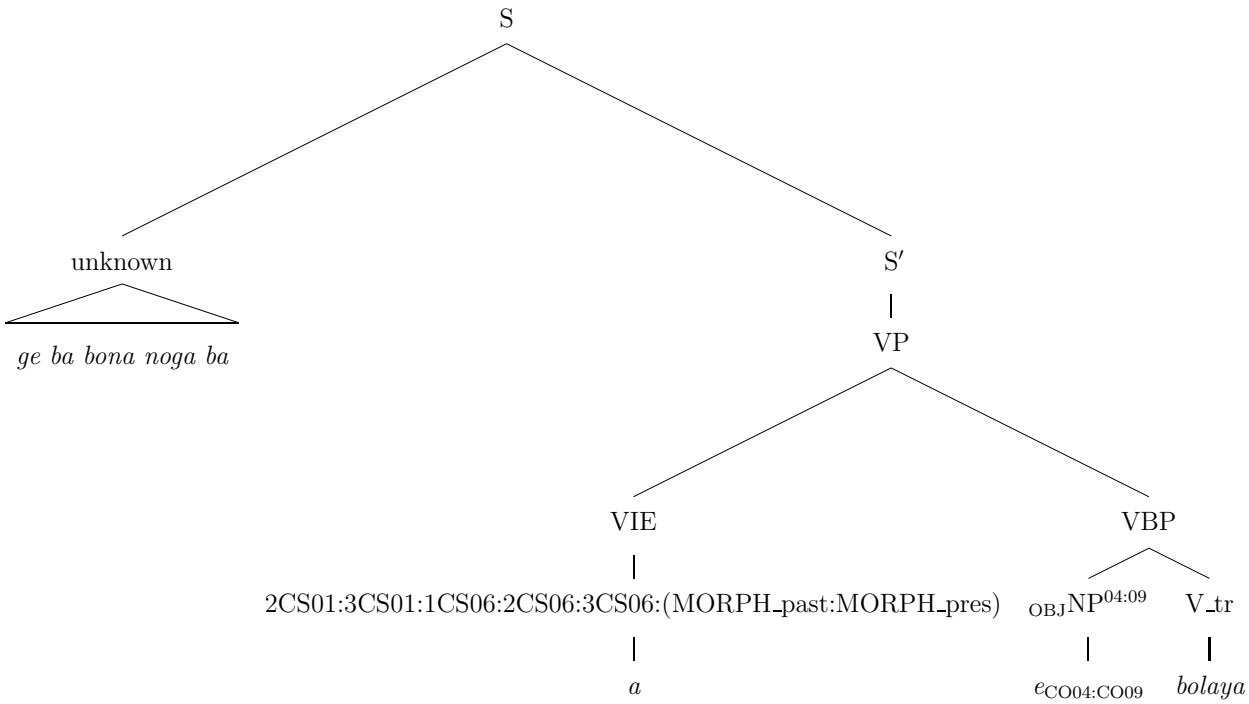

Figure 4.2: A partial analysis of (*ge ba bona noga ba*) *a e bolaya*'`subj-3rd(-cl1/-cl6)/pres/past obj-3rd-cl4:9` kill'

be either a present or past tense morpheme (another condition, namely that the sentence ends after the verb stem, is also fulfilled). The subject concord may in this case only be of the first set and the VP as a whole must be in the indicative mood (present tense, positive or perfect tense, negative), because the constellation contains a tense morpheme. Consider Table 4.4 and Figure 4.3.

Table 4.4: The remaining analyses when considering *ba* – indicative $^{\text{IND}}$VP

	VIE			**VBP**	
descr.	**zero-2**		**zero-1**	**zero**	**zero +1**
pres.pos.long		$1CS_{categ}$	MORPH_pres	VBP^{p}	$.
perf.neg. 4	ga_{MORPH_neg}	$1CS_{categ}$	MORPH_past	VBP	

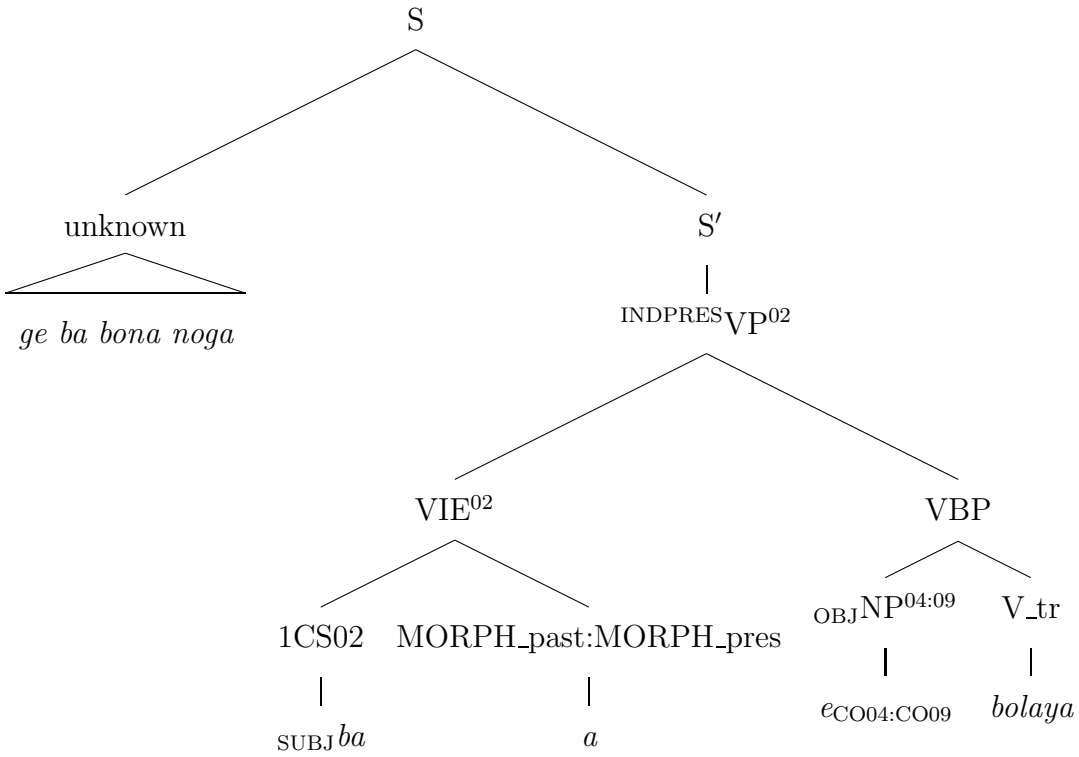

Figure 4.3: A first partial analysis of (*ge ba bona noga*) *ba a e bolaya* '`subj-3rd-cl2 pres obj-3rd-cl4:9` kill'

The token *ba*, however, could also be interpreted as an auxiliary verb[3] – if the following *a* is – as a subject concord – the first element of a VP, as demonstrated by the rules in Table 4.5. If the token should be considered to be a demonstrative concord, *ba* would be (part of) a $_{\text{subj}}$NP and would not be contained in the verbal phrase. In section 3.1 however, we described a general rule for subject-verb agreement in Northern Sotho stating that the subject concord and subject nominal must agree in their noun class. If *ba* represented a demonstrative (of class 2), it would not agree in its noun class with the supposed subject concord *a*, which is from either class 1 or 6. Therefore, the parser can rule out *ba* as CDEM02.

The next token under examination constitutes the unambiguous class 9 noun *noga* '(a) snake', as such it cannot be the subject of the assumed $^{\text{IND}}$VP, because the subject concord

[3]The token *ba* as an auxiliary is to be translated as 'furthermore' or as '(and) so', cf. De Schryver (2007)

Table 4.5: Possible analyses when considering *ba* – auxiliary

	$^{\text{AUX}}$VP		
tense	**elements**		**complement**
pres/perf.	$\mathbf{1CS}_{\text{categ}}$	**V_aux**	$\mathbf{VP}_{\text{categ}}$
future	$\mathbf{1CS}_{\text{categ}}$ $\boldsymbol{tlo/tla}_{\text{MORPH\_fut}}$	**V_aux**	$\mathbf{VP}_{\text{categ}}$
neg	$\boldsymbol{ga}_{\text{MORPH\_neg}}\mathbf{1CS}_{\text{categ}}$	**V_aux**	$\mathbf{VP}_{\text{categ}}$

it appears alongside is from class 2 (ba_{1CS02}). The fact that a noun appears, however, leads to resolving the remaining ambiguity of *a*. MORPH_pres is assigned, because the tag 'past tense morpheme' can only be assigned in the case of a negation morpheme (*ga*) preceding the subject concord *ba*. The noun may indeed also be a topicalized object (the ambiguity concerning the object concord $e_{\text{CO04:CO09}}$ is resolved in this case, cf. Figure 4.4). The second hypothesis entails that *noga* is part of a preceding sentence, not separated by punctuation from the current one under consideration, cf. Figure 4.5. A third and last hypothesis also needs considering, namely that *noga* may be the object of a preceding clause, which may be dependent on the main clause currently being analysed i.e. a situative, relative, consecutive, subjunctive or habitual (cf. Figure 4.6). In this case, it can also be assumed that the object concord of the main clause refers anaphorically to the object noun of the dependent clause, as both are of the same noun class.

In all of the described cases, however, we can already assume *(noga) ba a e bolaya* to constitute a full and complete $^{\text{ind}}\text{VP}^{02}$ of the present tense in the sense of '((a) snake) they kill it' that might agree with a subject nominal of noun class 2 still to appear. The other possibility is that the subject concord *ba* carries the subject function, here no subject nominal is expected anymore.

The verb/pronoun $bona_{\text{PROEMP02:PROPOSS02:V\_tr}}$ is the next item to be examined. Because it is followed by a noun, all three parts of speech categories are theoretically probable. (an NP may contain both, the POS-order PROEMP N and PROPOSS N, cf. section 3.8). However, the following noun is from class 9 and the pronominal word classes of *bona* are only found in

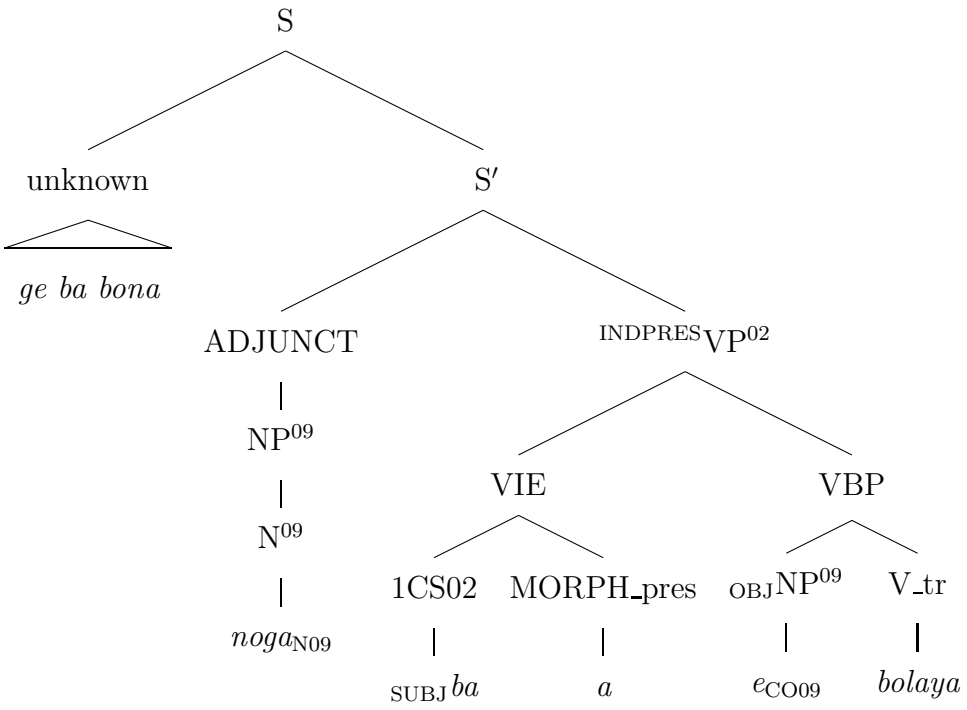

Figure 4.4: Hypothetical partial analysis 1 of (*ge ba bona*) *noga ba a e bolaya* 'snake `subj-3rd-cl2 pres obj-3rd-cl4:9` kill'

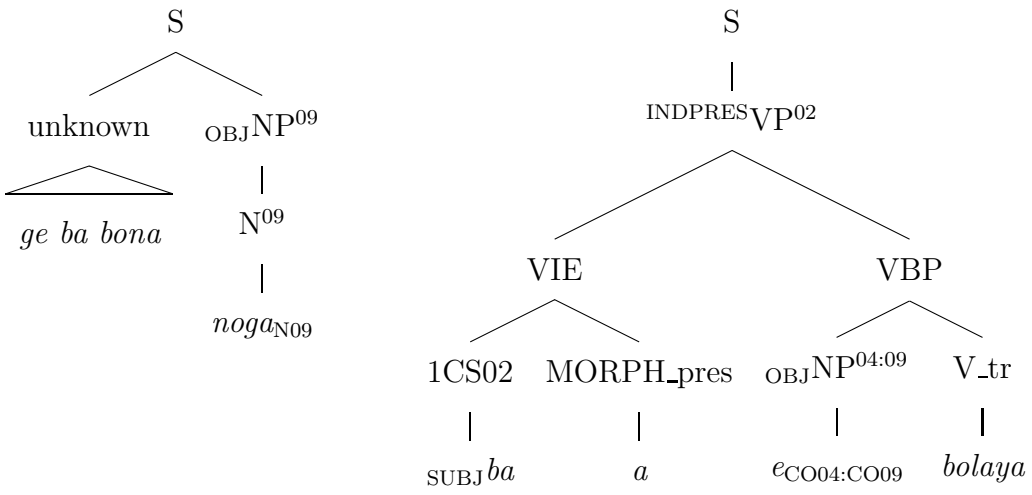

Figure 4.5: Hypothetical partial analysis 2 of (*ge ba bona*) *noga ba a e bolaya* 'snake `subj-3rd-cl2 pres obj-3rd-cl4:9` kill'

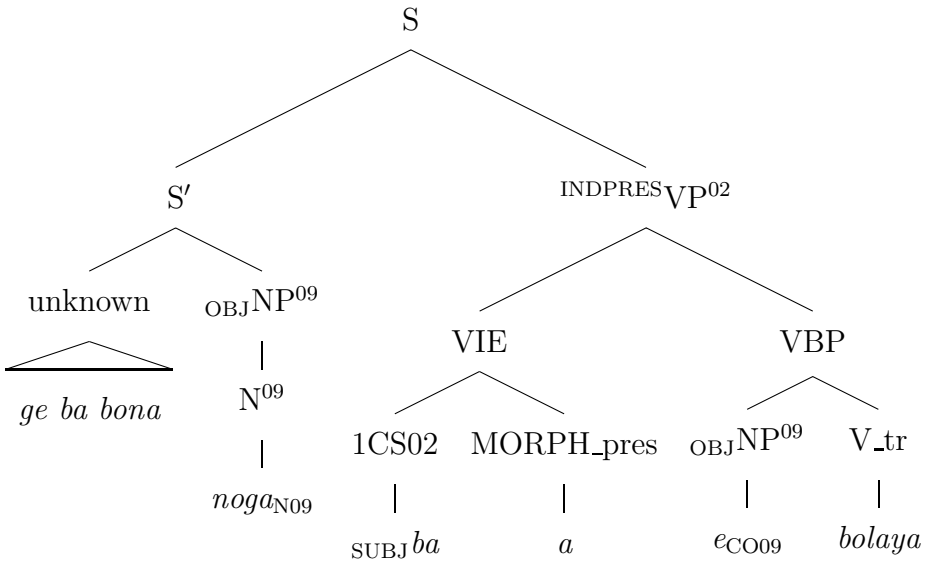

Figure 4.6: Hypothetical partial analysis 3 of (*ge ba bona*) *noga ba a e bolaya* 'snake `subj-3rd-cl2 pres obj-3rd-cl4:9` kill'

class 2. As a noun class agreement between all elements of an NP is mandatory, *bona* must therefore be determined as a transitive verb stem. The VBP-rules stated in paragraph 3.2.3 lead to the assumptions that either *noga* is the object of the transitive verb stem *bona* or that an object concord is to be expected to precede *bona*. Figures 4.7, 4.8, and 4.9 show the respective representations. For the analysis shown in Figure 4.7, note that the preceding element, *ba* is already assumed to be an object concord, as the constellation otherwise would not be legal (see further notes below).

When taking the next preceding token into account, the parser finds the ambiguous token *ba* again, which, as it precedes a transitive verb (which is followed by an object), can be identified by the parser as a subject concord (VCOP is ruled out as a main verb is present; a possessive or demonstrative cannot precede a predicative verb without a subject concord), therefore 1CS02:2CS02:3CS02 are assigned as possible annotations to *ba* and the hypothetical partial analysis shown in Figure 4.7 is abolished.

The structure defined so far is preceded by the last element found, the conjunction *ge*

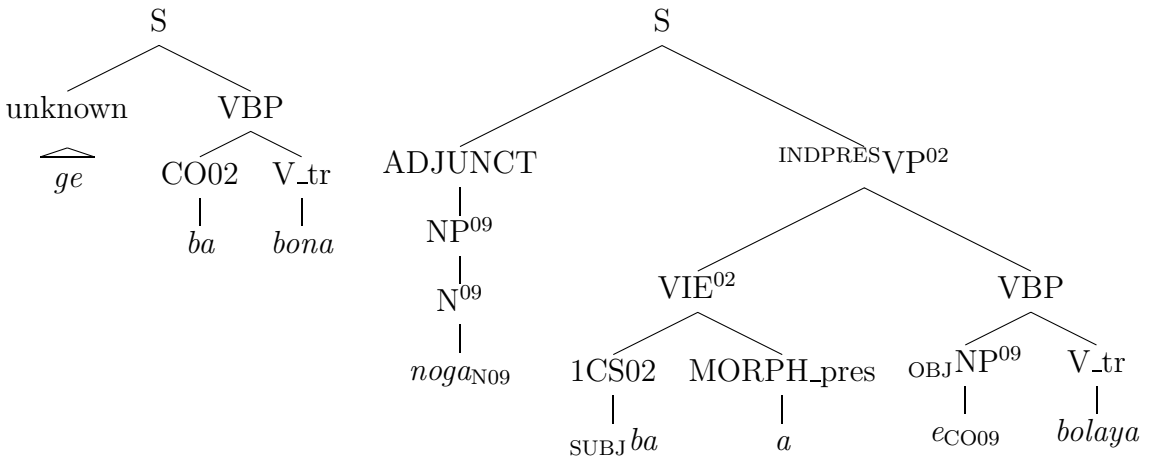

Figure 4.7: Hypothetical partial analysis 1 of (*ge*) *ba bona noga ba a e bolaya* 'obj-3rd-cl2 see (;) snake subj-3rd-cl2 pres obj-3rd-cl4:9 kill'

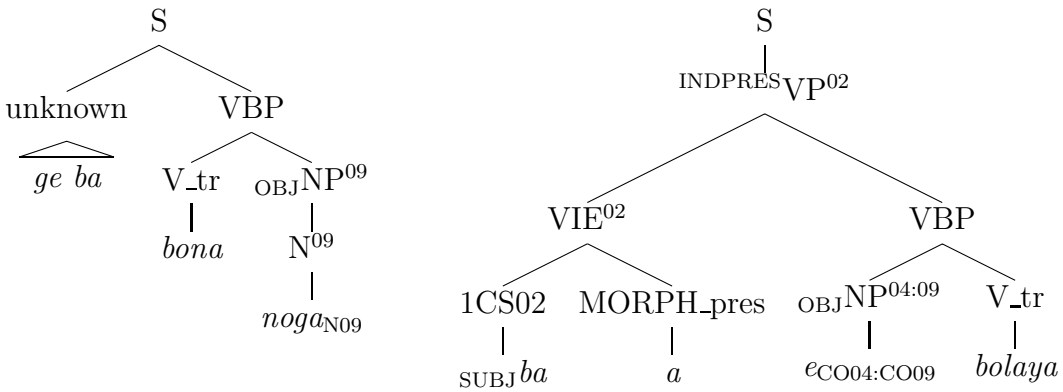

Figure 4.8: Hypothetical partial analysis 2 of (*ge ba*) *bona noga ba a e bolaya* 'see snake (;) subj-3rd-cl2 pres obj-3rd-cl4:9 kill'

'when' indicating that the hypothetical partial analysis 3 is correct and that a dependent clause can be identified (situative mood, cf. paragraph 3.2.6). Consequently, the element *ba* is identified as a subject concord of the second set, 2CS02. The object of the dependent clause, $noga_{\text{N09}}$ could probably be the antecedent for the object concord *e*. Therefore, as both are of the same class, it may be assumed that *e* anaphorically refers to this class 9 object. Figure 4.10 shows the resulting, unequivocal analysis.

This example demonstrated that utilising generalisations based on grammar rules reduces ambiguity significantly during analysis. Such generalisations – at a later stage – may also help to develop a linguistic sequence model of Northern Sotho sentences.

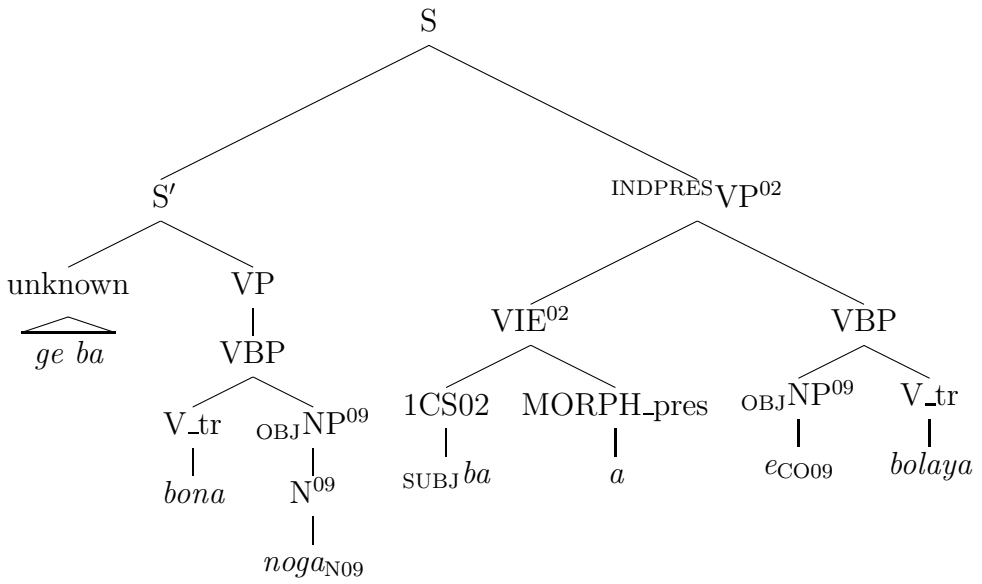

Figure 4.9: Hypothetical partial analysis 3 of (*ge ba*) *bona noga ba a e bolaya* 'see snake `subj-3rd-cl2 pres obj-3rd-cl4:9` kill'

In the following sections, the moods defined will be categorised by their features in order to find more such generalisations. The features considered are the selection of sets of subject concords; the utilization of tense morphemes; the negation constellations utilised by a mood and their respective combinations. Main verbs and copulative constellations are described separately because of the differences in their internal structures.

4.4 A basis for a data category inventory of main verbs

All main verb constellations as described in chapter 3[4], were examined for regular patterns in the distribution of features like verbal endings (abbreviated as Vend), negation morphemes or clusters thereof and subject concords. Table 4.6 illustrates the distribution of verbal endings. Tables 4.7 and 4.8 illustrate the selection of the three sets of subject concords (cf. Table 3.13), and the negated constellations of the described moods for the main verbs. Table 4.9 shows the distribution of morphemes indicating tenses.

[4]For the summaries, consider Tables 3.19 on page 104, 3.2.8 on page 119, and 3.2.11 on page 125.

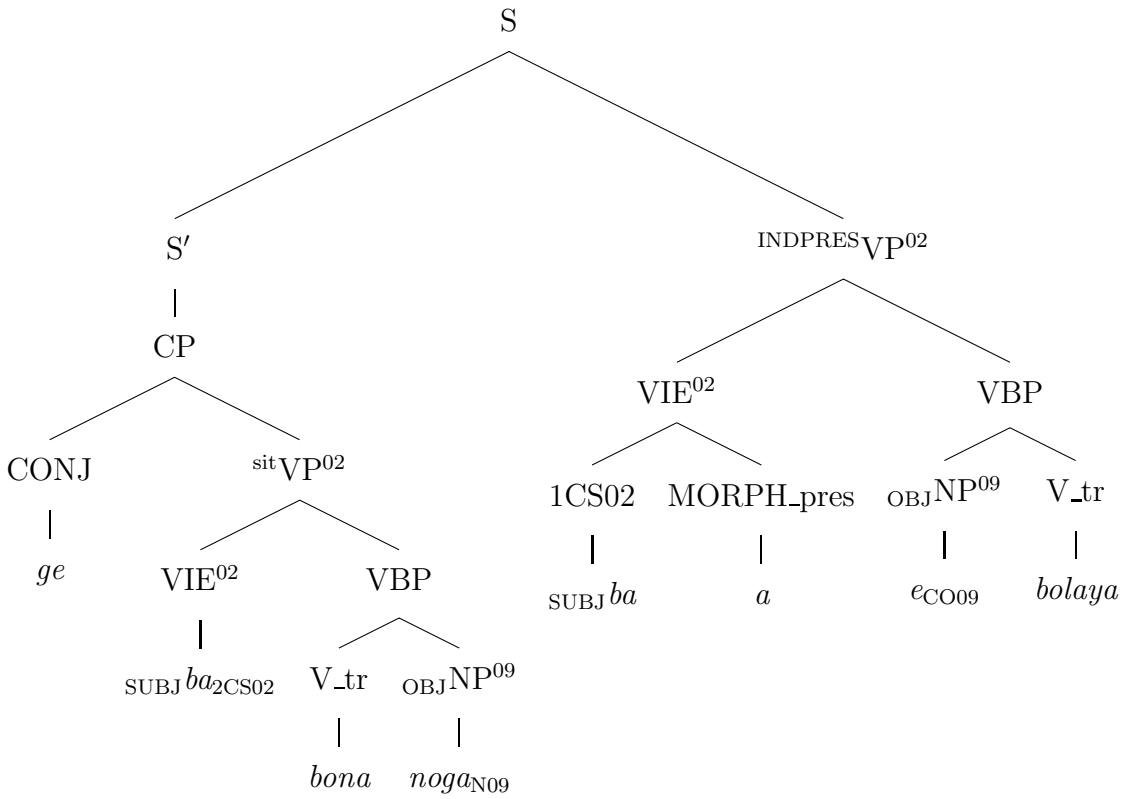

Figure 4.10: Resulting analysis of *ge ba bona noga ba a e bolaya* 'when *subj-3rd-cl2* see snake `subj-3rd-cl2 pres obj-3rd-cl9` kill'

From a right-to-left perspective[5], the Vend *–e* usually appears in negated predicative moods (with the exceptions of the subjunctive/habitual where it occurs in the positive case and of some non-predicative imperative constellations), however, there are a number of negated moods where Vend *–a* is used. Table 4.8 shows (from a left-to-right perspective) that each mood uses its own negation morpheme(s). The morpheme $ga_{\text{MORPH\_neg}}$ exclusively appears in indicative constellations, while *se* is used by the non-predicative (infinitive and imperative) and predicative dependent (consecutive, subjunctive and habitual) moods. However, this morpheme also appears in two alternatives of the negated perfect tense situatives. The negation $ka_{\text{MORPH\_pot}}$ $\text{se}_{\text{MORPH\_neg}}$ is contained in all negated future tense moods indicating that the future tense of Northern Sotho might be closely related to the potential, i.e. that it is of a more aspectual nature than the other tenses[6]. Interestingly, the negation *ka se* usually inhibits the appearance of a future tense morpheme with the one exception of the 2nd form of the relative future tense mood, where both occur. The past-tense morpheme in Table 4.9 only appears in one of the several possible negated perfect tense indicatives. Because neither non-predicative nor the dependent moods appear with tense morphemes, those moods are not mentioned in this table.
Combining the given information on 'verbal ending' with the attribute 'negation morphemes' and their clusters, i.e. simultaneously considering these two properties of VPs from right-to-left and from left-to-right may hint at a logic in their distribution, as $ga_{\text{MORPH\_neg}}$ only appears with Vend *–a* and $ga_{\text{MORPH\_neg}}$ $se_{\text{MORPH\_neg}}$ (and also $se_{\text{MORPH\_neg}}$ when appearing alone) exclusively with Vend *–e*. As said above, the negated future tense moods make use of the negated potential, i.e. the morpheme cluster $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$. The negated future tense forms of indicative and situative appear with Vend *–e*, while the negated future tense of the relative appears with Vend *–a*. The negation $sa_{\text{MORPH\_neg}}$ appears with both verbal endings, it could however be relevant that the combination $sa_{\text{MORPH\_neg}}$+Vend *–a* only appears in the perfect tense while $sa_{\text{MORPH\_neg}}$+Vend *–e* seems to represent present tense forms.

Generalisations on the basis of the subject concords that might appear on the leftmost position of VP, are not easy to determine, as their appearances in the moods do not seem to follow rules. However, the selection of subject concords seems to be solely dependent

[5]As a preferable direction for the horizontal parsing of Northern Sotho text has not yet become evident, the possible directions are explicitly described for each generalisation mentioned in this section.

[6]Note again (cf. paragraph 3.5.2) that the potential itself does not appear in a future tense form.

Table 4.6: Verbal endings of Northern Sotho moods

mood	**verbal ending**					
	−a	***−e***	***−ile***	***−a*** +'rel.'	***−e*** +'rel.'	***−ile+−a*** +'rel.'
imperative.pos	√					
imperative.neg		√				
infinitive.pos	√					
infinitive.neg		√				
ind.pres.pos.long	√					
ind.pres.pos.short	√					
ind.pres.neg		√				
ind.perf.pos			√			
ind.perf.neg.1	√					
ind.perf.neg.2		√				
ind.perf.neg.3	√					
ind.perf.neg.4	√					
ind.fut.pos	√					
ind.fut.neg		√				
sit.pres.pos	√					
sit.pres.neg		√				
sit.perf.pos			√			
sit.perf.neg1	√					
sit.perf.neg2	√					
sit.perf.neg3	√					
sit.fut.pos	√					
sit.fut.neg		√				
rel.pres.pos				√		
rel.pres.neg					√	
rel.perf.pos						√
rel.perf.neg1	√					
rel.perf.neg2		√				
rel.fut.pos1	√					
rel.fut.pos2				√		
rel.fut.neg1	√					
rel.fut.neg2				√		
rel.fut.neg3				√		
consecutive.pos	√					
consecutive.neg		√				
subj./habit.pos		√				
subj./habit.neg		√				

on the moods in which they appear. Set 1, for example (using *o* for class 1 subjects), dominates the indicative (positive) forms, while set 2 (using *a* for class 1 subjects) appears in all moods, either indicating the negated independent forms or the modifying constellations. Concerning Table 4.7, note that some constellations make use of several subject concords; an auxiliary for example is always surrounded by subject concords. For the sake of convenience, here is example (68) again as (79) to demonstrate this issue. In Table 4.7, the position of the subject concord in question appears as an ordinal number.

(79) *wena*$_{\text{PROEMPPERS\_2sg}}$ ***o***$_{\text{1CSPERS\_2sg}}$ *be*$_{\text{V\_aux}}$ ***o***$_{\text{2CSPERS\_2sg}}$ *šomile*$_{\text{V\_itr}}$
`emp-2nd-sg` `subj-2nd-sg` `past` `subj-2nd-sg` work
'you have worked'

In terms of word order, the negation morpheme *ga* and the cluster *ga se* are the only negation constellations that may precede the subject concord while in all other moods it is the subject concord that introduces the verbal phrase. The non-predicative moods do not entail subject concords.

Still, some exceptions (like the subject concord of set one appearing in an auxiliary constellation in the negated perfect tense of the situative (cf. paragraph 3.2.6)) inhibit the formulation of rules solely based on the first or last element(s) of the VIE. So far, one cannot find a relation between the distribution of subject concords and negations or verbal endings, neither.

Table 4.7: Feature selection of Northern Sotho moods: subject concords in slot zero-2

mood	**subject concords**		
	1CS$_{\text{categ}}$	**2CS$_{\text{categ}}$**	**3CS$_{\text{categ}}$**
ind.pres.pos.long	√		
ind.pres.pos.short	√		
ind.pres.neg		√	
ind.perf.pos	√		
ind.perf.neg.1			√
ind.perf.neg.2		√	
ind.perf.neg.3			√
ind.perf.neg.4	√		
ind.fut.pos	√		
ind.fut.neg		√	
sit.pres.pos		√	
sit.pres.neg		√	
sit.perf.pos		√	
sit.perf.neg.1		√(1.)	√(2.)
sit.perf.neg.2	√(2.)	√(1.)	
sit.perf.neg.3		√	
sit.fut.pos		√	
sit.fut.neg		√	
rel.pres.pos		√	
rel.pres.neg		√	
rel.perf.pos		√	
rel.perf.neg.1		√(1.)	√(2.)
rel.perf.neg.2		√(2x)	
rel.fut.pos.1		√	
rel.fut.pos.2		√	
rel.fut.neg.1		√	
rel.fut.neg.2		√	
rel.fut.neg.3		√	
consecutive.pos			√
consecutive.neg			√
subj./habit.pos		√	
subj./habit.neg		√	

Table 4.8: Feature selection of Northern Sotho moods: negation morphemes in slot zero-2

mood	**negation morphemes MORPH_neg**							
	ga	***se***	***ga se***		***ka se***		***sa***	***sego/ seng***
imperative.neg		√						
infinitive.neg		√						
ind.pres.neg	√							
ind.perf.neg.1			√					
ind.perf.neg.2			√					
ind.perf.neg.3	√							
ind.perf.neg.4	√							
ind.fut.neg					√			
sit.pres.neg							√	
sit.perf.neg.1		√						
sit.perf.neg.2		√						
sit.perf.neg.3							√	
sit.fut.neg					√			
rel.pres.neg							√	
rel.perf.neg.1								√
rel.perf.neg.2								√
rel.fut.neg.1					√			
rel.fut.neg.2					√			
rel.fut.neg.3					√			
consecutive.neg		√						
subj./habit.neg		√						

Table 4.9: Feature selection of Northern Sotho moods: slot zero-1

mood	**tense morphemes MORPH_**			
	_pres	**'past'**	**_fut**	**_fut**
	a	***a***	***tlo/tla***	***tlogo/tlago***
ind.pres.pos.long	√			
ind.pres.pos.short				
ind.pres.neg				
ind.perf.pos				
ind.perf.neg.1				
ind.perf.neg.2				
ind.perf.neg.3				
ind.perf.neg.4		√		
ind.fut.pos			√	
ind.fut.neg				
sit.pres.pos				
sit.pres.neg				
sit.perf.pos				
sit.perf.neg.1				
sit.perf.neg.2				
sit.perf.neg.3				
sit.fut.pos			√	
sit.fut.neg				
rel.pres.pos				
rel.pres.neg				
rel.perf.pos				
rel.perf.neg.1				
rel.perf.neg.2				
rel.fut.pos.1				√
rel.fut.pos.2			√	
rel.fut.neg.1				√
rel.fut.neg.2			√	
rel.fut.neg.3				

4.5 A basis for a data category inventory of copulas

The data categories of the copulative constellations as described in section 3.3 can be sorted according to the copula heading them. A second way to gain an overview of the data categories is to sort the cases according to the constellations preceding the copulas. In this section, both illustrations are contained.

4.5.1 A right-to-left perspective

A number of regularities appear when the occurrences of the different copulas appearing on the right hand side of the copulative constellations have been examined. For example, the copula *le* only appears in stative copulatives, and is always preceded by (a combination of) subject concords and – in the perfect tense – also auxiliaries. Only descriptive constellations contain class-specific concords, while identifying constellations are preceded by CSPCSN (either CSNEUT, i.e. the neutral, or CSPERS, i.e. the personal subject concords[7]). Tables 4.10 (containing *ba, eba, bile*), 4.11 (containing *be*), 4.12 (containing *bago, bego, bilego*), 4.13 (containing *le, lego*), 4.14 (containing the subject concords appearing as copulas $1VCOP_{categ}$ (set 1) or $2VCOP_{categ}$ (set 2)), 4.15 (containing *na, (e)na, nago*), and 4.16 (containing *se, sego/seng*) show summaries of the moods together with the features of the respective copulas, i.e. the element clusters preceding them.

A number of constellations are inherently ambiguous, however, as indicated in the section on copulatives (for example in paragraph 3.3.3), the complement of the copulative may indicate the correct analysis. Moreover, some of the constellations are introduced by a specific conjunction or by demonstrative concords. Such and other phrase-external elements also support the disambiguation process.

[7]The personal subject concords and the neutral subject concord were grouped as 'CSPCSN' in 3.31 on page 131, and we will use this and the other abbreviations again in the following tables.

Table 4.10: Distribution of *ba, eba, bile*

VCOP	**preceding elements**	**copulative**				
ba	CSPCSN	identifying	dynamic	indicative situative	pres.	pos.
	CSPCSN *sa*$_{\text{MORPH\_neg}}$	identifying	dynamic	situative	perf.	neg.
	CSPCSN *tlo/tla*$_{\text{MORPH\_fut}}$	identifying	dynamic	indicative situative	fut.	pos.
	CSPCSN *tlogo/tlago*$_{\text{MORPH\_fut}}$	identifying	dynamic	relative	fut.	pos.2
	1CS$_{\text{categ}}$	descriptive associative	dynamic	indicative	pres.	pos.
	1CS$_{\text{categ}}$ *tlo/tla*$_{\text{MORPH\_fut}}$	descriptive associative	dynamic	indicative	fut.	pos.
	2CS$_{\text{categ}}$	descriptive	dynamic	situative	pres.	pos.
	2CS$_{\text{categ}}$ *sa*$_{\text{MORPH\_neg}}$	descriptive associative	dynamic	situative	perf.	neg.
	2CS$_{\text{categ}}$ *tlo/tla*$_{\text{MORPH\_fut}}$	descriptive associative	dynamic	situative	fut.	pos.
	2CS$_{\text{categ}}$ *tlogo/tlago*$_{\text{MORPH\_fut}}$	descriptive	dynamic	relative	fut.	pos.2
	3CS$_{\text{categ}}$	identifying descriptive	dynamic	consecutive	–	pos.
	go$_{\text{MORPH\_cp15}}$			infinitive	–	pos.
eba	2CS$_{\text{categ}}$	associative	dynamic	situative	pres.	pos.
	–			imperative	–	pos.
bile	CSPCSN	identifying	dynamic	indicative situative	perf.	pos.
	1CS$_{\text{categ}}$	descriptive associative	dynamic	indicative	perf.	pos.
	2CS$_{\text{categ}}$	descriptive associative	dynamic	situative	perf.	pos.

Table 4.11: Distribution of *be*

VCOP	preceding elements	copulative				
be	CSPCSN	identifying	dynamic	subjunctive habitual	–	pos.
	CSPCSN $se_{\text{MORPH\_neg}}$	identifying	dynamic	subjunctive habitual	–	neg.
	CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	situative	pres.	neg.
	$ga_{\text{MORPH\_neg}}$ CSPCSN	identifying	dynamic	indicative	pres.	neg.
	CSPCSN $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	identifying	dynamic	indicative situative	fut.	neg.
	$2CS_{\text{categ}}$	descriptive	dynamic	subjunctive habitual	–	pos.
	$2CS_{\text{categ}}$ $se_{\text{MORPH\_neg}}$	descriptive	dynamic	subjunctive habitual	–	neg.
	$2CS_{\text{categ}}$ $sa_{\text{MORPH\_neg}}$	descriptive associative	dynamic	situative	pres	neg.
	$ga_{\text{MORPH\_neg}}$ $2CS_{\text{categ}}$	descriptive associative	dynamic	indicative	pres.	neg.
	$2CS_{\text{categ}}$ $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	descriptive associative	dynamic	indicative situative indicative situative	fut.	neg.
	$go_{\text{MORPH\_cp15}}$ $se_{\text{MORPH\_neg}}$			infinitive	–	neg.
	$se_{\text{MORPH\_neg}}$			imperative	–	neg.

Table 4.12: Distribution of *bago, bego, bilego*

VCOP	preceding elements	copulative				
bago	CSPCSN	identifying	dynamic	relative	pres.	pos.
	CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	relative	perf.	neg.
	CSPCSN $tlo/tla_{\text{MORPH\_fut}}$	identifying	dynamic	relative	fut.	pos.
	$2\text{CS}_{\text{categ}}$	descriptive associative	dynamic	relative	pres.	pos.
	$2\text{CS}_{\text{categ}}$ $sa_{\text{MORPH\_neg}}$	descriptive associative	dynamic	relative	perf.	neg.
	$2\text{CS}_{\text{categ}}$ $tlo/tla_{\text{MORPH\_fut}}$	descriptive associative	dynamic	relative	fut.	pos.
bego	CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	relative	pres.	neg.
	CSPCSN $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	identifying	dynamic	relative	fut.	neg.
	$2\text{CS}_{\text{categ}}$ $sa_{\text{MORPH\_neg}}$	descriptive associative	dynamic	relative	pres.	neg.
	$2\text{CS}_{\text{categ}}$ $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	descriptive associative	dynamic	relative	fut.	neg.
bilego	CSPCSN	identifying	dynamic	relative	perf.	pos.
	$2\text{CS}_{\text{categ}}$	descriptive associative	dynamic	relative	perf.	pos.

Table 4.13: Distribution of *le, lego*

VCOP	preceding elements	copulative				
le	CSPCSN	identifying	stative	situative	pres.	pos.
	CSPCSN $be_{\text{V\_aux}}$ CSPCSN	identifying	stative	indicative situative	perf.	pos.
	CSPCSN $bego_{\text{V\_aux}}$ CSPCSN	identifying	stative	relative	perf.	pos.
	$1\text{CS}_{\text{categ}}$ $be_{\text{V\_aux}}$ $2\text{CS}_{\text{categ}}$	descriptive	stative	indicative	perf.	pos.
	$2\text{CS}_{\text{categ}}$	descriptive	stative	situative	pres.	pos.
	$2\text{CS}_{\text{categ}}$ $be_{\text{V\_aux}}$ $2\text{CS}_{\text{categ}}$	descriptive	stative	situative	perf.	pos.
	$2\text{CS}_{\text{categ}}$ $bego_{\text{V\_aux}}$ $2\text{CS}_{\text{categ}}$	descriptive	stative	relative	perf.	pos.
lego	CSPCSN	identifying	stative	relative	pres.	pos.
	$2\text{CS}_{\text{categ}}$	descriptive	stative	relative	pres.	pos.

Table 4.14: Distribution of *na, (e)na, nago*

VCOP	**preceding elements**	**copulative**				
na	$1CS_{categ}$	associative	stative	indicative	pres.	pos.
	$2CS_{categ}$ se_{MORPH_neg}	associative	stative	situative	pres.	pos.
	ga_{MORPH_neg} $2CS_{categ}$	associative	stative	indicative	pres.	neg.
	$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$	associative	stative	relative	perf.	pos.
	$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$ se_{MORPH_neg}	associative	stative	relative	perf.	neg.
(e)na	$1CS_{categ}$ be_{V_aux} $2CS_{categ}$	associative	stative	indicative	perf.	pos.
	$1CS_{categ}$ be_{V_aux} $2CS_{categ}$ se_{MORPH_neg}	associative	stative	indicative	perf.	neg.
	$2CS_{categ}$	associative	stative	situative	pres.	pos.
	$2CS_{categ}$ be_{V_aux} $2CS_{categ}$	associative	stative	situative	perf.	pos.
	$2CS_{categ}$ be_{V_aux} $2CS_{categ}$ se_{MORPH_neg}	associative	stative	situative	perf.	neg.
nago	$2CS_{categ}$	associative	stative	relative	pres.	pos.
	$2CS_{categ}$ se_{MORPH_neg}	associative	stative	relative	pres.	neg.

Table 4.15: Distribution of $VCOP_{categ}$

VCOP	**preceding elements**	**copulative**				
1VCOP$_{categ}$	–	descriptive	stative	indicative	pres.	pos.
2VCOP$_{categ}$	ga_{MORPH_neg}	descriptive	stative	indicative	pres.	neg.
VCOP_pers	–	identifying	stative	indicative	pres.	pos.
VCOP_pers	ga_{MORPH_neg}	identifying	stative	indicative	pres.	neg.

Table 4.16: Distribution of *se, sego/seng*

VCOP	**preceding elements**	**copulative**				
se$_{\text{VCOP\_neg}}$	CSPCSN	identifying descriptive	stative	situative	pres.	neg.
	CSPCSN *be*$_{\text{V\_aux}}$ CSPCSN	identifying	stative	indicative situative	perf.	neg.
	CSPCSN *bego*$_{\text{V\_aux}}$ CSPCSN	identifying	stative	relative	perf.	neg.
	1CS$_{\text{categ}}$ *be*$_{\text{V\_aux}}$ 2CS$_{\text{categ}}$	descriptive	stative	indicative	perf.	neg.
	2CS$_{\text{categ}}$ *be*$_{\text{V\_aux}}$ 2CS$_{\text{categ}}$	descriptive	stative	situative	perf.	neg.
	2CS$_{\text{categ}}$ *bego*$_{\text{V\_aux}}$ 2CS$_{\text{categ}}$	descriptive	stative	relative	perf.	neg.
	ga$_{\text{MORPH\_neg}}$	identifying	stative	pres.	neg.	
sego/seng	CSPCSN	identifying	stative	relative	pres.	neg.
VCOP_neg	2CS$_{\text{categ}}$	descriptive	stative	relative	pres.	neg.

Table 4.17: Distribution of constellations solely containing copulas

preceding elements	**copulative**					**VCOP**
–	identifying	stative	indicative	pres.	pos.	VCOP_pers
–	descriptive	stative	indicative	pres.	pos.	$1VCOP_{categ}$
–			imperative	–	pos.	*eba*

4.5.2 A left-to-right perspective

Sorting the copulatives according to the element constellations preceding them (as in Tables 4.17 to 4.29) shows that there are a number of generalisations possible on the basis of the left-hand side of the copula; an observation that could support a left-to-right parsing strategy.

We begin with Table 4.17, illustrating constellations that solely contain copulas. Tokens named as VCOP are actually subject concords and their identification as copulas is only possible when taking the context on their righthand side into account: if no verb stem appears, the subject concord may be assumed to be a copulative.

The subject concord group CSPCSN (consisting of the set of all $CSPERS_{categ}$ and of CSNEUT) never appears as an element of the associative copulatives (cf. Table 4.18) while the combination CSPCSN and future tense morpheme moreover never appears in an indicative (cf. Table 4.19). On the other hand, if CSPCSN is followed by either the negation morpheme *se* or *sa* or the cluster *ka se* (as in Tables 4.20 and 4.21), the constellation is clearly an identifying dynamic copulative (but no indicative).

Table 4.22 shows that CSPCSN constellations containing auxiliaries always appear in the perfect tense of an identifying stative copulative. These constellations appear with the copula *le* in the positive and are all negated by the negated copula se_{MORPH_neg}. All VIEs beginning with a subject concord of the first set are indeed indicatives, however, they do not present identifying copulatives, as shown in Table 4.23. The latter is also true for the VIEs beginning with a subject concord of the second set, however, all modifying and the dependent moods are found in Table 4.24.

Table 4.18: Distribution of CSPERS and CSNEUT (CSPCSN)

preceding elements	**copulative**					**VCOP**
CSPCSN	identifying	stative	situative	pres.	pos.	*le*
CSPCSN	identifying	stative	situative	pres.	neg.	$se_{\text{VCOP\_neg}}$
CSPCSN	identifying	stative	relative	pres.	pos.	*lego*
CSPCSN	identifying	stative	relative	pres.	neg.	$sego/seng_{\text{VCOP\_neg}}$
CSPCSN	identifying	dynamic	indicative	pres.	pos.	*ba*
CSPCSN	identifying	dynamic	indicative	perf.	pos.	*bile*
CSPCSN	identifying	dynamic	situative	pres.	pos.	*ba*
CSPCSN	identifying	dynamic	situative	perf.	pos.	*bile*
CSPCSN	identifying	dynamic	relative	pres.	pos.	*bago*
CSPCSN	identifying	dynamic	relative	perf.	pos.	*bilego*
CSPCSN	identifying	dynamic	subjunctive	–	pos.	*be*
CSPCSN	identifying	dynamic	habitual	–	pos.	*be*
CSPCSN	descriptive	stative	situative	pres.	neg.	$se_{\text{VCOP\_neg}}$

Table 4.19: Distribution of CSPCSN and future tense morphemes

preceding elements	**copulative**					**VCOP**
CSPCSN $tlo/tla_{\text{MORPH\_fut}}$	identifying	dynamic	indicative	fut.	pos.	*ba*
CSPCSN $tlo/tla_{\text{MORPH\_fut}}$	identifying	dynamic	situative	fut.	pos.	*ba*
CSPCSN $tlo/tla_{\text{MORPH\_fut}}$	identifying	dynamic	relative	fut.	pos.	*bago*
CSPCSN $tlogo/tlago_{\text{MORPH\_fut}}$	identifying	dynamic	relative	fut.	pos.2	*ba*

Table 4.20: Distribution of CSPCSN followed by a negation morpheme

preceding elements	**copulative**					**VCOP**
CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	situative	pres.	neg.	*be*
CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	situative	perf.	neg.	*ba*
CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	relative	pres.	neg.	*bego*
CSPCSN $sa_{\text{MORPH\_neg}}$	identifying	dynamic	relative	perf.	neg.	*bago*
CSPCSN $se_{\text{MORPH\_neg}}$	identifying	dynamic	subjunctive	–	neg.	*be*
CSPCSN $se_{\text{MORPH\_neg}}$	identifying	dynamic	habitual	–	neg.	*be*

Table 4.21: Distribution of CSPCSN followed by *ka se*

preceding elements	**copulative**					**VCOP**
CSPCSN $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	identifying	dynamic	indicative	fut.	neg.	*be*
CSPCSN $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	identifying	dynamic	situative	fut.	neg.	*be*
CSPCSN $ka_{\text{MORPH\_pot}}$ $se_{\text{MORPH\_neg}}$	identifying	dynamic	relative	fut.	neg.	*bego*

bottomline

Table 4.22: Distribution of CSPCSN followed by auxiliary constellations

preceding elements	**copulative**					**VCOP**
CSPCSN $be_{\text{V\_aux}}$ CSPCSN	identifying	stative	indicative	perf.	pos.	*le*
CSPCSN $be_{\text{V\_aux}}$ CSPCSN	identifying	stative	indicative	perf.	neg.	$se_{\text{VCOP\_neg}}$
CSPCSN $be_{\text{V\_aux}}$ CSPCSN	identifying	stative	situative	perf.	pos.	*le*
CSPCSN $be_{\text{V\_aux}}$ CSPCSN	identifying	stative	situative	perf.	neg.	$se_{\text{VCOP\_neg}}$
CSPCSN $bego_{\text{V\_aux}}$ CSPCSN	identifying	stative	relative	perf.	pos.	*le*
CSPCSN $bego_{\text{V\_aux}}$ CSPCSN	identifying	stative	relative	perf.	neg.	$se_{\text{VCOP\_neg}}$

Table 4.23: Distribution of constellations beginning with $1CS_{\text{categ}}$

preceding elements	**copulative**					**VCOP**
$1CS_{\text{categ}}$	descriptive	dynamic	indicative	pres.	pos.	*ba*
$1CS_{\text{categ}}$	descriptive	dynamic	indicative	perf.	pos.	*bile*
$1CS_{\text{categ}}$	associative	dynamic	indicative	pres.	pos.	*ba*
$1CS_{\text{categ}}$	associative	dynamic	indicative	perf.	pos.	*bile*
$1CS_{\text{categ}}$	associative	stative	indicative	pres.	pos.	*na*
$1CS_{\text{categ}}$ $tlo/tla_{\text{MORPH\_fut}}$	descriptive	dynamic	indicative	fut.	pos.	*ba*
$1CS_{\text{categ}}$ $tlo/tla_{\text{MORPH\_fut}}$	associative	dynamic	indicative	fut.	pos.	*ba*
$1CS_{\text{categ}}$ $be_{\text{V\_aux}}$ $2CS_{\text{categ}}$	descriptive	stative	indicative	perf.	pos.	*le*
$1CS_{\text{categ}}$ $be_{\text{V\_aux}}$ $2CS_{\text{categ}}$	descriptive	stative	indicative	perf.	neg	$se_{\text{VCOP\_neg}}$
$1CS_{\text{categ}}$ $be_{\text{V\_aux}}$ $2CS_{\text{categ}}$	associative	stative	indicative	perf.	pos.	*(e)na*
$1CS_{\text{categ}}$ $be_{\text{V\_aux}}$ $2CS_{\text{categ}}$ $se_{\text{MORPH\_neg}}$	associative	stative	indicative	perf.	neg.	*(e)na*

Table 4.24: Distribution of $2CS_{\text{categ}}$

preceding elements	**copulative**					**VCOP**
$2CS_{\text{categ}}$	descriptive	stative	situative	pres.	pos.	*le*
$2CS_{\text{categ}}$	descriptive	stative	relative	pres.	pos.	*lego*
$2CS_{\text{categ}}$	descriptive	stative	relative	pres.	neg	$sego/seng_{\text{VCOP\_neg}}$
$2CS_{\text{categ}}$	descriptive	dynamic	situative	pres.	pos.	*ba*
$2CS_{\text{categ}}$	descriptive	dynamic	situative	perf.	pos.	*bile*
$2CS_{\text{categ}}$	descriptive	dynamic	relative	pres.	pos.	*bago*
$2CS_{\text{categ}}$	descriptive	dynamic	relative	perf.	pos.	*bilego*
$2CS_{\text{categ}}$	descriptive	dynamic	subjunctive	–	pos.	*be*
$2CS_{\text{categ}}$	descriptive	dynamic	habitual	–	pos.	*be*
$2CS_{\text{categ}}$	associative	stative	situative	pres.	pos.	*(e)na*
$2CS_{\text{categ}}$	associative	stative	relative	pres.	pos.	*nago*
$2CS_{\text{categ}}$	associative	dynamic	situative	pres.	pos.	*eba*
$2CS_{\text{categ}}$	associative	dynamic	situative	perf.	pos.	*bile*
$2CS_{\text{categ}}$	associative	dynamic	relative	pres.	pos.	*bago*
$2CS_{\text{categ}}$	associative	dynamic	relative	perf.	pos.	*bilego*

Table 4.25: Distribution of $2CS_{categ}$ followed by future tense morphemes

preceding elements		**copulative**					**VCOP**
$2CS_{categ}$ *tlo/tla*$_{MORPH\_fut}$	descriptive	dynamic	situative	fut.	pos.	*ba*
$2CS_{categ}$ *tlo/tla*$_{MORPH\_fut}$	descriptive	dynamic	relative	fut.	pos.	*bago*
$2CS_{categ}$ *tlogo/tlago*$_{MORPH\_fut}$	descriptive	dynamic	relative	fut.	pos.2	*ba*
$2CS_{categ}$ *tlo/tla*$_{MORPH\_fut}$	associative	dynamic	situative	fut.	pos.	*ba*
$2CS_{categ}$ *tlo/tla*$_{MORPH\_fut}$	associative	dynamic	relative	fut.	pos.	*bago*

As a counterpart to Table 4.19 (containing identifying copulatives), future tense morphemes following a subject concord of the second set indicate a descriptive or associative copulative, as illustrated in Table 4.25. In both cases (CSPCSN and $2CS_{categ}$), the copula *ba*$_{VCOP}$ appears in the indicative/situative, its relative form *bago* in the relative moods. The same stands for the copulatives containing a subject concord of the second set followed by the negation morphemes *sa*, *se*, or the cluster *ka se*, as in Table 4.26. Table 4.22, where the VIE was introduced by CSPCSN (followed by an auxiliary constellation), only contained identifying stative copulatives. Table 4.27 shows that such copulatives introduced by $2CS_{categ}$ indicate the descriptive and associative cases. There are only two cases beginning with the subject concord of the third set, $3CS_{categ}$; both identify positive dynamic consecutive moods, as in Table 4.28.

Lastly, Table 4.29 shows the cases where a copulative begins with the infinitive or negation morphemes. Of these, the infinitive and the imperative are clear-cut cases, because no other constellation can begin with these morphemes. The other cases contain all three main categories (identifying, descriptive, and associative) and both sub-categories (stative and dynamic) of copulatives. However, each of them are negated copulatives in the indicative mood and in the present tense.

Table 4.26: Distribution of $2CS_{categ}$ followed by negation morphemes and negation clusters

preceding elements	**copulative**					**VCOP**
$2CS_{categ}$ sa_{MORPH_neg}	descriptive	dynamic	situative	pres	neg.	*be*
$2CS_{categ}$ sa_{MORPH_neg}	descriptive	dynamic	situative	perf.	neg.	*ba*
$2CS_{categ}$ sa_{MORPH_neg}	descriptive	dynamic	relative	pres.	neg.	*bego*
$2CS_{categ}$ sa_{MORPH_neg}	descriptive	dynamic	relative	perf.	neg.	*bago*
$2CS_{categ}$ sa_{MORPH_neg}	associative	dynamic	situative	pres	neg.	*be*
$2CS_{categ}$ sa_{MORPH_neg}	associative	dynamic	situative	perf.	neg.	*ba*
$2CS_{categ}$ sa_{MORPH_neg}	associative	dynamic	relative	pres.	neg.	*bego*
$2CS_{categ}$ sa_{MORPH_neg}	associative	dynamic	relative	perf.	neg.	*bago*
$2CS_{categ}$ se_{MORPH_neg}	descriptive	dynamic	subjunctive	–	neg.	*be*
$2CS_{categ}$ se_{MORPH_neg}	descriptive	dynamic	habitual	–	neg.	*be*
$2CS_{categ}$ se_{MORPH_neg}	associative	stative	situative	pres.	pos.	*na*
$2CS_{categ}$ se_{MORPH_neg}	associative	stative	relative	pres.	neg.	*nago*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	descriptive	dynamic	indicative	fut.	neg.	*be*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	descriptive	dynamic	situative	fut.	neg.	*be*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	associative	dynamic	indicative	fut.	neg.	*be*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	associative	dynamic	situative	fut.	neg.	*be*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	descriptive	dynamic	relative	fut.	neg.	*bego*
$2CS_{categ}$ ka_{MORPH_pot} se_{MORPH_neg}	associative	dynamic	relative	fut.	neg.	*bego*

Table 4.27: Distribution of $2CS_{categ}$ followed by auxiliary constellations

preceding elements	**copulative**					**VCOP**
$2CS_{categ}$ be_{V_aux} $2CS_{categ}$	descriptive	stative	situative	perf.	pos.	*le*
$2CS_{categ}$ be_{V_aux} $2CS_{categ}$	descriptive	stative	situative	perf.	neg	se_{VCOP_neg}
$2CS_{categ}$ be_{V_aux} $2CS_{categ}$	associative	stative	situative	perf.	pos.	*(e)na*
$2CS_{categ}$ be_{V_aux} $2CS_{categ}$ se_{MORPH_neg}	associative	stative	situative	perf.	neg.	*(e)na*
$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$	descriptive	stative	relative	perf.	pos.	*le*
$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$	descriptive	stative	relative	perf.	neg	se_{VCOP_neg}
$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$	associative	stative	relative	perf.	pos.	*na*
$2CS_{categ}$ $bego_{V\_aux}$ $2CS_{categ}$ se_{MORPH_neg}	associative	stative	relative	perf.	neg.	*na*

Table 4.28: Distribution of constellations beginning with $3CS_{categ}$

preceding elements	**copulative**					**VCOP**
$3CS_{categ}$	identifying	dynamic	consecutive	–	pos.	*ba*
$3CS_{categ}$	descriptive	dynamic	consecutive	–	pos.	*ba*

Table 4.29: Constellations beginning with (negation) morphemes

preceding elements			**copulative**			**VCOP**
$go_{\text{MORPH\_cp15}}$			infinitive	–	pos.	*ba*
$go_{\text{MORPH\_cp15}}$ $se_{\text{MORPH\_neg}}$			infinitive	–	neg.	*be*
$se_{\text{MORPH\_neg}}$			imperative	–	neg.	*be*
$ga_{\text{MORPH\_neg}}$	identifying	stative	indicative	pres.	neg.	VCOP_pers
$ga_{\text{MORPH\_neg}}$	identifying	stative	indicative	pres.	neg.	$se_{\text{VCOP\_neg}}$
$ga_{\text{MORPH\_neg}}$	descriptive	stative	indicative	pres.	neg.	$2\text{VCOP}_{\text{categ}}$
$ga_{\text{MORPH\_neg}}$ CSPCSN	identifying	dynamic	indicative	pres.	neg.	*be*
$ga_{\text{MORPH\_neg}}$ $2\text{CS}_{\text{categ}}$	descriptive	dynamic	indicative	pres.	neg.	*be*
$ga_{\text{MORPH\_neg}}$ $2\text{CS}_{\text{categ}}$	associative	dynamic	indicative	pres.	neg.	*be*
$ga_{\text{MORPH\_neg}}$ $2\text{CS}_{\text{categ}}$	associative	stative	indicative	pres.	neg.	*na*

4.6 Conclusions

This chapter has shown that there are indeed some generalisations possible when examining elements or element clusters that are part of a VP. Unlike the main verb constellations, where only few signalling elements or element groups were found (e.g. the negation *ga se* solely appearing in the indicative mood), analyses of copulatives are supported by a number of indicating features, e.g. the subject concords of the second set ($2CS_{categ}$) followed by the negation morphemes *sa* or *se* that never appear in the indicative. A left-to-right parsing strategy of Northern Sotho verbal constellations might accelerate the disambiguation process.

In chapter 3, we defined the morphosyntactic description of - amongst other elements - Northern Sotho verbal phrases described in literature. From there, a hierarchical system of POS constellations was developed that describes these verbal categories in a top-down manner: sets of Northern Sotho verbal constellations are distinguished for each mood and tense, and a number of overviews are provided each summarising them (one example of such a summary is Table 3.26 on page 121). A partial implementation of Northern Sotho constellations according to these grammar fragment definitions of chapter 3 will be described in chapter 5.