

A computational implementation of the Northern Sotho infinitive

Gertrud Faaß

Institute for Information Science and Natural Language Processing,
University of Hildesheim, Hildesheim, Germany and
Department of African Languages, University of Pretoria,
Pretoria 0002, South Africa

D.J. Prinsloo

Department of African Languages, University of Pretoria,
Pretoria 0002, South Africa
danie.prinsloo@up.ac.za

The aim of this article is to describe the infinitive in Northern Sotho based on corpus data and the respective literature; so far, all share the same view: The infinitive is a noun (of class 15) and a verb at the same time – ‘it manifests both nominal as well as verbal features’ (Poulos & Louwrens, 1994:42). When implementing these constellations in a parser, however, a new perspective is found: to achieve its successful implementation, the infinitive must be defined as a verb on the one hand and as a noun of class 15 on the other, derived from this verb through nominalization (transposition). Instead of a subject concord, the verb stem in the infinitive is preceded by the respective class prefix.

Introduction

Traditional grammars and linguistic analysis of the infinitive in Northern Sotho reveal that linguists agree that the infinitive has the characteristics of both nouns and verbs. Unlike the other noun classes, the stem of the infinitive noun is not a nominal but a verbal stem and, unlike verbs, infinitives contain a noun class prefix (class prefix of noun class 15, henceforth ‘cp15’) and cannot include a subject concord. The combination of a noun class prefix and a verb stem in the infinitive proves the dual characteristics of the Northern Sotho infinitive; it is nominal as well as verbal from the outset, and this fact guides the descriptions of the infinitive in traditional grammars.

Linguists therefore describe the infinitive in comparison to other nouns and verbs and sensibly highlight the characteristics shared with other nouns and verbs and the instances where infinitives differ from these categories.

For this study, the focus will be on a description of the different infinitive constructions in terms of their morphology, syntax and semantics, although no attempt will be made to formally separate these categories in the description. The aim of the discussion in the next section is the subsequent build-up of a register of infinitives, which will be referred to as ‘infinitive constellations’ of which a selection will be made for computational descriptions in the subsequent sections.

In the second section, the infinitive is positioned and described within the two major categories of *verb* and *noun*. The section aims at an overall description in terms of the morphology, syntax and semantics of the infinitive. This is done mainly in terms of Lombard, Van Wyk and Mokgokong (1985), Louwrens (1991), Poulos and Louwrens (1994), Van Wyk, Groenewald, Prinsloo, Kock and Taljard (1992) and Ziervogel (1976), as a prerequisite for a subsequent register of the infinitives to be considered for computational description in the third section. In the fourth section, a partial implementation in the framework of lexical functional grammar is demonstrated.

(2c) Double transitive

<i>ke</i>	<i>nyaka</i>	<i>go</i>	<i>rekela</i>	<i>bana</i>	<i>malekere</i>
subj-1 st -p.sg	want	cp15	buy-appl	children	sweets
'I want to buy sweets for the children'					

As in all verbs, infinitives can include any of the object concords of the other nominal classes when the verb stem is transitive and the object concord occurs directly in front of the verb stem. Consider (3) as an infinitive containing the object concord of class 1.

(3)	<i>re</i>	<i>swanetše</i>	<i>go</i>	mo	<i>thuša</i>
	subj-1 st -p.pl.	must	cp15	obj-01	help
	'we must help him'				

Example (4) demonstrates that the reflexive prefix *i-* can also occur as part of the infinitive verb stem. See (6) for the occurrence of an adverb.

(4)	<i>mo</i>	<i>o</i>	<i>swanetše</i>	<i>go</i>	<i>itshwara</i>	<i>gabotse</i>
	here	subj-2 nd -p.-sg	must	cp15	behave-refl.	well
	'here you must behave yourself well'					

Like in all verbs, the infinitive can be negated by using the morphemes *se-* and (rarely) *sa-* with a verbal ending *-e* as in (5).

(5)	<i>go</i>	<i>se</i>	<i>thuše</i>	<i>moruti</i>
	<i>go</i>	<i>sa</i>	<i>thuše</i>	<i>moruti</i>
	cp15	neg	help	reverend
	'to not help the reverend'			

Infinitives can also be modified by adverbs as in (6).

(6)	<i>go</i>	<i>šoma</i>	<i>gabotse</i>
	cp15	work	well
	'to work well'		

Infinitives can occur in the position of auxiliary verbs in examples such as (7)¹

(7a)	<i>o</i>	<i>rata</i>	<i>go</i>	<i>hlwa</i>	<i>a</i>	<i>goga</i>	<i>peipi</i>
	subj-01	like	cp15	continue	subj-01	smoke	pipe
	'he likes to continually smoke pipe'						

(7b)	<i>Mna.</i>	<i>Kgapšane</i>	<i>ga se</i>	<i>a</i>	<i>hloka</i>	<i>go</i>	<i>hlwa</i>
	Mr.	Kgapšane	neg	subj-01	forget	cp15	continue
	<i>a</i>	<i>botša</i>	<i>phuthego</i>	<i>gore</i>	<i>e</i>	<i>theeletše</i>	
	subj-01	remind	congregation	that	subj-09	listen	
	'Mr. Kgapšane never forgot to consistently remind the congregation to listen'						

The infinitive can also act as a complementary verb in auxiliary verb groups (Lombard et al., 1985:160)². The following auxiliary verb stems select the infinitive: *leka* 'try', *gana* 'refuse', *tlwaetše* 'used to', *tseba* 'know', *rata* 'love, like', *swanetše* 'must', *kganyoga* 'desire', *nyaka* 'want' and *nyakile* 'nearly'. Example (8) is taken from Lombard et al. (1985:160). See Lombard et al. (1985:187) and Louwrens (1991:59) for others.

- (8) *dipudi di rata go ya tšhemong tša ja mabele*
 goats subj-10 like cp15 go field-loc subj-10 eat sorghum
 ‘the goats like to go to the field and then they eat the sorghum’

Infinitives occur in identifying, descriptive and associative dynamic copulatives in the positive as well as in the negative as shown in (9).

(9a) Identifying copulative

go ba motho
 cp15 be man
 ‘to be a human being’

go se be motho
 cp15 neg be man
 ‘not to be a human being’

(9b) Descriptive (and locational) copulative

go ba bohlae
 cp15 be clever
 ‘to be clever’

go se be bohlae
 cp15 neg be clever
 ‘not to be clever’

(9c) Associative copulative

go ba le mpša
 cp15 be with dog
 ‘to be with a dog’

go se be le mpša
 cp15 neg be with dog
 ‘not to be with a dog’

A register of the verbal constellations

In the previous paragraphs, it was shown that the Northern Sotho infinitive occurs in a number of verbal constellations; Table 2 shows a summary. We plan to write parser rules for at least some of the infinitive constellations (cf. section 3). As parser rules are not written on word level but on the part of speech level, it is necessary to assign such parts of speech to all tokens used³.

Table 2: A register of the verbal constellations

<i>Reg.no.</i>	<i>Note</i>	<i>Examples</i>
1	inf. alone	<i>go_{MORPH_cp15} reka_v dipuku_{N10}</i>
2	intrans.	<i>ba_{CS02} phethile_v go_{MORPH_cp15} sepela_v</i>
3	trans./caus.	<i>go_{MORPH_cp15} rekiša_v dikgomo_{N10}</i>
4	double trans./applied	<i>go_{MORPH_cp15} rekela_v bana_{N02} malekere_{N06}</i>
5	contains OC	<i>go_{MORPH_cp15} mo_{CO01} thuša_v</i>

6	reflexive /adverb	<i>go</i> _{MORPH_cp15} <i>itshwara</i> _V <i>gabotse</i> _{ADV}
7	negated	<i>go</i> _{MORPH_cp15} <i>se</i> _{MORPH_neg} / <i>sa</i> _{MORPH_neg} <i>thuše</i> _V <i>moruti</i> _{N01}
8	auxiliary	<i>o</i> _{CS01} <i>rata</i> _V <i>go</i> _{MORPH_cp15} <i>hlwa</i> _V <i>a</i> _{CS01} <i>goga</i> _V <i>peipi</i> _{N09}
9	ident. cop.	<i>go</i> _{MORPH_cp15} <i>ba</i> _{VCOP} <i>motho</i> _{N01}
10	neg. id. cop.	<i>go</i> _{MORPH_cp15} <i>se</i> _{MORPH_neg} <i>be</i> _{VCOP} <i>motho</i> _{N01}
11	descr. cop.	<i>go</i> _{MORPH_cp15} <i>ba</i> _{VCOP} <i>bohlale</i> _{N14}
12	neg. descr. cop.	<i>go</i> _{MORPH_cp15} <i>se</i> _{MORPH_neg} <i>be</i> _{VCOP} <i>bohlale</i> _{N14}
13	assoc. cop.	<i>go</i> _{MORPH_cp15} <i>ba</i> _{VCOP} <i>le</i> _{PART_con} <i>mpša</i> _{N09}
14	neg. assoc. cop.	<i>go</i> _{MORPH_cp15} <i>se</i> _{MORPH_neg} <i>be</i> _{VCOP} <i>le</i> _{PART_con} <i>mpša</i> _{N09}

Infinitive nouns

The infinitive is positioned within the nominal system of Northern Sotho as a noun in class 15 consisting of the noun class prefix *go* and a verbal stem. Van Wyk et al. (1992:7) give the noun class system of Northern Sotho as in Table 3.

Table 3: The noun class system of Northern Sotho

<i>Class no.</i>	<i>Class prefix</i>	<i>Example</i>	
1	mo-	<i>mosadi</i>	'woman'
2	ba-	<i>basadi</i>	'women'
1a	-	<i>rrangwane</i>	'paternal uncle'
2a	bo-	<i>borrhangwane</i>	'paternal uncles and company'
3	mo-	<i>mosela</i>	'tail'
4	me-	<i>mesela</i>	'tails'
5	le-	<i>lesogana</i>	'young man'
6	ma-	<i>masogana</i>	'young men'
7	se-	<i>seatla</i>	'hand'
8	di-	<i>diatla</i>	'hands'
9	n-/m-/-	<i>nku</i> <i>mpša</i> <i>kgoši</i>	'sheep' 'dog' 'captain'
10	din-/ dim-/ di-	<i>dinku</i> <i>dimpša</i> <i>dikgoši</i>	'(many) sheep' 'dogs' 'captains'
14	bo-	<i>bogobe</i>	'porridge'
(6)	ma-	<i>magobe</i>	'types of porridge'
15	go-	<i>go kitima</i>	'to run'
16	fa-	<i>fase</i>	'underside/below'
17	go-	<i>godimo</i>	'on top/above'
18	mo-	<i>morago</i>	'behind/at the back'

When considering the nominal characteristics of the infinitive, it should be stated clearly that it shows most of the typical characteristics normally associated with nouns such as the generation of concords and pronouns, and typical nominal meaning(s). It occurs in the known syntactic functions of nominals, i.e. as subject or object. The following paragraphs describe the most significant nominal constellations found in the literature and in corpora in no specific order.

Grammatical functions of the nominal infinitive

Verbs assign grammatical functions to other constellations, i.e. subjects and objects. Lombard et al. (1985:49) say that infinitives can indicate processes as nouns. In (10a) and (10b), the class prefix of class 15 and the verbal stem together act as a *subject noun* in the typical SVO sentence structure of Northern Sotho. This subject of the sentence generates the subject concord *go* (or *gwa*) of class 15. In (10c), the infinitive is moved to a post-verbal position with the subject concord of the infinitive class occurring in the initial sentence position acquiring its subject function [the infinitive itself is grammatically categorized as an adjunct in (10c)].

- (10a) *go bala go lapiša morutiši*
 cp15 read subj-15 exhaust teacher
- (10b) *go bala gwa lapiša morutiši*
 cp15 read subj-15 exhaust teacher
 ‘reading exhausts/tires the teacher’
- (10c) *go lapiša morutiši, go bala*
 subj-15 exhaust teacher cp15 read
 ‘it exhausts/tires the teacher, the reading’

In (11), the infinitive is used as an *object clause*. It once again consists of the nominal class prefix of class 15 and a verb stem. Its status as an object is confirmed by its ability to generate an object concord of class 15 as in (11b). Like in (10c), the infinitive is categorized as an adjunct here, while the object concord acquires the object function.

- (11a) *ba rata go bala*
 subj-02 love cp15 read
 ‘they love reading’
- (11b) *go bala ba a go rata*
 cp15 love subj-02 pres obj-15 love
 ‘reading, they love it’

Possessive constellations

An infinitive generates a class 15 possessive concord, *ga* as in (12a). It might also occur as an object of the possessive as in (12b).

- (12a) *go sepela ga gagwe go a nkgahliša*
 cp15 walk poss-15 poss.pro-01 cs15 pres 1st-p-sg.-make-happy
 ‘the way she walks (her walking) makes me happy’
- (12b) *(ke monna) wa go bolela*
 it-is man poss-01 cp15 talk
 ‘(It is the man) of talking’

Particle constellations

Lombard et al. (1985:159) enlist the set of particles with which the infinitive can combine. Consider the conjunctive particle, instrumental particle, and the copulative particle respectively in (13). Note that particles also appear adjunctively with verbs (and other constellations) hence this property of the infinitive can also be called a transient area. An occurrence with the conjunctive particle *le* is described in (13a), while (13b) shows an example with the instrumental particle *ka* and (13c) shows a copulative constellation.

(13a) *go* *bala* *le* *go* *ngwala*
 cp15 read and cp15 write
 ‘reading and writing’

(13b) *ka* *go* *ngwala*
 in cp15 write
 ‘in writing’

(13c) *go* *sepela* *ke* *go* *bona*
 cp15 travel is cp15 see
 ‘to travel is to experience’

Pronouns

Several pronouns in Northern Sotho represent infinitives. Consider the following extract for a number of noun classes in Table 4.

Table 4: Pronouns

<i>Class</i>	<i>Prefix</i>	<i>Emphatic</i>	<i>Possessive</i>	<i>Quantitative</i>
1	mo-	yena	gagwe	yohle
2	ba-	bona	bona	bohle
...				
9	n-	yona	yona	yohle
10	di-	tšona	tšona	tšohle
...				
15	go	gona	gona	gohle

Emphatic pronouns

Consider the following examples in (14) where the emphatic pronoun occurs with the infinitive, extending its meaning:

(14a) Absolute/emphatic pronoun
go *opela* *gona*
 cp15 sing emph.pron-15
 ‘as for it, the singing, (I mean ...)’
 (Poulos & Louwrens, 1994:44)

(14b) Quantitative pronoun
go *akaretša* *gohle*
 cp15 sum up quant.pron-15
 ‘the entire summary’ (to summarize all)

(14c) Possessive pronoun
dipoelo *tša* *gona*
 gains poss-15 poss.pron-15
 ‘the gains/rewards of it (e.g. of reading)’

Demonstratives

Consider the extract of demonstratives for a number of noun classes from Van Wyk et al. (1992:37 and 38) in Table 5.

Table 5: Demonstratives

Class	Prefix	Class of pronoun	Meaning of pronoun			
			'here'	'next to here'	'there'	'yonder'
1	<i>mo-</i>	1	<i>yo</i>	<i>yono</i>	<i>yoo</i>	<i>yola</i>
2	<i>ba-</i>	2	<i>ba</i>	<i>bano</i>	<i>bao</i>	<i>bale</i>
...						
9	<i>n-</i>	9	<i>ye</i>	<i>yeno</i>	<i>yeo</i>	<i>yela</i>
10	<i>di-</i>	10	<i>tše</i>	<i>tšeno</i>	<i>tšeo</i>	<i>tšela</i>
...						
15	<i>go-</i>	18	<i>mo</i>	<i>mono</i>	<i>moo</i>	<i>mola</i>

Note that according to Poulos and Louwrens (1994:43), infinitives use the demonstratives of class 18. In the PSC, however, examples of the use of the demonstrative *go*, which could belong to class 15, were found. See the section on adjectives below.

Examples of an infinitive appearing together with demonstratives are shown in (15a) to (15d), taken from Poulos and Louwrens (1994:43).

(15a) *go* *lwa* ***mo***
 cp15 fight dem-cl18
 'this fighting'

(15b) *go* *nyatša* ***mono*** *ga* *gagwe*

(15c) *go* *nyatša* ***moo*** *ga* *gagwe*

(15d) *go* *nyatša* ***mola*** *ga* *gagwe*
 cp15 disrespect dem-cl18 poss15 his/her
 'that disrespect of his/hers'

Demonstrative copulatives

Poulos and Louwrens (1994:88) also list the demonstrative copulatives of all noun classes of which Table 6 is an extract for the noun classes 1, 2, 9, 10 and 15.

Table 6: Demonstrative copulatives

Class	Prefix	Meaning			
		'here is'	'here (next to) is'	'there is'	'yonder is'
1	<i>mo-</i>	<i>šo</i>	<i>šono</i>	<i>šoo</i>	<i>šola/šole</i>
2	<i>ba-</i>	<i>šeba</i>	<i>šebano</i>	<i>šebao</i>	<i>šebala/šebale</i>
...					
9	<i>n-</i>	<i>še</i>	<i>šeno</i>	<i>šeo</i>	<i>šela/šele</i>
10	<i>di-</i>	<i>šidi</i>	<i>šidino</i>	<i>šidio</i>	<i>šidila/šidile</i>
...					
15	<i>go-</i>	<i>šefa</i>	<i>šefano</i>	<i>šefao</i>	<i>šefala/šefale</i>

In this case, the infinitive uses the demonstrative copulatives of class 16.

As mentioned above, some of these combinations are rare or unattested in the literature. Consider the examples in (16), which were found in the PSC in a number of occurrences.

- (16) *go* *šoma* *šefa*
 go *šoma* *šefano*
 go *šoma* *šefao*
 go *šoma* *šefala*
 go *šoma* *šefale*
 cp15 work dem-16
 ‘here/there is the work’

Adjectives

Class 15 adjectival constructions may occur with the infinitive. Compare Table 7 for an infinitive of class 15 in comparison with a class 1 noun with adjective stems. In this construction, the noun generates a demonstrative and its nominal prefix is used with the adjective stem. The last example in Table 7 shows an exception of the rule formulated by Poulos and Louwrens (1994:43): the infinitive does indeed occur with a demonstrative that could very well belong to class 15. Such constellations were found in a number of examples in the PSC.

Table 7: Adjectives

<i>Noun class 1</i>	<i>Demonstr. class 1</i>	<i>Class prefix class 1 + adjective stem</i>
<i>mosetsana</i> ‘a little girl’	<i>yo</i>	<i>monyane</i>
<i>Noun class 15</i>	<i>Demonstr. class 18</i>	<i>Class prefix class 15 + adjective stem</i>
<i>go tseba</i> ‘little (amount of) knowledge’	<i>mo</i>	<i>gonyane</i>
<i>go rata</i> ‘great love’	<i>mo</i>	<i>gogolo</i>
<i>go šupša tšona diphoofolo ka ge go ja</i> ‘the animals are shown that a huge meal (lots to eat) was there’	<i>mo</i>	<i>gogolo e be e le gona</i>
<i>go tanyega</i> ‘a strong attraction to Bubbles’	<i>go</i>	<i>gogolo mo go Bubbles</i>

Nominal derivations

Northern Sotho nouns occur with three types of suffixes, namely locative, diminutive and augmentative/feminine. No mention is made of infinitives occurring with the augmentative or feminine suffix *-gadi* and no examples could be found in the PSC. However, the locative and diminutive suffixes do indeed occur, cf. (17).

(17a) Locative suffix

<i>katlego</i>	<i>ya</i>	<i>diphesente</i>	<i>tše</i>	80	<i>mo</i>
success	poss-9	percentages	dem-10	80	dem-18
<i>go</i>	<i>swarweng</i>	<i>ga</i>	<i>babolai</i>		
cp15	catch-pass-loc	poss-loc	murderers		
‘increase of 80% in the capturing of murderers’					

(17b) Diminutive suffix

<i>ka baka la</i>	<i>go</i>	<i>ruteganyana</i>	<i>ga</i>	<i>gagwe</i>
because of	cp15	educate-dim	poss-15	poss.pron-15
‘because of his/her little education’				

A register of the nominal constellations

It was demonstrated above that the Northern Sotho infinitive indeed occurs in a number of nominal constellations; Table 8 shows a summary.

Like in Table 2, parts of speech are added to the orthographic words in Table 8 to assist in the writing of parser rules.

Table 8: A register of the nominal constellations

<i>reg.no.</i>	<i>Note</i>	<i>Example</i>
15	subject concord	<i>Go</i> _{MORPH_cp15} <i>balav</i> <i>go</i> _{CS15} <i>lapišav</i> <i>morutiši</i> _{N01}
16	subject concord	<i>Go</i> _{MORPH_cp15} <i>balav</i> <i>gwa</i> _{CS15} <i>lapišav</i> <i>morutiši</i> _{N01}
17	subject concord	<i>Go</i> _{CS15} <i>lapišav</i> <i>morutiši</i> _{N01} , <i>go</i> _{MORPH_cp15} <i>balav</i>
18	object concord	<i>Ba</i> _{CS02} <i>rata</i> _V <i>go</i> _{CO15} <i>balav</i>
19	object concord	<i>Go</i> _{MORPH_cp15} <i>balav</i> <i>ba</i> _{CS02} <i>a</i> _{MORPH_pres} <i>go</i> _{CO15} <i>rata</i> _V
20	possessive concord	<i>Go</i> _{MORPH_cp15} <i>sepela</i> _V <i>ga</i> _{CPOSS15} <i>gagwe</i> _{PROPOSS01} <i>go</i> _{CS15} <i>a</i> _{MORPH_pres} <i>nkgahliša</i> _(CO01+) _V
21	object of a possessive	(<i>Ke monna</i>) <i>Wa</i> _{CPOSS01} <i>go</i> _{MORPH_cp15} <i>bolela</i> _V
22	connective particle	<i>Go</i> _{MORPH_cp15} <i>balav</i> <i>le</i> _{PART_con} <i>go</i> _{MORPH_cp15} <i>ngwala</i> _V
23	instrumental particle	<i>Ka</i> _{PART_ins} <i>go</i> _{MORPH_cp15} <i>ngwala</i> _V
24	copulative particle	<i>Go</i> _{MORPH_cp15} <i>sepela</i> _V <i>ke</i> _{PART_cop} <i>go</i> _{MORPH_cp15} <i>bona</i> _V
25	emphatic pronoun	(<i>Go</i> _{MORPH_cp15} <i>opela</i> _V) <i>Gona</i> _{PROEMP15}
26	quantitative pronoun	<i>Go</i> _{MORPH_cp15} <i>akaretša</i> _V <i>gohle</i> _{PROQUANT15}
27	possessive pronoun	<i>Dipoelo</i> _{N10} <i>tša</i> _{CPOSS10} <i>gona</i> _{PROPOSS15}
28	demonstrative	<i>Go</i> _{MORPH_cp15} <i>lwa</i> _V <i>mo</i> _{CDEM18}
29	demonstrative	<i>Go</i> _{MORPH_cp15} <i>nyatša</i> _V <i>mono/mool/mola</i> _{CDEM15} <i>ga</i> _{CPOSS15} <i>gagwe</i> _{PROPOSS01}
30	demonstrative cop.	... <i>go</i> _{MORPH_cp15} <i>šoma</i> _V <i>šefa/šefano/šefao/šefala/šefale</i> _{CDEM_cop}
31	adjective	<i>Go</i> _{MORPH_cp15} <i>tseba</i> _V <i>mo</i> _{CDEM18} <i>gonyane</i> _{ADJ15}
32	adjective	<i>Go</i> _{MORPH_cp15} <i>tanyega</i> _V <i>go</i> _{CDEM15?} <i>gogolo</i> _{ADJ15}
33	nominal derivation	... <i>mo</i> _{PART_loc} <i>go</i> _{MORPH_cp15} <i>swarweng</i> _V <i>ga</i> _{CPOSSLOC} <i>babolai</i> _{N02}
34	diminutive derivation	<i>Ka</i> _{PART_ins} <i>baka</i> _{N05} <i>la</i> _{CPOSS05} <i>go</i> _{MORPH_cp15} <i>ruteganyana</i> _V <i>ga</i> _{CPOSS15} <i>gagwe</i> _{PROPOSS01}

A morpho-syntactic perspective on the infinitive

In this section, we will begin with a brief introduction into context-free grammar (CFG) rules. The design of such parser rules for the verbal infinitive will be followed by a section describing the rules covering the nominal infinitive. In other words, full acknowledgement to its verbal and nominal properties will be given.

Overview

Phrase structure grammar rules of CFGs are usually of the form $A \rightarrow \phi$, where A stands for a non-lexical category (a phrase) and ϕ for either a non-lexical category or a part of speech. The order of the elements appearing after the arrow is understood as fixed: In English, the noun phrase rule ‘NP \rightarrow Det N’ is transliterated as ‘a noun (nominal) phrase consisting of a determiner, followed by a noun’ (never a noun, followed by a determiner).

In modern grammars, such units are often extended with parameters: in English, with its rather fixed word order⁴, it is possible to assign a grammatical function to an element of a grammar rule in a specific position. Such may be formulated as ‘S \rightarrow NP_{subj} VP’ (a sentence consists of a noun phrase carrying the subject function followed by a verbal phrase). We assume the same to be valid for Northern Sotho, as it is known for its rather fixed word order. Additionally, constraints may be formulated on the validity of grammatical rules, see e.g. the parameter VEND below.

When parser rules for Northern Sotho infinitives are to be formulated, one should follow the head principle: ‘Every phrasal category contains a head’, according to Shapiro (1997), ‘the head and its phrasal counterparts share the

same properties'. Therefore, an infinitive constellation is first and foremost to be defined as a verbal phrase no matter if it occurs in a nominal or a verbal role, because it is headed by a verb.

In Northern Sotho, so far, no morphological analyser is available that would be capable of identifying linguistic words [Kotzé (2008) describes the respective challenges] therefore, Taljard, Faaß, Heid and Prinsloo (2008) designed a set of 'parts of speech'⁵ which describes orthographic tokens independently of their word status; all parser rules to describe infinitive constellations in this article are based on these units and on the head principle.

Following these two principles, a base form of the infinitive form could be generally described by the rule ' $V_{15} \rightarrow \text{MORPH_cp15 V}$ ', meaning that a verbal phrase (of class 15) consists of a class prefix of class 15, followed by any verb stem. Faaß (in Faaß, 2010a and 2010b), however, has introduced a more rational description of the verbs of Northern Sotho: the verb is to be split into two parts. The first consists of the verb stem and its arguments (or dependents), which form the core of the verbal phrase (VP), the *Basic Verbal Phrase* (VBP) while the second part precedes the VBP and contains morphemes describing tense/aspect sub-verb agreement information; this part is called the verbal inflectional element (VIE). Such a method makes sense as the constellations of the core elements, the VBP, are determined by the semantics of the verb stem while the VIE is linguistically solely dependent on the mood, tense (aspect), and polarity that the verb appears in. Rules for the VBP can thus be formulated once (describing verb stems and their possible grammatical arguments and adjuncts); the VBP then appears as a non-lexical category, which is the same as the VIE-rules formulated for each mood, tense (aspect), and polarity. Together they form a VP ($\text{VP} \rightarrow \text{VIE VBP}$).

The verbal register

When describing units occurring in rules, we position lexical constraints on the top left of the unit (such information is to be found in the lexicon⁶). Syntactic constraints are written on the top right of the unit [the parameter 'verbal end' (VEND) will be used below to add constraints]. Part of speech tags are found on the bottom right of a lexical category; in the case of a non-lexical category, the noun class is found whenever it is necessary for agreement reasons. Such annotations to elements appearing on the right hand side of the rules can be transmitted to the upper nodes (elements on the left hand side) by using placeholders, e.g. <categ> which is a placeholder for a specific noun class, e.g. 02 or PERS_2sg.

When a VBP stands alone, it represents the positive imperative mood (cf. Table 1; in the case of the infinitive, the VIE contains at least the class prefix of class 15, *go*. For a start, three parser rules describing such constellations lacking a subject-NP are to be defined: rule (S1) describes the sentence as solely consisting of a VP. The second rule, (VP1), describes a VP_{15} consisting of a VIE_{15} followed by any VBP. Note that the bottom right <categ> is extended with an index, 'i' appearing on both sides of the rule indicating that both are of the same class. In the case that the VIE will contain a class prefix of class 15 [rule (VIE1) applied], the (VP1)-rule will be interpreted as $\text{VP}_{15} \rightarrow \text{VIE}_{15}^{\text{VEND=a/e}} \text{VBP}$.

(S1)	S	→	VP
(VP1)	$\text{VP}_{\langle \text{categ-i} \rangle}$	→	$\text{VIE}_{\langle \text{categ-i} \rangle}^{\text{VEND=a/e}} \text{VBP}$
(VIE1)	pos VIE_{15}	→	MORPH_cp15

A detailed description of the VBP can be found in Faaß (2010b); for the purpose of this article, we only list these rules (VBP1–5).

(VBP1)	VEND=a/e VBP	→	VEND=a/e V	(intransitive verb stem)
(VBP2)	VEND=a/e VBP	→	VEND=a/e V NP	(transitive verb stem)
(VBP3)	VEND=a/e VBP	→	VEND=a/e V NP NP	(double trans. verb stem)
(VBP4)	VEND=a/e VBP	→	$\text{CO}^{\text{VEND=a/e}} \text{V}$	(object concord, trans. verb stem)
(VBP5)	VEND=a/e VBP	→	$\text{CO}^{\text{VEND=a/e}} \text{V NP}$	(object concord, double trans. verb stem)

As we will also have to cater for the case of a clausal complement (register 2), we add rules (VBP6) and (COMP1).

(VBP6)	$V_{END=a/e}VBP$	\rightarrow	$V_{END=a/e}V\ COMP$	(Verb requiring a clausal complement)
(COMP1)	COMP	\rightarrow	VP_{15}	(clausal complement containing an infinitive VP)

By applying these rules, register 1 to 5 can be analysed, e.g. see Figures 1 and 2 representing register 1 and 2.

Figure 1: Analysis of register 1: *go reka dipuku*

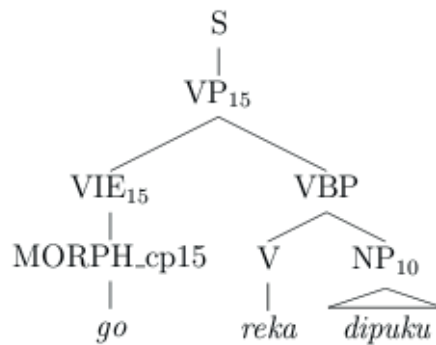
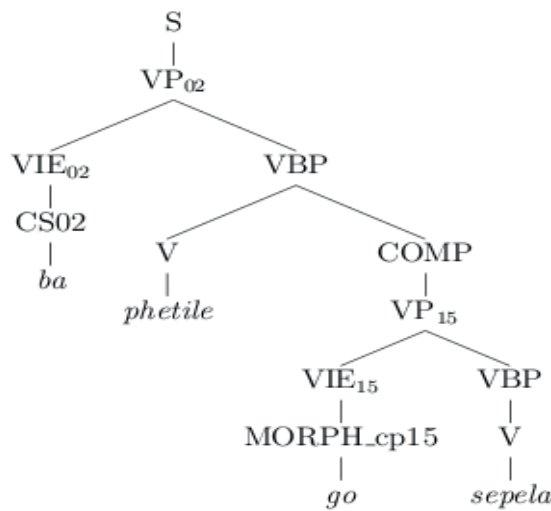


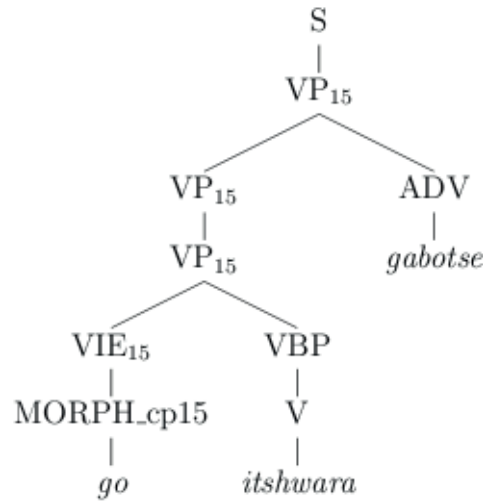
Figure 2: Analysis of register 2: *ba phetile go sepela*



To cater for the constellation of register 6, the adverbial extension, the VP-definition has to be extended, see (VP2) defining a general verbal phrase recursively to contain another that is extended by an adverb (ADV). Figure 3 shows the analysis of register 6 applying these rules that are kept general, as all verbal phrases may be extended by adverbs, not only the infinitive.

(VP2)	$VP_{<categ-i>}$	\rightarrow	$VP_{<categ-i>}\ ADV$
-------	------------------	---------------	-----------------------

Figure 3: Adverbial extension of a verbal phrase: *go itshwara gabotse*

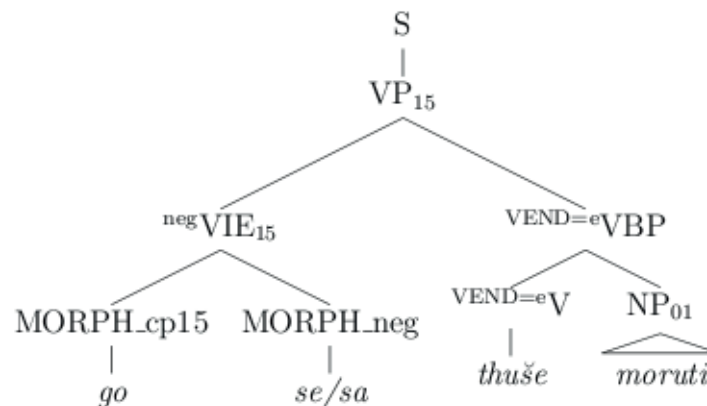


If infinitives are to be negated, the negation morpheme *se* (or *sa*, respectively) appears in the VIE after the class prefix [cf. rule (VIE2)]. Note that the verb stem has to end in *-e* in such a case, therefore a constraint referring to the verb stem has to be added. The parameter ‘verbal end’ (VEND) is added to the verbal stems in the lexicon so that the parser can find it and only allow the correct verb stems to be analysed with the rule. Another advantage of this parameter is that irregular/defective verb stems such as *re*, can be stored with the value ‘a’ for VEND, so that they may occur in all verbal phrases where this ending is expected.

(VP3) makes use of verb stems ending in *-e*. In order to avoid analyses where positive infinitives end in *-e*, we re-define rule (VP1) as (VP4). Figure 4 shows the analyses of register 7.

- (VIE2) $\text{negVIE}^{\text{inf}}$ \rightarrow MORPH_{cp15} ^{se/sa} MORPH_{neg}
- (VP3) VP_{15} \rightarrow negVIE_{15} ^{VEND=e} VBP
- (VP4) VP_{15} \rightarrow posVIE_{15} ^{VEND=a} VBP

Figure 4: The negated infinitive, register 7: *go se/sa thuše moruti*



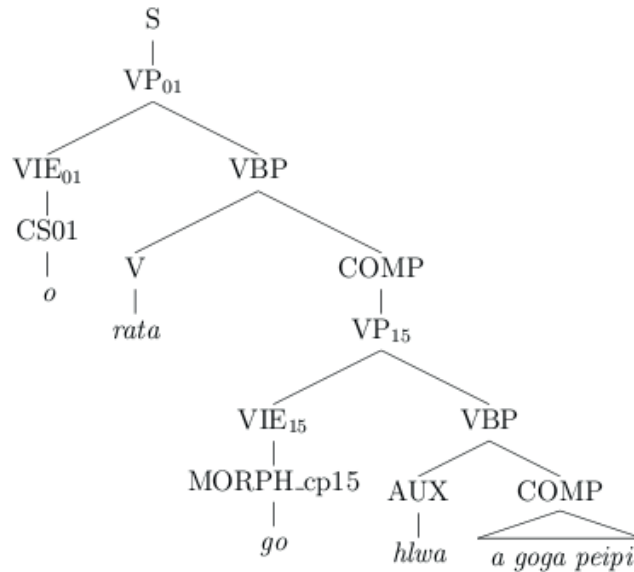
Auxiliaries of Northern Sotho take clausal complements, which might consist solely of verbal phrases. The example sentence of register 8 shows a predicative mood (*o rata ...*) as a main clause, the complementary clause of the auxiliary is also predicative and relates to this main sentence (see the agreement of the subject concords, both are of class 1). It would exceed the scope of this article to describe the rules to be defined for analysing these

predicative clauses, however, Figure 5 shows the analysis. For the sake of completeness, we add rule (VBP7) to our set, allowing for auxiliaries with their complements to appear.

(VBP7) $VP_{\langle \text{categ} \rangle} \rightarrow AUX\ COMP$

8	auxiliary	<i>o</i> _{CS01} <i>rata</i> _V <i>go</i> _{MORPH_cp15} <i>hlwa</i> _V <i>a</i> _{CS01} <i>goga</i> _V <i>peipi</i> _{N09}
---	-----------	--

Figure 5: An auxiliary constellation



Lastly, registers 9 to 12 describe copulative constellations that are all of identical structure except for containing a VCOP instead of a V. A parser should be enabled to analyse these, therefore another VBP-rule has to be added, cf. (VBP8).

(VBP8) $VBP \rightarrow VCOP\ NP$

The nominal register

We voluntarily left out a more detailed description of the NP assuming that in most cases a nominal will be contained. However, following the head principle, cf. Shapiro (1997), a verbal phrase such as the infinitive cannot be called an NP as long as no derivational process is assumed.

When examining languages such as German, one indeed finds similar processes: the German nominalized infinitive is assumed to present what is called the result of a transposition: From any verb in its base (infinitive) form, e.g. *schwimmen* ‘[to] swim’, a noun can be derived: *das Schwimmen* ‘the swimming’ without any obvious affixation.

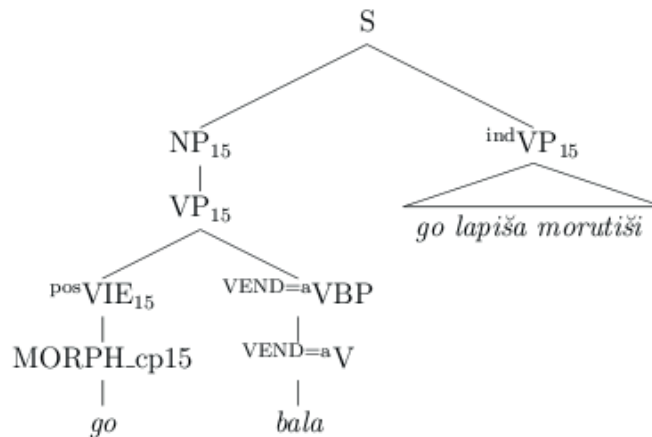
We therefore assume that the nominal Northern Sotho class 15 noun is the result of a transposition that is similar to this German nominalization process, and add rules (NP1) and (NP2) to our parser.

(NP1) $NP_{\langle \text{categ-i} \rangle} \rightarrow N_{\langle \text{categ-i} \rangle}$ (NPs of other noun classes)
 (NP2) $NP_{15} \rightarrow \text{int}VP_{15}$ (Infinitive verbal phrase)

As the sentences of the nominal register contain sentences consisting of a noun phrase followed by a verb phrase, we add rule (S2), enabling us to analyse infinitive verbal phrases as nominals, cf. Figure 6 (register 15). The NP is put into parenthesis because it is optional. Rule (S2) can thus also be read as (S1) and therefore replaces it.

$$(S2) \quad S \quad \rightarrow \quad (NP_{\langle \text{categ-i} \rangle}) VP_{\langle \text{categ-i} \rangle}$$

Figure 6: Analysis of register 15: *Go bala go lapiša morutiši*⁸



Register 20 and 21 demonstrate possessive constellations. These can be described with a recursive rule (NP3). Note that this rule is independent of the noun class appearing.

$$(NP3) \quad NP \quad \rightarrow \quad NP_{\langle \text{categ-i} \rangle} CPOSS_{\langle \text{categ-i} \rangle} NP_{\langle \text{categ-j} \rangle}$$

The generic rule (PP1) describes particle phrases. This rule can be applied to registers 22 to 24. To cater for the whole of the sentence (register 23 and 24), however, we need to add another rule allowing for these constellations to appear in sentences (S3). We again use parenthesis to indicate that (NP) and (PP) are optional. Hence, (S3) contains (S1) and (S2).

$$(PP1) \quad PP \quad \rightarrow \quad PART NP_{\langle \text{categ} \rangle}$$

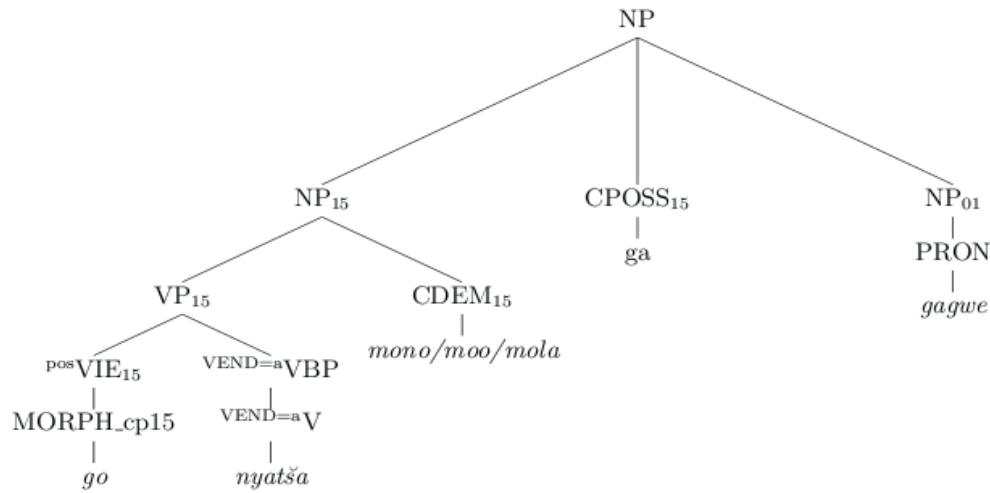
$$(S3) \quad S \quad \rightarrow \quad (NP_{\langle \text{categ-i} \rangle}) VP_{\langle \text{categ-i} \rangle} (PP)$$

Registers 25 to 29 cater for pronouns that may appear alone or accompany the nominal phrase. Again, rule (NP4) is valid for all NPs, independent of their noun class. We use the abbreviated label PRON to cater for all the above-mentioned pronouns; the noun itself is categorized as optional, as the pronoun can appear standalone.

$$(NP4) \quad NP \quad \rightarrow \quad (N_{\langle \text{categ-i} \rangle}) PRON_{\langle \text{categ-i} \rangle}$$

Figure 7 shows an analysis of register 29 to demonstrate the processability of the rules described.

Figure 7: Analysis of register 29: *Go nyatša mono/moo/mola ga gagwe*



Summary of the parsing rules designed

Table 9 summarizes all valid parsing rules designed in this section.

Table 9: Parsing rules

(S3)	S	→	(NP _{<categ-i>}) VP _{<categ-i>} (PP)
(VP2)	VP _{<categ-i>}	→	VP _{<categ-i>} ADV
(VP3)	VP _{<categ-i>}	→	negVIE _{<categ-i>} VEND=eVBP
(VP4)	VP ₁₅	→	posVIE ₁₅ VEND=aVBP
(VIE1)	posVIE ₁₅	→	MORPH_cp15
(VIE2)	negVIE _{inf}	→	MORPH_cp15 ^{se/sa} MORPH_neg
(VBP1)	VEND=a/eVBP	→	VEND=a/eV
(VBP2)	VEND=a/eVBP	→	VEND=a/eV NP
(VBP3)	VEND=a/eVBP	→	VEND=a/eV NP NP
(VBP4)	VEND=a/eVBP	→	CO VEND=a/eV
(VBP5)	VEND=a/eVBP	→	CO VEND=a/eV NP
(VBP6)	VEND=a/eVBP	→	VEND=a/eV COMP
(VBP7)	VBP	→	VCOP NP
(COMP1)	COMP	→	VP ₁₅
(NP1)	NP _{<categ-i>}	→	N _{<categ-i>}
(NP2)	NP ₁₅	→	infVP ₁₅
(NP3)	NP	→	NP _{<categ-i>} CPOSS _{<categ-i>} NP _{<categ-j>}
(NP4)	NP	→	N _{<categ-i>} PRON _{<categ-i>}
(PP1)	PP	→	PART NP _{<categ>}

An approach to parsing the Northern Sotho infinitive

Some of the infinitive constellations described in the literature are seldom found in the five million tokens corpus PSC, while others do not occur at all. Frequency of occurrence does not come to the fore in the traditional grammatical descriptions, while it plays a rather important role when implementing operational grammars, i.e. parsers. Here, only the constellations that are habitually used in the language should be implemented. In other words, the occurrences or lack of occurrences of constellations in corpora can at least give a hint if their implementation is justified.

In this section, we will show a possible way to analyse the infinitive in the framework of Lexical Functional Grammar (LFG). We will demonstrate an implementation of a few of the proposed structural and functional analyses, making use of the Xerox Linguistic Environment (XLE).

Lexical Functional Grammar: A brief introduction

LFG was designed by Kaplan and Bresnan (1982), who introduced this formal system to describe the grammar of a language. It supports the expression and explanation of generalizations that concern syntactic issues. It manages information on two levels: the lexicon, where semantic arguments are mapped to grammatical functions appearing at sentence level, and the syntactic rules that identify these functions with 'particular morphological and constituent structure configurations', cf. Kaplan and Bresnan (1982:174).

LFG represents linguistic structures on (at least) two levels: a constituent or c-structure and a functional or f-structure. The c-structure contains the system's knowledge on the constituents and their hierarchical relations; it is usually represented as a tree (see figures below). The f-structure represents the relations between the predicate and its arguments ('argument structure') in the form of a matrix containing attributes and values. These data may be generalized and used for e.g. machine translation: in the Northern Sotho sentence *monna o a sepela*, '(a/the) man walks', *monna* '(the/a) man' appears in the subject function, assigned by the verb *sepela* '[to] walk'. A (simplified) f-structure for this sentence is demonstrated by Figure 8a. The same structure appears in the two possible English translations of this sentence: 'a man walks' and 'the man walks', as shown in Figure 8b. Note that we do not define definiteness in Figure 8b; therefore, there is only one f-structure.

Figure 8a: LFG f-structure of *monna o a sepela*

PRED	'sepela <SUBJ>'	
SUBJ	[]	PRED 'monna' []

Figure 8b: f-structure of its English translation: 'the man walks'

PRED	'walk <SUBJ>'	
SUBJ	[]	PRED 'man' []

We have implemented a number of the verbal and nominal infinitive constellations with XLE, according to the recommendations of the pargram project, as described by Butt, Dyvik, Holloway King, Masuichi and Rohrer (2002). This section shows some results, i.e. c(onstituent) and f(unctional)-structures of the implemented grammar. They were successfully introduced and approved by active members of the pargram grammar group, cf. Faaß (2010c). For space reasons, only a selection of implemented rules described in the previous section is shown here.

Implementation

To implement the register, the XLE lexicon has to contain entries of all the units in the lexical category, the rules described above are to be defined in a more compact way. In this section, we will show some samples of our lexicon and of the rules implemented.

The lexicon

Typical verbal entries for different forms of *reka* '[to] buy' can look as follows:

reka Vtr * (^PRED)='reka <(^SUBJ)(^OBJ)>' (^VEND)=a⁹.
 reke Vtr * (^PRED)='reka <(^SUBJ)(^OBJ)>' (^VEND)=e.
 rekile Vtr * (^PRED)='reka <(^SUBJ)(^OBJ)>'
 (^VEND)=ile (^TNS-ASP TENSE) = past.

All three entries described in this lexicon are transitive (Vtr) and require a subject and an object to appear. We opted for using *reka* as the base form (another possible base form could be the root, *-rek-*). The first two entries do not define a tense feature as they may appear in any tense; the third, *rekile*, however, contains the past tense suffix *-ile* (*-il-e*) and is thus defined as occurring in this tense only. For reasons explained above, the verbal ending (VEND) is defined for all three accordingly.

A common noun entry, on the other hand, must contain information about the noun class. Metadata on the person and number of such an item would not be mandatory for a monolingual analysis of Northern Sotho; they were however added to enable the system to find corresponding matches in a goal language lexicon when doing machine translation into English at a later stage.

moruti N * (^PRED)='moruti' (^CLASS)=1 (^PERS)=3 (^NUM)=sg.

Lastly, an important part of all infinitive constellations, the class prefix of noun class 15, *go*, must be contained in the lexicon.

The unit *go* is highly ambiguous, therefore, there are a number of options described by its lexicon entry. As *go* is purely a functional item in the first four options, these entries do not contain a PRED-value, they describe *go* as the class prefix of class 15 (MORPH), and as a subject concord of class 15, the locative class, and the indefinite. The object concord *go* is homographous for the locative classes and class 15, the PRED-value is set to 'pro(nominal)' because the object concord is of a pronominal nature (it usually replaces an omitted object which is known in the discourse).

go MORPH (^TNS-ASP MOOD)=infinitive;
 CS (^SUBJ CLASS)=15;
 CS (^SUBJ CLASS)=indefinite;
 CS (^SUBJ CLASS)=LOC;
 CO *(^PRED)= 'pro' (CLASS)=15;
 CO *(^PRED)= 'pro' (CLASS)=LOC.

The rules section

XLE rules are implemented in a compact way, e.g. there may only be one VIE-rule; all possible constellations must be described as options (using 'l' as a delimiter). Our VIE-rules for the infinitive (VIEinf) are as follows:

```
VIEinf --> { MORPH: (^TNS-ASP MOOD)=infinitive
             | MORPH: (^TNS-ASP MOOD)=infinitive;
             { MORPH: (^NEG)=c se (^TNS-ASP POL)=neg
             | MORPH: (^NEG)=c sa (^TNS-ASP POL)=neg
             }
             }.
```

The first option of the VIEinf-rule above contains the lexicon entry of *go* as shown above (the first option there), the second and third show the two possible ways to negate an infinitive: *go se* and *go sa*. The negation morpheme is specifically described to exclude others (e.g. *ga*) from being analysed by the parser. All other VP and NP rules are implemented accordingly.

Some example analyses

We have implemented some of the parser rules described above. The following figures show the resulting analyses of register 1, 7, and 18 for demonstrating reasons in Figures 9 – 11.

Figure 9: XLE results for register 1: *go reka dipuku*

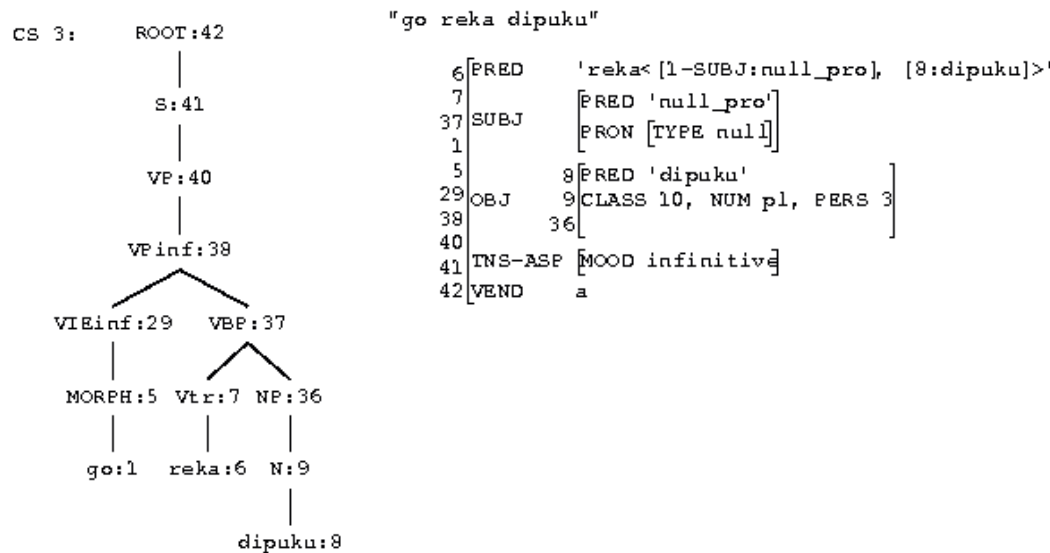


Figure 10: Analysis of register 7: *go se thuše moruti*

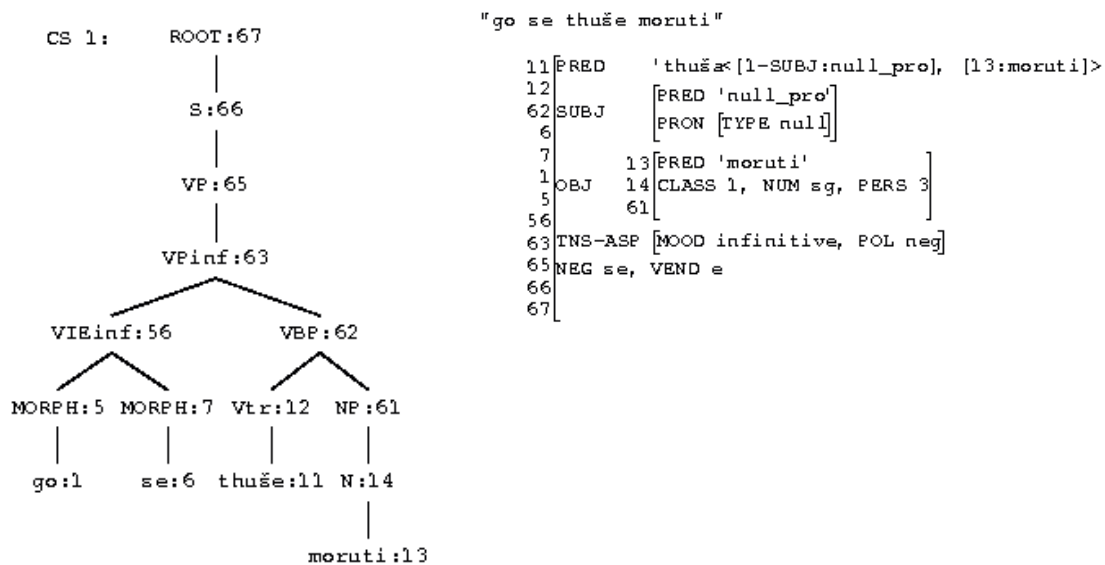
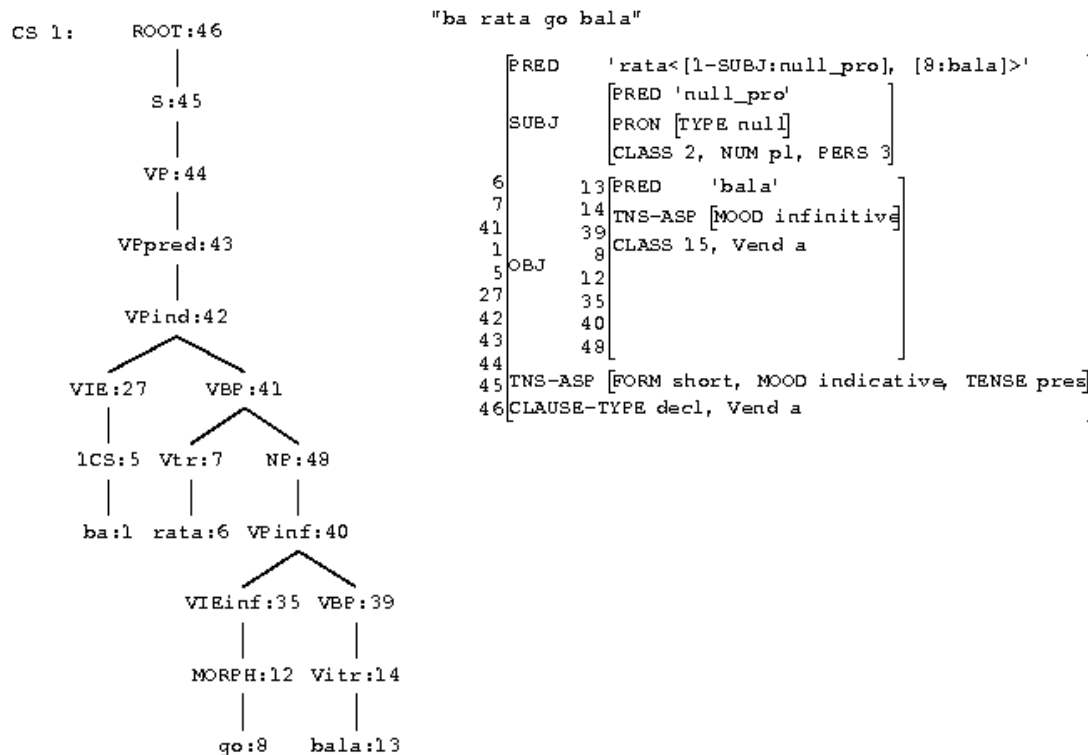


Figure 11: Analysis of register 18: *ba rata go bala*

Conclusions and future work

In the first two sections, the description of the infinitive differed from the way in which it was described in the traditional grammars of prominent scholars. This was done to compile a list of infinitive constellations to serve as a basis for subsequent morpho-syntactic analysis in the following section. It has been illustrated that infinitive constellations are rooted in the dual semantic character of the infinitive, namely its nominal and verbal characteristics. In most traditional grammars, infinitives are described in a rather haphazard way but an effort was made to compile a comprehensive list of infinitive constellations.

The internal structure of the infinitive is non-ambiguously verbal. Like the English gerund or the German nominalization resulting from a transposition, the infinitive may also appear as a noun phrase with a nominal function. We can thus assume that the Northern Sotho infinitive is to be treated similarly to the English gerund whenever it appears with a grammatical function assigned by the predicate of the sentence, i.e. as a noun phrase. Whenever it appears in a verbal function, i.e. as a predicate, it can be described as a verbal phrase.

We have implemented only parts of the grammar with the LFG, but plan for a more thorough and complete implementation of the rules found so far. In addition, we aim to make the implemented grammar fragment available as a web service. Currently, an evaluation is foreseen as a first step towards this goal.

Notes

1. The status of the infinitive as an auxiliary or a more strongly grammaticalized form of the infinitive with a dependent sub-clause in (7) is debatable but irrelevant to the aims of this study.
2. The use of the infinitive as a complementary verb is syntactically very close to a nominal.
3. The parts of speech utilized in this paper were described in Taljard, Faaß, Heid and Prinsloo (2008); they are assigned to orthographic tokens.

4. A sentence of Northern Sotho usually follows an SVO-order: subject, verb and object. The morph order can also be defined as stable in the verb especially, see e.g. Poulos and Louwrens (1994) who define inter alia Northern Sotho's verbal elements solely on the basis of morph order. This order can be changed for reasons of accentuation and pronominalization.
5. In terms of using the correct linguistic terminology, these 'parts of speech' might better be called 'morphemic units'.
6. We define a lexicon containing linguistic units together with their labels (which appear in the parser rules as lexical categories). With some units, additional metadata may appear, e.g. the parameter VEND with its respective value.
7. Register 2 shows a verb in the predicative mood (the positive infinitive of the present tense); defining parser rules describing this mood would exceed the scope of this article.
8. It would exceed the scope of this article to describe all the rules that are necessary for the indicative verbal phrases appearing in registers 15–19; see Faaß (2010b) for details on these phrases.
9. Note that such additional lexical information is usually encoded with templates (macros). For the sake of clarity, however, it is spelled out here.

References

- Butt, Miriam, Dyvik, Helge, Holloway King, Tracy, Masuichi, Hiroshi & Rohrer, Christian. 2002. The parallel grammar project. In *Proceedings of COLING-2002 Workshop on Grammar Engineering and Evaluation*:1–7. Available: <http://www2.parc.com/isl/groups/nltt/pargram/ref.html> Accessed on 2011/08/22.
- De Schryver, G.-M. & Prinsloo, D.J. 2000. The compilation of electronic corpora with special reference to the African languages. *Southern African Linguistics and Applied Language Studies* 18(1–4):89–106.
- Faaß, Gertrud. 2010a. A morpho-syntactic description of Northern Sotho as a basis for an automated translation from Northern Sotho to English. Unpublished doctoral thesis. Pretoria: University of Pretoria.
- Faaß, Gertrud. 2010b. The verbal phrase of Northern Sotho: A morpho-syntactic perspective. In *Proceedings of the 2nd workshop on African Languages Technology (AfLaT 2010)* at LREC, 18th May 2010, Valetta, Malta:37–42.
- Faaß, Gertrud. 2010c. A toy LFG grammar describing some of the morpho-syntactics of Northern Sotho verbs. Presented at the PARGRAM spring meeting, University of Constance, March 23, 2010.
- Kaplan, R.M. & Bresnan, Joan. 1982. Lexical-functional grammar: A formal system for grammatical representation. *The mental representation of grammatical relations*. Cambridge, MA: MIT Press:173–281.
- Kotzé, Petronella M. 2008. Northern Sotho grammatical descriptions: A challenge to the design of a tokeniser for the verbal segment. *South African Linguistics and Applied Language Studies* 26(2):197–208.
- Lombard, D.P., Van Wyk, E.B. & Mokgokong, P.C 1985. *Introduction to the grammar of Northern Sotho*. Pretoria: J.L. van Schaik.
- Louwrens, L.J. 1991. *Aspects of Northern Sotho grammar*. Pretoria: Via Afrika.
- Poulos, G. & Louwrens, L.J. 1994. *A linguistic analysis of Northern Sotho*. Pretoria: Via Afrika.
- Shapiro, L.P. 1997. Tutorial: An introduction to syntax. *Journal of Speech, Language and Hearing Research* 40(2):254–272.
- Taljad, E., Faaß, G., Heid, U. & Prinsloo, D.J. 2008. On the development of a tagset for Northern Sotho with special reference to the issue of standardization. *Literator, Journal of Literary Criticism, Comparative Linguistics and Literary Studies – Special Issue on Human Language Technologies* 29(1):111–137.
- Van Wyk, E.B., Groenewald, P.S., Prinsloo, D.J., Kock, J.H.M. & Taljad, E. 1992. *Northern Sotho for First-Years*. Pretoria: J.L. van Schaik.
- Ziervogel, D. 1976. *Handboek van Noord-Sotho*. Pretoria: Van Schaik.