

# A resource grammar for Zulu

Laurette Marais  
and Laurette Pretorius  
(with Lionel Posthumus)

# Outline

Zulu and its resources

Characteristic features

- Agglutinating morphology
- Nominal classification
- Concordial agreement

Tour of the Zulu RG

- Mostly the noun phrase
- Quick remarks on verbs

# Zulu

Bantu → Southern Bantu → Nguni

One of 11 official languages of South Africa

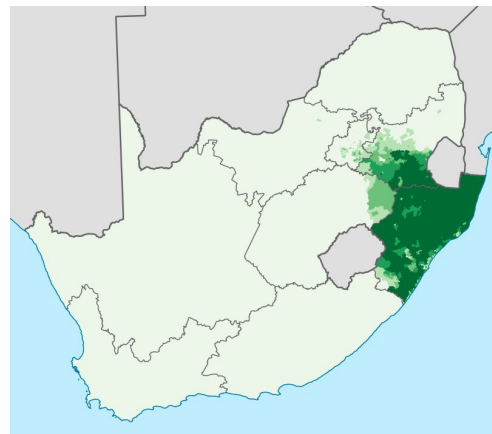
~11.5M home language speakers as of 2011

Largest home language in South Africa

Understood by 50% of South Africans

Second most widely spoken Bantu language after Swahili

9 174 Wikipedia articles (mostly stubs), 28 active registered users



# Zulu NLP resources

ZulMorph - XFST morphological analyser

WordNet project

NCHLT monolingual corpus - 1.4M sentences

JW300 corpus - 1.1M sentences

Autshumato Eng-Zul parallel corpus - 35k sentences

isizulu.net - not directly an NLP resource, but very useful

***Resource grammar WiP<sup>1</sup>***

1. <https://github.com/LauretteM/gf-rgl-zul>

# Characteristic features

Agglutinating morphology (conjunctive orthography)

Nominal classification

- 18 noun classes (depending on how you count)
- also called class gender

Concordial agreement of nouns with

- verbs, copulatives, relative clauses, adjectives, demonstratives...
- **u**Tatjana **u**zozuza **i**ndonda **ye**golide (*Tatjana will win the gold medal*)
- **ya** + **i**golide = **ye**golide

In GF, this means allomorphs and BIND (&+)

- uTatjana u &+ zo &+ zuza indonda ye &+ golide

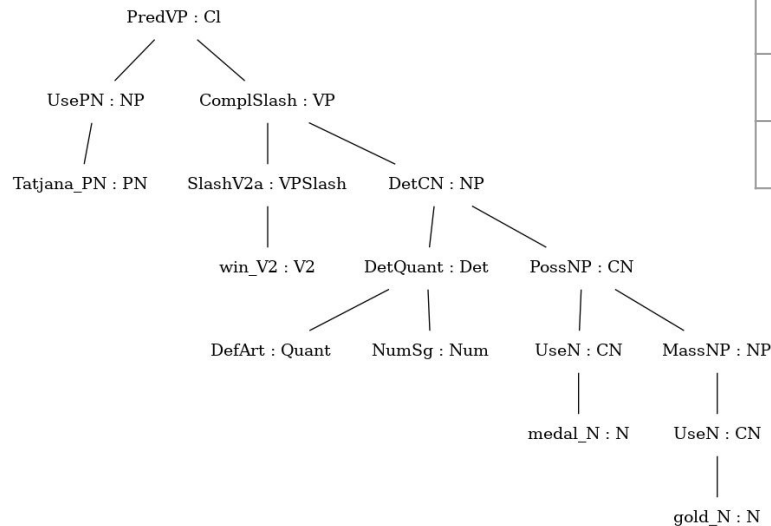
# Noun classes

Class	Typical nouns	Prefixes	Subject concords	Object concords
1-2	persons	umu- / ab(a)-	u, w, ka, e, a, wu / ba, b, be	m, mu / ba, b
1a-2a	proper names, titles	u- / o-		
3-4	nature	umu- / imi-	u, w, wu / i, y, yi	wu, w / yi, y
5-6	anatomy, birds, fruits, loan words...	i(li)- / ama-	li, l / a, , wa, e	li, l / wa, w
7-8	anatomy, attitudes, collectives...	isi- / izi-	si, s / zi, z	si, s / zi, z
9-10 (-6)	animals, anatomy, tools...	iN- / isiN-	i, y, yi / zi, z	yi, y / zi, z
11 (-10)	misc, diminutive significance	u(lu)-, lw-	lu, lw	lu, lw, l
14	abstract	ubu-, u-	bu, b	bu, b
15	infinitives	uku-, uk-, ukw-	ku, kw	ku, k
16, <b>17</b> , 18	locatives	<b>ku-</b> / [pha-, mu-]	<b>ku, kw</b>	<b>ku, k</b>

# Noun classes

Class	Prefixes	-ntu		-Zulu/zulu	
1-2	umu- / ab(a)-	umuntu, abantu	<i>person, people</i>	umZulu, abaZulu	<i>Zulu person, people</i>
1a-2a	u- / o-			uZulu, oZulu	<i>Mr Zulu, Mr Zulu and co</i>
3-4	umu- / imi-				
5-6	i(li)- / ama-			izulu, amazulu	<i>sky, heavens</i>
7-8	isi- / izi-	isintu	mankind	isiZulu	<i>Zulu language/culture</i>
9-10 (-6)	iN- / isiN-				
11 (-10)	u(lu)-, lw-				
14	ubu-, u-	ubuntu	humanity		
15	uku-, uk-, ukw-				
16, <b>17</b> , 18	<b>ku-</b> / [pha-, mu-]			phezulu	<i>above</i>

# Basic morphosyntax of Zulu

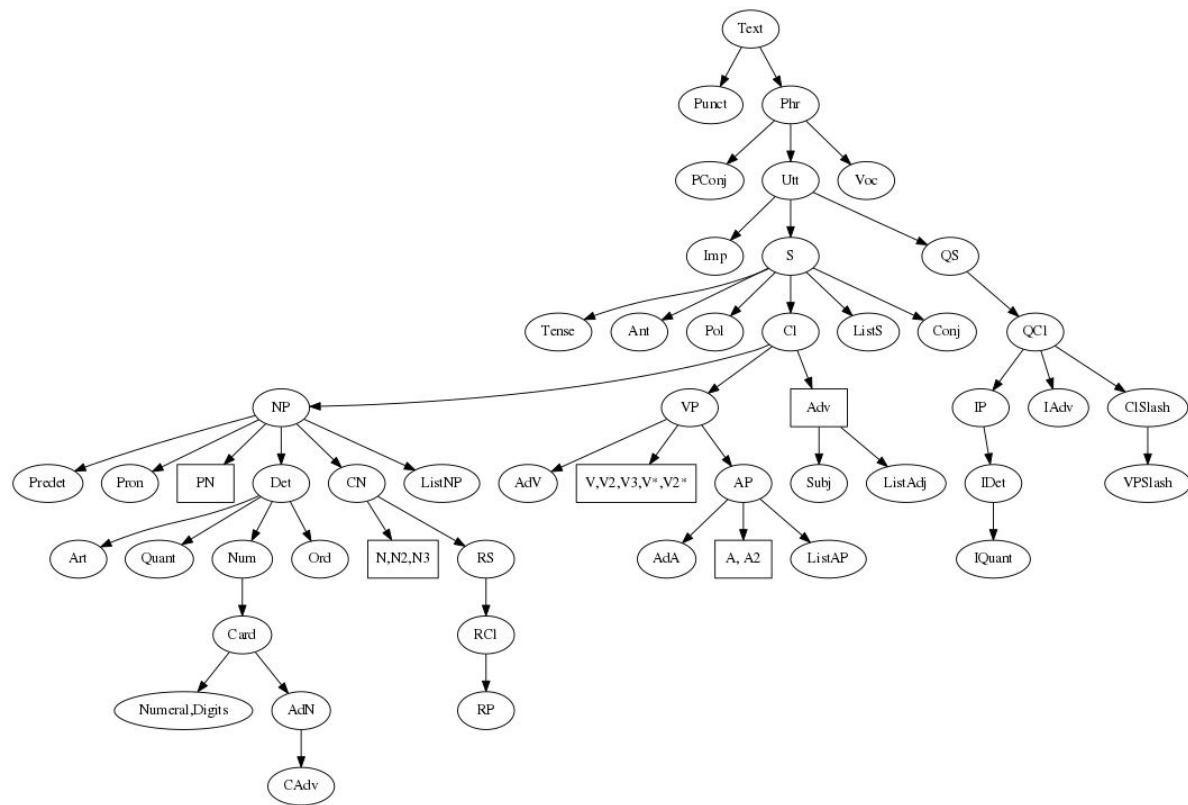


Present	Positive	uTatjana <b>uzuza</b> indondo <b>yegolide</b>
Recent past	Positive	uTatjana <b>uzuze</b> indondo <b>yegolide</b>
Remote past	Positive	uTatjana <b>wazuza</b> indondo <b>yegolide</b>
Remote past	Negative	uTatjana <b>akazuzanga</b> indondo yegolide
Immediate future	Positive	uTatjana <b>uzozuza</b> indondo <b>yegolide</b>
Immediate future	Negative	uTatjana <b>akazuzuza</b> indondo <b>yegolide</b>

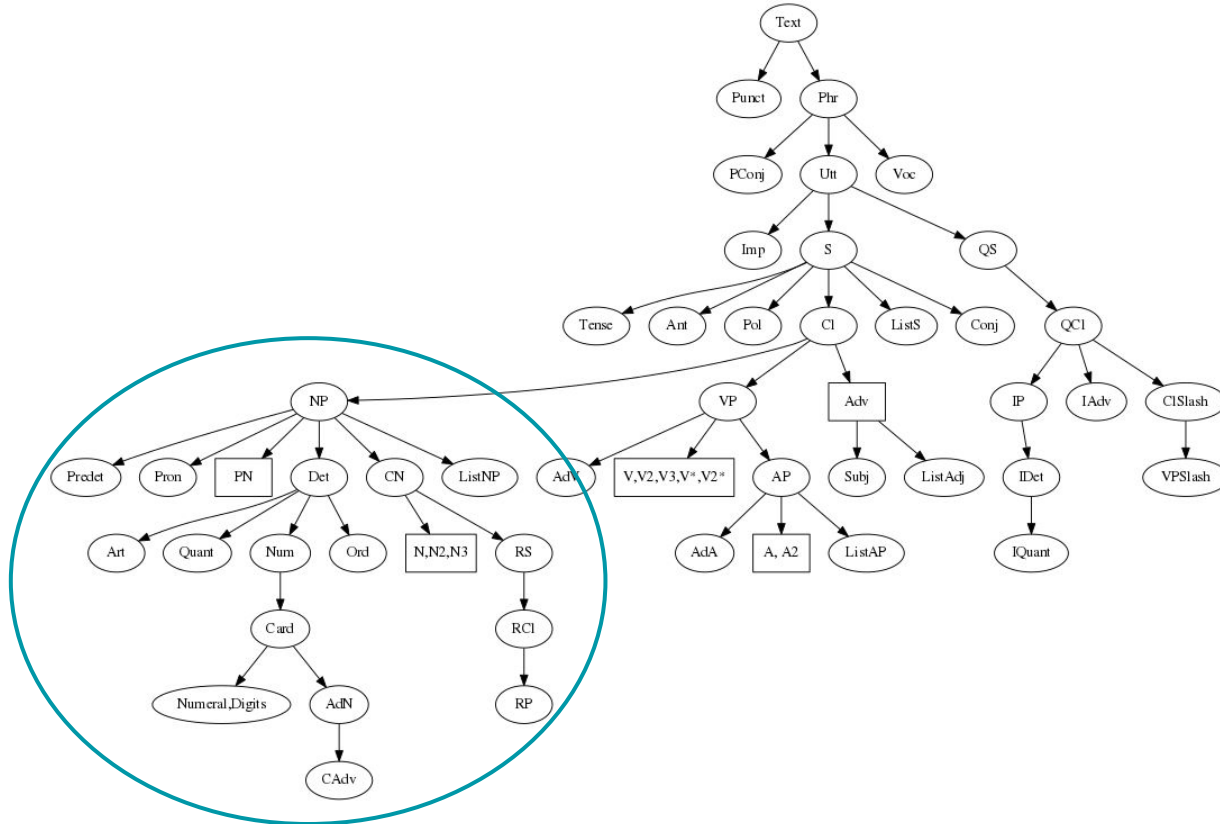
●	subject concord, class 1a
●	tense prefix
○	negative prefix
●	root + terminative suffix
●	possessive concord, class 9 (class 5)



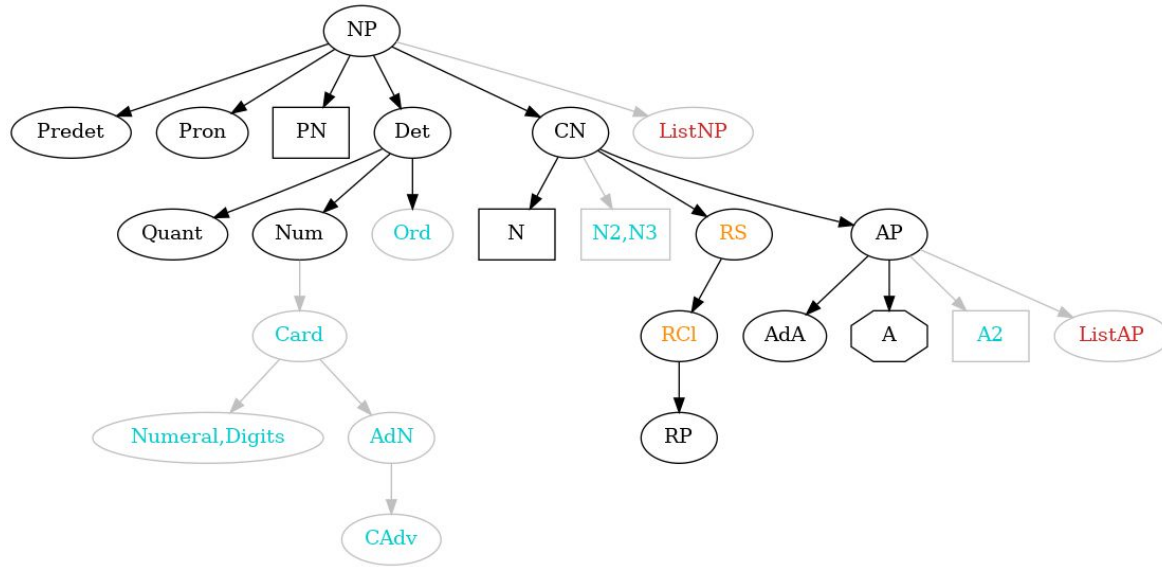
# Tour of the Zulu RG



# Noun phrases



# Noun phrase so far



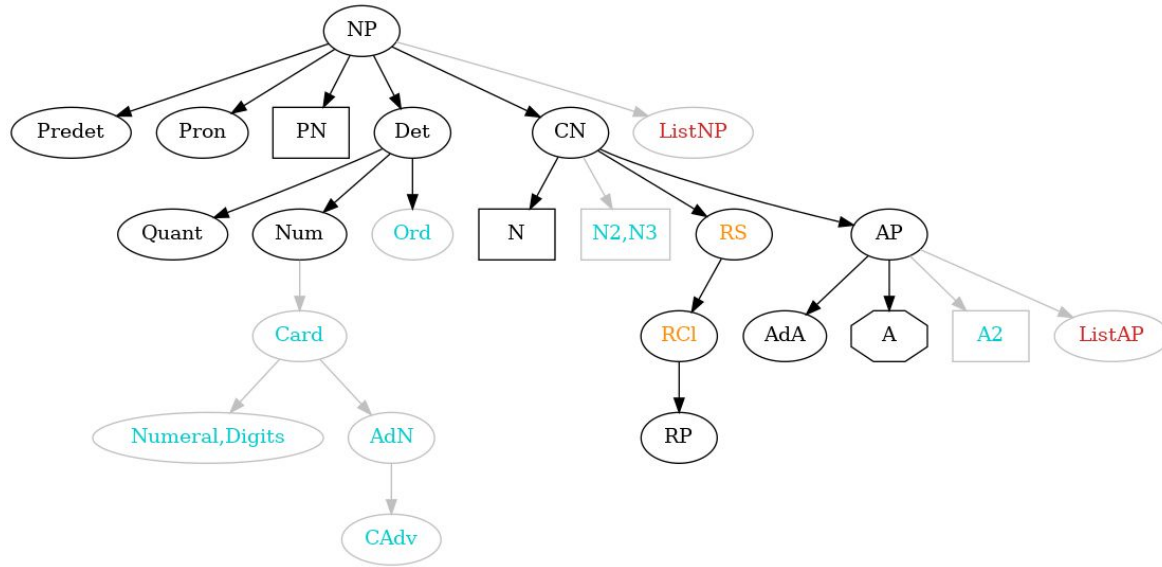
Black - mature

Orange - under development

Red - not yet implemented

Cyan - on ice

# Noun phrase so far



Black - mature-ish

Orange - under development

Cyan - on ice

Red - not yet implemented



# Nouns

Params	
ClassGender	C1_2   C1a_2a   C3_4   C5_6   C7_8   C9_10   C11_10   C9_6   C14   C15   C17
NForm	Full   Reduced   Poss   Loc
Agr	First Number   Second Number   Third ClassGender Number
Number	Sg   Pl

Lincat of N	
s	Number => NForm => Str
c	ClassGender

Lincat of CN	
s	Number => NForm => Str
mod	Number => Str
c	ClassGender

khundla_N ( <i>stadium</i> )		
s	Sg	Pl
Full	inkundla	izinkundla
Reduced	nkundla	zinkundla
Poss	nkundla	zinkundla
Loc	enkundleni	ezinkundleni
c	C9_10	

# Nouns

Params	
ClassGender	C1_2   C1a_2a   C3_4   C5_6   C7_8   C9_10   C11_10   C9_6   C14   C15   C17
NForm	Full   Reduced   Poss   Loc
Agr	First Number   Second Number   Third ClassGender Number
Number	Sg   Pl

Lincat of N	
s	Number => NForm => Str
c	ClassGender

Lincat of CN	
s	Number => NForm => Str
mod	Number => Str
c	ClassGender

indondo	<i>medal</i>
le ndondo	<i>this medal</i>
yona le ndondo	<i>this very medal</i>
le ndondo yona	<i>this medal, on the other hand</i>
le ndondo enkulu	<i>this big medal</i>
le ndondo yona enkulu	<i>this big medal, on the other hand</i>

# “Adjectives”

Zulu adjectives are a closed class.

There are three kinds of “adjectives” (qualificatives)

1. adjective stems
2. primitive relative stems
3. enumerative stems

Adjective	ngibona umgijimi omkhulu	<i>I see the big sprinter</i>
Relative	ngibona umgijimi oqotho	<i>I see the honest sprinter</i>
Enumerative	ngibona umgijimi mumbé	<i>I see the other sprinter</i>

```
adjConcLookup : Agr => Str =
  table {
    Third C1_2 Sg => "om" ;
    Third C1_2 Pl => "aba" ;
    Third C1a_2a Sg => "om" ;
    Third C1a_2a Pl => "aba" ;
    Third C3_4 Sg => "om" ;
    Third C3_4 Pl => "emi" ;
    Third C5_6 Sg => "eli" ;
    Third C5_6 Pl => "ama" ;
    Third C7_8 Sg => "esi" ;
    Third C7_8 Pl => "ezi" ;
    Third C9_10 Sg => "e" ;
    Third C9_10 Pl => "ezi" ;
    Third C11_10 Sg => "olu" ;
    Third C11_10 Pl => "ezi" ;
    Third C9_6 Sg => "e" ;
    Third C9_6 Pl => "ama" ;
    Third C14 _ => "obu" ;
    Third C15 _ => "oku" ;
    Third C17 _ => "oku" ;
    (First _ | Second _ ) => "om"
  } ;
```

# “Adjectives”

Params	
AForm	AF1   AF2   AF3
AType	AdjType   RelType   EnumType

Lincat of A, AP	
s	AForm => Str
t	AType

khulu_A ( <i>big</i> )		
s	AF1	khulu
	AF2	nkulu
	AF3	khulu
t	AdjType	



# “Adjectives”

Params	
AForm	AF1   AF2   AF3
AType	AdjType   RelType   EnumType

Lincat of A, AP	
s	AForm => Str
t	AType

khulu_A ( <i>big</i> )		
s	AF1	khulu
	AF2	nkulu
	AF3	khulu
t	AdjType	

qotho_R ( <i>honest</i> )		
s	AF1	qotho
	AF2	
	AF3	
t	RelType	

hle_A ( <i>beautiful</i> )		
s	AF1	hle
	AF2	nhle
	AF3	uhle
t	AdjType	

mbe_E ( <i>other</i> )		
s	AF1	mbe
	AF2	
	AF3	
t	EnumType	

```
aformN : Agr -> AForm = \agr ->
  case agr of {
    Third C1_2 Sg => AF3 ;
    Third C1a_2a Sg => AF3 ;
    Third C3_4 Sg => AF3 ;
    Third C7_8 Pl => AF2 ;
    Third C9_10 Sg => AF2 ;
    Third C9_10 Pl => AF2 ;
    Third C9_6 Sg => AF2 ;
    Third C11_10 Pl => AF2 ;
    _ => AF1
  } ;
```

# “Adjectives”

Params	
AForm	AF1   AF2   AF3
AType	AdjType   RelType   EnumType

Lincat of A, AP	
s	AForm => Str
t	AType

khulu_A ( <i>big</i> )		
s	AF1	khulu
	AF2	nkulu
	AF3	khulu
t	AdjType	

hle_A ( <i>beautiful</i> )		
s	AF1	hle
	AF2	nhle
	AF3	uhle
t	AdjType	

```
aformN : Agr -> AForm = \agr ->
  case agr of {
    Third C1_2 Sg => AF3 ;
    Third C1a_2a Sg => AF3 ;
    Third C3_4 Sg => AF3 ;
    Third C7_8 Pl => AF2 ;
    Third C9_10 Sg => AF2 ;
    Third C9_10 Pl => AF2 ;
    Third C9_6 Sg => AF2 ;
    Third C11_10 Pl => AF2 ;
    _ => AF1
  } ;
```

ngibona isiziba esihle

*I see the beautiful pool*

ngibona indondo enhle

*I see the beautiful medal*

ngibona umgijimi omuhle

*I see the beautiful sprinter*

# Quantifiers

IndefArt, DefArt : Quant

this\_Quant, that\_Quant, *yonder\_Quant* : Quant

Params	
QuantDef	Article Definiteness   Demonstrative Distance
Definiteness	Indef   Def
Distance	Dem1   Dem2   Dem3

dem_pron	Dem1	Dem2	Dem3
Third C1_2 Sg	lo	lowo	loya
Third C1_2 Pl	laba	labo	labaya
Third C3_4 Sg	lo	lowo	loya
Third C3_4 Pl	le	leyo	leya
Third C5_6 Sg	leli	lelo	leliya
Third C5_6 Pl	la	lawo	lawaya
Third C7_8 Sg	lesi	leso	lesiya
Third C7_8 Pl	lezi	lezo	leziya
Third C9_10 Sg	le	leyo	leya
Third C9_10 Pl	lezi	lezo	leziya
...	...	...	...

# Quantifiers

IndefArt, DefArt : Quant

this\_Quant, that\_Quant, *yonder\_Quant* : Quant

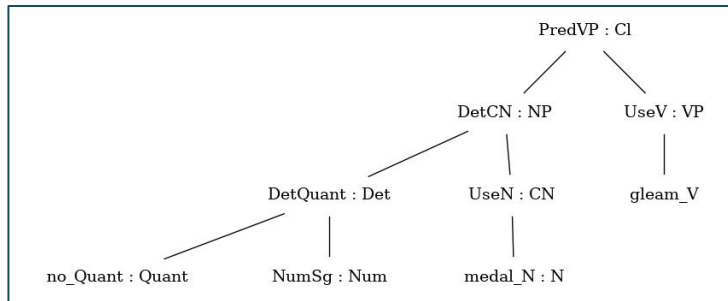
Params	
QuantDef	Article Definiteness   Demonstrative Distance
Definiteness	Indef   Def
Distance	Dem1   Dem2   Dem3

Lincat of Quant	
s	Str
qdef	QuantDef

indondo iyakhazimula	<i>a/the medal gleams</i>
le ndondo iyakhazimula	<i>this medal gleams</i>
leyo ndondo iyakhazimula	<i>that medal gleams</i>
leya ndondo iyakhazimula	<i>yonder medal gleams</i>

Where is no\_Quant?

# Quantifiers

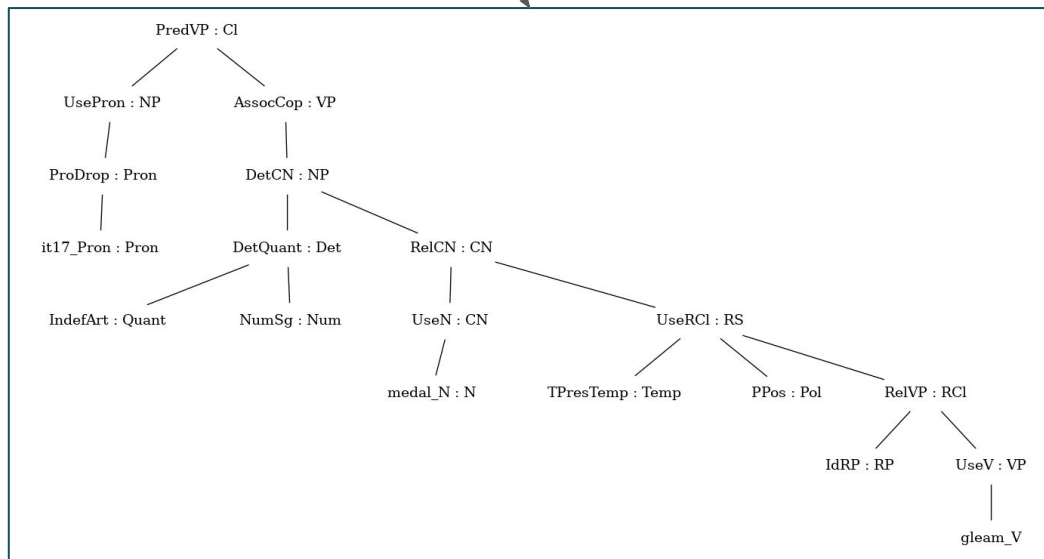


*no medal gleams*

akunandondo ekhazimulayo

*there is no medal that gleams*

*Literally: it is not with a medal that  
gleams*



# Pronouns

Lincat of Pron	
s	NForm => Str
agr	Agr
proDrop	Bool
empty	Str

i_Pron		
s	Full	mina
	Reduced	mi
	Poss	mi
	Loc	kimi
agr	First Sg	
proDrop	False	
empty		

youSg_Pron		
s	Full	wena
	Reduced	we
	Poss	kho
	Loc	kuwe
agr	Second Sg	
proDrop	False	
empty		

it17_Pron		
s	Full	khona
	Reduced	kho
	Poss	kho
	Loc	kukho
agr	Second Sg	
proDrop	False	
empty		

# Pre- and postdeterminers

Lincat of Predet	
s	Str
isPost	Bool

only_Predet	
s	dwa
isPost	True

all_Predet	
s	nke
isPost	True

all_pre_Predet	
s	nke
isPost	False

PredetNP : Predet -> NP -> NP

*EmphCN : CN -> CN            (or maybe EmphNP : NP -> NP ?)*

*ContrastCN : CN -> CN*

izinhlabi **zonke** zincintisane

uTatjana **yedwa** ubhukude kahle kakhulu

**yena** uTatjana uzuze indondo yegolide

izinhlabi **zona** zimbe bezidumele

***all** the swimmers competed*

***only** Tatjana swam very well*

***the same** Tatjana won the gold medal*

*the other swimmers, **on the other hand**, were disappointed*

# Possessives and relative clauses

The open classes used most frequently for determination/quantification/modification are

- nouns (PossNP)
- verbs (RelCN)

English	Zulu
most_Predet (DetCN (DetQuant IndefArt NumPl) (UseN swimmer_N))	PossNP (UseN majority_N) (DetCN (DetQuant IndefArt NumPl) (UseN swimmer_N))
most swimmers	iningi lezinhlambi ( <i>majority of the swimmers</i> )
AdjCN (PositA fast_A) (UseN swimmer_N)	RelCN (UseN swimmer_N) (UseRCI TPresTemp PPos (RelVP IdRP (UseV be_fast_V)))
fast swimmer	inhlambi eshesayo ( <i>swimmer that is-fast</i> )



# Noun phrases in GF

Lincat of NP	
s	NForm => Str
mod	Str
predet_pre	Str
predet_post	Str
agr	Agr
qdef	QuantDef
proDrop	Bool
isPron	Bool
empty	Str

word order tbd

?

ngiyabona le ndondo	<i>I see this medal</i>
angiboni le ndondo	<i>I don't see this medal</i>
ngiyabona indondo	<i>I see a/the medal</i>
angiboni indondo	<i>I don't see the medal</i>
angiboni ndondo	<i>I don't see a medal</i>

## Lincat of VP

...	...
comp	Polarity => Str

# Noun phrases in GF

PredetNP all_pre_Predet (DetCN (DetQuant that_Quant NumPl) (AdjCN (PositA big_A) (UseN stadium_N)))				
s	Full	Reduced	Poss	Loc
	izinkundla	zinkundla	zinkundla	ezinkundleni
mod	ezinkulu			
predet_pre	zo &+ nke			
predet_post				
agr	Third C9_10 Pl			
qdef	Demonstrative Dem1			
proDrop	False			
isPron	False			
empty				

*I saw **all** **those** **big** stadiums*

ngibone **zonke** **lezi** zinkundla **ezinkulu**

*I went to **all** **those** **big** stadiums*

ngiye ?

*I went to **the** stadiums*

ngiye ezinkundleni

*I went to **all** **those** **big** stadiums*

ngiye ezinkundleni **lezi** **zonke** **ezinkulu**

# Noun phrases in GF

PredetNP all_pre_Predet (DetCN (DetQuant that_Quant NumPl) (AdjCN (PositA big_A) (UseN stadium_N)))				
s	Full	Reduced	Poss	Loc
	izinkundla	zinkundla	zinkundla	ezinkundleni
mod	ezinkulu			
predet_pre	zo &+ nke			
predet_post				
agr	Third C9_10 Pl			
qdef	Demonstrative Dem1			
proDrop	False			
isPron	False			
empty				

*I saw **all** **those** **big** stadiums*

ngibone **zonke** **lezi** zinkundla **ezinkulu**

*I went to **all** **those** **big** stadiums*

ngiye ?

*I went to **the** stadiums*

ngiye ezinkundleni

*I went to **all** **those** **big** stadiums*

**ngiye ezinkundleni lezi zonke ezinkulu**

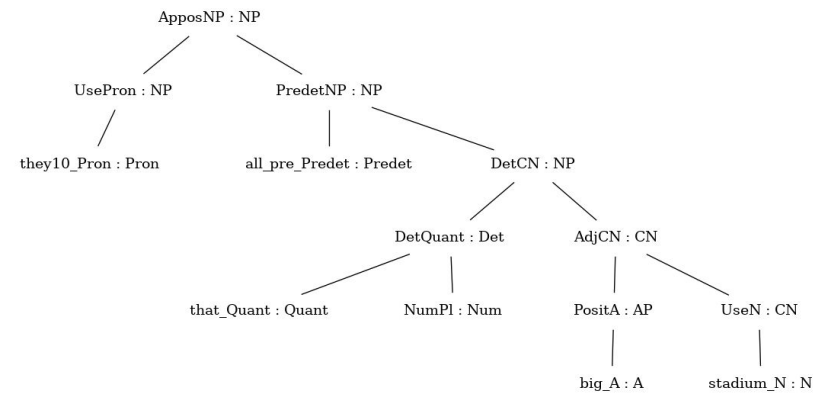
ngiye kuzo **zonke lezi** zinkundla **ezinkulu**

*I went to **them**, **all** **those** **big** stadiums*

# Noun phrases in GF

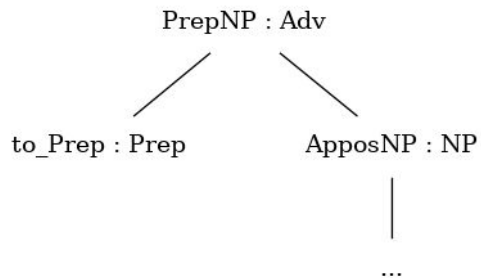
PredetNP all_pre_Predet (DetCN (DetQuant that_Quant NumPl) (AdjCN (PositA big_A) (UseN stadium_N)))				
s	Full	Reduced	Poss	Loc
	izinkundla	zinkundla	zinkundla	ezinkundleni
mod	ezinkulu			
predet_pre	zo &+ nke			
predet_post				
agr	Third C9_10 Pl			
qdef	Demonstrative Dem1			
proDrop	False			
isPron	False			
empty				

ApposNP : NP -> NP -> NP



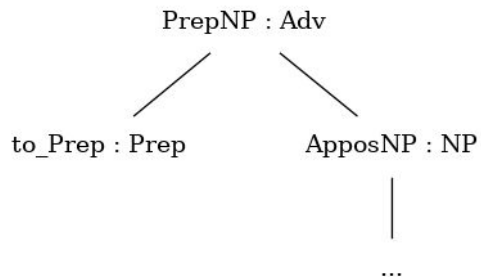
*I went to all those big stadiums*  
 ngiye ezinkundleni lezi zonke ezinkulu  
 ngiye kuzo zonke lezi zinkundla ezinkulu  
*I went to them, all those big stadiums*

# Locatives (“adverbs”)



kuzo zonke lezi zinkundla ezinkulu  
*to them, all those big stadiums*

# Locatives (“adverbs”)



kuzo zonke lezi zinkundla ezinkulu

*to them, all those big stadiums*

*in them, all those big stadiums*

*on them, all those big stadiums*

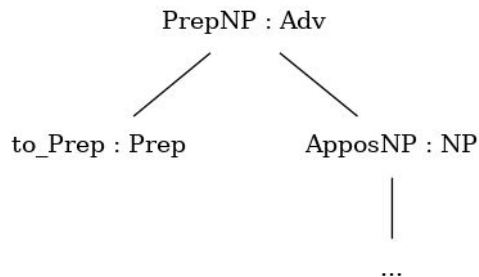
*below them, all those big stadiums*

*phakathi*

*phezulu*

*phansi*

# Locatives (“adverbs”)



kuzo zonke lezi zinkundla ezinkulu

*to them, all those big stadiums*

*in them, all those big stadiums*

*on them, all those big stadiums*

*below them, all those big stadiums*

*phakathi*

*phezulu*

*phansi*

*LocNPAdv : NP -> Adv*

esitejini

*(to, in, on, below) the podium*

*KwaNPAdv : NP -> Adv*

kwa-Tatjana

*at Tatjana's place*

*InstrNPAdv : NP -> Adv*

ngamusa

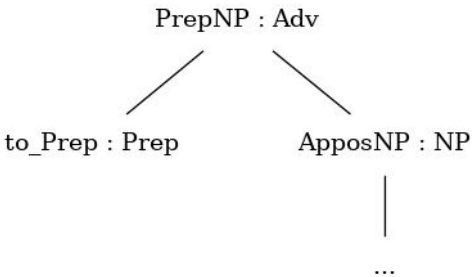
*graciously / with grace*

*KwaAdvNPAdv : Adv -> NP -> Adv*

ngaphezu kwesiteji

*on top of the podium*

# Locatives (“adverbs”)



kuzo zonke lezi zinkundla ezinkulu  
*to them, all those big stadiums*  
*in them, all those big stadiums*      *phakathi*  
*on them, all those big stadiums*      *phezulu*  
*below them, all those big stadiums*      *phansi*

<i>LocNPAdv : NP -&gt; Adv</i>	esitejini	<i>(to, in, on, below) the podium</i>
<i>KwaNPAdv : NP -&gt; Adv</i>	kwa-Tatjana	<i>at Tatjana's place</i>
<i>InstrNPAdv : NP -&gt; Adv</i>	ngamusa	<i>graciously / with grace</i>
<i>KwaAdvNPAdv : Adv -&gt; NP -&gt; Adv</i>	ngaphezu kwesiteji	<i>on top of the podium</i>

above_Adv	
s	ngaphezu
reqLocS	False



# Quick verb morphology

Params	
RForm	R_a   R_ile   R_e   R_i   R_anga
RInit	RA   RE   RI   RO   RU   RC
Syl	SylMono   SylMult
Voice	Active   Passive

Lincat of V, V2, VS	
s	RForm => Str
r	RInit
syl	Syl
voice	Voice

bon_V2 (see)					
s	R_a	R_ile	R_e	R_i	R_anga
	bona	bonile	bone	boni	bonanga
r	RC				
syl	SylMult				
voice	Active				

th_V2 (say)					
s	R_a	R_ile	R_e	R_i	R_anga
	thi	thile	the	thi	thanga
r	RC				
syl	SylMono				
voice	Active				

# Quick verb morphology

Params	
DMood	Princ   Part
Aspect	Null   Prog   Excl
BasicTense	PerfTense   PastTense   PresTense   FutTense
ZTense	Absolute BasicTense   Relative BasicTense BasicTense
VForm	VFIndic DMood Polarity BasicTense Aspect   VFPot DMood Polarity Aspect   VFSubj Polarity
VPType	CopIdent   CopAssoc   CopDescr   CopEq   VNPCompl   NoComp   VSCompl   AdvComp

Opers	
rform	VForm -> Bool -> RForm = \vform,longform -> ...
tensePref, negPref, negPrefNga, negPrefNge	VForm -> Str = \vform -> ...
subjConc	VForm -> Agr -> Bool -> Str = \vform,agr,prevow -> ...
objConc	Agr -> RInit -> Syl -> Str = \agr,rinit,syl -> ...

```
PredVP np vp = case vp.vptype of {  
  CopIdent => cl_with_id_cop_predicate np vp ;  
  CopAssoc => cl_with_ass_cop_predicate np vp ;  
  CopDescr => cl_with_descr_cop_predicate np vp ;  
  CopEq => cl_with_eq_cop_predicate np vp ;  
  AdvComp => cl_with_adv_comp_predicate np vp ;  
  _ => cl_with_verb_predicate np vp  
} ;
```

a &+  
ka &+  
zu &+  
zuza  
= akazuzuza

# Conclusion

Noun phrase is nearing completion

Next up: revamp TAMP

Current challenges

- If wrong, how wrong?
- Ambiguity, overgeneration, compiling

Future challenges

- Multilingual lexicon/GF WordNet
- Chunk-like parsing