Outline

- Introduction
- Qualificatives in Zulu
- Problem
- Zulu Resource Grammar
- Solution
- Evaluation
- Discussion
- Conclusion

Introduction: Zulu

- Zulu → Nguni → Southern Bantu → Bantu
- One of 11 official languages of South Africa
- Home language of 25% of South Africans (12M in 2011)

- Agglutinative morphology
- Nominal classification (nouns exhibit "class gender")
- Concordial agreement
  - **u**muntu **omu**hle (1) *nice person*
  - **in**daba **en**hle (9) *nice story*
  - **izi**hlala **ezin**hle (8) *nice trees*

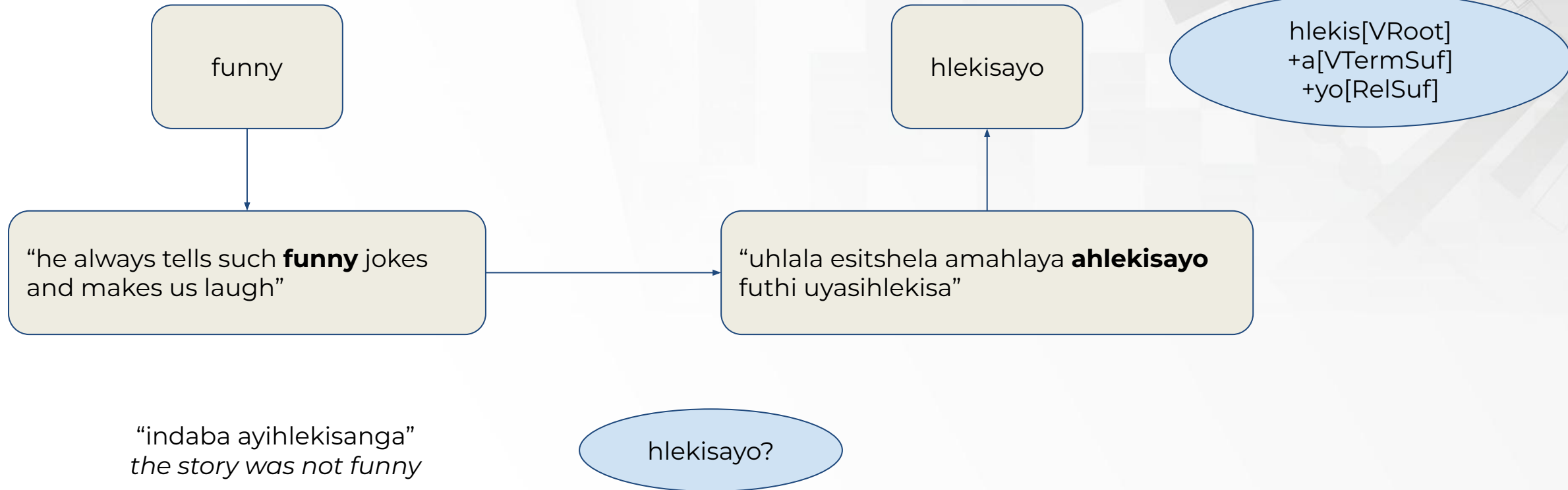CSIR
Touching lives through innovation

Introduction: AfWN

- PWN for NLP inspired AfWN (ZWN) for NLP (Bosch et al, since 2008)
  - ZWN
    - n: 4907
    - v: 1523
    - a: 1590 (1338)

- Why the expand model?
  - Under-resourced languages
  - Shared underlying structure

- Developers of AfWN confronted with the problem of adjectives
  - Often complex constructions based on nouns and verbs

Qualification in Zulu

- English adjectives seldom map to Zulu adjectives

- Qualificatives
  - Adjectives (small closed class) **(big, small, tall etc)**
  - *Descriptive possessive* **(of gold → golden)**
  - Verbal relative **(which does not see → blind)**
  - Copulative relative **(which is a lie → false, which has noise → noisy)**

- Mojapelo (2014): "[t]he challenge ... is that while they [qualificatives] are all meaning equivalents of the same English word category, they straddle a number of morphosyntactic categories in Northern Sotho, which nevertheless share a semantic function."

CSIR
Touching lives through innovation

Problem: Qualificatives in the ZWN

funny

"he always tells such **funny** jokes and makes us laugh"

"uhlala esitshela amahlaya **ahlekisayo** futhi uyasihlekisa"

hlekisayo

hlekis[VRoot]
+a[VTermSuf]
+yo[RelSuf]

"indaba ayihlekisanga"
the story was not funny

hlekisayo?

class gender → tense → polarity → attr/pred → long/short → **380+** forms

CSIR
Touching lives through innovation

Problem: diverse morphosyntax

English sense	ZWN written form	Past negative predicative (class 9)	Essence?
nice	hle	**ibingeyin**hle	hle [ADJ]
funny	hlekis**ayo**	**ayi**hlekis**anga**	hlekis [V]
blind	**nga**bon**i**	**i**bon**ile** / **i**bon**e**	bon [V] + NEG
false	ngamanga	**ibinge**ngamanga	nga + amanga [COP]
believable	n**o**kukholwa	**ibinge**nakukholwa	na + (u)kukholwa [COP assoc]

All these forms are predicate based qualificatives → there is some **verb phrase** implicit in every written form!

A sophisticated computational solution is required to effectively deal with the complexity of the adjective-like qualificatives in the ZWN.

CSIR
Touching lives through innovation

Problem: surface forms

	Form	Tense	Pol	Zulu	English
Class 1	Attributive	Pres	Pos	umuntu o**ngaboni**	the person who is blind
			Neg	umuntu o**bon**a(yo)	the person who is not blind
		Past	Pos	umuntu o**ngabon**anga	the person who was blind
			Neg	umuntu o**bon**ile(yo)	the person who was not blind
		Future	Pos	umuntu o**nga**zuku**bon**a	the person who will be blind
			Neg	umuntu ozo**bon**a	the person who will not be blind
	Predicative	Present	Pos	umuntu aka**boni**	the person is blind
			Neg	umuntu uya**bon**a	the person is not blind
			Pos	umuntu aka**bon**anga	the person was blind
			Neg	umuntu u**bon**(il)e	the person was not blind
			Pos	umuntu akazuku**bon**a	the person will be blind
			Neg	umuntu uzo**bon**a	the person will not be blind

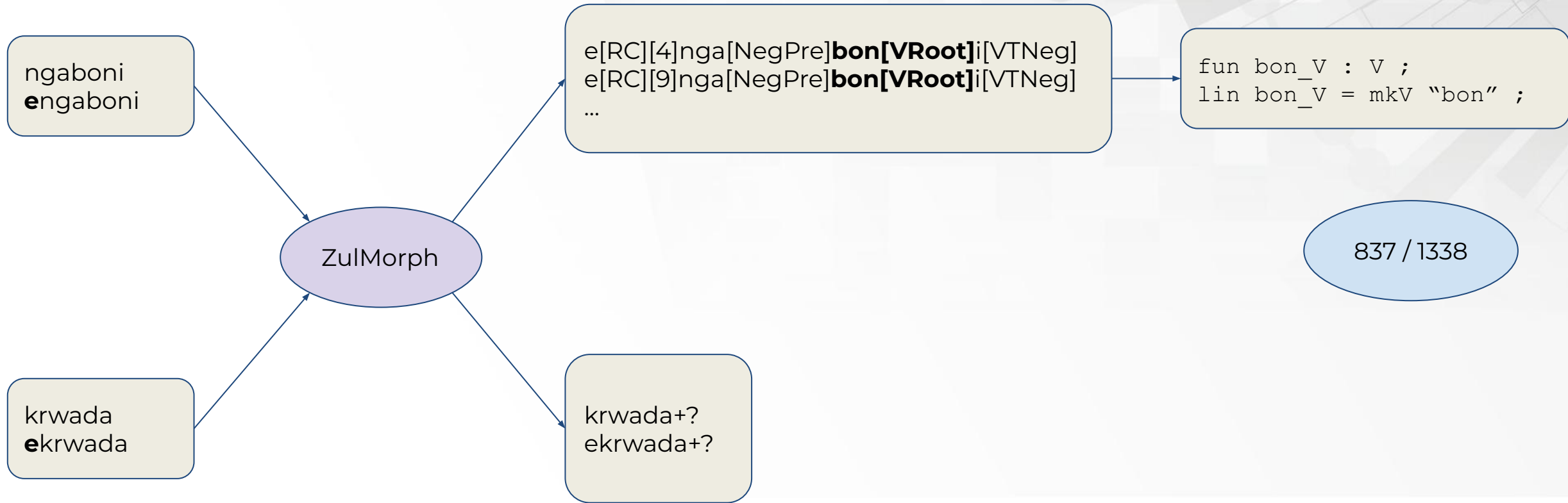Examples of Zulu qualificatives derived from the Zulu written form **ngaboni** ('blind')

Zulu GF Resource Grammar

- Grammatical Framework resource grammar

  - Implements morphosyntax of Zulu

  - First available general purpose computational grammar for Zulu (indeed, any African language)

  - GF runtime enables parsing (text to tree) and linearising (tree to text)

CSIR
Touching lives through innovation

Solution: overview

- **Recover** implicit morphosyntactic constructions from the written forms by parsing them using the ZRG

- Provide functionality to **generate** and **analyse** full forms of these constructions

- Mostly **automatic** process which can be reused
  - for future versions of the ZWN
  - for the other languages in the AfWN once resource grammars for these languages are available
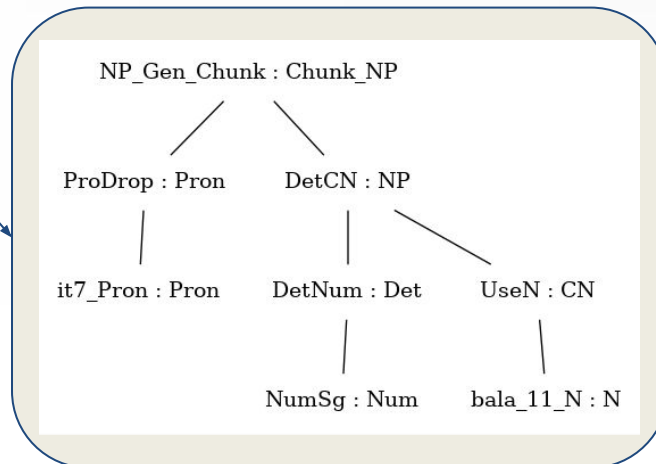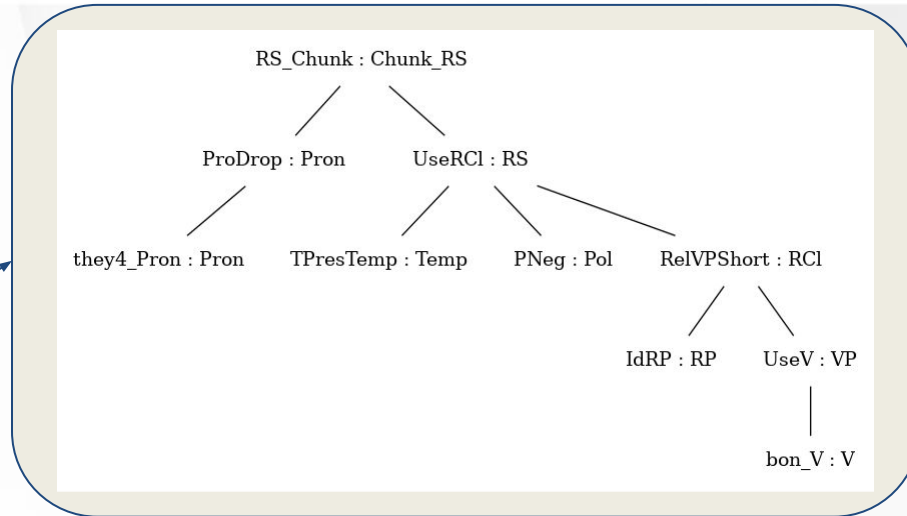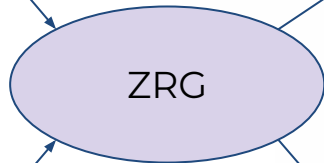
CSIR
Touching lives through innovation

Solution: preparing to parse

ngaboni
engaboni

krwada
ekrwada

ZulMorph

e[RC][4]nga[NegPre]**bon[VRoot]**i[VTNeg]
e[RC][9]nga[NegPre]**bon[VRoot]**i[VTNeg]
…

```
fun bon_V : V ;
lin bon_V = mkV "bon" ;
```

837 / 1338

krwada+?
ekrwada+?

Solution: parsing written forms

ngaboni
engaboni

sobala

ZRG

RS_Chunk : Chunk_RS
├── ProDrop : Pron
│   └── they4_Pron : Pron
└── UseRCl : RS
    ├── TPresTemp : Temp
    ├── PNeg : Pol
    └── RelVPShort : RCl
        ├── IdRP : RP
        └── UseV : VP
            └── bon_V : V

NP_Gen_Chunk : Chunk_NP
├── ProDrop : Pron
│   └── it7_Pron : Pron
└── DetCN : NP
    ├── DetNum : Det
    │   └── NumSg : Num
    └── UseN : CN
        └── bala_11_N : N

Parses	Number	%
RS	628	0.75
NP	104	0.12
confident	(628 + 104) = 732	0.88
not confident	105	0.12

Solution: mapping

English sense	ZWN written form	Function	VP	Polarity
nice	hle	nice_1_A	CopAP (PositA hle_A)	Pos
funny	hlekisayo	funny_1_A	UseV hlekis_V	Pos
blind	ngaboni	blind_1_A	UseV bon_V	Neg
false	ngamanga	false_1_A	CopNP (DetCN (DetNum NumPl) (UseN anga_6_N))	Pos
believable	nokukholwa	believable_1_A	CopNPAssoc (Deverb15 PPos (UseV kholw_V))	Pos

Solution: adjective grammar

```
lincat
  ZWN_Adjective = Str ;
  ZWN_S = S ;
  ZWN_Pron = Pron ;
  ZWN_NP = NP ;
  ZWN_APred = { vp : VP ; pol : ZPol } ;
  ZWN_Temp = Temp ;
  ZWN_Pol = { id : MonoLexLangZul.Pol ; di : MonoLexLangZul.Pol } ;

lin
  PAdjective s = s.s ;
  AAdjective np = np.s!NFull ;

  PredicVP temp pol pron apred = UseCl temp (polarity pol apred.pol) (PredVP (UsePron (ProDrop pron)) apred.vp) ; --

  AttribVP temp pol pron apred = RelNP (UsePron (ProDrop pron)) (UseRCl temp (polarity pol apred.pol) (RelVP IdRP apred.vp)) ;

  abject_1_A =       { vp = UseV enyanyek_V ; pol = ZPos } ;
  able_1_A =         { vp = CopNPAssoc (DetCN (DetNum NumSg) (UseN khono_5_6_N)) ; pol = ZPos } ;
  aboriginal_1_A = { vp = CopNPAssoc (DetCN (DetNum NumSg) (UseN velo_9_10_N)) ; pol = ZPos } ;
  aboriginal_2_A = { vp = UseV qal_V ; pol = ZPos } ;
  aboriginal_3_A = { vp = CopNPAssoc (DetCN (DetNum NumPl) (UseN ndulo_5_6_N)) ; pol = ZPos } ;
  abrupt_1_A =       { vp = UseV qubulis_V ; pol = ZPos } ;
  absolute_1_A =     { vp = UseVStative phelel_V ; pol = ZPos } ;
  absolved_1_A =     { vp = UseVStative khululw_V ; pol = ZPos } ;
  abstract_1_A =     { vp = CopNPAssoc (Deverb15 PPos (UseV bambek_V)) ; pol = ZNeg } ;
  abstruse_1_A =     { vp = UseVStative endel_V ; pol = ZPos } ;
  abundant_1_A =     { vp = CopAP (PositA ningi_A) ; pol = ZPos } ;
  acceptable_1_A = { vp = UseV mukelek_V ; pol = ZPos } ;
  accessible_1_A = { vp = UseV ngenek_V ; pol = ZPos } ;
  accessible_2_A = { vp = UseV finyelelek_V ; pol = ZPos } ;
  accidental_1_A = { vp = UseV enzekel_V ; pol = ZPos } ;
  accurate_1_A =     { vp = UseV nemb_V ; pol = ZPos } ;
  accusative_1_A = { vp = UseV mangalel_V ; pol = ZPos } ;
  achromatic_1_A = { vp = CopAP (PositA mhlophe_R) ; pol = ZPos } ;
  acquitted_1_A =   { vp = UseVStative khululw_V ; pol = ZPos } ;
```
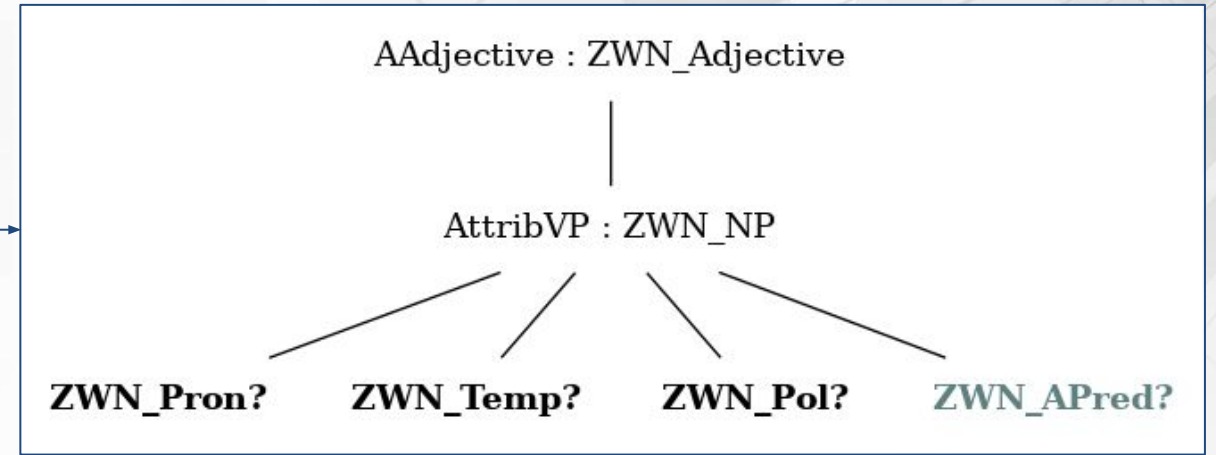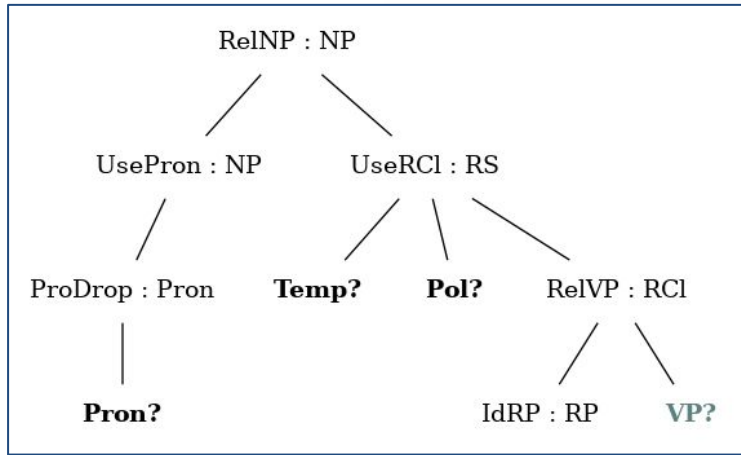
CSIR
Touching lives through innovation

Solution: adjective grammar
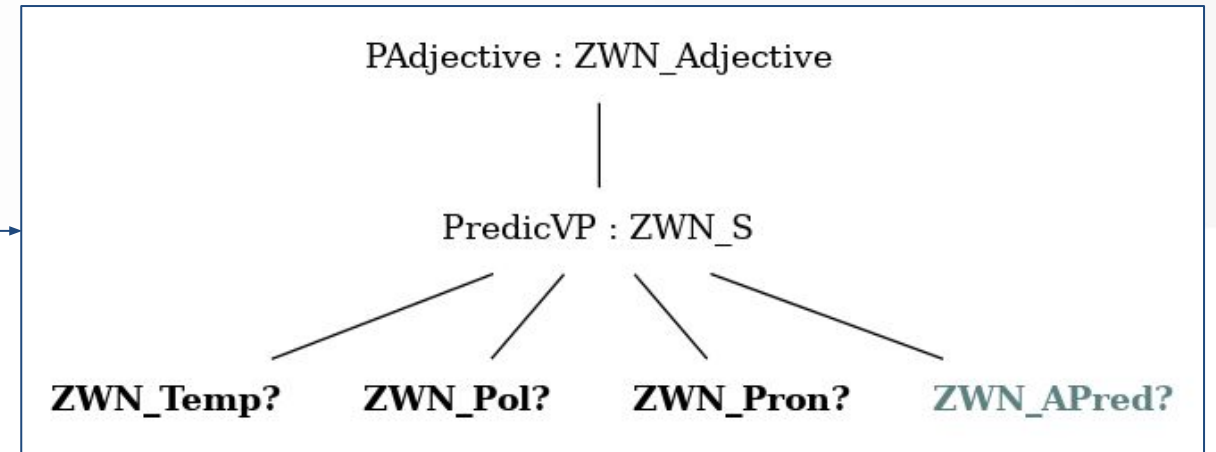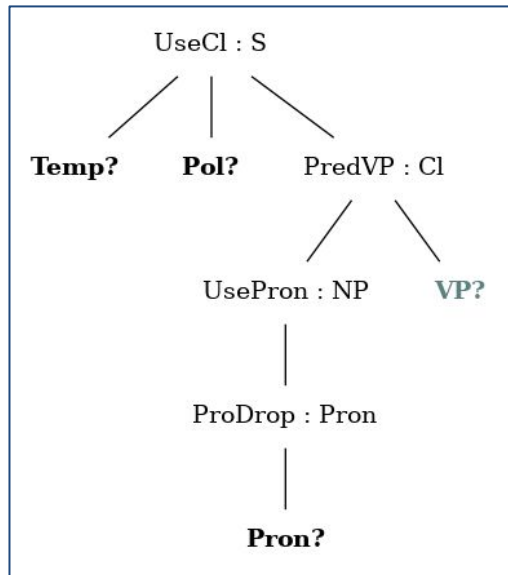
Solution: wrapper tool

```
gf-afwn$ python3 afwn_adjectives.py generate Pres Pos 1 Attr blind

Tense        Polarity    Class    Form       Long/short Qualificative
--------------------------------------------------------------------
Pres         Pos         1        Attr                  ongaboni
```

```
gf-afwn$ python3 afwn_adjectives.py generate Fut Pos 5 Pred thundering

Tense        Polarity    Class    Form       Long/short Qualificative
--------------------------------------------------------------------
Fut          Pos         5        Pred                  lizomangalisa kakhulu
```

CSIR
Touching lives through innovation

Solution: wrapper tool

```
gf-afwn$ python3 afwn_adjectives.py generate ? ? 1 ? blind
```

Tense	Polarity	Class	Form	Long/short	Qualificative
Fut	Neg	1	Attr		ozobona
Fut	Neg	1	Pred		uzobona
Fut	Pos	1	Attr		ongazukubona
Fut	Pos	1	Pred		akazukubona
Past	Neg	1	Attr	long	obonileyo
Past	Neg	1	Attr	short	obone
Past	Neg	1	Pred	long	ubonile
Past	Neg	1	Pred	short	ubone
Past	Pos	1	Attr		ongabonanga
Past	Pos	1	Pred		akabonanga
Pres	Neg	1	Attr	long	obonayo
Pres	Neg	1	Attr	short	obona
Pres	Neg	1	Pred	long	uyabona
Pres	Neg	1	Pred	short	ubona
Pres	Pos	1	Attr		ongaboni
Pres	Pos	1	Pred		akaboni
RemFut	Neg	1	Attr		oyobona
RemFut	Neg	1	Pred		uyobona
RemFut	Pos	1	Attr		ongayukubona
RemFut	Pos	1	Pred		akayukubona
RemPast	Neg	1	Attr		owabona
RemPast	Neg	1	Pred		wabona
RemPast	Pos	1	Attr		ongabonanga
RemPast	Pos	1	Pred		akabonanga

17

Solution: wrapper tool

```
gf-afwn$ python3 afwn_adjectives.py analyze ayizukungeneka

Tense        Polarity    Class    Form       Long/short Adjective
-----------------------------------------------------------------
Fut          Neg         4        Pred                  accessible
Fut          Neg         9        Pred                  accessible


gf-afwn$ python3 afwn_adjectives.py analyze obona

Tense        Polarity    Class    Form       Long/short Adjective
-----------------------------------------------------------------
Pres         Neg         1        Attr       short      blind
Pres         Neg         1a       Attr       short      blind
Pres         Neg         3        Attr       short      blind
Pres         Pos         1        Attr       short      conscious
Pres         Pos         1a       Attr       short      conscious
Pres         Pos         3        Attr       short      conscious
```

Evaluation

- How to evaluate?
- Use ZulMorph to facilitate evaluation

- 48 / 50 = 96%

Discussion

- Adjectives are a small percentage of "adjectives" in ZWN

- "Adjectives" in the ZWN are most often verbs

Qualificative property	% of Total
Verb	61.5%
UseV, UseVStative, ComplV2	
Associative copulative	18.6%
CopNPAssoc	
Locative copulative	2.2%
CopLocative	
Adjective/primitive relative	2.0%
CopAP	
Identifying copulative	1.6%
CopNP	
Negative	9.3%
PNeg	

Conclusion

- Zulu resource grammar allows us to take full advantage of morphologically complex written forms in the ZWN

- Repeatable process
  - Future versions
  - Other word categories
  - More languages

- https://github.com/LauretteM/gf-afwn

CSIR
Touching lives through innovation

Siyabonga!