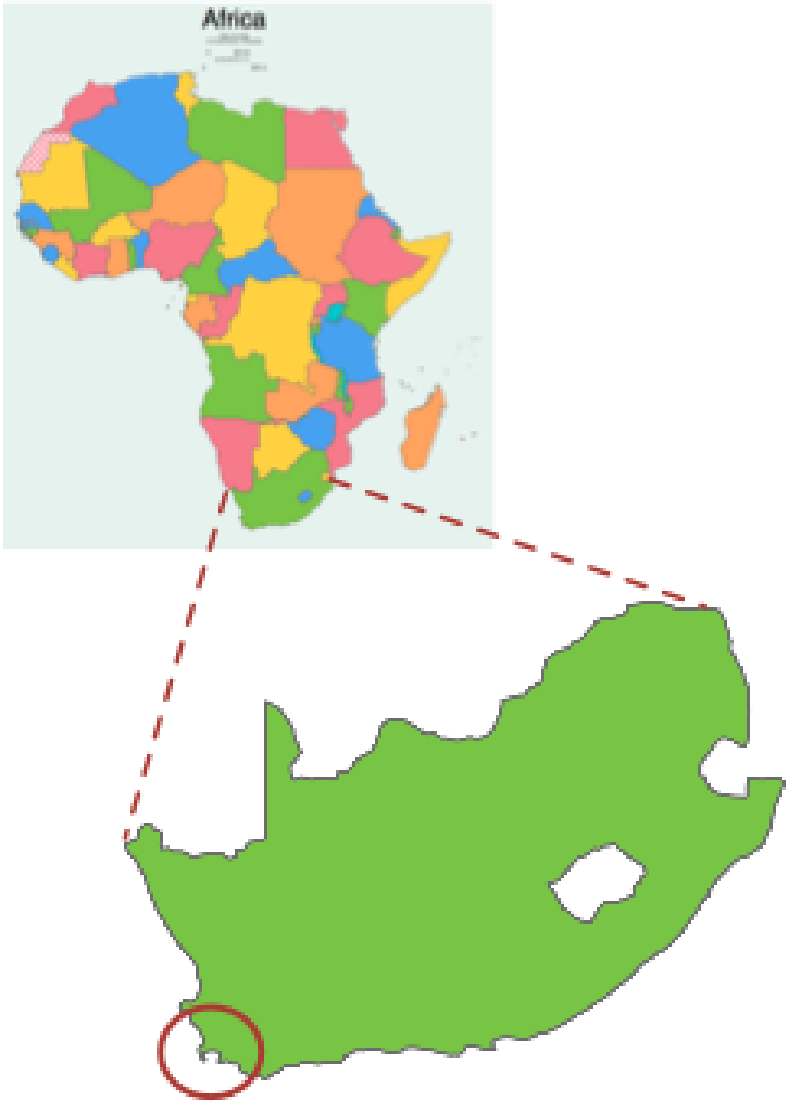
Ontology verbalisation for African languages

C. Maria Keet

Department of Computer Science
University of Cape Town, South Africa
mkeet@cs.uct.ac.za

Business Intelligence & Semantic Web Colloquium 2019
Yaoundé, Cameroon, 2 December 2019

CS@UCT



CS@UCT

CS@UCT

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Motivation

- Doing business, government services provision, etc in one's own language, beyond English and French
- (The "untapped billion", in FAANG's terminology)
- Requires tools with African languages in the interface, not just some 'pretty pictures and icons'
- The Business Intelligence entails analysing data and presenting the outcomes ourselves, also textually, for a local or regional audience
- Need to transform structured data and structured knowledge into text
- Structured input is represented in, a.o.: XML, RDF, OWL (or SQL, JSON, or excel/OO spreadsheets)

Motivation: example areas for knowledge-to-text

- Electronic health records and patient discharge notes generation
  - e.g., SNOMED CT, OpenMRS localisation
  - "The patient has as symptom fever and dizziness"
  - "The patient must drink water when taking the pills"
    "If the patient takes the pills, then he must drink water"
- Getting the relevant business logic into your app
  - Requirements engineering, data analysis (i.e., knowledge acquisition for modelling)
- Querying the data with conceptual queries in OBDA
  - "Show me all employees who are not working on a project"
- And many other areas; e.g., question generation

Motivation: example areas for knowledge-to-text

- Electronic health records and patient discharge notes generation
  - e.g., SNOMED CT, OpenMRS localisation
  - "The patient has as symptom fever **and** dizziness"
  - "The patient must drink water **when** taking the pills"
    "**If** the patient takes the pills, **then** he must drink water"
- Getting the relevant business logic into your app
  - Requirements engineering, data analysis (i.e., knowledge acquisition for modelling)
- Querying the data with conceptual queries in OBDA
  - "Show me all employees who are not working on a project"
- And many other areas; e.g., question generation

Motivation: example areas for knowledge-to-text

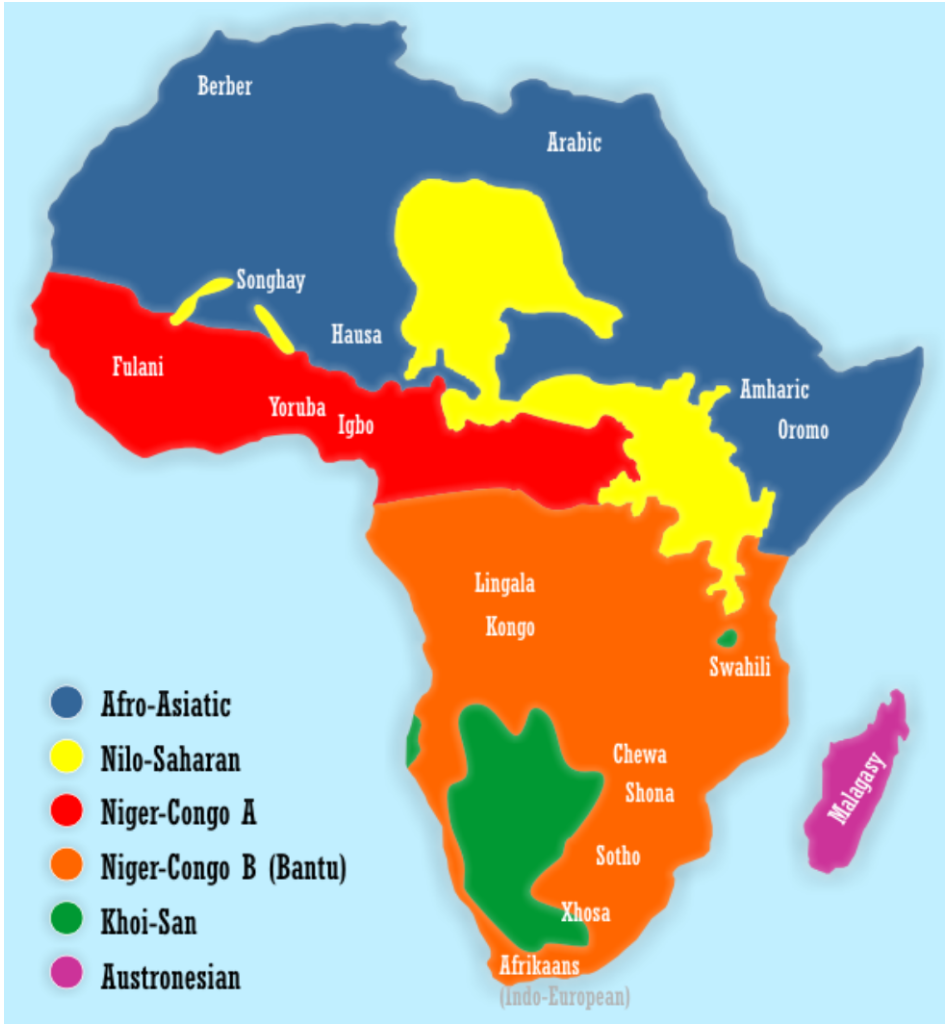
- Electronic health records and patient discharge notes generation
  - e.g., SNOMED CT, OpenMRS localisation
  - "The patient has as symptom fever **and** dizziness"
  - "The patient must drink water **when** taking the pills"
    "**If** the patient takes the pills, **then** he must drink water"
- Getting the relevant business logic into your app
  - Requirements engineering, data analysis (i.e., knowledge acquisition for modelling)
- Querying the data with conceptual queries in OBDA
  - "Show me **all** employees who are **not** working on a project"
- And many other areas; e.g., question generation

Outline

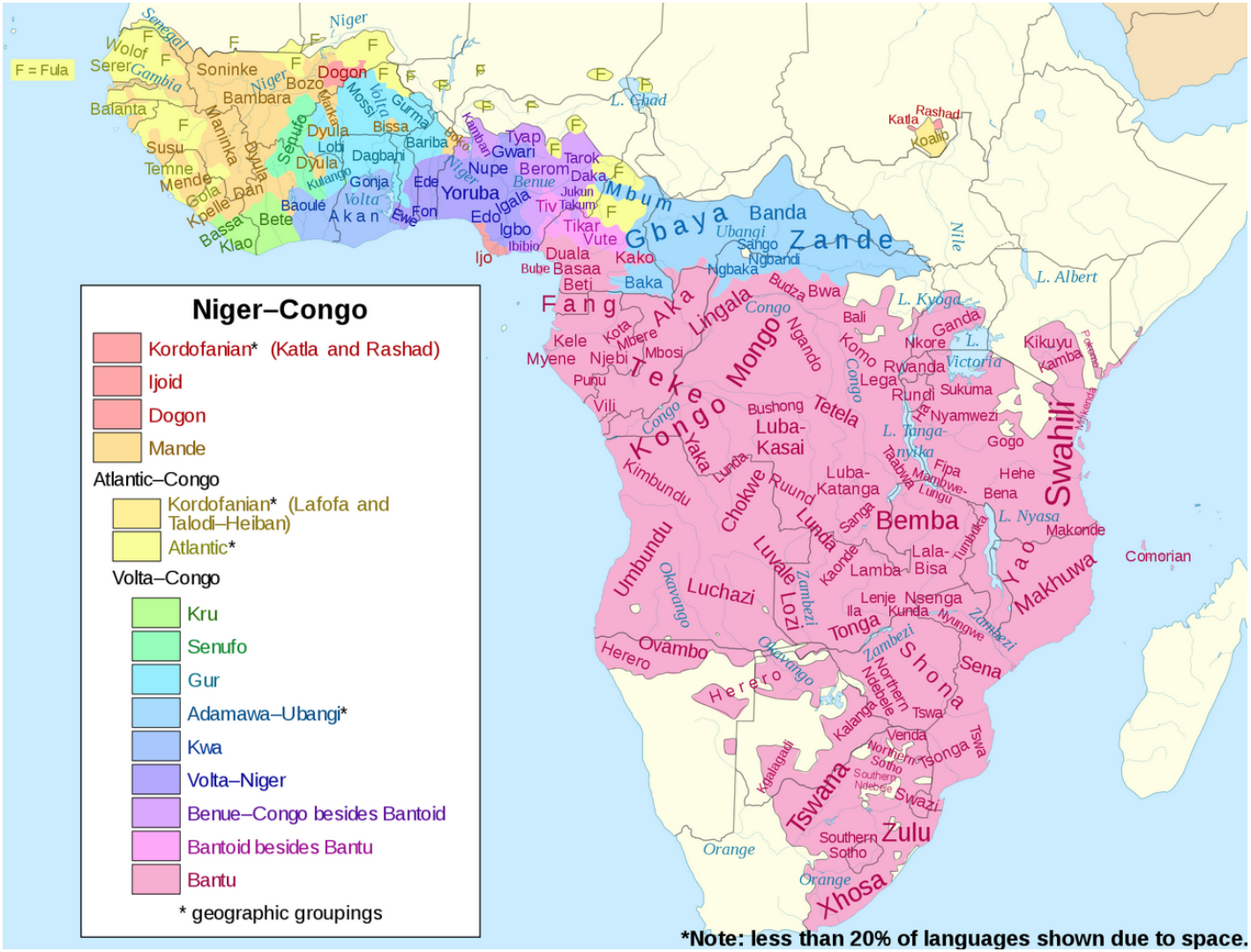
1. Motivation
   - Context
   - Language 'crash course'
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Basics

1500-2000 African languages (6 main groups) spoken by 1.2 billion people



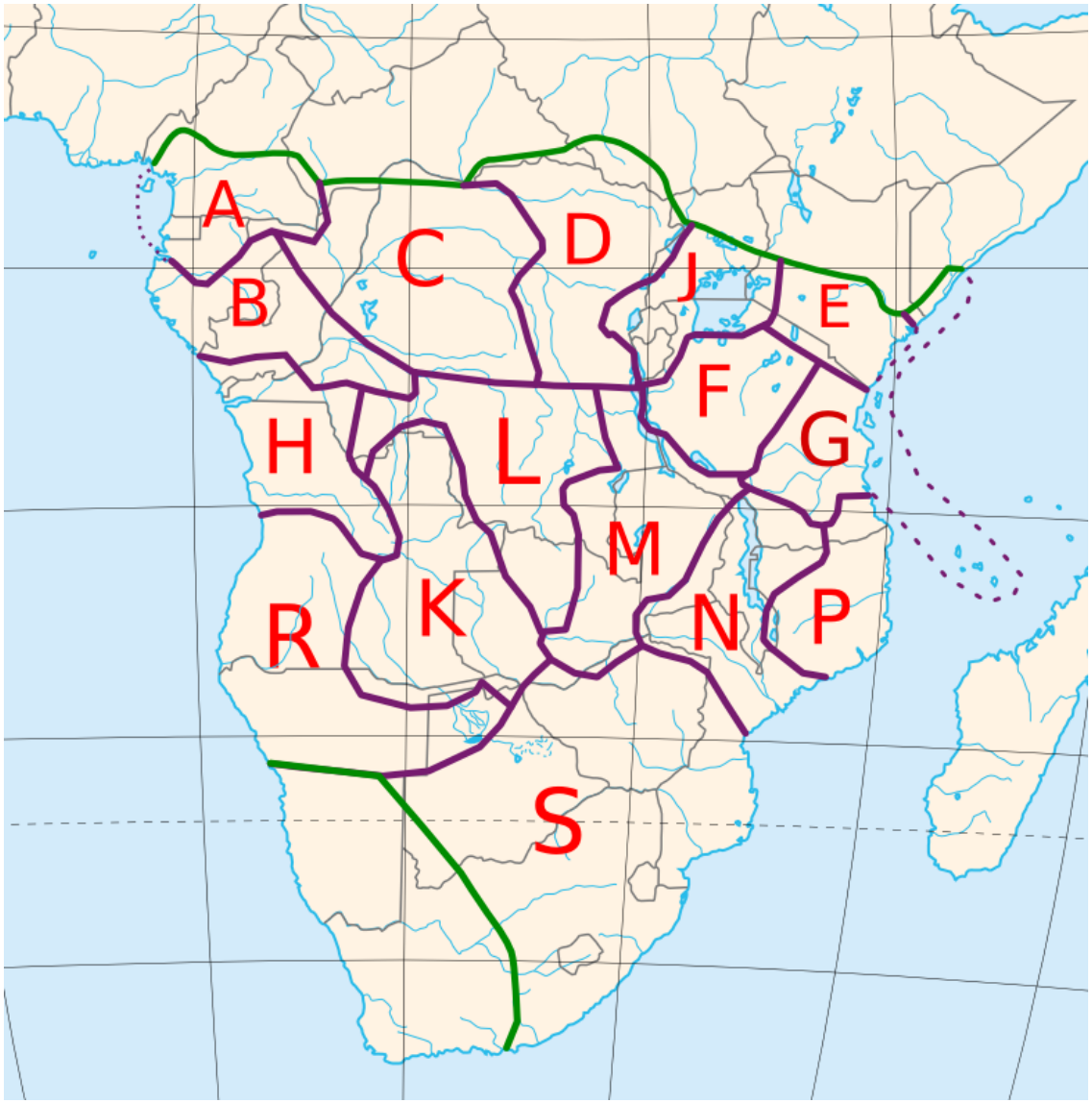
Sub-Sahara


F = Fula
Niger–Congo
Kordofanian* (Katla and Rashad)
Ijoid
Dogon
Mande
Atlantic–Congo
Kordofanian* (Lafofa and Talodi–Heiban)
Atlantic*
Volta–Congo
Kru
Senufo
Gur
Adamawa–Ubangi*
Kwa
Volta–Niger
Benue–Congo besides Bantoid
Bantoid besides Bantu
Bantu
* geographic groupings
*Note: less than 20% of languages shown due to space.

Sub-Sahara: Bantu languages

- Bantu languages: group of languages spoken in Sub-Saharan Africa
- *Bantu* means 'human'; bit of a laden term, but still used in linguistics
- Number of languages varies by who counts ($>$ 200 at least)
- Organised in so-called Guthrie zones

Guthrie Zones



Note on languages in Cameroon

- Official languages: English, French
- Recognised regional languages, includes:
  - Fula: in the Senegambian branch of Niger-Congo
  - Ewondo: trade language, in the Benu-Congo branch (and in Guthrie zone A)
  - Pidgin English, Camfranglais
- Other (about 250); and Southern-Bantu: Beti (1.7 million, includes Ewondo), Basaa (230,000), and Duala (350,000)

(according to wikipedia, d.d. 23-11-'19)

Relevant core characteristics (1/2)

- System of noun classes
  - Each noun is classified into a noun class
  - Meinhof identified 23 noun classes; not all of them used, varies by language; some refinements
  - Singular and plural pairings
  - There's semantics: e.g., NC1 for humans, NC9 for animals, NC15 infinitive nouns

NC	AU	PRE	Stem (example)	Meaning	Example (isiZulu)	
1 2	u- a-	m(u)- ba-	-fana -fana	humans and other animates	umfana abafana	boy boys
1a 2a	u- o-	- -	-baba -baba	kinship terms and proper names	ubaba obaba	father fathers
3a (2a)	u- o-	- -	-shizi -shizi	nonhuman	ushizi oshizi	cheese cheeses
3 4	u- i-	m(u)- mi-	-fula -fula	trees, plants, non-paired body parts	umfula imifula	river rivers
5 6	i- a-	(li)- ma-	-gama -gama	fruits, paired body parts, and natural phenomena	igama amagama	name names
7 8	i- i-	si- zi-	-hlalo -hlalo	inanimates and manner/ style	isihlalo izihlalo	chair chairs
9a (6)	i- a-	- ma-	-rabha -rabha	nonhuman	irabha amarabha	rubber rubbers
9 10	i(n)- i-	- zi(n)-	-ja -ja	animals	inja izinja	dog dogs
11 (10)	u- i-	(lu)- zi(n)-	-thi -thi	inanimates and long thin objects	uthi izinthi	stick sticks
14 15	u- u-	bu- ku-	-hle -cula	abstract nouns infinitives	ubuhle ukucula	beauty to sing
17		ku-		locatives, remote/ general		locative

Relevant core characteristics (2/2)

- Many of the languages are *agglutinating*
  - i.e., what are separate words in, say, English are 'components' of a word

Ex: `titukakimureeterahoganu` (Runyankore, Uganda)
'We have never ever brought it to him'

Relevant core characteristics (2/2)

- Many of the languages are *agglutinating*
  - i.e., what are separate words in, say, English are 'components' of a word

Ex: `titukakimureeterahoganu` (Runyankore, Uganda)
'We have never ever brought it to him'
`ti tu ka ki mu reet er a ho ga nu`
neg-(NC2 SC)-RM-(NC7 SC)-(NC1 SC)-VR-App-FV-Loc-Emp-Dec

Relevant core characteristics (2/2)

- Many of the languages are *agglutinating*
  - i.e., what are separate words in, say, English are 'components' of a word

Ex: `titukakimureeterahoganu` (Runyankore, Uganda)
'We have never ever brought it to him'
`ti tu ka ki mu reet er a ho ga nu`
neg-(NC2 SC)-RM-(NC7 SC)-(NC1 SC)-VR-App-FV-Loc-Emp-Dec

- System of concordial agreement (more about that soon)

Illustrative examples of some consquences (isiZulu)

- 'and', enumerative: *na-*, phonologically conditioned
  - Ex: milk and butter: *ubisi <u>ne</u>bhotela* (-a+i-=-e-)
  - Ex: butter and milk: *ibhotela <u>no</u>bisi* (-a+u-=-o-)

Illustrative examples of some consquences (isiZulu)

- 'and', enumerative: *na-*, phonologically conditioned
  - Ex: milk and butter: *ubisi nebhotela* (-a+i-=-e-)
  - Ex: butter and milk: *ibhotela nobisi* (-a+u-=-o-)
- copulative (to be): depends on first letter of noun: *ng-* for a-, o-, u-, else *y-*
  - Ex: is a dog: *yinja*
  - Ex: is a grandmother: *ngugogo*
- 'is not a': combine NEG SC with PRON, both depend on noun class
  - Ex: an animal is not a plant: *isilwane asiwona umuthi*
  - Ex: a plant is not an animal: *umuthi awusona isilwane*

Illustrative examples of some consquences (isiZulu)

- 'and', enumerative: *na-*, phonologically conditioned
  - Ex: milk and butter: *ubisi nebhotela* (-a+i-=-e-)
  - Ex: butter and milk: *ibhotela nobisi* (-a+u-=-o-)
- copulative (to be): depends on first letter of noun: *ng-* for a-, o-, u-, else *y-*
  - Ex: is a dog: *yinja*
  - Ex: is a grandmother: *ngugogo*
- 'is not a': combine NEG SC with PRON, both depend on noun class
  - Ex: an animal is not a plant: *isilwane asiwona umuthi*
  - Ex: a plant is not an animal: *umuthi awusona isilwane*
- Other verbs: concordial agreement ($\sim$ conjugation) based on noun class
  - Ex: The human eats: *umuntu udla*
  - Ex: The dog eats: *inja idla*

Concordial agreement—example (isiZulu, South Africa)

Abafana abancane bazozithenga izincwadi ezinkulu
aba-fana **aba**-ncane **ba**- zo- **zi**- thenga **izi**-ncwadi e-**zi**-nkulu
2.boy **2**.small **2.SUBJ**-FUT-**10.OBJ**-buy **10**.book REL-**10**.big
'The little boys will buy the big books'

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

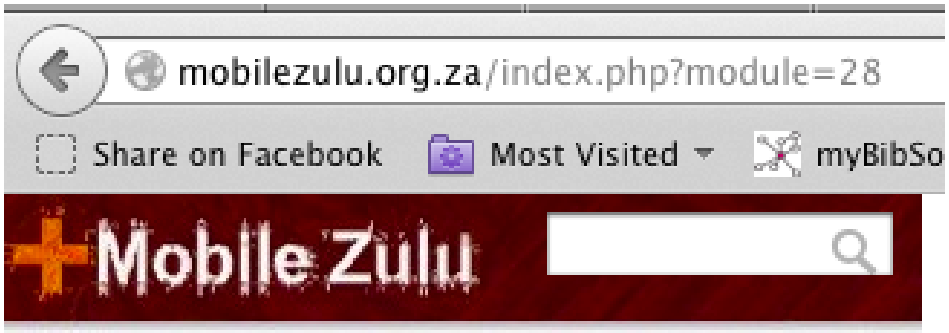
Outline

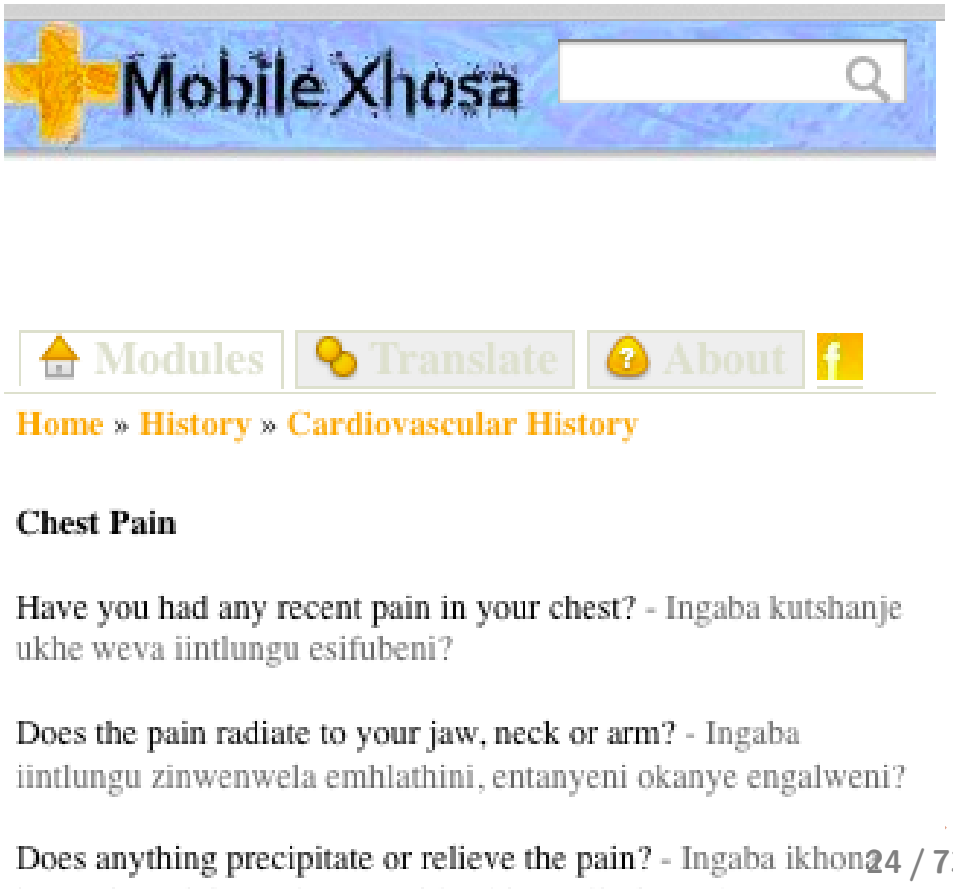
1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Short answer

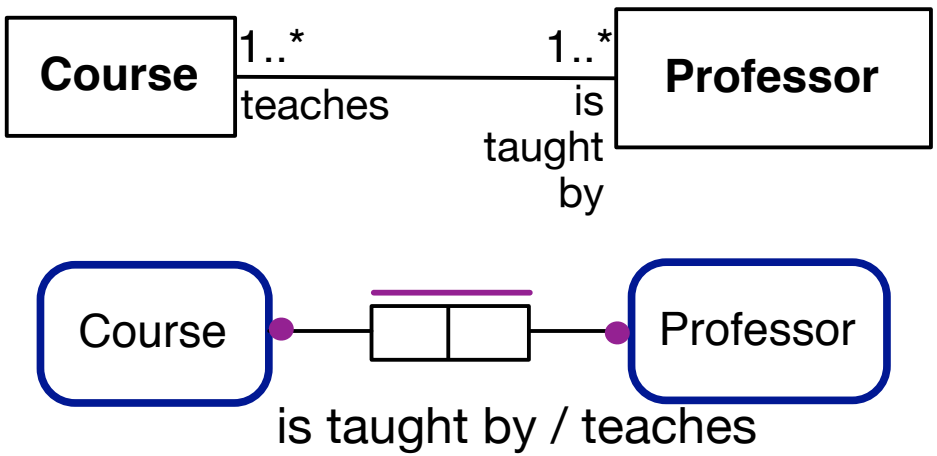
- **C**controlled **N**atural **L**anguage: constrain the grammar/vocabulary of a natural language
- **N**atural **L**anguage **G**eneration: generate natural language text from structured data, information, or knowledge

Ex: S. Moolla's mobile healthcare app with **canned text**





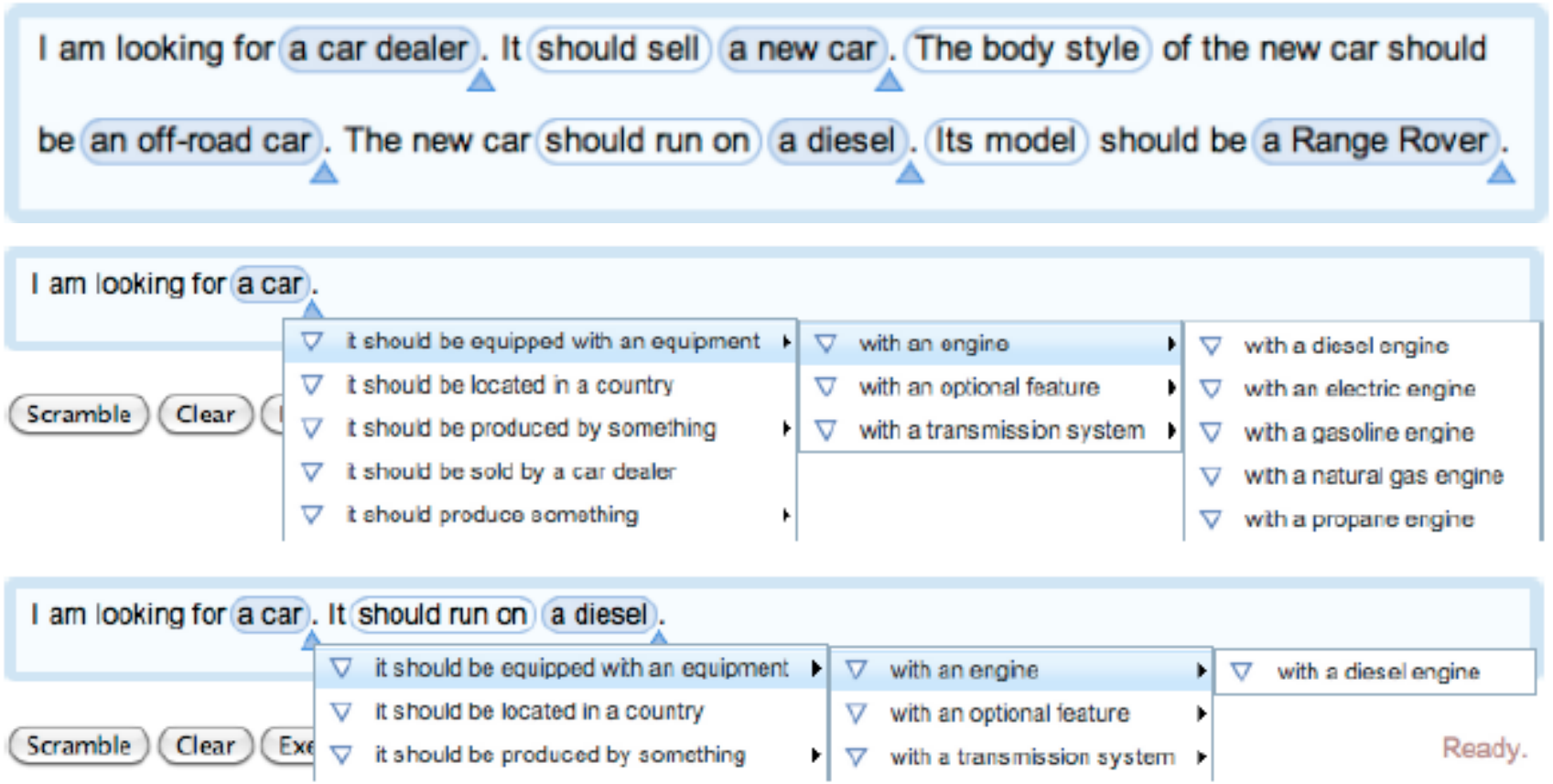
Ex: Business rules and conceptual data models with *static* templates



Each Course is taught by **at least one** Professor

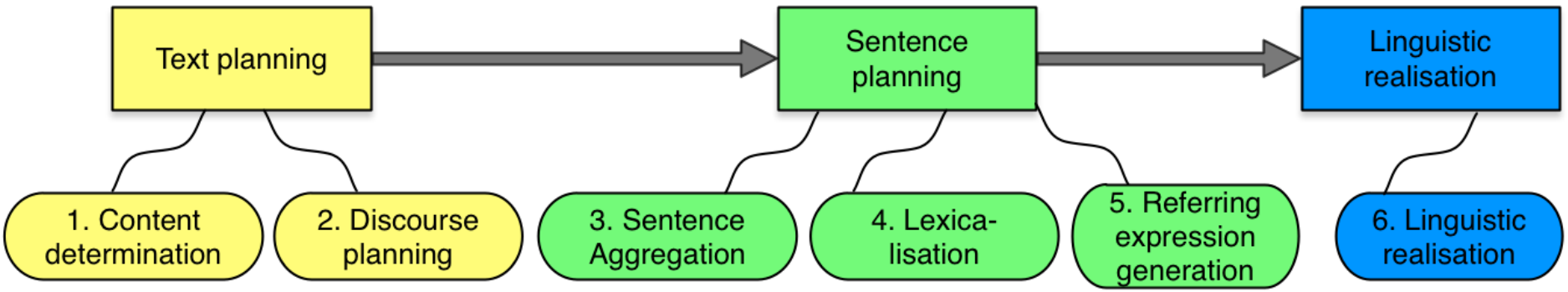
Each Professor teaches **at least one** Course

Ex.: Query formulation with Quelo [Franconi et al.(2010)] with *context-sensitive* **templates**



Pictures from: Quelo @ The IESD Challenge 2012

The 'NLG pipeline'



1. What structured data/info/ knowledge do you want to put into NL sentences?
2. In what order should it be presented?

3. Which messages to put together into a sentence?
4. Which words and phrases will it use for each domain concept and relation?
5. Which words or phrases to select to identify domain entities?

6. Use grammar rules to produce syntactically, morphologically, and orthographically correct (and is also meaningful)

NLG, principal approaches to generate the text

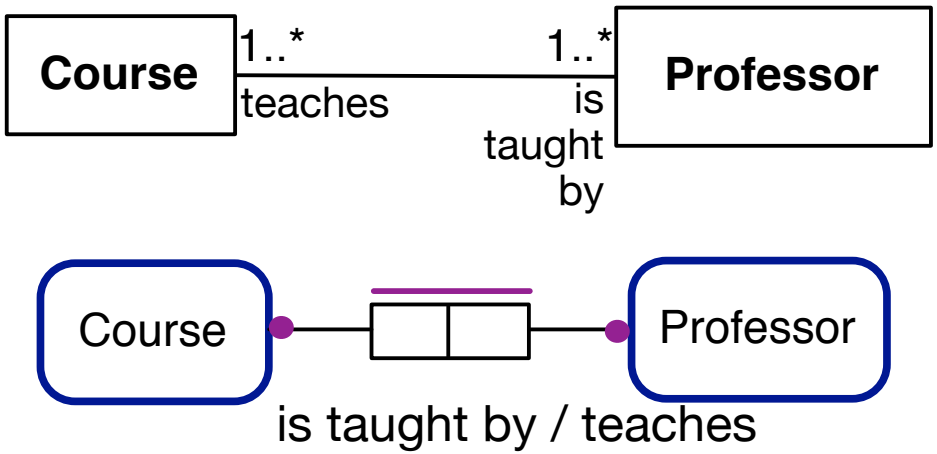
- Canned text
- Templates
  - Notably for English [Fuchs et al.(2010), Schwitter et al.(2008), Third et al.(2011), Curland and Halpin(2007)],
  - but also other languages [Jarrar et al.(2006)]
- Grammar engines, such as [Kuhn(2013)], Grammatical Framework (`http://www.grammaticalframework.org/`), SimpleNLG

NLG, principal approaches to generate the text

- Canned text
- Templates
  - Notably for English [Fuchs et al.(2010), Schwitter et al.(2008), Third et al.(2011), Curland and Halpin(2007)],
  - but also other languages [Jarrar et al.(2006)]
- Grammar engines, such as [Kuhn(2013)], Grammatical Framework (`http://www.grammaticalframework.org/`), SimpleNLG

$\Rightarrow$ Hand-crafted or ML/neural-based

Business rules/conceptual data models and logic reconstruction



BR: **Each** Course is taught by **at least one** Professor

FOL: $\forall x\ (\text{Course}(x) \rightarrow \exists y\ (\text{is\_taught\_by}(x, y) \wedge \text{Professor}(y)))$

DL: $\text{Course} \sqsubseteq \exists\ \text{is\_taught\_by}.\text{Professor}$

- (i.e., a mandatory constraint / existential quantification)

Example of templates

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory] Cada</Text>
 <Object index="0"/>
 <Text>debe</Text>
 <Role index="0"/>
 <Text>al menos un(a)</Text>
 <Object index="1"/>
</Constraint>
```

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory]  Each</Text>
 <Object index="0"/>
 <Text>must</Text>
 <Role index="0"/>
 <Text>at least one</Text>
 <Object index="1"/>
</Constraint>
```

for a large fragment of ORM, and 11 languages [Jarrar et al.(2006)]

Example of templates

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory] Cada</Text>
 <Object index="0"/>
 <Text>debe</Text>
 <Role index="0"/>
 <Text>al menos un(a)</Text>
 <Object index="1"/>
</Constraint>
```

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory]  Each</Text>
 <Object index="0"/>
 <Text>must</Text>
 <Role index="0"/>
 <Text>at least one</Text>
 <Object index="1"/>
</Constraint>
```

for a large fragment of ORM, and 11 languages [Jarrar et al.(2006)]

Example of templates

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory] Cada</Text>
 <Object index="0"/>
 <Text>debe</Text>
 <Role index="0"/>
 <Text>al menos un(a)</Text>
 <Object index="1"/>
</Constraint>
```

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory]  Each</Text>
 <Object index="0"/>
 <Text>must</Text>
 <Role index="0"/>
 <Text>at least one</Text>
 <Object index="1"/>
</Constraint>
```

for a large fragment of ORM, and 11 languages [Jarrar et al.(2006)]

Example of templates

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory] Cada</Text>
 <Object index="0"/>
 <Text>debe</Text>
 <Role index="0"/>
 <Text>al menos un(a)</Text>
 <Object index="1"/>
</Constraint>
```

```
<Constraint xsi:type="Mandatory">
 <Text> -[Mandatory]  Each</Text>
 <Object index="0"/>
 <Text>must</Text>
 <Role index="0"/>
 <Text>at least one</Text>
 <Object index="1"/>
</Constraint>
```

for a large fragment of ORM, and 11 languages [Jarrar et al.(2006)]

NL Grammars, illustration

$$
\begin{aligned}
\textit{Sentence} &\longrightarrow \textit{NounPhrase} \mid \textit{VerbPhrase} \\
\textit{NounPhrase} &\longrightarrow \textit{Adjective} \mid \textit{NounPhrase} \\
\textit{NounPhrase} &\longrightarrow \textit{Noun} \\
&\ldots \\
\\
\textit{Noun} &\longrightarrow \textit{car} \mid \textit{train} \\
\textit{Adjective} &\longrightarrow \textit{big} \mid \textit{broken} \\
&\ldots
\end{aligned}
$$

(and complexity of the grammar)

Outline

1. Motivation
   - Context
   - Language 'crash course'
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Question

- Is this template-based approach useable for Bantu (or Niger-Congo) languages, be they agglutinating or not?

Question

- Is this template-based approach useable for Bantu (or Niger-Congo) languages, be they agglutinating or not?
  - Short answer: No

Question

- Is this template-based approach useable for Bantu (or Niger-Congo) languages, be they agglutinating or not?
  - Short answer: No
- Tasks:
  - For structured input: use a practically useful language with tool support already (Sem Web tech)
  - Start with basics for a grammar engine (develop the new algorithms)
  - Pick an appealing sample domain (e.g., health)
  - Do it in a way so as to benefit both ICT and linguists

Question

- Is this template-based approach useable for Bantu (or Niger-Congo) languages, be they agglutinating or not?
  - Short answer: No
- Tasks:
  - For structured input: use a practically useful language with tool support already (Sem Web tech)
  - Start with basics for a grammar engine (develop the new algorithms)
  - Pick an appealing sample domain (e.g., health)
  - Do it in a way so as to benefit both ICT and linguists
- First language to experiment with: isiZulu
- Details in [Keet and Khumalo(2014b), Keet and Khumalo(2014a), Keet and Khumalo(2017)]
- Turned out that results are transferrable to other agglutinating Bantu languages (some results obtained with Runyankore [Uganda], Chichewa [Malawi], and isiXhosa [South Africa], and bootstrapapbility)

A logic foundation for isiZulu knowledge-to-text

- Roughly OWL 2 EL
- OWL 2 EL is a W3C-standardised profile of OWL 2
- Tools, ontologies in OWL 2 (notably SNOMED CT)

$\mathcal{ALC}$ syntax

- *Concepts* denoting entity types/classes/unary predicates/universals, including top $\top$ and bottom $\bot$;
- *Roles* denoting relationships/associations/n-ary predicates/properties;
- *Constructors*: and $\sqcap$, or $\sqcup$, and not $\neg$; quantifiers 'for all' $\forall$ and 'there exists' $\exists$
- *Complex concepts* using constructors: Let C and D be concept names, R a role name, then
  - $\neg C$, $C \sqcap D$, and $C \sqcup D$ are concepts, and
  - $\forall R.C$ and $\exists R.C$ are concepts
- *Individuals*
- e.g., $Lion \sqsubseteq \exists eats.Herbivore \sqcap \forall eats.Herbivore$

$\mathcal{ALC}$ semantics

- *domain of interpretation*, and an *interpretation*, where:
  - Domain Δ is a non-empty set of objects
  - Interpretation: $\cdot^{\mathcal{I}}$ is the *interpretation function*, domain $\Delta^{\mathcal{I}}$
    - $\cdot^{\mathcal{I}}$ maps every concept name A to a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
    - $\cdot^{\mathcal{I}}$ maps every role name R to a subset $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
    - $\cdot^{\mathcal{I}}$ maps every individual name a to elements of $\Delta^{\mathcal{I}}$: $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
  - Note: $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$ and $\bot^{\mathcal{I}} = \emptyset$
- $(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
- $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\forall R.C)^{\mathcal{I}} = \{x \mid \forall y.R^{\mathcal{I}}(x,y) \rightarrow C^{\mathcal{I}}(y)\}$
- $(\exists R.C)^{\mathcal{I}} = \{x \mid \exists y.R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)\}$

Universal Quantification

- Consider here only the universal quantification at the start of the concept inclusion axiom ('nominal head')
- 'all'/'each' uses *-onke*, prefixed with the oral prefix of the noun class of that first noun (OWL class/DL concept) on lhs of $\sqsubseteq$

(U1) Boy $\sqsubseteq$ ...

wonke umfana ...	('each boy...'; *u-* + *-onke*)
bonke abafana ...	('all boys...'; *ba-* + *-onke*)

(U2) Phone $\sqsubseteq$ ...

lonke ifoni ...	('each phone...'; *li-* + *-onke*)
onke amafoni ...	('all phones...'; *a-* + *-onke*)

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

NC	QC_{oral}	QC (all) -onke	QC_{nke}	NEG SC	PRON	RC	QC_{dwa}	EC
1	u-onke →	wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke →	bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke →	wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke →	bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke →	wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke →	bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke →	wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke →	yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke →	lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke →	onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke →	sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke →	zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke →	yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke →	onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke →	yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke →	zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke →	lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke →	zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke →	bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke →	konke	zo-	aku-	khona	oku-	zo-	ku-

Subsumption

- Two different ways of carving up the nouns to determine which rules apply: semantic and syntactic
- Need to choose between
  - singular and plural
  - with or without the universal quantification voiced
  - generic or determinate

(S1) `MedicinalHerb` $\sqsubseteq$ `Plant`

ikhambi ngumuthi	('medicinal herb is a plant')
amakhambi yimithi	('medicinal herbs are plants')
wonke amakhambi ngumuthi	('all medicinal herbs are a plant')

(S2) (generic)

(S3) (determinate)

Possible subsumption patterns

a. N_1 <copulative *ng*/*y* depending on first letter of N_2>N_2.

b. <plural of N_1> <copulative *ng*/*y* depending on first letter of plural of N_2><plural of N_2>.

c. <All-concord for NC_x>onke <plural of N_1, being of NC_x> <copulative *ng*/*y* depending on first letter of N_2>N_2.

Complement/disjointness (adding negation

- Need to choose between
  - singular and plural, and with or without the universal quantification voiced
- Copulative is omitted
- Combines the negative subject concord (NEG SC) of the noun class of the first noun (*aku-*) with the pronomial (PRON) of the noun class of second noun (*-yona*)

(SN1) $\texttt{Cup} \sqsubseteq \neg\texttt{Glass}$

indebe akuyona ingilazi ('cup not a glass')

zonke izindebe aziyona ingilazi ('all cups not a glass')

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

NC	QC (all) $QC_{oral+onke}$	QC_{nke}	NEG SC	PRON	RC	QC_{dwa}	EC
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

Possible negation (disjointness) patterns

a. <N_1 of NC_x> <NEG SC of NC_x><PRON of NC_y> <N_2 of NC_y>.

b. <All-concord for NC_x>onke <plural N_1, being of NC_x> <NEG SC of NC_x><PRON of NC_y> <N_2 with NC_y>.

Existential Quantification

(E1) Giraffe $\sqsubseteq \exists$eats.Twig

yonke indlulamithi idla ihlamvana elilodwa ('each giraffe eats at least one twig')

zonke izindlulamithi zidla ihlamvana elilodwa ('all giraffes eat at least one twig')

a. <All-concord for NC_x>onke <pl. N_1, is in NC_x> <conjugated verb> <N_2 of NC_y> <RC for NC_y><QC for NC_y>dwa.

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

NC	QC (all)		NEG SC	PRON	RC	QC_{dwa}	EC
	$QC_{oral+onke}$	QC_{nke}					
1	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
1a	u-onke → wonke	wo-	aka-	yena	o-	ye-	mu-
2a	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3a	u-onke → wonke	wo-	aka-	wona	o-	ye-	mu-
(2a)	ba-onke → bonke	bo-	aba-	bona	aba-	bo-	ba-
3	u-onke → wonke	wo-	awu-	wona	o-	wo-	mu-
4	i-onke → yonke	yo-	ayi-	yona	e-	yo-	mi-
5	li-onke → lonke	lo-	ali-	lona	eli-	lo-	li-
6	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
7	si-onke → sonke	so-	asi-	sona	esi-	so-	si-
8	zi-onke → zonke	zo-	azi-	zona	ezi	zo-	zi-
9a	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
(6)	a-onke → onke	o-	awa-	wona	a-	wo-	ma-
9	i-onke → yonke	yo-	ayi-	yona	e-	yo-	yi-
10	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
11	lu-onke → lonke	lo-	alu-	lona	olu-	lo-	lu-
(10)	zi-onke → zonke	zo-	azi-	zona	ezi-	zo-	zi-
14	ba-onke → bonke	bo-	abu-	bona	obu-	bo-	bu-
15	ku-onke → konke	zo-	aku-	khona	oku-	zo-	ku-

Example

- $\forall x\ (\text{Professor}(x) \rightarrow \exists y\ (\text{teaches}(x, y) \wedge \text{Course}(y)))$
- $\text{Professor} \sqsubseteq \exists\ \text{teaches}.\text{Course}$
- **Each** Professor teaches **at least one** Course

Example

- $\forall x\ (\text{uSolwazi}(x) \rightarrow \exists y\ (\text{ufundisa}(x, y) \wedge \text{Isifundo}(y)))$
- $\text{uSolwazi} \sqsubseteq \exists\ \text{ufundisa}.\text{Isifundo}$
- ?

$$\forall x\ (\text{uSolwazi}(x) \rightarrow \exists y\ (\text{ufundisa}(x, y) \wedge \text{Isifundo}(y)))$$

$$\text{uSolwazi} \sqsubseteq \exists\ \text{ufundisa}.\text{Isifundo}$$

$\forall x\ (\text{uSolwazi}(x) \rightarrow (x, y) \wedge \text{Isifundo}(y)))$

$\text{uSolwazi} \sqsubseteq \exists\ \text{ufund}$

look-up NC

pluralise

for-all

NC	AU	PRE
1	u-	m(u)-
2	a-	ba-
1a	u-	-
2a	o-	-
3a	u-	-
(2a)	o-	-
3	u-	m(u)-
4	i-	mi-
5	i-	(li)-
6	a-	ma-
7	i-	si-
8	i-	zi-
9a	i-	-
(6)	a-	ma-
9	i(n)-	-
10	i-	zi(n)-
11	u-	(lu)-
(10)	i-	zi(n)-
14	u-	bu-
15	u-	ku-
17		ku-

NC	QC (all) $QC_{oral+onke}$
1	u-onke → wonke
2	ba-onke → bonke
1a	u-onke → wonke
2a	ba-onke → bonke
3a	u-onke → wonke
(2a)	ba-onke → bonke
3	u-onke → wonke
4	i-onke → yonke
5	li-onke → lonke
6	a-onke → onke
7	si-onke → sonke
8	zi-onke → zonke
9a	i-onke → yonke
(6)	a-onke → onke
9	i-onke → yonke
10	zi-onke → zonke
11	lu-onke → lonke
(10)	zi-onke → zonke
14	ba-onke → bonke
15	ku-onke → konke

Bonke oSolwazi

$$\forall x\ (\text{uSolwazi}(x) \rightarrow \exists y\ (\text{ufundisa}(x, y) \wedge \text{Isifundo}(y)))$$

$$\text{uSolwazi} \sqsubseteq \exists \text{ufundisa}.\text{Isifundo}$$

... for relevant NC. Here:

AlgoConjugate

ngi-
u-
u-
si-
ni-
ba-

Bonke oSolwazi bafundisa

$$\forall x\ (\text{uSolwazi}(x) \rightarrow \exists y\ (\text{ufundisa}(x, y) \wedge \text{Isifundo}(y)))$$

$$\text{uSolwazi} \sqsubseteq \exists\ \text{ufundisa}.\text{Isifundo}$$

Bonke oSolwazi bafundisa Isifundo

$\forall x\ (\text{uSolwazi}(x) \rightarrow \exists y\ (\ldots) \wedge \ldots))$

uSolwazi $\sqsubseteq \exists$ ufundisa.I

look-up NC

get RC

get QC

add -dwa

NC	AU	PRE
1	u-	m(u)-
2	a-	ba-
1a	u-	-
2a	o-	-
3a	u-	-
(2a)	o-	-
3	u-	m(u)-
4	i-	mi-
5	i-	(li)-
6	a-	ma-
7	i-	si-
8	i-	zi-
9a	i-	-
(6)	a-	ma-
9	i(n)-	-
10	i-	zi(n)-
11	u-	(lu)-
(10)	i-	zi(n)-
14	u-	bu-
15	u-	ku-
17		ku-

RC	QC_{dwa}
o-	ye-
aba-	bo-
o-	ye-
aba-	bo-
o-	ye-
aba-	bo-
o-	wo-
e-	yo-
eli-	lo-
a-	wo-
esi-	so-
ezi	zo-
e-	yo-
a-	wo-
e-	yo-
ezi-	zo-
olu-	lo-
ezi-	zo-
obu-	bo-
oku-	zo-

Bonke oSolwazi bafundisa lsifundo esisodwa

Evaluation

- Typical way of evaluating: ask linguists and/or intended target group
- Questions depend on what you want to know; e.g.,
  - Does the text capture the semantics adequately?
  - Must it really be grammatically correct or is understandable also acceptable?
  - Compared against alternate representation (figures, tables) or human-authored text?

Evaluation

- Typical way of evaluating: ask linguists and/or intended target group
- Questions depend on what you want to know; e.g.,
  - Does the text capture the semantics adequately?
  - Must it really be grammatically correct or is understandable also acceptable?
  - Compared against alternate representation (figures, tables) or human-authored text?
- Survey, asked linguists and non-linguists for their preferences
- 10 questions pitting the patterns against each other
- Online, with isiZulu-localised version of Limesurvey

Evaluation – interesting results

- Linguist agreed more among each other than the 'non-linguists'
- More agreement for the shorter sentences
- Open questions on 'deep Zulu' vs 'township Zulu', level of education in isiZulu, dialects
  - Sociolinguistics is not our task to investigate, but it may affect human evaluation results w.r.t. quality, grammaticality, naturalness

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Figuring out the present tense

1. Verb, and start of the grammar:
   $v \rightarrow$ *pre vr post* a *wh* | *npre vr post* i *wh* | *ppre vr* e | *vr st* a | *excl s cont o vr post* a
2. Prefix (subject and object concord, tense, mode, and aspect):
   pre $\rightarrow$ *s* | *s m* | *s t m* | *s asp m* | *s o* | *s m o* | *s t m o* | *s asp m o*
3. Negative prefix (negation; e.g. 'does not' eat):
   npre $\rightarrow$ *ns* | *ns m* | *ns t m* | *ns asp m* | *ns o* | *ns m o* | *ns t m o* | *ns asp m o*
4. Postfix, begin the "CARP" extensions:
   post $\rightarrow$ *c* | *c a* | *c a r* | *c a p* | *c r* | *c r p* | *c p* | *c a r p* | *a* | *a r* | *a r p* | *a p* | *r* | *r p* | *p* | ε
5. List of subject concords and negative subject concords (terminals for conjugation):
   s $\rightarrow$ ngi | u | si | ni | ba | i | li | a | zi | lu | bu | ku | ε
   ns $\rightarrow$ angi | awu | aka | ali | asi | ayi | alu | abu | aku | ani | aba | awa | azi | ε
6. List of mod:
   m $\rightarrow$ a | e | ka | ma | nga | ε
7. List of tense (present (ε) and continuous (ya)tense; incomplete):
   t $\rightarrow$ ya | ε
8. List of aspect (additional rules omitted in this first iteration):
   asp $\rightarrow$ sa | se | be | ile | ε
9. List of object concords:
   o $\rightarrow$ ngi | si | ku | ni | m | ba | wu | yi | li | wa | zi | lu | bu | ε
10. Causative:
    c $\rightarrow$ is
11. Applicative:
    a $\rightarrow$ el
12. Reciprocative:
    r $\rightarrow$ an
13. Passive (with phonological conditioning options):
    p $\rightarrow$ iw | w
14. Politeness (own prefix system and a FV=e):
    ppre $\rightarrow$ *pl s*
    pl $\rightarrow$ aw | awu | mawu | ε | ma
15. Stative (insertion of the *-ek-* between the VR and the FV):
    st $\rightarrow$ ek
16. Wh-questions (in the post-final slot and are added at the end of the verb, being *-ni* 'what'/'who'/ 'why'/'how', *-nini* 'when', and *-phi* 'where'.):
    wh $\rightarrow$ ni | nini | phi | ε
17. 'Double aspect'/exclusive (with *excl* $\subset$ *asp*)
    excl $\rightarrow$ se
18. Continuous tense (with *cont* $\subset$ *t*):
    cont $\rightarrow$ ya
19. Lexicon of verb roots:
    vr $\rightarrow$ ab | ... | zwib

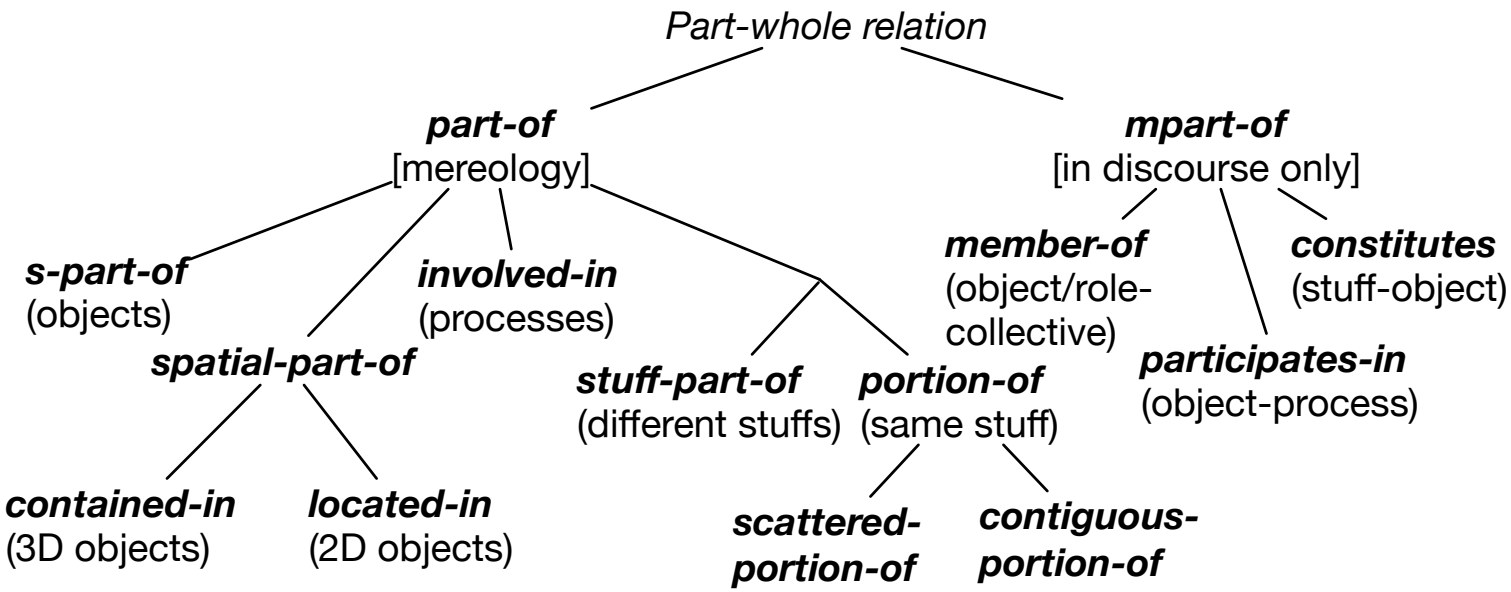
Extensions: part-whole relations

- Part-whole relations are used widely in medical and healthcare ontologies
- Many different types (23 in OpenGalen)
- Would that be convenient 1:1 translations?

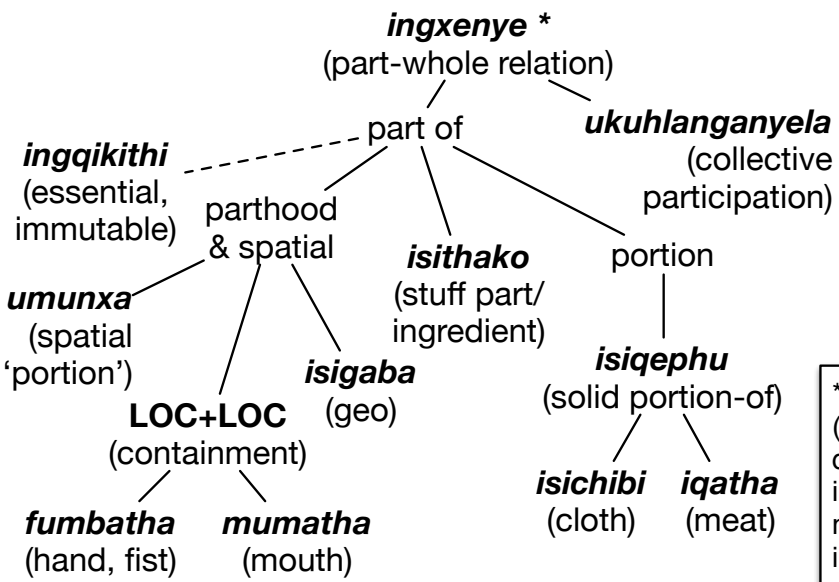
Extensions: part-whole relations

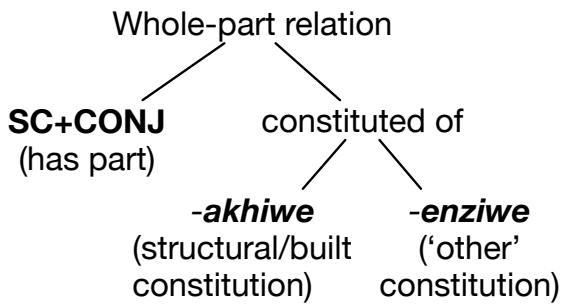
- Part-whole relations are used widely in medical and healthcare ontologies
- Many different types (23 in OpenGalen)
- Would that be convenient 1:1 translations?
  - No. both less and more specific ones: ontological differences
  - Other complications with verbs and prepositions
  - Details in: [Keet and Khumalo(2016)] [Keet(2017)] [Keet and Khumalo(2018)]

Part-whole relations: main differences [Keet and Khumalo(2018)]



Part-whole relations: main differences [Keet and Khumalo(2018)]





* includes, explicitly, at least: involvement (between processes), stuff parts (between different amounts of matter), participation of individual objects (cf. collectives) in events, and membership, and those subsumed by *ingxenye* in this figure.

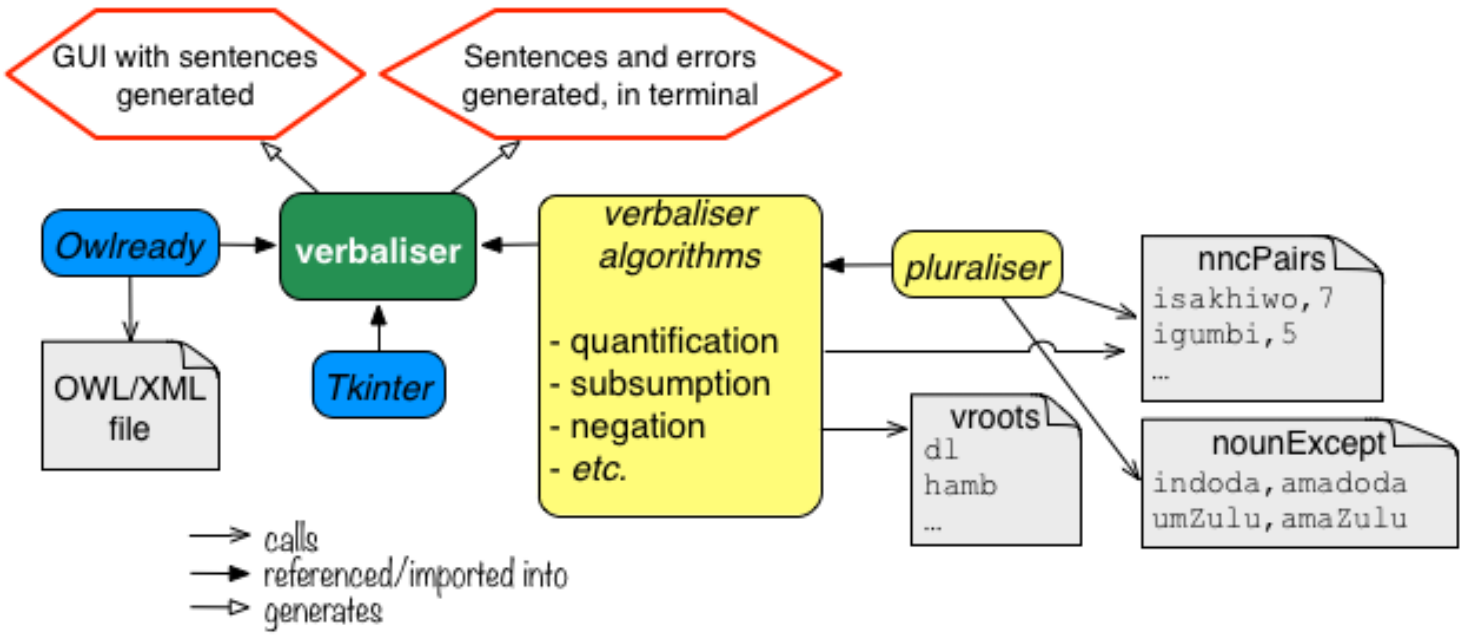
Extensions: part-whole relations

- 'part' *ingxenye* + 'of' <PC for NC of *ingxenye* that's then phonologically conditioned with noun of the whole>
  - e.g.: 'part of a human'
    ingxenye + *ya* + *umuntu*
    ingxenye yomuntu

Extensions: part-whole relations

- 'part' *ingxenye* + 'of' <PC for NC of *ingxenye* that's then phonologically conditioned with noun of the whole>
  - e.g.: 'part of a human'
    ingxenye + *ya* + *umuntu*
    ingxenye yomuntu
- 'contained in': locative affixes on the object that plays the container role
  - Each bolus of food is contained in some stomach
  - 'bolus of food' *indilinga yokudla* (nc9)
  - 'stomach' *isisu* (nc7)
  - 'is contained in' : SC-EP-LOC-Whole-LOCSUF
  - zi-s-e-sis-wini (phonological conditioning: e+i=e and u+ini=wini)
  - *Zonke izindilinga zokudla zisesiswini esisodwa*

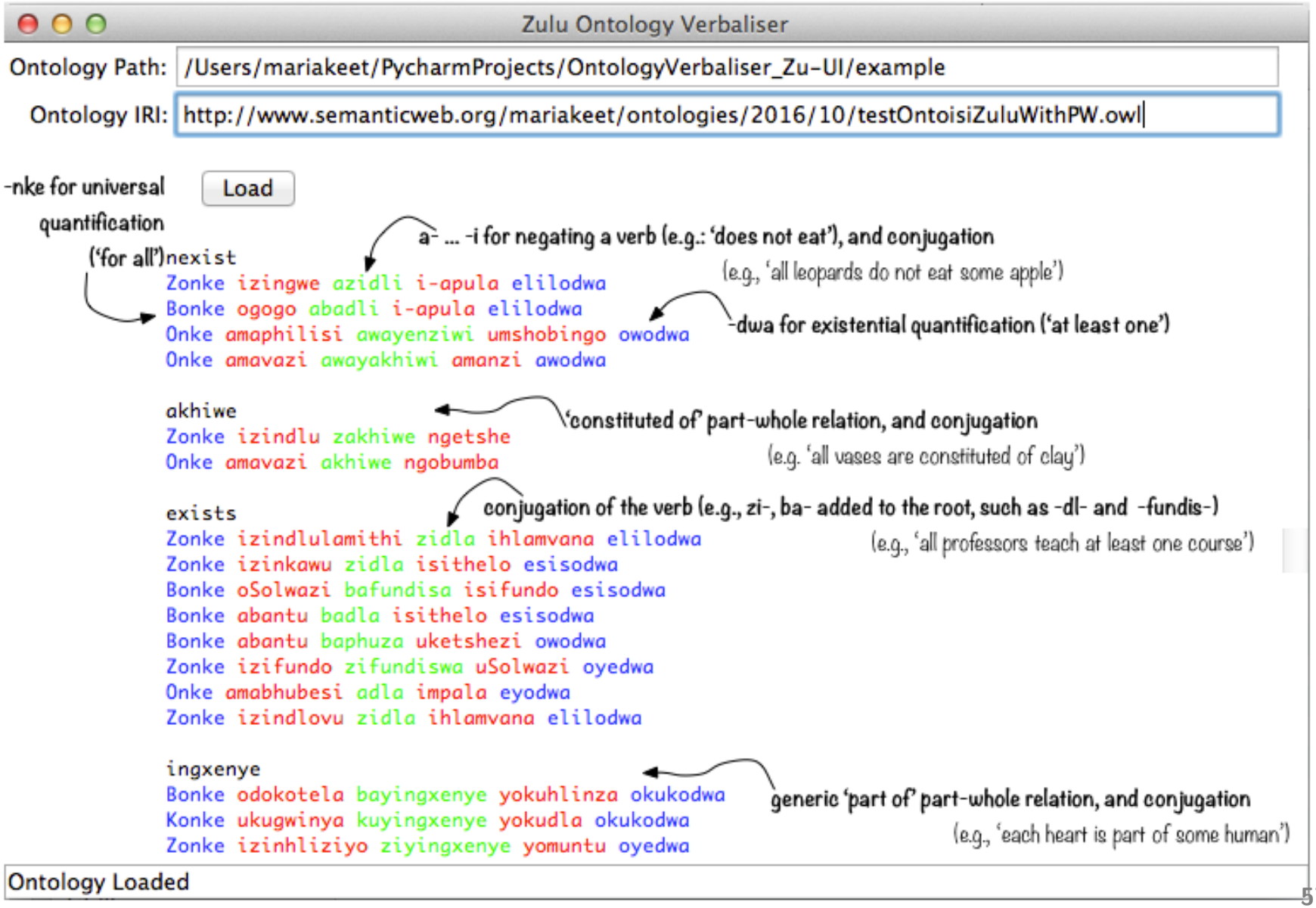
Tool: isiZulu verbaliser design



Tool: section of the OWL ontology

```
<SubClassOf>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#okubomvu"/>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#umbala"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#uMnumzana"/>
    <ObjectSomeValuesFrom>
        <ObjectProperty IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#ingxenye"/>
        <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#iNgqungquthela"/>
    </ObjectSomeValuesFrom>
</SubClassOf>
<SubClassOf>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#uSolwazi"/>
    <ObjectSomeValuesFrom>
        <ObjectProperty IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#fundisa"/>
        <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#isifundo"/>
    </ObjectSomeValuesFrom>
</SubClassOf>
<SubClassOf>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#ucezu_isinkwa"/>
    <ObjectSomeValuesFrom>
        <ObjectProperty IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#isiqephu"/>
        <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#isinkwa"/>
    </ObjectSomeValuesFrom>
</SubClassOf>
<SubClassOf>
    <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#udokotela"/>
    <ObjectSomeValuesFrom>
        <ObjectProperty IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#ingxenye"/>
        <Class IRI="http://www.meteck.org/files/ontologies/isiZulutestontologyPW.owl#ukuhlinza"/>
    </ObjectSomeValuesFrom>
</SubClassOf>
```

Tool: isiZulu verbaliser output



Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Initial results

- Tried that in detail with Runyankore [Byamugisha et al.(2016)]: it's faster than starting from scratch; (also shown by [Bosch et al.(2008)] for morphological analysers)
- Multilingual pluraliser, with a new table for the noun classes to make it deterministic choices for computation [Byamugisha et al.(2018)]
- Trying to understand morphological and verb similarities as proxies for possibly [easy/not-easy] to bootstrap from/to [Keet(2016), Mahlaza and Keet(2018)]
- Assessing bootstrappability between vs across Guthrie zones w.r.t. ontology verbalisation; zones indeed are not a good predictor [Byamugisha(2019)]

A few practical 'loose ends'

- Where to best store the NC info needed for verbalisation?
- What if your language doesn't have an ISO language tag?
- (There are more engineering questions to make it work)

A few practical 'loose ends'

- Where to best store the NC info needed for verbalisation?
  - Ontolex-Lemon is good for declarative information, not for rules
  - New annotation model [Keet and Chirema(2016)]
- What if your language doesn't have an ISO language tag?

- (There are more engineering questions to make it work)

A few practical 'loose ends'

- Where to best store the NC info needed for verbalisation?
  - Ontolex-Lemon is good for declarative information, not for rules
  - New annotation model [Keet and Chirema(2016)]
- What if your language doesn't have an ISO language tag?
  - Create your own!
  - e.g., with MoLA [Gillis-Webber et al.(2019)]
- (There are more engineering questions to make it work)

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

The NLG algorithms can be used elsewhere

- Paper-based language learning exercises
- Exercise books have a lot of exercises on ‘give plural noun’, ‘complete verb’ etc

The NLG algorithms can be used elsewhere

- Paper-based language learning exercises
- Exercise books have a lot of exercises on ‘give plural noun’, ‘complete verb’ etc
- Our algorithms already can do that!
- Reuse the algorithms to pluralise and conjugate
- Proof of concept tool, tried to use both NLP (corpus, POS tagger) and the grammar engine of NLG

Examples of the CNL it uses

- Pluralise subject
  - Q: * *Umfowethu bayaphuza*
  - A: *Abafowethu bayaphuza*
    [prefixSG+stem] [PLSC+VerbRoot+FV]
    [prefixPL+stem] [PLSC+VerbRoot+FV]

Examples of the CNL it uses

- Pluralise subject
  - Q: * *Umfowethu bayaphuza*
  - A: *Abafowethu bayaphuza*
    [prefixSG+stem] [PLSC+VerbRoot+FV]
    [prefixPL+stem] [PLSC+VerbRoot+FV]
- Negate the verb
  - Q: *Batotoba*
  - A: *Abatotobi*
    [PLSC+VerbRoot+FV]
    [PLNEGSC+VerbRoot+NEGFV]

Examples of the CNL it uses

- Pluralise subject
  - Q: * *Umfowethu bayaphuza*
  - A: *Abafowethu bayaphuza*
    [prefixSG+stem] [PLSC+VerbRoot+FV]
    [prefixPL+stem] [PLSC+VerbRoot+FV]
- Negate the verb
  - Q: *Batotoba*
  - A: *Abatotobi*
    [PLSC+VerbRoot+FV]
    [PLNEGSC+VerbRoot+NEGFV]
- Possible to combine components for new exercises

  [prefixSG+stem] [SGSC+VerbRoot+FV] [prefixSG+stem]
  [prefixPL+stem] [PLNEGSC+VerbRoot+NEGFV] [prefixPL+stem]
  - Q: *umfowethu usula inkomishi* '(my) brother washes the cup'
  - A: *abafowethu abasuli izinkomishi* '(my) brothers do not wash the cups'

Outline

1. Motivation
   - Context
   - Language ‘crash course’
2. Rule-based NLG
   - What is CNL, NLG?
   - Generating basic sentences
   - Extending basic sentences
3. On broadening and generalising results
   - Other languages
   - Reuse of the algorithms
4. Summary

Summary

- Explorations in controlling the language and generating sentences → improved understanding of issues, insights on what works (and what not)
- Templates inapplicable to isiZulu due to its grammar (OWL verbalisation), hence a tailor-made grammar engine
- NLG algorithms generic and modularised in the sense that they can be reused in other tools (CALL exercises)
- Not addressed much now, but no less important: underresourced language

Collaborators

- IsiZulu Linguist: Langa Khumalo
- Current/former students: Dr. Joan Byamugisha, Catherine Chavula, Nikhil Gilbert, Francis Gillis-Webber, Zola Mahlaza

References I

Sonja Bosch, Laurette Pretorius, and Axel Fleisch.

Experimental bootstrapping of morphological analysers for nguni languages.
Nordic Journal of African Studies, 17(2):66–88, 2008.

J. Byamugisha, C.M. Keet, and B. DeRenzi.

Bootstrapping a runyankore cnl from an isizulu cnl.
In B. Davis et al., editors, *5th Workshop on Controlled Natural Language (CNL'16)*, volume 9767 of *LNAI*, pages 25–36. Springer, 2016.
25-27 July 2016, Aberdeen, UK.

J. Byamugisha, C. M. Keet, and B. DeRenzi.

Pluralizing nouns across agglutinating Bantu languages.
In *27th International Conference on Computational Linguistics (COLING'18)*, pages 2633–2643. ACL, 2018.
20-26 August, 2018, Santa Fe, New Mexico, USA.

Joan Byamugisha.

Ontology Verbalization in Agglutinating Bantu Languages: A Study of Runyankore and Its Generalizability.
Phd thesis, Department of Computer Science, November 2019 2019.

M. Curland and T. Halpin.

Model driven development with NORMA.
In *Proceedings of the 40th International Conference on System Sciences (HICSS-40)*, pages 286a–286a. IEEE Computer Society, 2007.
Los Alamitos, Hawaii.

References II

Enrico Franconi, Paolo Guagliardo, and Marco Trevisan.

An intelligent query interface based on ontology navigation.
In *Workshop on Visual Interfaces to the Social and Semantic Web (VISSW'10)*, 2010.
Hong Kong, February 2010.

Norbert E. Fuchs, Kaarel Kaljurand, and Tobias Kuhn.

Discourse Representation Structures for ACE 6.6.
Technical Report ifi-2010.0010, Department of Informatics, University of Zurich, Zurich, Switzerland, 2010.

F. Gillis-Webber, S. Tittel, and C. M. Keet.

Dimensions affecting representation styles in ontologies.
In *1st Iberoamerican conference on Knowledge Graphs and Semantic Web (KGSWC'19)*, volume 1029 of *CCIS*, pages 1–16. Springer, 2019.
24-28 June 2019, Villa Clara, Cuba.

Mustafa Jarrar, C. Maria Keet, and Paolo Dongilli.

Multilingual verbalization of ORM conceptual models and axiomatized ontologies.
Starlab technical report, Vrije Universiteit Brussel, Belgium, February 2006.

C. M. Keet.

An assessment of orthographic similarity measures for several african languages.
Technical Report Arxiv.org 1608.03065, University of Cape Town, August 2016.
`http://arxiv.org/abs/1608.03065`.

References III

C. M. Keet.

Representing and aligning similar relations: parts and wholes in isizulu vs english.
In J. Gracia, F. Bond, J. McCrae, P. Buitelaar, C. Chiarcos, and S. Hellmann, editors, *Language, Data, and Knowledge 2017 (LDK'17)*, volume 10318 of *LNAI*, pages 58–73. Springer, 2017.
19-20 June, 2017, Galway, Ireland.

C. M. Keet and T. Chirema.

A model for verbalising relations with roles in multiple languages.
In E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali, editors, *Proceedings of the 20th International Conference on Knowledge Engineering and Knowledge Management (EKAW'16)*, volume 10024 of *LNAI*, pages 384–399. Springer, 2016.
19-23 November 2016, Bologna, Italy.

C. M. Keet and L. Khumalo.

Toward a knowledge-to-text controlled natural language of isiZulu.
Language Resources and Evaluation, 51(1):131–157, 2017.

C. M. Keet and L. Khumalo.

On the ontology of part-whole relations in Zulu language and culture.
In S. Borgo and P. Hitzler, editors, *10th International Conference on Formal Ontology in Information Systems 2018 (FOIS'18)*, volume 306 of *FAIA*, pages 225–238. IOS Press, 2018.
17-21 September, 2018, Cape Town, South Africa.

C. Maria Keet and Langa Khumalo.

Toward verbalizing logical theories in isiZulu.
In B. Davis, T. Kuhn, and K. Kaljurand, editors, *Proceedings of the 4th Workshop on Controlled Natural Language (CNL'14)*, volume 8625 of *LNAI*, pages 78–89. Springer, 2014a.
20-22 August 2014, Galway, Ireland.

References IV

C. Maria Keet and Langa Khumalo.

Basics for a grammar engine to verbalize logical theories in isiZulu.
In A. Bikakis et al., editors, *Proceedings of the 8th International Web Rule Symposium (RuleML'14)*, volume 8620 of *LNCS*, pages 216–225. Springer, 2014b.
August 18-20, 2014, Prague, Czech Republic.

C. Maria Keet and Langa Khumalo.

On the verbalization patterns of part-whole relations in isizulu.
In *9th International Natural Language Generation conference (INLG'16)*, pages 174–183. ACL, 2016.
5-8 September, 2016, Edinburgh, UK.

Tobias Kuhn.

A principled approach to grammars for controlled natural languages and predictive editors.
Journal of Logic, Language and Information, 22(1):33–70, 2013.

Zola Mahlaza and C. Maria Keet.

Measuring verb similarity using binary coefficients with application to isixhosa and isizulu.
In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT'18)*, pages 65–71, New York, NY, USA, 2018. ACM.

R. Schwitter, K. Kaljurand, A. Cregan, C. Dolbear, and G. Hart.

A comparison of three controlled natural languages for OWL 1.1.
In *Proc. of OWLED 2008 DC*, 2008.
Washington, DC, USA metropolitan area, on 1-2 April 2008.

References V

Allan Third, Sandra Williams, and Richard Power.
OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies.
poster/demo paper, Open Unversity UK, 2011.
10th International Semantic Web Conference (ISWC'11), 23-27 Oct 2011, Bonn, Germany.

The figures on slides 11, 12, and 14 are from Wikipedia.

Thank you!

Questions?

Online information:
GeNi project details: `http://www.meteck.org/files/geni/`
My homepage: `http://www.meteck.org`
OE textbook: `https://people.cs.uct.ac.za/~mkeet/OEbook/`