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
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-Saharan Africa

**Motivation**

*Language ‘crash course’*

*Note: less than 20% of languages shown due to space.*
Sub-Saharan: 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
<table>
<thead>
<tr>
<th>NC</th>
<th>AU</th>
<th>PRE</th>
<th>Stem (example)</th>
<th>Meaning</th>
<th>Example (isiZulu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>u-</td>
<td>a-</td>
<td>m(u)-fana</td>
<td>humans and other animates</td>
<td>umfana abafana</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-fana</td>
<td>boy</td>
<td>boys</td>
</tr>
<tr>
<td>1a</td>
<td>u-</td>
<td>o-</td>
<td>-baba</td>
<td>kinship terms and proper names</td>
<td>ubaba obaba</td>
</tr>
<tr>
<td>2a</td>
<td>-</td>
<td>-</td>
<td>-baba</td>
<td>father</td>
<td>fathers</td>
</tr>
<tr>
<td>3a</td>
<td>u-</td>
<td>o-</td>
<td>-shizi</td>
<td>nonhuman</td>
<td>ushizi oshizi</td>
</tr>
<tr>
<td>(2a)</td>
<td></td>
<td></td>
<td>-shizi</td>
<td>cheese</td>
<td>cheeses</td>
</tr>
<tr>
<td>3</td>
<td>u-</td>
<td>i-</td>
<td>m(u)-fula</td>
<td>trees, plants, non-paired body parts</td>
<td>umfula imifula</td>
</tr>
<tr>
<td>4</td>
<td>i-</td>
<td>mi-</td>
<td>-fula</td>
<td>river</td>
<td>rivers</td>
</tr>
<tr>
<td>5</td>
<td>i-</td>
<td>a-</td>
<td>(li)-gama</td>
<td>fruits, paired body parts, and natural phenomena</td>
<td>igama amagama</td>
</tr>
<tr>
<td>6</td>
<td>a-</td>
<td>ma-</td>
<td>-gama</td>
<td>name</td>
<td>names</td>
</tr>
<tr>
<td>7</td>
<td>i-</td>
<td>i-</td>
<td>si-hlalo</td>
<td>inanimates and manner/style</td>
<td>isihlalo izihlalo</td>
</tr>
<tr>
<td>8</td>
<td>i-</td>
<td>zi-</td>
<td>-hlalo</td>
<td>chair</td>
<td>chairs</td>
</tr>
<tr>
<td>9a</td>
<td>i-</td>
<td>a-</td>
<td>-rabha</td>
<td>nonhuman</td>
<td>irabha amarabha</td>
</tr>
<tr>
<td>(6)</td>
<td></td>
<td>ma-</td>
<td>-rabha</td>
<td>rubber</td>
<td>rubbers</td>
</tr>
<tr>
<td>9</td>
<td>i(n)-</td>
<td>-</td>
<td>ja</td>
<td>animals</td>
<td>inja izinja</td>
</tr>
<tr>
<td>10</td>
<td>i-</td>
<td>-</td>
<td>-ja</td>
<td>dog</td>
<td>dogs</td>
</tr>
<tr>
<td>11</td>
<td>u-</td>
<td>i-</td>
<td>(lu)-thi</td>
<td>inanimates and long thin objects</td>
<td>uthi izinthi</td>
</tr>
<tr>
<td>(10)</td>
<td></td>
<td></td>
<td>-thi</td>
<td>stick</td>
<td>sticks</td>
</tr>
<tr>
<td>14</td>
<td>u-</td>
<td>i-</td>
<td>bu-hle</td>
<td>abstract nouns</td>
<td>ubuhle ukucula</td>
</tr>
<tr>
<td>15</td>
<td>u-</td>
<td>-</td>
<td>ku-cula</td>
<td>infinitives</td>
<td>beauty to sing</td>
</tr>
<tr>
<td>17</td>
<td>ku-</td>
<td></td>
<td></td>
<td>locatives, remote/ general</td>
<td>locative</td>
</tr>
</tbody>
</table>
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’
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’

\[\text{ti tu ka ki mu reet er a ho ga nu}\]

\[\text{neg-(NC2 SC)-RM-(NC7 SC)-(NC1 SC)-VR-App-FV-Loc-Emp-Dec}\]
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 consequences (isiZulu)

- ‘and’, enumerative: na-, phonologically conditioned
  - Ex: milk and butter: *ubisi nebhotela* (-a+i=-e-)
  - Ex: butter and milk: *ibhotela nobisi* (-a+u=-o-)

Other verbs: concordial agreement (∼ conjugation) based on noun class
- Ex: The human eats: *umuntu u dla*
- Ex: The dog eats: *inja i dla*
Illustrative examples of some consequences (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 consequences (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 (∼ conjugation) based on noun class
  - Ex: The human eats: *umuntu udl*a
  - 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


‘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
What is CNL, NLG?

Short answer

- **C**ontrolled **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**

Mobile Zulu

---

Home » History » Cardiovascular History

**Chest Pain**

Have you had any recent pain in your chest? - Uke waba nobuhlungu esifubeni maduzane?

Does the pain radiate to your jaw, neck or arm? - Engabe ubuhlungu bakho bujikeleza emihlathini, emqaleni nomalase ngizikhumbula?

Does anything precipitate or relieve the pain? - Ingabe ikhona into eyenza ubuhlungu buqhubeke noma eyehlisa ubuhlungu?

**Dyspnoea**

**Mobile Xhosa**

---

Home » History » Cardiovascular History

**Chest Pain**

Have you had any recent pain in your chest? - Ingaba kutshanje ukhe weva iintlungu esifubeni?

Does the pain radiate to your jaw, neck or arm? - Ingaba iintlungu zinwenwela emihlathini, entanyeni okanye engalwena?

Does anything precipitate or relieve the pain? - Ingabe ikhona into eyenza iintlungu buqhubeke noma eyehlisa iintlungu?
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
  - 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
  - but also other languages [Jarrar et al.(2006)]
- Grammar engines, such as [Kuhn(2013)], Grammatical Framework (http://www.grammaticalframework.org/), SimpleNLG
  ⇒ 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: $\forall x \ (\text{Course}(x) \rightarrow \exists y \ (\text{is\_taught\_by}(x, y) \land \text{Professor}(y)))$

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

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

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

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

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

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

\[
\begin{align*}
\text{Sentence} & \rightarrow \text{NounPhrase} \mid \text{VerbPhrase} \\
\text{NounPhrase} & \rightarrow \text{Adjective} \mid \text{NounPhrase} \\
\text{NounPhrase} & \rightarrow \text{Noun} \\
\text{Noun} & \rightarrow \text{car} \mid \text{train} \\
\text{Adjective} & \rightarrow \text{big} \mid \text{broken} \\
\text{...} & \\
\text{(and complexity of the grammar)} & \\
\end{align*}
\]
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?
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

First language to experiment with: isiZulu

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)
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

Turned out that results are transferrable to other agglutinating Bantu languages (some results obtained with Runyankore [Uganda], Chichewa [Malawi], and isiXhosa [South Africa], and bootstrapability)
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)
**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., $\text{Lion} \sqsubseteq \exists \text{eats.Herbivore} \sqcap \forall \text{eats.Herbivore}$
**ALC semantics**

- **domain of interpretation**, and an **interpretation**, where:
  - Domain $\Delta$ 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 $\bot^\mathcal{I} = \emptyset$

- $(-C)^\mathcal{I} = \Delta^\mathcal{I}\setminus C^\mathcal{I}$
- $(C \cap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$
- $(C \cup D)^\mathcal{I} = C^\mathcal{I} \cup D^\mathcal{I}$
- $(\forall R. C)^\mathcal{I} = \{x \mid \forall y. R^\mathcal{I}(x, y) \rightarrow C^\mathcal{I}(y)\}$
- $(\exists R. C)^\mathcal{I} = \{x \mid \exists y. R^\mathcal{I}(x, y) \land C^\mathcal{I}(y)\}$
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)
<table>
<thead>
<tr>
<th>NC</th>
<th>QC (oral + onke)</th>
<th>QC (nke)</th>
<th>NEG SC</th>
<th>PRON</th>
<th>RC</th>
<th>QC_dwa</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>(2a)</td>
<td></td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>(6)</td>
<td></td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td>(10)</td>
<td></td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
<tr>
<td>NC</td>
<td>QC&lt;sub&gt;oral&lt;/sub&gt;</td>
<td>QC (all)</td>
<td>QC&lt;sub&gt;nke&lt;/sub&gt;</td>
<td>NEG SC</td>
<td>PRON</td>
<td>RC</td>
<td>QC&lt;sub&gt;dwa&lt;/sub&gt;</td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>----------------</td>
</tr>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>3</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>bona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
</tbody>
</table>
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 ⊑ 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. $N_1 <\text{copulative} ng/y \text{ depending on first letter of } N_2 > N_2$.  
b. $<\text{plural of } N_1 > <\text{copulative} ng/y \text{ depending on first letter of plural of } N_2 > <\text{plural of } N_2 >$.  
c. $<\text{All-concord for NC}_x > \text{onke} <\text{plural of } N_1 , \text{being of NC}_x > <\text{copulative} ng/y \text{ depending on first letter of } N_2 > N_2$. 
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)

\[(SN1) \text{ Cup } \nsubseteq \neg \text{ Glass} \]

inobe akuyona ingilazi

izonke izindebe aziyona ingilazi

\(\text{‘cup not a glass’}\)

\(\text{‘all cups not a glass’}\)
<table>
<thead>
<tr>
<th>NC</th>
<th>QC (all)</th>
<th>NEG SC</th>
<th>PRON</th>
<th>RC</th>
<th>QCdwa</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCoral+onke</td>
<td>QCnke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
</tr>
<tr>
<td>2a</td>
<td>(2a) ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yon-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yon-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
</tr>
<tr>
<td>10</td>
<td>(10) zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
</tr>
</tbody>
</table>
## Rule-based NLG
Generating basic sentences

<table>
<thead>
<tr>
<th>NC</th>
<th>QC&lt;sub&gt;oral+onke&lt;/sub&gt;</th>
<th>QC&lt;sub&gt;nke&lt;/sub&gt;</th>
<th>NEG SC</th>
<th>PRON</th>
<th>RC</th>
<th>QC&lt;sub&gt;nda&lt;/sub&gt;</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abo-</td>
<td>abo-</td>
<td>ba-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abo-</td>
<td>abo-</td>
<td>ba-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td></td>
<td>(2a)</td>
<td>bo-</td>
<td>abo-</td>
<td>abo-</td>
<td>ba-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awoa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td></td>
<td>(10)</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
</tbody>
</table>
The table below represents the generation of basic sentences using Rule-based NLG (Natural Language Generation).

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

a. \(<N_1 \text{ of } NC_x> <\text{NEG SC of } NC_x> <\text{PRON of } NC_y> <N_2 \text{ of } NC_y>\).

b. \(<\text{All-concord for } NC_x> \text{onke } <\text{plural } N_1, \text{ being of } NC_x> <\text{NEG SC of } NC_x> <\text{PRON of } NC_y> <N_2 \text{ with } NC_y>\).
Existential Quantification

(E1) Giraffe ⊑ ∃ eats. Twig
    yonke indlulamithi idla ihlamvana elilodwa
    zonke izindlulamithi zidla ihlamvana elilodwa

('each giraffe eats at least one twig')
('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.
<table>
<thead>
<tr>
<th>NC</th>
<th>QCoral+onke</th>
<th>QCnke</th>
<th>NEG SC</th>
<th>PRON</th>
<th>RC</th>
<th>QCdwa</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
<tr>
<td>NC</td>
<td>QC (all)</td>
<td>NEG SC</td>
<td>PRON</td>
<td>RC</td>
<td>QCdwa</td>
<td>EC</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>-------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>wona</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>(2a)</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>(6)</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td>(10)</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
<tr>
<td>NC</td>
<td>QC (oral+onke)</td>
<td>QC (nke)</td>
<td>NEG SC</td>
<td>PRON</td>
<td>RC</td>
<td>QC dwa</td>
<td>EC</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
<td>----------</td>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>--------</td>
<td>----</td>
</tr>
<tr>
<td>1</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>1a</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>aka-</td>
<td>yena</td>
<td>o-</td>
<td>ye-</td>
<td>mu-</td>
</tr>
<tr>
<td>2a</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>aba-</td>
<td>bona</td>
<td>aba-</td>
<td>bo-</td>
<td>ba-</td>
</tr>
<tr>
<td>3</td>
<td>u-onke → wonke</td>
<td>wo-</td>
<td>awu-</td>
<td>wona</td>
<td>o-</td>
<td>wo-</td>
<td>mu-</td>
</tr>
<tr>
<td>4</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yon-</td>
<td>mi-</td>
</tr>
<tr>
<td>5</td>
<td>li-onke → lonke</td>
<td>lo-</td>
<td>ali-</td>
<td>lona</td>
<td>eli-</td>
<td>lo-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>7</td>
<td>si-onke → sonke</td>
<td>so-</td>
<td>asi-</td>
<td>sona</td>
<td>esi-</td>
<td>so-</td>
<td>si-</td>
</tr>
<tr>
<td>8</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>9a</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>6</td>
<td>a-onke → onke</td>
<td>o-</td>
<td>awa-</td>
<td>wona</td>
<td>a-</td>
<td>wo-</td>
<td>ma-</td>
</tr>
<tr>
<td>9</td>
<td>i-onke → yonke</td>
<td>yo-</td>
<td>ayi-</td>
<td>yona</td>
<td>e-</td>
<td>yo-</td>
<td>yi-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>11</td>
<td>lu-onke → lonke</td>
<td>lo-</td>
<td>alu-</td>
<td>lona</td>
<td>olu-</td>
<td>lo-</td>
<td>lu-</td>
</tr>
<tr>
<td>10</td>
<td>zi-onke → zonke</td>
<td>zo-</td>
<td>azi-</td>
<td>zona</td>
<td>ezi-</td>
<td>zo-</td>
<td>zi-</td>
</tr>
<tr>
<td>14</td>
<td>ba-onke → bonke</td>
<td>bo-</td>
<td>abu-</td>
<td>bona</td>
<td>obu-</td>
<td>bo-</td>
<td>bu-</td>
</tr>
<tr>
<td>15</td>
<td>ku-onke → konke</td>
<td>zo-</td>
<td>aku-</td>
<td>khona</td>
<td>oku-</td>
<td>zo-</td>
<td>ku-</td>
</tr>
</tbody>
</table>
Example

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

- $\forall x \ (uSolwazi(x) \rightarrow \exists y \ (ufundisa(x, y) \land Isifundo(y)))$
- $uSolwazi \sqsubseteq \exists ufundisa.Isifundo$
- $?$
∀x (uSolwazi(x) → ∃y (ufundisa(x, y) ∧ Isifundo(y)))

uSolwazi ⊆ ∃ ufundisa.Isifundo
### Generating basic sentences

**Rule-based NLG**

- **Bonke oSolwazi**

#### Look-up NC

- **pluralise**

#### For-all

<table>
<thead>
<tr>
<th>TC</th>
<th>NC</th>
<th>NC</th>
<th>QC</th>
<th>QC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1a</td>
<td>u-</td>
<td>m(u)-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2a</td>
<td>a-</td>
<td>ba-</td>
</tr>
<tr>
<td></td>
<td>3a</td>
<td>u-</td>
<td>-</td>
<td>3a</td>
</tr>
<tr>
<td></td>
<td>(2a)</td>
<td>o-</td>
<td>-</td>
<td>(2a)</td>
</tr>
<tr>
<td>4</td>
<td>i-</td>
<td>m(u)-</td>
<td>4</td>
<td>i-onke → wonke</td>
</tr>
<tr>
<td>5</td>
<td>i-</td>
<td>(li)-</td>
<td>5</td>
<td>li-onke → lonke</td>
</tr>
<tr>
<td>6</td>
<td>a-</td>
<td>ma-</td>
<td>6</td>
<td>a-onke → onke</td>
</tr>
<tr>
<td>7</td>
<td>i-</td>
<td>si-</td>
<td>7</td>
<td>si-onke → sonke</td>
</tr>
<tr>
<td>8</td>
<td>i-</td>
<td>zi-</td>
<td>8</td>
<td>zi-onke → zonke</td>
</tr>
<tr>
<td>9</td>
<td>i(n)-</td>
<td>-</td>
<td>9</td>
<td>i-onke → yonke</td>
</tr>
<tr>
<td>(6)</td>
<td>i-</td>
<td>ma-</td>
<td>(6)</td>
<td>a-onke → onke</td>
</tr>
<tr>
<td>10</td>
<td>i-</td>
<td>zi(n)-</td>
<td>10</td>
<td>zi-onke → zonke</td>
</tr>
<tr>
<td>11</td>
<td>u-</td>
<td>(lu)-</td>
<td>11</td>
<td>lu-onke → lonke</td>
</tr>
<tr>
<td>(10)</td>
<td>i-</td>
<td>zi(n)-</td>
<td>(10)</td>
<td>zi-onke → zonke</td>
</tr>
<tr>
<td>14</td>
<td>u-</td>
<td>bu-</td>
<td>14</td>
<td>ba-onke → bonke</td>
</tr>
<tr>
<td>15</td>
<td>u-</td>
<td>ku-</td>
<td>15</td>
<td>ku-onke → konke</td>
</tr>
<tr>
<td>17</td>
<td>ku-</td>
<td>-</td>
<td>17</td>
<td>-</td>
</tr>
</tbody>
</table>
$\forall x \ (uSolwazi(x) \rightarrow \exists y \ (ufundisa(x, y) \land Isifundo(y)))$

$uSolwazi \subseteq \exists y (ufundisa(y))$

... for relevant NC. Here:

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

AlgoConjugate

Bonke oSolwazi bafundisa
∀x (uSolwazi(x) → ∃y (ufundisa(x, y) ∧ Isifundo(y)))

uSolwazi ⊆ ∃ ufundisa Isifundo

Bonke oSolwazi bafundisa Isifundo
Bonke oSolwazi bafundisa Isifundo 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:
   \[ v \rightarrow \text{pre vr post a wh} | \text{npre vr post i wh} | \text{ppre vr e} | \text{vr st a} | \text{excl s cont o vr post a} \]
2. Prefix (subject and object concord, tense, mode, and aspect):
   \[ \text{pre} \rightarrow s | s m | s t m | s asp m | s o | s m o | s t m o | s asp m o \]
3. Negative prefix (negation; e.g. ‘does not’ eat):
   \[ \text{npre} \rightarrow ns | ns m | ns t m | ns asp m | ns o | ns m o | ns t m o | ns asp m o \]
4. Postfix, begin the “CARP” extensions:
   \[ \text{post} \rightarrow c | c a | c a r | c a p | c r | c r p | c p | c a r p | a | a r | a r p | a p | r | r p | p | \varepsilon \]
5. List of subject concords and negative subject concords (terminals for conjugation):
   \[ s \rightarrow ngi | u | si | ni | ba | i | li | a | z | lu | bu | ku | \varepsilon \]
   \[ ns \rightarrow angi | awu | aka | ali | asi | ayi | alu | abu | aku | ani | aba | awa | azi | \varepsilon \]
6. List of mod:
   \[ m \rightarrow a | e | ka | ma | nga | \varepsilon \]
7. List of tense (present (\varepsilon) and continuous (ya)tense; incomplete):
   \[ t \rightarrow ya | \varepsilon \]
8. List of aspect (additional rules omitted in this first iteration):
   \[ \text{asp} \rightarrow sa | se | be | ile | \varepsilon \]
9. List of object concords:
   \[ o \rightarrow ngi | si | ku | ni | m | ba | wu | yi | li | wa | zi | lu | bu | \varepsilon \]
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=\varepsilon):
    \[ \text{ppre} \rightarrow pl s \]
    \[ pl \rightarrow aw | awu | mawu | \varepsilon | ma \]
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’):
    \[ wh \rightarrow ni | nini | phi | \varepsilon \]
17. ‘Double aspect’/exclusive (with excl \subset asp)
    \[ \text{excl} \rightarrow se \]
18. Continuous tense (with cont \subset t):
    \[ \text{cont} \rightarrow ya \]
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 OpenGalcn)
- **Would that be convenient 1:1 translations?**
  - No. both less and more specific ones: ontological differences
  - Other complications with verbs and prepositions
Part-whole relations: main differences
[Keet and Khumalo(2018)]

Part-whole relation

- part-of
  - s-part-of
    - contained-in (3D objects)
    - located-in (2D objects)
  - spatial-part-of
  - involved-in (processes)

- mpart-of
  - member-of (object/role-collective)
  - constitutes (stuff-object)
  - participates-in (object-process)

- stuff-part-of
  - different stuffs
- portion-of
  - same stuff

- located-in
  - scattered-portion-of
- constituent-of
  - contiguous-portion-of
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 * in this figure.
Extensions: part-whole relations

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

Diagram:

- **Owlready**
  - OWL/XML file
  - calls
  - referenced/imported into
generates

- **Tkinter**

- **verbaliser**
  - verbaliser algorithms
    - quantification
    - subsumption
    - negation
    - etc.

- **pluraliser**
  - nncPairs
    - isakhiwo, 7
    - igumbi, 5
    - ...
  - nounExcept
    - indoda, amadoda
    - umZulu, amaZulu

- GUI with sentences generated
- Sentences and errors generated, in terminal
Tool: section of the OWL ontology
Tool: isiZulu verbaliser output

- **nke** for universal quantification
- **a** for negating a verb (e.g., 'does not eat'), and conjugation
- **dwa** for existential quantification ('at least one')

- **akhiwe**
  - Zonke izingwe azidli i-apula elilodwa
  - Bonke ogogo abadli i-apula elilodwa

- **exists**
  - Zonke izindlulamithi zidla ihlamvana elilodwa
  - Zonke izinkawu zidla isithelo esisodwa

- ** constituent of** part-whole relation, and conjugation
  - Zonke izifundo zifundiswa uSolwazi oydwa

- **ingxenye**
  - Bonke odokotela bayingxenye yokuhlinza okukodwa

- **generic ‘part of’** part-whole relation, and conjugation
  - Zonke izinhliziyo ziyingxenye yomuntu oydwa
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  

  \[
  \text{[prefixSG+stem]} \ [\text{PLSC+VerbRoot+FV}] \\
  \text{[prefixPL+stem]} \ [\text{PLSC+VerbRoot+FV}]
  \]

- Negate the verb
  
  Q:  Batotoba  
  A:  Abatotobi  

  \[
  \text{[PLSC+VerbRoot+FV]} \\
  \text{[PLNEGSC+VerbRoot+NEGFV]}
  \]
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
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.

J. Byamugisha, C.M. Keet, and B. DeRenzi.
Bootstrapping a runyankore cnl from an isizulu cnl.

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.

M. Curland and T. Halpin.
Model driven development with NORMA.
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.
24-28 June 2019, Villa Clara, Cuba.

Mustafa Jarrar, C. Maria Keet, and Paolo Dongilli.
Multilingual verbalization of ORM conceptual models and axiomatized ontologies.

C. M. Keet.
An assessment of orthographic similarity measures for several african languages.
Reference III

C. M. Keet.
Representing and aligning similar relations: parts and wholes in isiZulu vs English.

C. M. Keet and T. Chirema.
A model for verbalising relations with roles in multiple languages.
19-23 November 2016, Bologna, Italy.

C. M. Keet and L. Khumalo.
Toward a knowledge-to-text controlled natural language of isiZulu.

C. M. Keet and L. Khumalo.
On the ontology of part-whole relations in Zulu language and culture.
17-21 September, 2018, Cape Town, South Africa.

C. Maria Keet and Langa Khumalo.
Toward verbalizing logical theories in isiZulu.
20-22 August 2014, Galway, Ireland.
C. Maria Keet and Langa Khumalo.
Basics for a grammar engine to verbalize logical theories in isiZulu.
August 18-20, 2014, Prague, Czech Republic.

C. Maria Keet and Langa Khumalo.
On the verbalization patterns of part-whole relations in isizulu.
5-8 September, 2016, Edinburgh, UK.

Tobias Kuhn.
A principled approach to grammars for controlled natural languages and predictive editors.

Zola Mahlaza and C. Maria Keet.
Measuring verb similarity using binary coefficients with application to isixhosa and isizulu.

A comparison of three controlled natural languages for OWL 1.1.
Washington, DC, USA metropolitan area, on 1-2 April 2008.
Allan Third, Sandra Williams, and Richard Power.  
OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies.  
poster/demo paper, Open University UK, 2011.  

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