Basics for a grammar engine to verbalize logical theories in isiZulu

C. Maria Keet[1] and Langa Khumalo[2]

[1]Department of Computer Science, University of Cape Town, South Africa,
mkeet@cs.uct.ac.za
[2]Linguistics Program, School of Arts, University of KwaZulu-Natal, South Africa,
Khumalol@ukzn.ac.za

Abstract. The language isiZulu is the largest in South Africa by numbers of first language speakers, yet, it is still an underresourced language. In this paper, we approach the grammar piecemeal from a natural language generation approach, and viewed from a potential utility for verbalizing OWL ontologies as a tangible use case. The elaborate rules of the grammar show that a grammar engine and dictionary is essential even for basic verbalizations in OWL 2 EL. This is due to, mainly, the 17 noun classes with embedded semantics and the agglutinative nature of isiZulu. The verbalization of basic constructs requires merging a prefix with a noun and distinguishing an 'and' between a list and linking clauses.

1 Introduction

South Africa has hitherto seen limited investment in human language technologies and computational linguistics, especially for its 9 official African languages. Large companies, such as Google and Microsoft, do pick the low-hanging fruit with localizations of the user interfaces of their software. The South African Department of Science and Technology demands for its potential outputs, notably with its "National Recordal System" (NRS) project by its National Indigenous Knowledge Systems Office. NRS software infrastructure was launched in 2013 and requires not just a standard document system [7] but for full usage, it requires an 'intelligent' one [1] that can handle multilingualism in, among others, document search and annotation, and in model development of the knowledge that is to be stored in the NRS. Systems with relevant functionalities for the NRS exist elsewhere for multiple languages in Europe, e.g., the multilingual and collaborative systems by [2,10], or a CNL-mediated query interface (e.g., [6]). This is to quite an extent thanks to large FP7 projects, such as Monnet [http://www.monnet-project.eu] for foundational aspects and applied projects such as Organic.Lingua [http://www.organic-lingua.eu]. No such resources exist for promoting the 9 official African languages in South Africa, yet such system requirements for, among others, the NRS, demand for both NLP and NLG technologies for those languages. Here, we focus on NLG for isiZulu, which is the first ("home") language for about 23% of the population (±10 million), about half of the population in South Africa can speak it, and it has several closely related languages, such as isiXhosa.

Unlike for NLP and corpus building [16,18], no NLG results exist for any of the languages in the Nguni language group, of which isiZulu is a member. One could consider Google Translate, which has English–isiZulu since October 2013, but it cannot handle articles and quantification (among other things), and its technology is proprietary and inaccessible. There are mainly old and outdated grammar books and Doke's seminal work on the general description of the isiZulu morphology [4,5] has remained an important reference for linguistic work not only in isiZulu but in Southern Bantu languages; this makes it challenging to commence defining grammars similar to Kuhn [12]. It will take many years and resources to fill this gap. Here, we start with some basics that should aid both linguists and information systems development. To this end, we take language constructs of a practical logic language with low expressiveness, such as the OWL 2 EL profile [14], as a starting point and extant approaches for other languages and systems. Concerning such practical verbalizations of logical theories, there are verbalization options within English [17], implementations in different systems, such as for the Semantic Web (ACE [8]) and for conceptual data models (e.g., monolingual [3] and multilingual ORM [9]), and we assume that a multilingual ontology is in place, perhaps managed through the *Lemon* model [13]. For isiZulu, it appears that the grammar rules are quite complex, and we summarise those for subsumption, disjointness, existential and universal quantification, and conjunction. There are two particular features of isiZulu that have a major effect on verbalizations, which are that the semantics of the noun (more precisely: the category of the entity it refers to) and the quantifiers in an axiom influence the verbalization patterns.

The remainder of this paper is organised as follows. Section 2 describes some basic aspects of isiZulu, and Section 3 presents the main results on verbalization patterns for simple taxonomic subsumption, disjointness (negation), conjunction, and quantification. We reflect in Section 4 and conclude in Section 5.

2 Some very basic aspects of isiZulu

IsiZulu is a highly agglutinating language with a complex morphology. As is emblematic of Bantu languages, isiZulu has a system of noun classes. Every noun belongs to a noun class. The class is often identifiable from the noun prefix that is attached to the noun, and it governs the agreement of all words that modify the noun, as well as of predicates of which the noun is a subject. Object agreement may also be marked on the predicate.

There is more than one convention for labeling and referring to these classes, most of which are essentially numbering systems. We will use Meinhof's (1948) classification system, which is used in most scholarly works and permits comparison of corresponding classes across Bantu languages, all of which lack at least some of the classes posited for proto-Bantu. Most noun classes are set off into pairs in isiZulu such that most nouns have a singular form in one class and a

Table 1. Zulu noun classes, with examples. The noun class prefix of classes 1 and 3 is conditioned by the morphology of the stem to which it attaches: *-mu-* before monosyllabic stems and *-m-* for other stems. The *n* of the noun prefixes of noun class 9 and 10 fuses with the following consonant forming prenasalized consonants; NC: Noun class, AU: augment, PRE: prefix.

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

plural form in another; the classes are summarised in Table 1. The morphological structure of a noun in isiZulu typically takes the shape of the tree structure.

For the most part, the semantics of a noun plays a role in determining what noun class a word falls in; their deeper meanings as well as shift and colloquial use are being investigated (e.g., [15]), and is summarised in column 5 of Table 1. Most noun stems belong to only one noun class pair, but exceptions exist (e.g., *-ntu*). Noun class prefixes can also be used to form new nouns from other noun stems and other stems, like noun class 15 that creates infinitives out of verbal stems. The vast majority of the nouns in noun class 14 is derived as well: the prefix *-bu-* forms abstract nouns from other noun stems and adjective stems. Class 17 is a non-productive locative class with the noun prefix *ku-*. IsiZulu lacks classes 12 and 13, which are found in other Bantu languages.

The complexity of the morphology of isiZulu is compounded by the fact that a number of prefixes have allomorphic forms. This is a consequence of the fact that isiZulu proscribes vowel sequencing, so that a prefix whose canonical form is *nga-* will have an allomorph *ng-* before roots that begin with vowels. Furthermore, many morphemes are homographs, so that the prefix *nga-* could represent either the potential mood morpheme or a form of the negative that occurs in subordinate clauses; and the sequence *ng-* could be the allomorph of either of these, or of a number of homographic morphemes *ngi-*, which represent the first person singular in various moods. Besides these phonologically conditioned allomorphs, there are also morphologically conditioned ones; e.g., the locative prefix *e-* has an allomorph *o-* that occurs in certain morphological circumstances [18] (p1023). Nominal morphology triggers agreement, as is shown in the example:

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'

The fact that the subject *abafana* ('boys') is of noun class 2 is reflected both in the agreement prefix on the adjective *abancane* ('small') and in the subject agreement on the verb. The noun class 10 feature of the object *izincwadi* ('books') is reflected in the class 10 agreement on the adjective *ezinkulu* ('big') and in the object marker on the verb. A selection of such agreements, called concords, is included in Table 2. The normal word order is Subject Verb Object (SVO) but there is attested variation since post verbal subjects are also common.

It is imperative to further state that isiZulu also has a very complex verbal morphology. The verbs can be conjugated in five different tenses (remote past, recent past, present, immediate future and remote future) as well as for various aspects and moods. The verb usually agrees with the subject and sometimes with the object in person and number (as shown in the example above) and in 3rd person for noun class as well. To account for this, a verb form can consist of many morphemes. Such complex morphology characteristic of most Bantu languages presents a lot of challenges in the attempts to develop computational technologies in isiZulu.

3 Verbalization patterns and algorithms

We obviously cannot cover all the grammar rules, and will focus only on the—from a logic viewpoint—seemingly 'simple' constructs, being subsumption, conjunction, negation, and quantification. This fits roughly with the OWL 2 EL profile (plus negation), that has a nice use-case scenario: upon localizing SNOMED CT, the axioms can then be verbalised in isiZulu and be used in healthcare applications. We will take examples from the general domain, however, so as not to complicate matters with medicine, and we assume a suitable multilingual encoding of the ontology, and use the Description Logics notation for conciseness.

Universal Quantification. We consider here only the universal quantification at the start of the concept inclusion axiom, such as for verbalizing taxonomic subsumption for atomic classes and the typical 'forall-some' construction, or, in linguistic terms, the nominal head. In isiZulu, the 'all'/'each' uses *-onke*, which

Table 2. Zulu noun classes with a selection of 'concords'. NC: Noun class; QC: quantitative concord; NEG SC: negative subject concord, PRON: pronominal; RC: relative concord; EC: enumerative concord; oral: oral prefix (see also AU and PRE in Table 1).

NC	**QC (all)**		**NEG SC**	**PRON**	**RC**	**QC$_{\mathbf{dwa}}$**	**EC**
	QC$_{\mathbf{oral+onke}}$	**QC$_{\mathbf{nke}}$**					
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-

is prefixed with the oral prefix (see AU and PRE in Table 1) of the noun class of that first noun—i.e., a named OWL class/DL concept on the left-hand side of $\sqsubseteq$ in the ontology—and modified based on what the prefix was; e.g.:

(U1) `Girl` $\sqsubseteq$ ...
<u>wonke</u> umfana ... ('<u>each</u> girl...'; *u-* + *-onke*)
<u>bonke</u> abafana ... ('<u>all</u> girls...'; *ba-* + *-onke*)

This looks laborious, but it can be simplified computationally. The oral prefixes are stable for each noun class, so one can pre-compute the complete list of nominal heads (column 2 in Table 2) and carry out a simple look-up of the term when generating the verbalization. Whether singular or plural should be used depends on the context, and will be addressed below and in Algorithms 1 and 3.

Subsumption. There are different ways of carving up the nouns to determine which rules apply for verbalizing subsumption. One can use either the living/non-living thing distinction into which nouns are grouped, but we postulate that a purely syntactic approach may be feasible, which is easier to implement computationally. The latter requires one to select the right copulative ('is a'), which is based on the first letter of the noun of the superclass, being *ng* for nouns starting with a-, o-, or u-, or else *y*. In addition, among generic and determinate verbalization, the generic is chosen. For instance:

(S1) `Giraffes` $\sqsubseteq$ `Animals`
izindlulamithi yizilwane (‘giraffes are animals’; animals: *izilwane*)
(S2) `MedicinalHerb` $\sqsubseteq$ `Plant`
ikhambi ngumuthi (*umuthi*: (medicinal) plant)

The general pattern that emerges is as follows: *<noun1> <ng/y depending on first letter of noun2><noun2>*. This holds for when the subsumption is not followed by negation. If it is followed by negation, then the verbalization for subsumption and negation are combined into one term and the copulative is omitted. This can be with or without including the quantifiers in the verbalization. For instance:

(SN1) `Cup` $\sqsubseteq \neg$`Glass`
zonke izindebe aziyona ingilazi (‘all cups are not a glass’)

Here, we address only the negation in the context of the subsumption symbol. The *azi-* is the negative subject concord (NEG SC) for the noun class of the noun (name of the OWL class) on the left-hand side of the subsumption (noun class 10 for *izindebe*), and the *-yona* is the part indicating the pronomial (PRON) for the noun of the class on the right-hand side of the subsumption (*ingilazi* is in noun class 9a), which is then adjusted for each class; see Table 2. Thus, the pattern for simple disjointness is: *<QC-all of* NC_x*> <plural of noun1 with* NC_x*> <NEG SC of* NC_x*><PRON of* NC_y*> <noun2 with* NC_y*>*. The high-level algorithm for simple class subsumption and disjointness for isiZulu is included as Algorithm 1, which is more elaborate compared to the ‘is a’ and ‘is not a’ in English verbalization templates. We leave the more complicated cases, like $\forall R.C \sqsubseteq \exists S.(D \sqcap E)$, for future work, as well as negation in other contexts.

Conjunction. The ‘and’ in the sense of a list of things uses *na*, but this changes into (a + i =) *ne* or (a + u =) *no*, depending on the first letter of the noun that follows it, and this *no* or *ne* is then a prefix to the second noun that drops its first letter (always a vowel); e.g. (C1). Conjunction as connective of clauses uses a different term for ‘and’, being *kanye* or *futhi*; e.g., (C2).

(C1) `Milk` $\sqcap$ `Butter`
Ubisi nebhotela (*Ubisi* + *na* + *Ibhotela*)
(C2) ... $\exists$`has_filling.Cream` $\sqcap \exists$`has_Icing.Lemon_flavour` ...
...kune zigcwalisa ukhilimu kanye nezinye uqweqwe olunambitheka_ulamula...

That is, the pattern for the enumerative-and is *<noun1> <na/ne/no depending on noun2><noun2 minus first character>*, and for the connective-and it is *<first clause> <kanye> <second clause>*. Algorithm 2 first recognises whether it is a listing of atomic classes or several axioms—check the first element after the $\sqcap$: if it is an OWL object or data property (relationship or attribute), then use the connective-and, else an enumeration-and—and if the former, then it checks the first letter of the second word to choose the *na*/*ne*/*no*.

Algorithm 1 Determine the verbalization of simple taxonomic subsumption

$\mathcal{C}$ set of classes, language $\mathcal{L}$ with $\sqsubseteq$ for subsumption and $\neg$ for negation; variables: A axiom, NC_i nounclass, $c_1, c_2 \in \mathcal{C}$, a_1 term, a_2 letter; functions: $getFirstClass(A)$, $getSecondClass(A)$, $getNC(C)$, $pluralizeNoun(C, NC_i)$, $checkNegation(A)$, $getFirstChar(C)$, $getNSC(NC_i)$, $getPNC(NC_i)$.
Require: axiom A with a $\sqsubseteq$ has been retrieved
$c_1 \leftarrow getFirstClass(A)$ {get subclass}
$c_2 \leftarrow getSecondClass(A)$ {get superclass}
$NC_1 \leftarrow getNC(c_1)$ {determine noun class by augment and prefix or dictionary}
$NC_2 \leftarrow getNC(c_2)$ {determine noun class by augment and prefix or dictionary}
if $checkNegation(A) = true$ **then**
    $NC_1' \leftarrow$ lookup plural nounclass of NC_1 {from known list}
    $c_1' \leftarrow pluralizeNoun(c_1, NC_1')$
    $a_1 \leftarrow$ lookup quantitative concord for NC_1' {from quantitative concord (QC(all)) list}
    $n \leftarrow getNSC(NC_1')$ {get negative subject concord for c_1'}
    $p \leftarrow getPNC(NC_2)$ {get pronomial for c_2}
    RESULT $\leftarrow$ ' a_1 c_1' np c_2. ' {verbalise the disjointness}
else
    $a_2 \leftarrow getFirstChar(c_2)$ {retrieve first letter of c_2}
    select case
        a_2 = 'i' **then**
            RESULT $\leftarrow$ ' c_1 yc_2 ' {verbalise as taxonomic subsumption with y}
        a_2 = {'a', 'o', 'u'} **then**
            RESULT $\leftarrow$ ' c_1 ngc_2 ' {verbalise as taxonomic subsumption with ng}
        $a_2 \notin$ {'a', 'i', 'o', 'u',} **then**
            RESULT $\leftarrow$ 'this is not a well-formed isiZulu noun'
    end select case
end if
return RESULT

Existential Quantification. There are multiple aspects to the verbalization, and we focus here only on the existential quantification, not the verb, due to additioanl complexities of verb tenses and the prepositions that are typically put in the name of the object property in the ontology or conceptual data model. Choices are discussed in [11], and we show here only the final outcome, using the *-dwa* option. For instance:

(E1) Professor $\sqsubseteq$ $\exists$teaches.Module ('all professors teach <u>at least one</u> module')
bonke oSolwazi bafundisa isifundo <u>esisodwa</u>

The *esisodwa* in (E1) is composed of the relative concord (RC), which is determined by the noun class system, that is attached to the quantitative concord (QC) and then suffixed with the quantitative suffix *-dwa*; e.g.: *esi* (RC7) + *so* (QC7) + *dwa*. The RC and QC for each noun class is fixed, and is included in Table 2. Overall, the following pattern is obtained: *<QC-all of NC_x> <pl. noun1 of NC_x> [conjugated verb] <noun2 of NC_y> <RC for NC_y><QC for NC_y>dwa*; This is presented in Algorithm 3.

Algorithm 2 Determine the verbalization of conjunction in an axiom

$\mathcal{R}$ is the set of relationships, $\mathcal{A}$ of attributes, $\mathcal{C}$ of classes, and language $\mathcal{L}$ uses $\sqcap$ to denote conjunction; variables: e_2, c_1 a letter, A axiom; functions: $getNextVocabularyElement(A)$, $getFirstChar(e_2)$.

Require: axiom with a $\sqcap$ has been retrieved and position in string is known

$e_2 \leftarrow getNextVocabularyElement(A)$ {retrieve element after the $\sqcap$}
if $e_2 \in \mathcal{R} \cup \mathcal{A}$ **then**
    RESULT $\leftarrow$ ' kanye ' {verbalise $\sqcap$ as kanye}
else
    if $e_2 \in \mathcal{C}$ **then**
        $c_1 \leftarrow getFirstChar(e_2)$ {retrieve first letter of e_2}
        select case
            c_1 = 'i' **then**
                $e_2^- \leftarrow$ drop c_1 from e_2
                RESULT $\leftarrow$ ' nee_2^- ' {verbalise $\sqcap$ with ne- prefix}
            c_1 = 'u' **then**
                $e_2^- \leftarrow$ drop c_1 from e_2
                RESULT $\leftarrow$ ' noe_2^- ' {verbalise $\sqcap$ with no- prefix}
            c_1 = 'a' **then**
                $e_2^- \leftarrow$ drop c_1 from e_2
                RESULT $\leftarrow$ ' nae_2^- ' {verbalise $\sqcap$ with na- prefix}
            $c_1 \notin$ {'i', 'u', 'a'} **then**
                RESULT $\leftarrow$ 'this is not a well-formed isiZulu noun'
        end select case
    else
        RESULT $\leftarrow$ 'this is not a well-formed axiom'
    end if
end if
return RESULT

4 Discussion

For grammatically less complicated languages that have an isolating morphology, such as English, verbalization templates are known to be an effective way to tackle the problem, and may even suffice. This approach breaks down for grammatically richer languages [9], and for isiZulu, we have, so far, not found a single case where a plain template suffices. The insufficiently structured grammar rules in the outdated documentation made it also clear that committing to a comprehensive specification of the isiZulu grammar in such a way as to be computationally useful and correct (e.g., by using the Grammatical Framework [http://www.grammaticalframework.org/]), will take a substantial amount of resources. Such resources are not available at present, yet something has to be done for multilingual knowledge repositories that are adequate in the multilingual society in South Africa. Despite that no software has been presented in this paper, we hope to have provided some motivational use cases for investigation, which is benefiting both isiZulu linguistics and ICT in general, and introduced some interesting new challenges for the verbalization of logical theories in gram-

Algorithm 3 Determine the verbalization of existential quantification with object property (first, basic, version)

$\mathcal{C}$ set of classes, language $\mathcal{L}$ with $\sqsubseteq$ for subsumption and $\exists$ for existential quantification; variables: A axiom, NC_i noun class, $c_1, c_2 \in \mathcal{C}$, $o \in \mathcal{R}$, a_1 a term; r_2, q_2 concords; functions: $getFirstClass(A)$, $getSecondClass(A)$, $getNC(C)$, $pluralizeNoun(C, NC_i)$, $getRC(NC_i)$ $getQC(NC_i)$.

Require: axiom A with a $\sqsubseteq$ and a $\exists$ on the rhs of the inclusion has been retrieved

$c_1 \leftarrow getFirstClass(A)$ {get subclass}

$c_2 \leftarrow getSecondClass(A)$ {get superclass}

$o \leftarrow getObjProp(A)$ {get object property}

$NC_1 \leftarrow getNC(c_1)$ {determine noun class by augment and prefix or dictionary}

$NC_2 \leftarrow getNC(c_2)$ {determine noun class by augment and prefix or dictionary}

$NC_1' \leftarrow$ lookup plural nounclass of NC_1 {from known list}

$c_1' \leftarrow pluralizeNoun(c_1, NC_1')$

$a_1 \leftarrow$ lookup quantitative concord for NC_1' {from quantitative concord (QC(all)) list}

$o' \leftarrow AlgoConjugate(o, NC_1)$ {call algorithm $AlgoConjugate$ to conjugate o}

$r_2 \leftarrow getRC(NC_2)$ {get relative concord for c_2}

$q_2 \leftarrow getQC(NC_2)$ {get quantitative concord for c_2 from the QC_{dwa}-list}

RESULT $\leftarrow$ ‘ a_1 c_1' o' c_2 $r_2 q_2$dwa. ’ {verbalise the simple axiom}

return RESULT

matically rich languages. We will continue to extend the algorithms, add more, and implement them.

The algorithms may also be of use for machine translation. For instance, Google Translate English-isiZulu translates, e.g., “mix the sugar and milk and butter” as “*hlanganisa ushukela nobisi ibhotela*” (translation d.d. 14-1-2014), which misses the second conjunction in the enumeration, whereas a ushukela $\sqcap$ ubisi $\sqcap$ ibhotela with Algorithm 2 obtains the correct verbalisation/translation (*ushukela nobisi nebhotela*). Similarly, “all giraffes eat twigs” is translated as “*yonke izindlulamithi udle amahlumela*” (translation d.d. 14-1-2014), but *izindlulamithi* is in noun class 10, not 9, so it goes with *zonke* instead, not Google Translate's *yonke*. This can be correctly verbalised following Algorithm 1, line 9.

An aspect of further investigation is the implementability of subsumption with the living/non-living thing distinction compared to the syntax-based shortcut, as it is not clear yet whether a syntax-based criteria holds for other cases when a distinction is made between living and non-living things. Such annotations will be less than assigning noun classes to each term. Also, this means there has to be some way to encode such multilingual information, which may be possible by extending the *Lemon* model [13] or putting it in a designated annotation field.

5 Conclusions

Verbalizing ontologies in isiZulu requires more than a template-based approach for each construct investigated. We provided novel verbalization patterns for simple subsumption, disjoint classes, conjunction, and basic options with quantification. The main features complicating verbalization in isiZulu were the 17 noun classes with embedded semantics in the term, the agglutinative nature of isiZulu, and contextual knowledge about the position of the symbol in the axiom.

There are many avenues for further works on the verbalization rules, with more variations on the basic axioms, more construct, and conjugation. There are also questions concerning how to make the ontology multilingual so that it covers the aspects that need to be recorded to facilitate verbalization.

References

1. Alberts, R., Fogwill, T., Keet, C.M.: Several required OWL features for indigenous knowledge management systems. In: Proc. of OWLED'12. CEUR-WS, vol. 849 (2012), 27-28 May, Crete, Greece
2. Bosca, A., Dragoni, M., Francescomarino, C.D., Ghidini, C.: Collaborative management of multilingual ontologies. In: Buitelaar, P., Cimiano, P. (eds.) Towards the Multilingual Semantic Web, p. (in press). Springer (2014)
3. Curland, M., Halpin, T.: Model driven development with NORMA. In: Proc. of HICSS-40. pp. 286a–286a. IEEE Computer Society (2007), los Alamitos, Hawaii
4. Doke, C.: Text Book of Zulu Grammar. Witwatersrand University Press (1927)
5. Doke, C.: Bantu Linguistic Terminology. London: Longman, Green and Co (1935)
6. Dongilli, P., Franconi, E.: An Intelligent Query Interface with Natural Language Support. In: Proc. of FLAIRS'06. Melbourne Beach, Florida, USA (May 2006)
7. Fogwill, T., Viviers, I., Engelbrecht, L., Krause, C., Alberts, R.: A software architecture for an indigenous knowledge management system. In: Indigenous Knowledge Technology Conference 2011 (2011), windhoek, Namibia, 2-4 November 2011
8. Fuchs, N.E., Kaljurand, K., Kuhn, T.: Discourse Representation Structures for ACE 6.6. Tech. Rep. ifi-2010.0010, Dept of Informatics, University of Zurich, Switzerland (2010)
9. Jarrar, M., Keet, C.M., Dongilli, P.: Multilingual verbalization of ORM conceptual models and axiomatized ontologies. Starlab technical report, Vrije Universiteit Brussel, Belgium (February 2006)
10. Kaljurand, K., Kuhn, T., Canedo, L.: Collaborative multilingual knowledge management based on controlled natural language. Semantic Web J. (submitted (2013))
11. Keet, C., Khumalo, L.: Toward verbalizing logical theories in isizulu. In: Proc. of CNL'14. p. (accepted). LNAI, Springer (2014), 20-22 August 2014, Galway, Ireland
12. Kuhn, T.: A principled approach to grammars for controlled natural languages and predictive editors. Journal of Logic, Language and Information 22(1), 33–70 (2013)
13. McCrae, J., et al.: The Lemon cookbook. Tech. rep., Monnet Project (2012)
14. Motik, B., Grau, B.C., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C.: OWL 2 Web Ontology Language Profiles. W3C recommendation, W3C (27 Oct 2009)
15. Ngcobo, M.N.: Zulu noun classes revisited: A spoken corpus-based approach. South African Journal of African Languages 1, 11–21 (2010)
16. Pretorius, L., Bosch, S.E.: Enabling computer interaction in the indigenous languages of South Africa: The central role of computational morphology. ACM Interactions March + April, 56 (2003)
17. Schwitter, R., et al.: A comparison of three controlled natural languages for OWL 1.1. In: Proc. of OWLED'08DC (2008), washington, DC, USA, 1-2 April 2008
18. Spiegler, S., van der Spuy, A., Flach, P.A.: Ukwabelana – an open-source morphological Zulu corpus. In: Proc. of COLING'10. pp. 1020–1028. ACL (2010)