Is *lemon* Sufficient for Building Multilingual Ontologies for Bantu Languages?

Catherine Chavula and C. Maria Keet

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

Abstract. The current enormous amount of data on the Semantic Web and its increasing uptake raises the question of how this data can be accessed in several languages. OWL provides limited support for multilingualism through the use of an annotation property. However, it is known that more expressive models are required for linguistically demanding applications. Among the possible solutions, Lexicon Model for Ontologies (*lemon*) enables associating linguistic information with ontology elements by separating the lexical from the ontological layer. This paper investigates whether *lemon* is sufficient for specifying multilingual ontologies for Bantu languages. Specifically, the paper: (i) identifies the requirements for building lexica in *lemon* format for Bantu languages; (ii) describes the results in overcoming some of the challenges, notably concerning noun classes; and (iii) presents some open issues that will have to be addressed to increase usability of *lemon*.

1 Introduction

Multilingual ontologies are required to provide access to ontology-based information in the languages of the users. However, most ontologies are available in English, i.e., ontology elements are named with English terms, which, at least, brings afore the requirement to localise these ontologies to other natural languages. For example, vocabularies for the Semantic Web such as Friend-Of-A-Friend (FOAF) [4] and GoodRelations [19] have annotations in English only but are widely used to annotate resources on the Web. OWL provides support for multilingualism using annotation properties such as `rdfs:label` and `rdfs:comment`; e.g., a lexicalisation of the class foaf:Person can be given in English, Chichewa, and isiZulu through adding annotations as shown in Fig. 1. However, the amount of linguistic annotation that can be included in this labelling system is limited and most multilingual applications require more data such as Part of Speech (POS) and grammatical features, among others. Moreover, ontologies are for representing knowledge, and such linguistic data need not to be included in an ontology. Several approaches that separate the ontological layer from the terminological layer have been proposed [6, 8, 23] and the W3C Community Group ontolex-lemon submission is under way[1]. Notably, the LExicon Model for ONtologies (*lemon*) [22, 23] separates the ontological layer and linguistic layer, and

[1] `http://www.w3.org/community/ontolex/wiki/Final_Model_Specification`

```
<rdfs:Class rdf:about="http://xmlns.com/foaf/0.1/Person">
<rdfs:label xml:lang = "en"> person </rdfs:label>
<rdfs:label xml:lang = "ny"> munthu </rdfs:label>
<rdfs:label xml:lang = "zu"> umuntu </rdfs:label>
</rdfs:Class>
```

Fig. 1. The class foaf:Person with language labels for English (`en`), Chichewa (`ny`) and isiZulu (`zu`).

is gaining momentum in adoption for multilingual ontologies. In *lemon*, each ontology element is associated with an entry in a separate lexical resource, which in turn is annotated with linguistic data. In this manner, the ontology provides the semantics of terms in a lexical resource while the entries provide the lexicalisation of the ontology elements. This looks like a promising solution to problems identified for Indigenous Knowledge Management Systems [1] as well as possible ontology-driven applications in the ICT4D area and ontology verbalisation [21].

Bantu languages are characterised by a comprehensive noun class and concordial agreement system among terms. A noun class determines the affixes on nouns in that noun class and other elements; e.g. *umfula* ('river') is in noun class 3, where *-fula* is the stem and *um-* the prefix for that noun class. Each noun class has its own concords for the noun and lexical categories such as adjectives and verbs. This is emblematic for all Bantu languages that have between 10 and 23 noun classes, depending on the language. This paper investigates whether *lemon* can be used to model these characteristics to create multilingual ontologies with Bantu languages terms. The general issue on representing lexical information is addressed, which requires a new ontology module that has noun class as a grammatical category—the new noun class system ontology—and the use of the *lemon* morphology module is described with elements of that ontology. The development of a lexicon for FOAF [4] and GoodRelations [19] in Chichewa are used to examine implementability of the requirements for multilingual ontologies to accommodate Bantu languages. To ensure the evaluation and proposed solution is not fitted to Chichewa only, isiZulu is also considered, which is in a different sub-family of Bantu languages.

The remainder of this paper is structured as follows. Section 2 describes requirements for ontology-based applications with Bantu languages and Section 3 discusses related work. Section 4 describes the process of enriching a domain ontology with lexical information using the *lemon* model. Section 5 discusses the challenges in the process of developing the resources and Section 6 concludes and presents ideas for future work.

2 Linguistic Requirements for Ontologies in Bantu Languages

Bantu languages are characterised by complex morphosyntactic features due to a Noun Class System (NCS) and a system of concordial agreement: each noun belongs to a noun class (nc) and each class has its collection of affixes, which then also determines the agreement markers (grammatical concord) on related lexical categories such as adjectives and verbs. For instance, *ubuntu* is in nc:14 (*ubu-*+*-ntu*) and *umuntu* in nc:1 (*umu-*+*-ntu*) in isiZulu, and *munthu* (*mu-*+*-nthu*) in nc:1 in Chichewa. While each noun class typically has some semantics (e.g., nc:1 for humans and other animates), the semantics of the noun classification in different Bantu languages is still a topic of investigation [27]. Bantu noun classes are identified using Arabic Numerals based on different classification methods and naming schemes. Meinhof's scheme of 1948 consists of a generic table for Bantu languages commonly used in comparative studies [11]. The classes are grouped in pairs of singular and plural forms with their associated prefixes. The number of classes varies among languages, with some languages exhibiting similarities in the prefixes; see Table 1 for examples. A linguistic requirement for multilingual ontologies of Bantu languages is thus to have a way to annotate the nc of the term whose sense is denoted by the OWL class.

In addition, an ontology-based task such as ontology verbalisation requires information about a class's nc to determine what combination of prefixes to add to the verb in the name of the object/data property [21]. For example, the foaf:knows property has as domain and range foaf:Person, and in order to verbalise a fact using these vocabulary elements, the nc for person as well as its associated prefixes such as tense need to be available. Lexicalising the fact, 'John knows Jim' in Chichewa would be *John amadziwa Jim* (agreement marker underscored), but if the domain was of a different class (not a person), then the agreement marker for that other nc has to be used. For instance, eats with verb stem *-dla*: when a giraffe (in nc:9) eats something it is *idla* and for a person (nc:1) it is *udla*. That is, an object or data property is not simply named with a verb in third person singular, as is deemed good practice in ontology engineering, but the term depends on the domain and range and its use in an axiom.

As all peculiarities of Bantu languages cannot be covered here, this study uses Chichewa and isiZulu. Chichewa, a dialect of the Nyanja language, is spoken by over 12 Million people in Malawi [25], and isiZulu is among the Nguni languages of South Africa spoken as L1 language by over 10 million people. According to the Guthrie classification [18] of Bantu languages into zones based on language characteristics, Chichewa is in zone N, unit N31, and isiZulu in zone S, unit S42.

3 Related Work

Research into multilingual ontologies investigates models for representing a lexical/terminological layer for ontologies and so far three architectures have been proposed [13], namely: (i) using multilingual OWL annotation property, such as `rdfs:label` and `rdfs:comment`; (ii) mapping ontology elements designed for different cultures and languages, e.g., EuroWordNet[2] and BabelNet [12]; and (iii) using external lexical resources to linguistically enrich ontologies [7, 6, 8, 9,

[2] `www.illc.uva.nl/EuroWordNet/`

Table 1. Chichewa [2] and isiZulu noun classes (NC). Prefixes are added to the stem to form a word (grouped by singular and plural); ∅: the class does not have an affix.

NC	Prefix (Chichewa)	Prefix (isiZulu)	Examples	Meaning
1	mu-	um-,umu-	munthu	person
2	a-	aba-	anthu	people
1a		u-	ubaba	father
2a		o-	obaba	fathers
1a	∅		galu	dog
2	a-		agalu	dogs
3a		u-	ushizi	cheese
(2a)		o-	oshizi	cheeses
3	mu-	um-,umu-	mudzi	village
4	mi-	imi-	midzi	villages
5	∅	i-, ili-	tsamba	leaf
6	ma-	ama-	masamba	leaves
7	chi-	isi-	chipatso	fruit
8	zi-	izi-	zipatso	fruits
9a		i-	irabha	rubber
(6)		ama-	amarabha	rubbers
9	∅	i- , in-	njoka	snake
10	∅	izi- , izin-	njoka	snakes
11		u-, ulu-	uthi	stick
(10)		izi-, izin-	izinthi	sticks
12	ka-		kagalimoto	small car
13	ti-		timagalimoto	small cars
14	u-		ulendo	journey
(6)	ma-		maulendo	journeys
14		ubu-	ubuhle	beauty
15	ku-	uku-	kuwerenga	to read
16	pa-		pamsika	round the market
17	ku-	ku-	kumsika	at the market
18	mu-		mumsika	in the market

26]. The last approach provides a means of modelling morphosyntactic features required by more linguistically demanding tasks such as ontology-based information access, and is the focus of this paper. Regarding models for building lexical resources, there have been different approaches for publishing lexical data, and WordNet is the most popular English lexical database [15], which is organised around semantic relationships between words called synsets (i.e., synonym sets). The WordNet model has also been applied to other languages and efforts to create a global WordNet are underway[3]. The WordNet model, however, is limited in working with environments of diverse requirements [23] and Bantu languages have a complex morphosyntactic structure which cannot be modelled in the WordNet structure. The Lexical Markup Framework (LMF) [16], an ISO stan-

[3] http://globalwordnet.org/wordnets-in-the-world/

dard for representing lexica in XML/UML, provides a format for data exchange and interoperability. However, it is difficult to fully exploit lexica in LMF format since the semantics of the model are not formalised. Models that associate lexical information with ontological elements such as LexInfo [8], LexOnto [9], Linguistic Information Repository (LIR) [26], LingInfo [6] and *lemon* [23] have been proposed. *Lemon* is informed by LMF, LIR and LexInfo, and provides a means for separating the lexical and semantic layer of the ontology. A *lemon* lexicon defines how ontological elements are realised in a particular natural language. The *lemon* design is based on a premise that a word sense relates to the conceptualisation of the world which can be specified in an ontology. It provides a rich model for modelling semantic multilingual knowledge thanks to its modular design and extensibility. It consists of a core module and five task-specific modules, namely: Morphology, Syntax and Mapping, Phrase structure, Variation and Linguistic Description [24]. *lemon* has been used with many languages, notablys English [28] and German [10], and a collection of 50 languages in BabelNet 2.0 [12]. To promote interoperability among the linguistic resources, agreed-upon grammatical categories are used. For example, LMF and *lemon* advocate using general language ontologies such as GOLD ontology [14] and the ISOcat data category registry [5].

Zooming in on lexical data models for Bantu languages, Bosch et al. [3] proposed to represent South African Bantu languages' complex structure using sub-entries. A typical entry is organised into head and body tags, with head containing the stem or root and body containing syntactic and morphological information about the stem. From the entries in [3], it can be seen that repetition is used to model morphological processes such as inflection. However, repetition of NCS data in the lexicon can make the size of the lexicon to grow very big as the same information is repeated within each entry. The *lemon* morphology module provides an economical way of modelling highly inflectional languages.

4 Building Lemon Lexica in Bantu Languages

Before immediate usage, several practical modelling and design choices have to be made, which are described first. This is followed by an assessment of the practical feasibility of using *lemon* with Bantu languages, by lexicalising the FOAF and GoodRelations ontologies in Chichewa.

4.1 General Modelling Aspects

Ontology-based applications for Bantu languages require morphosyntactic data, notably: (i) defining a NCS with associated prefixes and associating the NCS with the lexical entries; (ii) defining rules for verbs and adjectives to ensure agreement with the noun class; and (iii) writing rules for agglutination process.

Modelling the NCS. The first requirement on handling the NCS can be accommodated by *lemon*'s extensible approach that promotes the use of externally defined properties to annotate lexica. *lemon* advocates the usage of linguistic ontologies such as GOLD [14] and linguistic data repositories such as ISOcat [5] and user defined properties. However, the properties defined in GOLD and ISOcat do not fully meet the requirement. GOLD has a concept gold:ArabicNumeralGender, reflecting that the NCS has been proposed by some Bantu linguists [11] as a grammatical gender system. The ISOcat so-called "data categories" has an isocat:otherGender that is intended to be used in the same manner analogous to gender that has genders as instances. However, these properties do not have their analogous instances for the 23 Bantu noun classes and thus do not provide a means of labelling the entries. As *lemon*'s LexInfo ontology for linguistic annotation imports ISOcat, it does not provide for this either. Moreover, it is arguable whether the NCS is the same as gender as is commonly understood (masculine etc.; see also [11]), as, e.g., isiZulu nc:9 is for animals and nc:14 for abstract nouns, i.e., semantic groups that have nothing to do with that interpretation of gender. Different schemes have been developed to specify the Bantu NCS for individual Bantu languages. Instead of encoding each scheme, we note that the Meinhof (1948) one is used among linguists as standard for defining noun classes; hence, this is used here as well, for it facilitates cross-language comparisons and use.

Because of the shortcomings of the extant linguistic resources, an ontology based on the Bantu languages' noun class system was developed to allow the annotation of ontology elements with noun class information. The goal of the ontology was to specify the conceptualisation of nouns in the sense of Bantu languages structure. In order to achieve this, the ncs:partOfSpeech class was introduced, which subsumes ncs:Noun, and a ncs:Property class was added that subsumes ncs:MorphoSyntacticProperty to capture the characteristics of nouns in this domain. The ncs:Gender class is made a sibling class of ncs:NounClass, which are subsumed by ncs:MorphoSyntacticProperty. In this regard, other classes for classifying nouns that do not relate to the NCS and nominal morphology can be added without interfering with the conceptualisation of the NCS. The Bantu languages NCS concepts based on the Meinhof noun classification scheme were specified with the assumption that all specific noun class classifications of Bantu languages can be aligned to this scheme. The Meinhof classification has 23 classes labelled mostly with arabic numerals only. However, classes recognised later after the Meinhof scheme are tabulated as subdivisions of Meinhof classes to freeze the number of classes. Therefore, the numbering scheme of Arabic numerals is sometimes augmented with an alphabetical letter, such as class 1a, 2b and 3a to signify that the classes had elements of class 1, 2, and 3, respectively (but is disjoint from it). In order to capture all these complexities, classes labelled with Arabic Numeral classes as well as their derivatives were added as separate disjoint classes.

Object properties were defined in the ncs ontology to specify the relationships between the classes in the ontology as conceptualised in the domain of Bantu Languages. The properties ncs:hasNounclass, ncs:hasNumber, ncs:hasPlural and ncs:hasSingular were introduced. Additionally, restrictions were added to specify the constraints on the relationship as well as relationship characteristics. For example, ncs:hasPlural is used to specify that the classes can only have specific classes as their plural and vice versa using ncs:hasSingular and that these classes are inverse of each other. Fig. 2 depicts some of the classes and an annotation, and the ontology is accessible online at `http://www.meteck.org/files/ontologies/`. Following this, the nouns of the names of the classes in an ontology can now be annotated with their noun class.

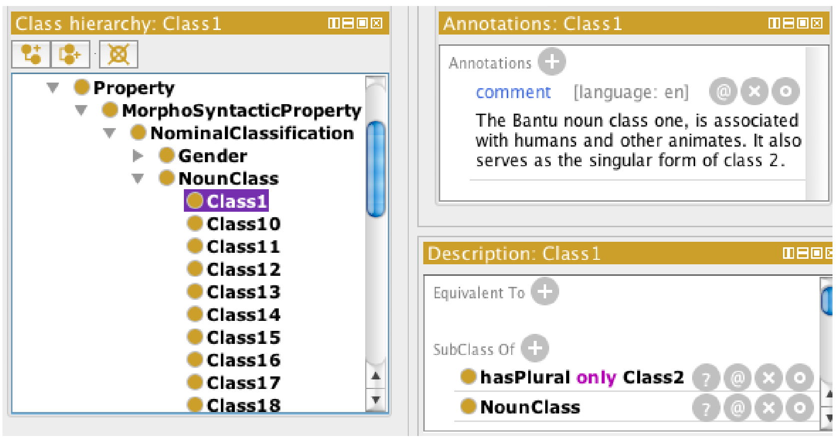

Fig. 2. Bantu languages Noun Classes as subclasses of NounClass class.

Specifying Rules for Word Variation. The NCS determines how words of grammatical categories such as nouns are inflected by adding prefixes to stems. The *lemon* morphology module uses transformational rules to account for the inflection of words in a particular language using PERL-like regular expressions. The nominal morphology in Bantu languages is based on the prefixes of noun classes while other syntactic categories like verbs and adjectives depend on agreement markers prescribed by the noun classes. For example, a verb stem in Chichewa can have over 20 different forms generated through the concatenation of agreement markers and other syntactic components to roots, in a similar way as the eats example for isiZulu in the previous section. Although rules can be written once for each language using the *lemon* morphology module, the transformation system for the arbitrary case is very complex and would both overburden the *lemon* system and render such rules useless outside the setting of multilingual ontologies. No computational version of such generative rules exist yet (except initial theoretical results [20]), and in that light it will be more effective to develop a separate generic grammar engine that can be plugged in. Nevertheless, it may be feasible to write a subset of the rules to generate lexica for 'limited' ontologies, like those that are mainly taxonomies and those where object/data properties have their domain and range declared in detail. An example is shown in Fig. 3 for partial noun morphology for nc:1 and nc:2 using the *lemon* morphology module. The rule is partial, as the prefix can differ also depending on the stem.

```
:NYNC1_2 a lemon:MorphPattern ;
lemon:transform [
lemon:rule "mu~" ;
lemon:generates [
lexinfo:number lexinfo:singular;
ncs:hasNounClass ncs:class1
]],[
lemon:rule "a~" ;
lemon:generates [
lexinfo:number lexinfo:plural;
ncs:hasNounClass ncs:class2]].
```

Fig. 3. Chichewa nc:1 and nc:2 morphology represented using *lemon* rules.

Agglutination. Bantu languages are highly agglutinative, to the extent that a word can be composed of over five constituents. For comparatively well-studied languages within the Bantu language family, a few rules can be written based on the Morphology module of *lemon*. However, some of the aspects of the Bantu languages morphology cannot be handled using the proposed approach which favours concatenation morphology. Due to space limitations, we do not discuss those details here. The shortest and simplest example for an object property (verb) is included in Fig. 5, ignoring such issues as past tense, variation due to the domain being of a different noun class, and the habit in 'English OWL ontologies' to include a preposition in the name of the object property. It is not doable in the general case as a form variation generator and novel methods need to be developed to accommodate such phenomena for object properties. More complications exist if one were to partially verbalise writing an axiom, as in Protégé's "class expression editor", as even conjunction 'and' can be glued to the second class, modifying it [21] (e.g., *ushizi* becomes *noshizi* for 'and cheese').

4.2 FOAF and GoodRelations in Chichewa

The lexica for the selected ontologies were written manually due to unavailability of language resources and tools for the two languages.

The translation of each ontology element was collected and a further analysis was done to determine how it can be lexicalised in *lemon* format. The lexicalisation ended up to be a 1:1 correspondence between the lexical entry and the English terms of ontology elements. Due to the absence of some terms in Chichewa, more specifically for technical terms, some of the elements were not lexicalised: no equivalences were provided for FOAF phrases foaf:'sha1sum of a personal mailbox URI name', foaf:DNAchecksum (a joke), and foaf:openid. Secondly, due to unavailability of equivalent terms, the translation of most of the vocabulary elements turned out to be phrase elements, which they essentially are in English as well;

e.g. the foaf:PersonalProfileDocument class consists of three English words that is a phrase as well when lexicalised in Chichewa. An example of a fully lexicalised entry is shown in Fig. 4 for foaf:Person (*munthu*), specifying its part of speech, plural form and sense referenced through the FOAF `foaf:Person`.

```
lemon:entry :munthu.
:munthu a lemon:LexicalEntry ;
lemon:canonicalForm [lemon:writtenRep "munthu"@ny;
lexinfo:number lexinfo:singular
ncs:hasNounClass ncs:class1] ;
lemon:otherForm [lemon:writtenRep "anthu"@ny;
lexinfo:number lexinfo:plural;
ncs:hasNounClass ncs:class2];
lexinfo:partOfSpeech lexinfo:noun;
lemon:sense [lemon:reference foaf:Person].
```

Fig. 4. Chichewa foaf:Person entry, using *lemon*, LexInfo, and the ncs ontology.

Most of the FOAF object properties consisted of multi-word phrases and their lexicalisations turned out to be complete sentences in a natural language and has the issue of variation depending on the noun class of the noun of the name of the OWL class associated to it. Where possible, the *lemon* Morphology module was used to write the rules for the verbs, as illustrated in Fig. 5 for foaf:knows (*dziwa*). This was doable, as both the domain and range of foaf:knows are declared to be foaf:Person, therewith greatly reducing the set of possible rules for this object property. Data properties in the FOAF vocabulary are associated with aspects of people such as birthday, first name and surname, which were easier to translate. Overall, the FOAF *lemon* lexicon in Chichewa covers 90% of the classes and individuals of the original English FOAF, and 80% of its properties.

The GoodRelations ontology is an e-commerce vocabulary for product, price, store and company data and is widely used on the web for annotating resources and has the support of Yahoo and Google [19]. The ontology is highly domain specific and lexicalisation was a big challenge, partly due to the absence of suitable terms in Chichewa. Only 25% of the entities were lexicalised in Chichewa, which mostly included nouns of classes that belong to noun classes that were not already covered in the FOAF vocabulary.

5 Discussion

The primary challenge of building *lemon* lexica for ontologies in Bantu languages is how to handle the NCS and morphological structure of the Bantu languages. Modelling the NCS as a gender classification is unsatisfactory and the gender grammatical category in ISOcat data category registers and GOLD is irrelevant. Hence, the requirement to build an ontology module to support this aspect.

```
:NYNC1_2_verb_stem a lemon:MorphPattern ;
lemon:transform
       :present_transform ,
       :agreement_transform.
:present_transform
       lemon:rule "ma~" ;
       lemon:rule "i~/ma~";
lemon:nextScope :agreement_transform.
lemon:generates [lexinfo:tense lexinfo:present].
:agreement_transform
       lemon:rule "a~" ;
lemon:generates [lexinfo:number lexinfo:plural],
[lexinfo:number lexinfo:singular]].
:dziwa a lemon:LexicalEntry.
:dziwa lemon:pattern :NYNC1_2_verb_stem.
lemon:abstractForm [lemon:writtenRep "dziwa"@ny;
lexinfo:partOfSpeech lexinfo:verb];
lemon:sense [lemon:reference foaf:knows].
```

Fig. 5. A collection of forms of *dziwa* denoting foaf:knows.

Bantu languages have a complex morphological structure based on the NCS. However, *lemon* morphology module provides a limited rule encoding feature. An option to separate the rules from the lexica so as to foster their reusability would be beneficial. Other challenges not limited to Bantu Languages include:

- Limitations of OWL. OWL has too limited support for internationalization and ontology localization for deep semantic processing in building true multilingual ontologies.
- Limitations of the Lexical-Semantic Relationship. The direct relation of the ontology with the lexical resources cannot yet be fully exploited in applications that require referencing and non 1:1 mappings (e.g., there are 19 different translations for 'part' in isiZulu). The variation module is limited to model this aspect, and effects on inferences are unclear.
- Heterogeneity of Linguistic Annotations. GOLD and ISOcat use different formalisms for modelling linguistic annotations, and in most cases lexica can use properties from different communities which cannot be aligned, creating some confusion and it is a source of incompatibilities. Most of the resources are not based on OWL DL, so no reasoning on its properties can be done.
- Poor tool support and documentation. As *lemon* is being used and is gaining momentum, tools have to be developed that can be incorporated into existing Semantic Web platforms. More documentation also would be helpful.

Bantu languages have their own language specific challenges. There is still some work to be done in the field of linguistics on the semantic classification of nouns and morphology. Additionally, no satisfactory methods for word generation has yet been proposed. Previous studies in other Bantu languages have shown that the regular expression methods are limited, and the morphology module of *lemon* may serve less for Bantu languages lexica [17]. In addition, there is relatively little syntactic knowledge and language resources to enable other avenues of research in this area [17]. *lemon* can currently be used in building lexica for small general ontologies that are in general domains or those relevant for the application context.

6 Conclusion and future work

This paper has presented some challenges of building *lemon* lexica in Bantu languages. It required the building of a noun class system ontology module, `ncs.owl`, to allow annotation for noun classes. While rules can be encoded, the complexity of rules for Bantu languages makes it a challenge for *lemon*. Localisation of FOAF and GoodRelations into Chichewa have been experimented with, resulting in a near-full coverage of FOAF, mainly through hard-coding the classes and using only comparatively short rules for the object properties. Some remaining challenges were outlined. The ontology and FOAF and GoodRelations lexicalizations in Chichewa are available at `www.meteck.org/files/ontologies/`. The next step focuses on the use of multilingual ontologies for tasks such as ontology verbalisation, of which the first results have been obtained [20, 21] and multilingual access to data as well as evaluate other models such as ontolex-lemon.

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, p. 12p (2012), 27-28 May, Heraklion, Crete, Greece
2. Bentley, M., Kulemeka, A.: Chichewa. Lincom Europa, München (2001)
3. Bosch, S., Pretorious, L., Jones, J.: Towards Machine-Readable Lexicons for South African Bantu languages. Nordic Journal of African Studies 16(2), 131–145 (2007)
4. Brickley, D., Libby, M.: FOAF Vocabulary Specification 0.99 . `http://xmlns.com/foaf/spec/` (2014)
5. Broeder, D., et al.: A data category registry-and component-based metadata framework. In: Proc. of LREC'10. ELRA (2010), 19-21 May,Valletta, Malta
6. Buitelaar, P., Sintek, M., Kiesel, M.: LingInfo:Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: Proc. of LREC'06. ELRA (2006), 22-28 May, Genoa, Italy
7. Buitelaar, P., Sintek, M., Kiesel, M.: A multilingual/multimedia lexicon model for ontologies. In: Proc. of ESWC'06, LNCS, vol. 4011, pp. 502–513. Springer (2006), budva, Montenegro
8. Cimiano, P., Buitelaar, P., McCrae, J., Sintek, M.: Lexinfo: A declarative model for the lexicon-ontology interface. Web Semant. 9(1), 29–51 (2011)
9. Cimiano, P., Haase, P., Herold, M., Mantel, M., Buitelar, P.: Lexonto: A model for ontology lexicons for ontology-based NLP. In: Proc. of OntoLex07 Workshop held in conjunction with ISWC'07. ISW (2007), 11 Nov, Busan, South-Korea
10. Cimiano, P., Nagel, D., Unger, C.: Exploiting ontology lexica for generating natural language texts from RDF data. In: Proc. of the 14th European Workshop on NLG. pp. 10–19. ACL (2013), sofia, Bulgaria, 8-9 Aug 2013

11. Cobert, G., Mtenje, A.: Gender Agreement in Chichewa. Studies in African Linguistics 18(1), 1–38 (1987)
12. Ehrmann, M., Cecconi, F., Vannella, D., McCrae, J., Cimiano, P., Navigli, R.: Representing multilingual data as linked data: the case of BabelNet 2.0. In: Proc. of LREC'14. pp. 401–408. ELRA (2014), reykjavik, Iceland, May 26-31
13. Espinoza, M., Montiel-Ponsoda, E., Gómez-Pérez, A.: Ontology localization. In: Proc. of K-CAP'09. pp. 33–40. ACM (2009), 1-4 Sept, 2009, Redondo Beach, California, USA
14. Farrar, S., Langendoen, D.: Markup and the GOLD ontology. In: Proc. of workshop on digitizing and annotating text and field recordings. pp. 845–863. LSA Institute (2003), july 11th -13th, Michigan, USA
15. Fellbaum, C. (ed.): WordNet: An electronic lexical database. MIT press, Cambridge (1998)
16. Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M.: Lexical Markup Framework. In: Proc. of LREC'06. vol. 17, pp. 233–236. ELRA (2010), genoa - Italy, 24-26 may
17. Getao, K., Miriti, E.: Computational Modelling in Bantu Language. Advances in Systems Modelling and ICT Applications 2(2), 128–138 (2006)
18. Guthrie, M.: Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages. No. v. 1-2, Gregg (1971)
19. Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In: Proc. of EKAW'08. LNCS, vol. 5268, pp. 332–347. Springer (2008)
20. Keet, C., Khumalo, L.: Basics for a grammar engine to verbalize logical theories in isiZulu. In: Proc. of RuleML'14. LNCS, vol. 8620, pp. 216–225. Springer (2014), 18-20 August, Prague, Czech Republic
21. Keet, C., Khumalo, L.: Toward verbalizing logical theories in isiZulu. In: Proc. CNL'14. LNAI, vol. 8625, pp. 78–89. Springer (2014), 20-22 August, Galway, Ireland
22. McCrae, J., Spohr, D., Cimiano, P.: Linking lexical resources and ontologies on the semantic web with lemon. In: Proc. of ESWC'11. LNCS, vol. 6643, pp. 245–259. Springer (2011)
23. McCrae, J., et al.: Interchanging lexical resources on the Semantic Web. Language Resources and Evaluation 46(4), 701–719 (May 2012)
24. McCrae, J., et al.: The lemon cookbook. Technical report, Monnet Project (June 2012), www.lemon-model.net
25. Mchombo, S.: Chichewa. In: Spencer, A., Zwicky, A.M. (eds.) The Handbook of Morphology. Blackwell Publishers (2001)
26. Montiel-Ponsoda, E., Aguado, G., Gómez-Pérez, A., Peters, W.: Modelling multilinguality in ontologies. In: Coling 2008: Companion volume - Posters and Demonstrations. pp. 67–70 (2008), manchester,UK, August
27. Ngcobo, M.N.: Zulu noun classes revisited: A spoken corpus-based approach. South African Journal of African Languages 1, 11–21 (2010)
28. Unger, C., McCrae, J., Walter, S., Winter, S., Cimiano, P.: A lemon lexicon for DBpedia. In: Proc. of the NLP & DBpedia workshop. CEUR-WS, vol. 1064, pp. 21–25 (2013), sydney, Australia