

Preliminary steps toward an ontology for noun classes in Niger-Congo languages (Abstract)

C. Maria Keet^{1,*}

¹Department of Computer Science, University of Cape Town, 18 University Avenue, Rondebosch 7701, Cape Town, South Africa

The noun class system is emblematic of the Niger-Congo language family, whose hundreds of languages are spoken predominantly in Sub-Saharan Africa. Each language in this family has between 11 to 23 noun classes according to Meinhof’s classification [1] that is used in linguistics and illustrated in Table 1. Each noun is categorised in a noun class, and it is generally accepted, though not uncontested, that each noun class has a particular meaning [2]. For instance, animals go in noun class 9 (singular) or 10 (plural), liquid stuffs in class 5 or 6, and long thin objects are categorised in noun class 11. These noun classes govern agreement among tokens throughout the sentence, extensively affecting various parts of speech. Therefore, it is crucial to understand them to develop and improve natural language processing tasks, such as ontology verbalisation [3], and prospective further development of, among others, an African WordNet [4] that in turn can then be used to enhance NLP tasks.

How the semantics of the noun classes relate to Ontology and ontologies is unclear. At present, there are many more questions with illustrative hints than answers. From a theoretical perspective, under the assumption that the noun class system is not mere syntax, these questions intersect with ontology, cognitive science, and linguistics and are posed both at the word-level and the class system-level.

Word-level questions and challenges They include why a certain noun is in a particular noun class, how to categorise new words or loanwords that take sociolinguists into account to make sense of the categorisation [6], and formal and conceptual blending [7]. For instance, *inswelaboya* ‘criminal’ (noun class 9, isiZulu) is not categorised in noun class 1 where humans and their roles and professions typically reside, because a criminal “lacks the value of humanity”, and in, e.g., *umthethosisekelo* ‘constitution’ (noun class 3, isiZulu), the “source” noun that appears first in the blend determines the noun class of the blend, having combined *umthetho* ‘law’ (noun class 3) and *isisekelo* ‘base/foundation’ (noun class 7) [7].

Second, the noun class can be used to deduce the meaning of a noun. For instance, ‘journey’ is *ulendo* in Chichewa (noun class 14) and *uhambo* in isiZulu (noun class 11): based on the semantics of noun classes, in Chichewa the idea of a journey takes centre stage, since 14 is for so-called abstract nouns, whereas in isiZulu it is the route taken, since 11 is for long thin or

JOWO 2024

*Corresponding author.

✉ mkeet@cs.uct.ac.za (C. M. Keet)

ORCID 0000-0002-8281-0853 (C. M. Keet)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Table 1

Generalisation of the semantics of the kinds of entities typically found in that noun class (NC). Examples are taken from isiZulu (classes 1-11, 14, 15), Chichewa (12,13,16-18), Hunde (19), Runyankore (20,21), and Luganda (22,23). (Source: adapted from [5].)

NCs	Semantics (generalised)	Examples
1	People and kinship	<i>umfana</i> (nc1) 'boy'
2		<i>abafana</i> (nc2) 'boys'
3	Plants, nature, some parts of the body	<i>umuthi</i> (nc3) 'tree'
4		<i>imithi</i> (nc4) 'trees'
5	Fruits, liquids, parts of the body, loan words, paired things	<i>ijikijolo</i> 'raspberry'
6		<i>amajikijolo</i> 'raspberries'
7	Inanimate objects	<i>isihlalo</i> 'chair'
8		<i>izihlalo</i> 'chairs'
9	Loan words, tools, and animals	<i>indlovu</i> 'elephant'
10		<i>izindlovu</i> 'elephants'
11	Long thin stringy objects, languages, inanimate objects	<i>ucingo</i> 'wire'
(10)		<i>izingcingo</i> 'wires'
12	Diminutives	<i>kagalimoto</i> 'small car'
13		<i>timagalimoto</i> 'small cars'
14	Abstract concepts	<i>ubuhle</i> 'beauty'
15	Infinitive nouns	<i>ukucula</i> 'to sing'
16	Locative classes	<i>pamsika</i> 'round the market'
17		<i>kumsika</i> 'at the market'
18		<i>mumsika</i> 'in the market'
19	Diminutives	<i>hyùndù</i> 'a little bit of porridge'
20	Augmentative and pejorative	<i>ogusajja</i> 'big ugly man'
21		<i>agasajja</i> 'big ugly men'
22		<i>gubwa</i> 'mutt' (pejorative of dog)
23	Locative class	<i>eka</i> 'at home'

stringy objects. They are different senses that a simple bidirectional dictionary without noun class information will not be able to detect.

Noun class system questions The groupings used for noun classes, such as human, non-human, object vs mass vs collective, abstract, natural phenomena vs utensils, trees, and non-paired body parts, clearly allude to notions of foundational, core, and domain ontologies, which raises many questions. Among others: what holds of that intuition? Are these semantics the same across the Niger-Congo B (Bantu) family or only among subgroups, such as the Nguni group, or neither? Does the system combine linguistic and cognitive aspects, or linguistic and ontological, or all three, or can their contributive component-parts be identified? Might it be the case that none of the extant ontologies fit with the salient categories because there is some underlying Afro-centric foundational ontology distinct from the likes of BFO, DOLCE, UFO, and YAMATO that were developed on other continents?

The noun class system semantics has been investigated mainly from a linguistics perspective (e.g., [2, 8]). Ngcobo's structured "continuum" (attributed to Hendrikse and Poulos 1994) [2] omits noun class 16 (possibly a typo), confuses perdurants with abstract entities, stative entities, such as '(being) at home', with relations, and it is presented as being discrete for the noun

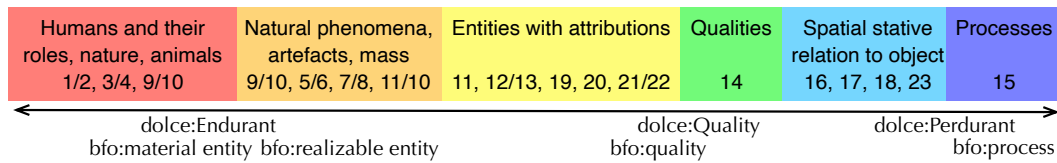


Figure 1: Revised continuum organisation of noun classes and with indicative entities from the DOLCE and BFO ontologies. The first four groups (fltr) are on a continuum, whereas the last two on the rhs have crisp boundaries.

classes in six groups but with only four “cognitive categor[ies]” and “word-like categor[ies]”. A revised preliminary foundational ontology-informed restructuring of Ngcobo’s proposal is shown in Fig. 1, which also indicates what contributes to the continuum (a.o., noun classes 9 and 10) and maps categories to all six groups of noun classes. It is still a generalisation with exceptions, however, and it is already known that there are slight differences between languages (e.g., compare Table 1 to the one in [3]).

Current work therefore concerns research into developing so-called application ontologies for a language of interest rather than one core ontology for all NCB languages, in line with the approaches of [9, 10].

References

- [1] C. Meinhof, *Grundzüge einer Vergleichenden Grammatik der Bantusprachen*, Reimer, Berlin, 1906.
- [2] M. N. Ngcobo, Zulu noun classes revisited: A spoken corpus-based approach, *South African Journal of African Languages* 1 (2010) 11–21.
- [3] C. M. Keet, L. Khumalo, Toward a knowledge-to-text controlled natural language of isiZulu, *Language Resources and Evaluation* 51 (2017) 131–157.
- [4] S. Bosch, M. Griesel, Strategies for building wordnets for under-resourced languages: The case of African languages, *Literator* 38 (2017) 12.
- [5] J. Byamugisha, C. M. Keet, B. DeRenzi, Pluralizing nouns across agglutinating Bantu languages, in: *Proc. of 27th Int Conf on Computational Linguistics (COLING’18)*, ACL, 2018, pp. 2633–2643. 20-26 Aug, 2018, Santa Fe, New Mexico, USA.
- [6] M. Ngcobo, Loan words classification in isiZulu: The need for a sociolinguistic approach, *Language Matters: Studies in the Languages of Africa* 44 (2013) 21–38.
- [7] T. M. Buthelezi, Exploring the role of conceptual blending in developing the extension of terminology in isizulu language, *Alternation* 15 (2008) 181–200.
- [8] P. Denny, C. A. Creider, *Noun classes and categorization*. *Typological Studies in Language* 7, Amsterdam: John Benjamins, 1986, pp. 217–241.
- [9] C. Chavula, C. M. Keet, An orchestration framework for linguistic task ontologies, in: *Proc. of 9th Metadata and Semantics Research Conference (MTSR’15)*, volume 544 of *CCIS*, Springer, 2015, pp. 3–14. 9-11 Sept, 2015, Manchester, UK.
- [10] C. Chiarcos, M. Sukhareva, OLiA – ontologies of linguistic annotation, *Semantic Web Journal* 6 (2015) 379–386.