

# On the Ontology of Part-Whole Relations in Zulu Language and Culture

C. Maria KEET<sup>a,1</sup> and Langa KHUMALO<sup>b</sup>

<sup>a</sup>University of Cape Town, South Africa

<sup>b</sup>University of KwaZulu-Natal, South Africa

**Abstract.** Parthood and attendant part-whole relations enjoy interest in ontology authoring for various subject domains, as well as in, e.g., NLP to understand text. The list of common part-whole relations is occasionally slightly modified for languages other than English. For isiZulu, it was shown that there are not always 1:1 mappings and, moreover, dictionaries list many more translations for parthood and part-whole relations. This complicates selecting the semantically appropriate ones for localising ontologies or aligning local ontologies to other ones. It also raises the question whether the ‘common’ part-whole relations are really that common. We aim to investigate the extant part-whole relations in isiZulu and determine their ontological status. We harvested a lexicon of 81 terms from dictionaries, which was reduced to 31 through several iterations of refinement, of which 13 were formalised and aligned to well-known part-whole relations. It showed that in some cases distinctions are made—and for which words exist—that have not been included before in part-whole relations, yet in other cases it is more coarse-grained; e.g., a parthood for portions of cloth, for objects properly contained in the mouth, and for regions with a part-region that has a fiat boundary and objects located in it.

**Keywords.** Mereology, Meronymy, IsiZulu

## 1. Introduction

Parthood, and, more generally, part-whole relations, are well-known to play a central role in ontology authoring across multiple subject domains and have been investigated both in philosophy and ontology engineering. Here, we refer to both mereological theories proper starting from ground mereology (e.g. [34]) and the language and cognition-inspired lists and taxonomies that took off since Winston et al’s paper [37] that is typically of greater interest for domain ontology developers than the properties of parthood. These common relations include, among others, *involvement* as a parthood between processes, *containment* as a parthood of regions occupied by objects, and *membership* of objects or the roles they play in a collective. They are just as relevant for localisation and internationalisation of ontologies; e.g., to localise SNOMED CT [31] for its use with electronic health record systems such as OpenMRS [27] that is relatively popular in Sub-Saharan Africa, and to describe more precisely (Southern African) architecture [11]. Such ontology-driven information systems in a local language can assist with, e.g., gen-

<sup>1</sup>Department of Computer Science, University of Cape Town, South Africa, mkeet@cs.uct.ac.za.

erating patient discharge notes in one's language to improve medical treatments in order to ameliorate language barriers in health care [15] and preserve indigenous knowledge.

This does bring afore the need for properly representing part-whole relations in the local languages, as medical terminologies, such as the FMA [30], SNOMED CT [31], and OpenGALEN [29] are replete with part-whole relations. Such common relations were assessed on their counterpart in isiZulu—the most-widely spoken language in South Africa by first language speakers—in the context of natural language generation [22], which turned out not to be a 1:1 translation and mapping exercise [20]. Some differences also seem to exist for Chinese [4], Turkish [38], and Spanish [6]. For isiZulu, the most interesting one ontologically was the meronymic *participation* relation, which distinguishes between a single object vs. a collective being part of an event [22]. That preliminary result already raised the question of whether the ontology for part-whole relations really would be different for isiZulu-speaking people or whether it is mostly just terminology after all. The English-isiZulu dictionary [7] list 18 entries under 'part' alone, but most terms do not have a definition, dictionaries are imprecise, and it is well-known that there is a difference between language and terms on the one hand and ontology on the other. Yet, if ontology is universal and one for all, including the common part-whole relations, then there should not be ontological differences but be merely one of terminology. However, based on this information, the expectation (or hypothesis) is that refinements will be encountered that also may be useful for ontology engineering in general. The main questions that guide this investigation are thus:

1. Which part-whole relations have been named in isiZulu, and to what extent are they not only lexically but also semantically distinct?
2. Can all those part-whole relations be mapped with equivalence relations to the common part-whole relations?
3. For those that cannot be mapped with equivalence relations: is the difference in meaning ontologically possibly interesting for ontology engineering?
4. Is there something different as gleaned from isiZulu part-whole relations that is useful in improving the theoretical appreciation of part-whole relations?

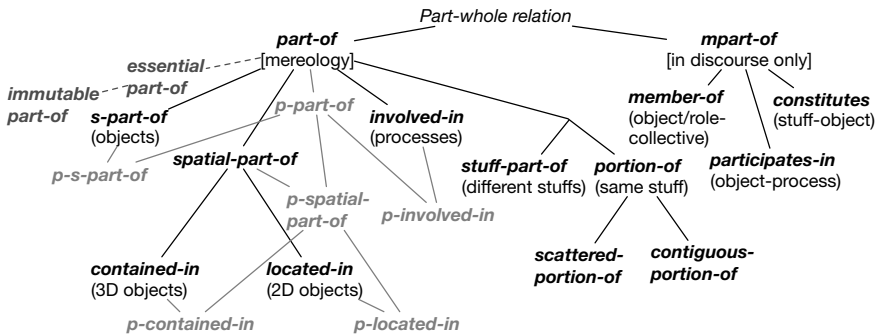
We aim to answer these questions in this paper. The approach is a combination of evidence gathering and theoretical analysis. We harvested common isiZulu terms for 'part' and similar terms from the dictionary, and clarified and analysed them in several iterations. A selection of 13 terms/relations were formalised and aligned with subsumption or equivalence to common part-whole relations. They were also checked against a subset of the isiZulu National Corpus. The main outcome is that there are both more precise part-whole relations and certain distinctions are not made—hence, equivalence and subsumption mappings—where the former can be of interest to refine extant ontologies, such as a parthood that is actually two chained parthood relations and very specific ones, such as portions for meat. There are further distinctions, such as parts with an identity vs without, which deserve further scrutiny. This is, to the best of our knowledge, the first systematic investigation into whether there may be other part-whole relations in languages or cultures other than English and the countries where it is spoken predominantly.

The remainder of this paper is structured as follows. We first consider related work in Section 2. The procedure of approach is briefly described in Section 3 with the principal results in Section 4. We discuss in Section 5 and conclude in Section 6.

## 2. Related work

There exists ample literature on the various mereological theories and their properties, such as whether antisymmetry is really needed or whether strong or weak supplementation is better [34]. We take those as a given, and instead focus on the ‘multitude’ aspect of part-whole relations, i.e., the multiple part-whole relations that have been proposed in the literature and used in ontologies, conceptual models, linguistics, and NLP (among many: [12,21,26,32,36,37]), and are being declared in both domain and foundational ontologies (see [19] for examples). This ‘multitude’ approach has resulted in a stable list of common part-whole relations, which has been structured in a hierarchy that is shown informally in Fig. 1. Its first main distinction is between parthood *sensu* mereology and part-whole relations in natural language utterances only (meronymy) [21]. With mereology, we refer to the primitive *parthood* relation that is antisymmetric, reflexive, and transitive [34], whereas meronymic relations are non-transitive or intransitive and where ‘part’ is used loosely, as in, e.g., “each soccer player is part of [i.e., member-of] a soccer team”. Its second main distinction rests on the notion that constraining a relation’s domain and range means it is a more precise representation of its intended meaning, where applicable [21,28,36]. For instance, *involvement* is a parthood relation that is constrained to relating processes to its sub-processes (more precisely [21]: DOLCE’s *perdurants*).

The respective proper parthood versions are shown in grey in Fig. 1. There are various possible extensions and variations. For instance, essential and immutable parthood that require some modality and mereotopology in the *spatial parthood* branch so that one can avail of tangential and non-tangential location.



**Figure 1.** Summarised sketch of the part-whole taxonomy and informal descriptions of their domain and range, extended from [21]. See text for explanation.

One may prefer different surface readings/labels, such as made of instead of constituted of or the respective names in another language, which is not of interest ontologically. What is relevant to note, is that these relations have been proposed in research done by people from multiple countries and cultures who speak multiple natural languages, so one might be tempted to generalise and claim there is some universality to them.

Linguistically motivated ontological analyses for languages other than English on their use of part-whole relations to confirm universality of such common part-whole relations are sparse. Vieu and Aurnague [36] focus on French, yet their scope on function and parthood with “entities-as-a-lexical-type” for the Component Integral Whole parthood

still matches *structural parthood* (s-part-of in Fig. 1) [36]. There are some works on mereology/meronymy for Arabic, Chinese, and Turkish that focus on relation extraction from text documents, but this is also within the aforementioned typical set of part-whole relations or a subset thereof [1,4,38], rather than an ontological analysis of the relations. Cao et al. [4] ‘refined’ constitution with an Element-Object relation—e.g., calcium as part of milk—where the element is an atomic element “for convenient verification” without discussing whether it is semantically or linguistically distinguishable from the others. Yıldız et al [38] stated explicitly that a subset only was relevant, notably excluding the spatial part-whole relations, yet in the details, they did distinguish between constituted of and made of, with the former having a ‘built’ flavour to it and the latter intended as a generic constitution. Finally, we also started from the typical set of part-whole relations, observing some differences for isiZulu [22] (discussed briefly in [20]): there are refinements in some cases and the lack thereof for others, such as distinguishing participation for objects vs. collectives.

Parts and part-whole relations in natural language have been investigated for less-widely spoken languages, including African languages [5]. Its chapter on Haya (spoken in Tanzania) focuses on possessor deletion and promotion in the sentence, rather than any part-whole relation [14] and the linguistic realisation of describing body parts in Ewe (Ghana) refers only implicitly to part-whole relations, such as ‘the cover of the book’ [2] rather than constructions such as ‘the cover is part of the book’. To the best of our knowledge, there is no inventarisation of part-whole relations in any of the Sub-Saharan African languages, let alone an ontological analysis thereof other than the informal analysis of the common relations (as in Fig. 1) in [22].

The paucity of ontological analyses on interaction between language and ontology for languages other than English does not mean that there would be no scope for it. For instance, there are at least seven entries in WordReference for ‘part’ in Spanish<sup>2</sup> and Climent proposed a basic categorisation for anything partitive in Spanish based on whether it is bounded or not and whether it is an individual or not, thus including part-whole relations involving individuals, groups, masses, or aggregates [6]. German may be interesting as well, as even just the translation of ‘part’ from English to German<sup>3</sup> yielded 918 noun entries, 58 adjectives, and 72 verbs. These huge numbers are largely due to the fact that compound nouns, descriptions, or concepts in English are single nouns in German; e.g., a part for construction is *Bauteil* (from *Bau-* ‘construction’ *-teil* ‘part’). The ‘part of’ in a sentence is less elaborate, although grammatically more involved than in English due to its system of gender and case. Thus, while we focus on isiZulu in the remainder of the paper, the same methodological steps could be used for other languages.

### 3. Procedure

In order to reduce the possibility of shoe-horning isiZulu terms and conceptualisations into those reported in the literature, the procedure of analysis was specified upfront. Also, its detail here may facilitate reuse (replace ‘isiZulu’ with one’s language of choice).

<sup>2</sup><http://www.wordreference.com/es/translation.asp?tranword=part>; last accessed: 22-3-2018.

<sup>3</sup><https://dict.leo.org/englisch-deutsch/part>; last accessed: 22-3-2018.

1. Create an isiZulu corpus of verbal lexicon from an isiZulu dictionary. This includes: looking up the common terms for ‘part’ and similar terms in the English-isiZulu dictionary, in both directions. For each entry:
  - (a) Write down the term and a description of the meaning of the term.
  - (b) Determine whether it is a part-whole relation, at least broadly construed; if not, add them to the ‘discarded’ list;
  - (c) Check the English entries under the identified candidates (i.e., *ingxenye* and similar terms) and revise the preliminary list, if applicable.
2. Categorise the potential part-whole relations obtained in Step 1b by similar informal meanings.
3. Refine descriptions where necessary based on that categorisation and remove any term that does not make the cut after all.
4. Create a formal definition (or at least a logic-based characterisation) for each part-whole relation and relate each one to a part-whole relation described in ontology literature, where possible. For each of the relations where this fails, determine reason(s) and identify underlying pattern, if any.
5. Query the isiZulu National Corpus (INC) on the term’s total use and a section for detailed annotation on number, relevance, and agreement with theoretical analysis (concordance search only, as only technologically feasible option).

The materials used are principally the *Scholar’s Zulu Dictionary* [7], assisted occasionally in the first round by [8] to verify older/outmoded meanings and by [isizulu.net](http://isizulu.net) to cross-check translations in case of doubt. The step-wise reduction and term analyses were documented in a spreadsheet in successive sheets to foster traceability of motivations and decisions. The 31 million-tokens INC [23] is stored in Wordsmith Tools and consists mainly of novels and news items; the section for detailed analysis consists of 36 novels by female authors that is also used for another (ongoing) experiment. The analysis was carried out by the authors, with one (LK) being a specialist in isiZulu linguistics with some knowledge of ontologies, and the other (CMK) vv.

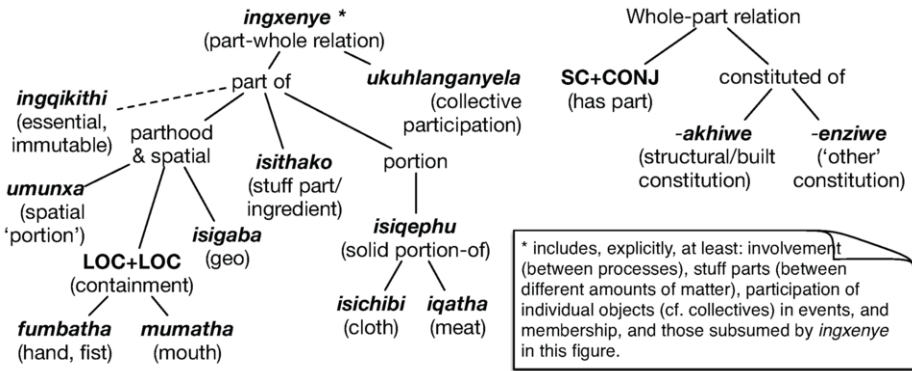
#### 4. Parts and Wholes in Zulu Language, Culture, and Conceptualisation

We first describe the results from harvesting a lexicon of terms that, according to the dictionary, have something to do with part or whole, and the elimination of those that do not after all (Section 4.1). Thirteen terms were selected for formalisation (Section 4.2) and assessment against the INC (Section 4.3). The data and analyses are available from <http://www.meteck.org/files/pwZUonto.xlsx>.

##### 4.1. Harvesting and reduction of number of terms

First, the dictionary entries that we looked at were, in summary:

- English→isiZulu: ‘part’ has 18 entries in isiZulu; ‘portion’ has 11 isiZulu terms; ‘quantity’ has 8 isiZulu terms; ‘piece’ lists 19 isiZulu terms, ‘pinch’ lists 6; ‘contain’ and ‘component’ each lists 4 isiZulu terms.
- isiZulu→English: the principal part-whole relation *ingxenye*, as well as, among others, *umnunzo*, *isigamu*, and others that were harvested in the previous step.



**Figure 2.** Tentative partial taxonomy of linguistically-motivated part-whole relations; the isiZulu terms are written in bold italics and a keyword is added for illustrative purpose; LOC+LOC and SC+CONJ: there is no single term for them as the surface realisation is constructed depending on the noun class of the noun that participates in the axiom (for SC) or the noun's orthography and phonological conditioning (CONJ and LOC).

The raw list resulting from this exercise consists of 81 unique isiZulu terms. They were, where immediately obvious, annotated with a description and a tentative status, of which 41 were put on the 'discarded' list. The discarded ones can be divided roughly into four categories: 1) terms such as *-aba* 'share', *-ahlukanisa* 'separate', and *-vithiza* 'break to pieces', which is about creating parts rather than relating parts to wholes, 2) words that were artefacts of English compound nouns or idioms, such as the entry refinement 'piece of paper' (*ipheshana*) listed under 'piece' and 'I for my part' (*mina ngokwami*) listed under 'part', which are linguistically related in English but not ontologically, 3) many quantities, portions, and pieces were discarded as they referred to standalone or size of quantities rather than subquantities of something else (e.g., *ubungako* is a quantity in the sense of hugeness), and 4) simply wrong, such as *isibhamu* 'firearm' in the 'piece' entry, or only distantly related, such as *ifa* 'inheritance' listed under 'portion' (it assumes several people each will receive a portion of what the deceased left behind).

The remaining 40 were further annotated with an indicative category of the kind of part, such as relating stuffs or regions, whether it still concerns with how the part comes about, whether there is a temporal aspect to it, and their POS category (noun or verb) and noun class if it is a noun, and further descriptions on their more precise meaning. This resulted in the list being reduced to 28 entries and an additional three that were missed in the original assessment (*umunxa*, *akhiwe*, and *enziwe* from [22]). The discarded ones were discarded for a myriad of reasons. For instance, *-hlakazekile* refers to the state of being scattered as a result from breaking or dispersing a whole and refers to a relation among the parts (e.g., the pearls from the broken necklace), *-xhumelela* 'piece together' is about making a whole, and *indima* '(take) part of responsibility' (in, say, raising a child) refers to expectations associated to the role played by a person in the activity.

The final reduction to the part-whole relations that will be formalised and assessed against the INC was guided by two considerations: those that are deemed important and expected to return many instances in the INC, such as *ingxenye*, and those that, at first impression at least, seem overly specific, such as *iqatha* that applies to portions of meat only. Structuring the informal characterisation of the final selection of the terms leads to a small taxonomy for further analysis, which is depicted in Fig. 2.

## 4.2. Formal characterisation

As we want to know commonalities and differences in part-whole relations, we rely as much as possible on existing formalisations and theories. Therefore, we first describe some preliminaries to keep the paper sufficiently self-contained before moving on to the formal characterisation of the putative part-whole relations depicted in Fig. 2.

### 4.2.1. Preliminaries

Several putative relations require constraints on their domain or range (relata), such as Collective and Mouth. In order to be precise in the meaning of the relata, a foundational ontology has to be chosen. Foundational ontologies have been assessed on their parthood theories [10] and on the part-whole relations and relata [19]. These assessments showed that neither is a perfect fit. As the taxonomy of part-whole relations in [21] uses DOLCE categories [24], we use it as well for possible compatibility and comparison.

The putative relations depicted in Fig. 2 already indicate that a formalisation likely will require second-order logic, because the stuffs and portions need it in order to state that the stuffs involved are either different kinds of stuff (stuff part) or the same kind of stuff (portion) [18]<sup>4</sup>. They also suggest that refinements for the spatial aspects may be needed into a full-fledged mereotopological theory (notably *LOC+LOC/containment*), which also requires second order logic [33]. While this does not look promising computationally, one could later truncate the formalisation to more widely implemented languages like OWL 2. The aim here is to capture the meaning as precisely as possible.

We present relevant definitions and axioms investigated elsewhere that our formalisation for part-whole relations in isiZulu require directly. For mereological *parthood*, denoted with  $p$  in the axioms, we use Ground Mereology [34] as conservative commitment, which is a primitive relation that is reflexive, antisymmetric, and transitive, and *proper parthood* ( $pp$ ) is then defined in terms of parthood (axioms omitted). Motivated by conceptual modelling and domain ontologies, parthood has two refinements regarding the spatial aspect, being *containment* for 3-dimensional objects occupying some region (Eq. 1) and *location* for 2-dimensional geographical objects (Eq. 2), even though ontologically this is strictly not necessary [21]. The region refers to DOLCE's region ( $R$ ) and the particular objects located at those regions to DOLCE's enduring ( $ED$ ), which are related through *has3D* and *has2D*, respectively, which are compact shorthand relations standing for the same notion as DOLCE's approach of qualities and qualia.

$$\forall x, y (ci(x, y) \leftrightarrow p(x, y) \wedge R(x) \wedge R(y) \wedge \exists z, w (has3D(z, x) \wedge has3D(w, y) \wedge ED(z) \wedge ED(w))) \quad (1)$$

$$\forall x, y (li(x, y) \leftrightarrow p(x, y) \wedge R(x) \wedge R(y) \wedge \exists z, w (has2D(z, x) \wedge has2D(w, y) \wedge ED(z) \wedge ED(w))) \quad (2)$$

Observe that this does not imply that those objects are related also by structural parthood. For their proper contained/located in version, one simply can substitute *parthood* for *proper parthood* in the definitions above and name the relation  $pci$  and  $pli$ , respectively.

For portions and stuff parts—in language typically denoted with mass nouns—we avail of the Stuff Ontology [17,18]. A *stuff part* ( $sp$ ) is a *proper part* ( $pp$ ) between different kinds of Stuff (Eq. 3) and its inverse, *hassp*, is defined in the usual way; e.g., the

<sup>4</sup>another candidate may be the portions of [9], but it resorts to many-sorted logics, hence not much of an improvement from the viewpoint of computational use.

alcohol that is part of wine. This is contrasted with *portion* (po) where the stuffs are of the same kind (Eq. 4), which may be contiguous, cpo (Eq. 5) with  $t$  time points, like the upper half of the cake, or scattered, spo (Eq. 6), such as the wine in the wine glass taken from the wine in the bottle.

$$\forall x, y \exists S, S' (sp(x, y) \leftrightarrow pp(x, y) \wedge S(x) \wedge S'(y) \wedge \text{Stuff}(S) \wedge \text{Stuff}(S') \wedge S \neq S') \quad (3)$$

$$\forall x, y \exists^1 S (po(x, y) \leftrightarrow pp(x, y) \wedge S(x) \wedge S(y) \wedge \text{Stuff}(S)) \quad (4)$$

$$\forall x, y \exists t (cpo(x, y, t) \leftrightarrow po(x, y, t) \wedge pci(x, y, t)) \quad (5)$$

$$\forall x, y \exists t, t' (spo(x, y, t) \leftrightarrow cpo(x, y, t') \wedge \neg cpo(x, y, t) \wedge t' < t) \quad (6)$$

Stuff has as one of its subtypes *MixedStuff*, which is a stuff that has at least two stuff parts that are different kinds of stuff; e.g., cake that has butter and flour as ingredients. In shorthand notation with “ $\exists^{\geq 2}y$ ” denoting that the  $y$ s are distinct (which follows directly from  $sp$  in Eq. 3), it can be defined as  $\forall x (\text{MixedStuff}(x) \leftrightarrow \text{Stuff}(x) \wedge \exists^{\geq 2}(\text{hassp}(x, y)))$ . Entities like a solid and/or heterogeneous mixture (e.g., wood) then simply can be defined as a subclass where the state is solid or made up of different pure or mixed stuffs (see [17] for further details).

Finally, to distinguish the non-transitive part-whole relation of the meronymic but not mereological part-whole relation, we use a placeholder name/relation for structuring purposes, i.e., that is not intended to be used, called  $mp$ , with for *participation* and *constitution* the following common specifications [21], where PD is perdurant, POB physical object, and M amount of matter from DOLCE.

$$\forall x, y (pi(x, y) \leftrightarrow mp(x, y) \wedge ED(x) \wedge PD(y)) \quad (7)$$

$$\forall x, y (co(x, y) \leftrightarrow mp(y, x) \wedge POB(y) \wedge M(x)) \quad (8)$$

#### 4.2.2. Formalisation

We proceed with the putative part-whole relations down in the left-hand hierarchy first.

*Ingxeny* (n.) is the generic ‘catch all’ part. This includes not only mereological *parthood* [34] and several more specific ones identified in [21], being *involvement* between processes and *stuff part* and those subsumed by it in the figure, but also *participation* of individual objects (vs. collectives) in events, and *membership* of objects (or the roles they play) in a collective, as has been analysed before [20,22]. One thus cannot be sure that the relation’s transitivity will generate only the intended deductions in a particular ontology because of different possible categories of domain and range that inadvertently can be mixed up. Using the generic, catch-all part-whole relation *part-whole* (pw) as primitive for this, then, in first-order logic notation, we have  $\forall x, y (\text{ingxeny}(x, y) \leftrightarrow \text{pw}(x, y))$ . For instance,

(1) *inhliziy*o *iyingxeny* *yomuntu*  
in-hliziy o i-y-ingxeny yo-muntu  
9.-heart 9.SC-is-part 7.PC-human  
‘a heart is part of a human’

(2) *uMnumzana* *uyingxeny* *yedili*  
uM-numzana u-y-ingxeny ye-dili  
1.-Sir 1.SC-is-part 7.PC-dinner party  
‘the Sir participates in the dinner party’

*Ukuhlanganyela* (v.) denotes participation of specifically a collective in an event where the members of the collective act in unison, such as the electorate participating in an election [22] or an operating team participating in an operation, and would, in natural language text, be used inflected (e.g., *Wonke umphakathi uhlanganyele okheth-*



*weni olulodwa*, with *umphakathi* ‘electorate’ and *ukhetho* ‘election’). Ontologically, it requires one to constrain the domain of *participation* with the notion of Collective. Unlike SUMO and GIST, DOLCE does not have a ‘collective’ category, but it could be added as  $\forall x(\text{Collective}(x) \rightarrow \text{SOB}(x))$ , i.e., as a subclass of social object. This additional constraint makes *ukuhlanganyela* thus more specific than the usual *participation* relation; hence, we obtain  $\forall x, y(\text{ukuhlanganyela}(x, y) \leftrightarrow \text{pi}(x, y) \wedge \text{Collective}(x) \wedge \text{PD}(y))$  cf. the domain ED (endurant) for pi (Eq. 7).

*Isiqephu* (n.) is used for the regular notion of portion (Eq. 4), but the kind of stuff that the portion is made of is solid or ‘solid-like’ stuff only; e.g., that each slice (*ucezu*) of bread (*isinkwa*) is a portion of some bread (*Zonke izicezu zesinkwa ziyisiqephu sesinkwa esisodwa*). The notion of ‘solid-like’ is ambiguous. For instance, it includes blood—as in: a sample of blood as portion of the blood (of the patient)—which is a viscous liquid in its natural state but becomes solid once in contact with the air due to coagulation, and, conversely, the lick of the ice cream as portion of the ice cream, which arguably may have melted into a liquid state when licked. We have not been able to clarify the precise cut-off point ontologically by pure theoretical analysis and shall return to this in the next section where it is queried against the corpus. Either way, it may or may not be a scattered portion and the stuff may or may not be a mixture. The minimum that can be said is that it denotes a sub-property of *portion*. Because the ‘solid-like’ is yet to be determined, we currently do not include it in the axiom:  $\forall x, y(\text{isiqephu}(x, y) \rightarrow \text{po}(x, y) \wedge \text{hasState}(x, z) \wedge \text{hasState}(y, z) \wedge \text{Solid}(z))$ .

*Iqatha* (n.) is used for a scattered portion for solids that are portions of meat only. This thus amounts to a straight-forward refinement of the *spo* relation of [18] (Eq. 6) with a more precise domain and range, where  $\forall x(\text{Meat}(x) \rightarrow \text{SolidHeterogeneousMixture}(x))$  from the Stuff Ontology [17] so that we obtain  $\forall x, y(\text{iqatha}(x, y) \leftrightarrow \text{spo}(x, y) \wedge \text{Meat}(x) \wedge \text{Meat}(y))$ . Note that, as meat is a solid stuff (i.e., it has a *hasState* that is *Solid*) and can be scattered only, it follows that  $\forall x, y(\text{iqatha}(x, y) \rightarrow \text{isiqephu}(x, y))$ .

*Isichibi* (n.) is a straightforward refinement of scattered portion as well, alike the *iqatha* relation, but then restricted to Cloth, which is also a *SolidHeterogeneousMixture*, hence,  $\forall x, y(\text{isichibi}(x, y) \leftrightarrow \text{spo}(x, y) \wedge \text{Cloth}(x) \wedge \text{Cloth}(y))$  and it also follows that  $\forall x, y(\text{isichibi}(x, y) \rightarrow \text{isiqephu}(x, y))$ .

*Isithako* (n.) refers to subquantity in the sense of ingredient that is an input so it makes another stuff, such as food, medicine, and paint, which implies that the whole stuff is a mixture and thus *stuff part* applies. It suggests this applies to human-made mixtures only, or at least agent-created, due to the *making of* the mixture, but such potential additional properties of the whole deserves further scrutiny regarding its usage. So, conservatively, we have at least  $\forall x, y(\text{isithako}(x, y) \rightarrow \text{sp}(x, y) \wedge \text{MixedStuff}(y))$ .

*Isigaba* (n.) is used for provinces/districts etc., i.e., geographical entities, and thus is ontologically equivalent to *location* in the broadest sense in that it does not explicitly refer to something being tangentially located in another region or not. Thus,  $\forall x, y(\text{isigaba}(x, y) \leftrightarrow \text{li}(x, y))$ . Observe that this is different from individual named entities, for which locative affixes may be used (as in, e.g., *iGlenwood iseThekwini* ‘[the

particular quarter] Glenwood is located in *iTheku* (Durban)’).

*LOC+LOC* ‘containment’. The *containment* relation does exist in isiZulu but it does not have an immutable word for it. Instead, the entity that plays the whole is affixed with locatives to indicate it is the one containing the part-entity [22]. For instance, for a bolus of food that is contained in the stomach it changes *isisu* ‘stomach’ to *-sesiswini* ‘contained in the stomach’, for some object contained in the computer, it changes *ikhompyutha* ‘computer’ to *-sekhompyutheni* ‘contained in the computer’, and ‘contained in the envelope’ changes *imvilophu* into *-semvilophini*. The locatives to indicate a (possibly temporary) container role of the whole in these examples are *-e- ... -wini*, *-e- ... -eni*, and *-e- ... -ini*, respectively, as a result of phonological conditioning of the locatives. This does not fit with ontologies where things have to have either an immutable name or an identifier with a fixed label. In [20], this was addressed by labelling the relation an arbitrary sequence of letters (*in casu*, ffff). Using this in the formalisation, then  $\forall x, y(\text{ffff}(x, y) \leftrightarrow \text{ci}(x, y))$ , for ontologically they denote the same thing.

*Mumatha* (v.) and *fumbatha* (v.) are more constrained versions of *containment*, for it is a proper containment and it holds for entities contained in the mouth, resp. hand or fist, only. Mouth, Hand and Fist are not the sort of entity for a foundational ontology, but they can be aligned to DOLCE’s non-agentive physical object, i.e., as  $\forall x(\text{Mouth}(x) \rightarrow \text{NAPO}(x))$ . From these two characteristics, the formalisation then follows trivially as  $\forall x, y(\text{mumatha}(x, y) \leftrightarrow \text{pci}(x, y) \wedge \text{Mouth}(y))$  and  $\forall x, y(\text{fumbatha}(x, y) \leftrightarrow \text{pci}(x, y) \wedge (\text{Hand}(y) \vee \text{Fist}(y)))$ .

*Umunxa* (n.) refers to a contiguous ‘portion’ of some meaningful area, like the portion of the kitchen where the kitchen utensils are and the area where the fireplace is in the hut as in, e.g., *Onke amaziko angumunxa wexhiba* with *iziko* ‘fireplace’ and *ixhiba* ‘hut’. This goes counter to the ontological notion of portion that are for stuffs only. Instead, it is about the area/region and the object(s) located at it, hence, spatial parts or containment, as that the region occupied by the kitchen utensils is contained in the region occupied by the kitchen. *Umunxa* does not extend to arbitrary containment of regions with objects nor of regions alone, i.e., it does not apply to, e.g., the bottom 1/3 of the whiskey glass where the amount of whiskey is nor to the north-east quarter of a circle. Having examined multiple examples, the main distinction that can be discerned is that there is a region with a fiat boundary that contains some object(s) with the regions they occupy that is smaller than the whole part-region. Thus, those objects are not located at that *entire* region (say,  $r_1$ ), as is assumed with *containment*, but are in a region,  $r_2$ , that is a proper part of that region  $r_1$ ; hence,  $r_2$  is a proper part of  $r_1$  and  $r_1$  is a proper part of the whole region ( $r_3$ ) like the kitchen or hut. Given the commitment to ground mereology and the low relevance of  $r_1$ ’s boundary cf. the part-within-a-part chain of relations, we formalise only the latter core feature. To be explicit with the constraints on the relata:  $\forall x, y(\text{umunxa}(x, y) \rightarrow \text{pp}(x, z) \wedge \text{R}(x) \wedge \text{R}(z) \wedge \exists w(\text{has3D}(x, w) \wedge \text{ED}(w)) \wedge \text{pp}(z, y) \wedge \text{R}(y) \wedge \exists v(\text{has3D}(v, y) \wedge \text{ED}(v)))$ . Its underspecified, but conceptually easier, shorthand notation is  $\forall x, y, z(\text{umunxa}(x, y) \rightarrow \text{pp}(x, z) \wedge \text{pp}(z, y))$ .

*Inqikithi* (n.) is used for both essential and immutable part, and is thus orthogonal to the other parthood relations. We omit its extensive formalisation here (be it if it were

to be cast in necessities, as attempted in [13], or in a temporal logic, as in [3]).

Lastly, the meronymic *constitution* relation is used in isiZulu only in the whole-part reading direction, like vases that are constituted of clay, for which there are two options: *-akhiwe* (v.) and *-enziwe* (v.). The former is used for built or constructed sort of things—e.g., houses are made of stone (*Zonke izindlu zakhiwe ngetshe*)—and the latter is used for other things, such as pills that are made of starch [22]. It is beyond the scope of part-whole relations to investigate the precise nature of the relata when it does not readily fit an existing one, which they do not. The ontological distinction between the two is about the way how the entity that plays the whole has come into existence, but it remains to be seen what exactly that difference is (e.g., human-made applies to both). Formally, then, the generic *enziwe* can be weakly characterised with  $\forall x, y(\text{enziwe}(x, y) \rightarrow \text{co}(x, y))$ , rather than an equivalence, for it excludes the *-akhiwe* cases. For *-akhiwe* to be a proper relation ontologically, the whole has to be constrained to some elusive ‘built thing’. Whilst unsatisfactory, the clearest way to indicate there is a difference between the two is to constrain it with a to-be-defined Built artefact to indicate at least the intuition of it, hence  $\forall x, y(\text{akhiwe}(x, y) \rightarrow \text{co}(x, y) \wedge \text{BuiltArtefact}(x))$ . We will discuss this further below.

#### 4.3. Evaluation against the INC

The whole corpus yielded the following number of hits of words (i.e., e.g., *ingxenye* ‘part’ was queried but not, say, *ziyixxenye* ‘are part’ for nouns in noun class 10): *ingxenye*: 132; *ukuhlanganyela*: 95; *isiqephu*: 269; *iqatha*: 194; *isichibi*: 3; *isithako*: 27; *isigaba*: 3002; *-mumatha*: 0; *-fumbatha*: 0; *umunxa*: 105; *ingqikithi*: 239; *-akhiwe*: 153; *-enziwe*: 267. Querying the selected section of the INC did, for the most part, not yield the results that were hoped for, whose reasons will be discussed in Section 5. The core results are available in the online raw data and the aggregate statistics are shown in Table 1. The main observations are: 1) *iqatha* is indeed restricted to meat; 2) *ingqikithi* is always used as essential part-whole relation and shows a plethora of examples that will be useful for future research; 3) *umunxa*’s and *isigaba*’s notion of region has been used in a much broader sense than the examples of physical regions given in the previous section, but may still fall within their respective formal definition; 4) *ingxenye* is indeed used for part in the broadest sense, and 5) alternate uses of the meaning, such as *isiqephu* also meaning ‘section’ and *akhiwe*, appeared mainly in alternate sentence constructions, such as *akhiwe yimisindo yenkulumo* ‘built by speech sounds’, due to the broader meaning of the verb (*-akha*). Discarding the false positives, then it can be concluded that no relevant concordance result violated the definitions of Section 4.2.

## 5. Discussion

The characterisation of the selected part-whole relations, for as much as currently could be understood, already showed that there are differences between the ‘common’ part-whole relations and those encountered in isiZulu and held up on ontological analysis; compare Fig. 1 with Fig. 2. It is unsurprising that there are several equivalences, notably for the generic *part-whole* relation (*ingxenye*) and for *location* and *containment*. Some of the relations might be too specific even for domain ontologies, such as *mumatha*

**Table 1.** Concordance results from the random section of the INC. *n*: total hits, relevant: term used in sense of part; match: used as specified in the formalisation.

relation	<i>n</i>	relevant	match	relation	<i>n</i>	relevant	match
ingxenye	79	62	62	isithako	0	0	0
ingqikithi	16	16	16	isigaba	84	43	13
ukuhlanganyela	2	2	2	isichibi	0	0	0
umunxa	8	0	0	iqatha	2	2	2
fumbatha	0	0	0	akhiwe	10	0	0
mumatha	0	0	0	enziwe	2	0	0
isiqephu	5	0	0				

(mouth), *fumbatha* (hand), and *isichibi* (cloth). Yet, it does bear a resemblance to those terms for specific parts in German, like the aforementioned *Bauteil* (among many), and perhaps also with YAMATO's modelling approach of having very many sub-relations for 'has part', such as has-arm, has-brain, and has-mouth [25]. It may be worth examining how to manage these sort of very specific relations in ontologies in a systematic way. The distinction between objects vs collectives participating in an event (*ukuhlanganyela*) is fundamental, as is *umunxa* that is a parthood relation characterised by composition of two parthood relations. The *akhiwe/enziwe* issue between 'built' vs. 'other' objects that are made of some stuff may exist elsewhere as well, as suggested for Turkish [38], and therefore merits further investigation.

Regarding the terms not analysed here, it should be possible to refine *membership* cf. lumping it together with *ingxenye*, but the scope of the current candidate, *ilungu*, is not fully clear: it refers to 'member of council' but it may also hold for parliament, community, family, organization, and similar, or: being part of an institution or group, which, however, is still more restricted than *membership* that also applies to non-human groups, such as herds. Another avenue draws in identity; e.g., *isihlephu*, where the scattered part has an identity of its own, such as the ear of a cup that has broken off (but *isihlephu* does not apply to a chip of the cup) and *-yimvithimvithi* where the parts/pieces are such that the whole is no longer recognisable, such as the pieces of the glass that has shattered. These part-whole relation issues are the subject of ongoing and future work.

Our intention with a corpus-based evaluation of the theoretical analysis turned out to be ahead of the current technologies. The main limitation of using concordance is simple string matching, but there are multiple permutations of the words due to the agglutinating nature of isiZulu and the deep prepositions. For instance, 'is part' with a noun from noun class 1 (that plays the part) results in *uyingxenye* in the text, but with a noun of noun class 10 it is *ziyingxenye*; there are 17 noun classes. The 'of' of 'part of' is realised as a phonologically conditioned possessive concord that is added to the noun of the object that plays the whole; e.g., *-ingxenye* 'of [a/the/at least one/some] human' is *ya + umuntu = yomuntu* and 'of orchestra' is *ya + i-okhestra = ye-okhestra*. Each noun class has its own possessive concord and three phonological conditioning rules. This is more complex for the verbs due to agglutination and inflection; e.g., *mumatha* would appear as, e.g., *uswidi umumethwe emlomeni* 'the sweet is contained in the mouth'. Including all permutations for all relations is not possible with the current limited technologies for isiZulu NLP, yet searching for only the noun or infinitive returns too many false positives, such as idioms, unrelated compound nouns, and unrelated sentence constructions, just like it would in English (e.g., 'I for my part', 'private parts', and 'you must participate!', respectively).

With a POS tagger, one would be able to detect noun(phrase)-verb-noun(phrase) patterns, but there is no isiZulu POS tagger that works and can be integrated with the Wordsmith tools that the INC is locked into. Thus, a corpus-driven approach to elucidate more about part-whole relations in isiZulu, in analogy with results obtained in NLP elsewhere (e.g., [16]), will require more resource development first.

There remains a question as to why the part-whole taxonomies are different. IsiZulu belongs to a different linguistic classification (Nguni) from English (Germanic language family), which may presuppose different cultural-linguistic groups as well. For instance, in a study of culture and personality [35], terminology and word clusters used were shown to be linked to culture, in particular regarding social and relational aspects that were shown to be more pronounced in Africa than in Western conceptualisations. A conclusive answer is yet to be found.

## 6. Conclusions

Novel insights were obtained on part-whole relations in isiZulu, by having used a bottom-up lexicon-based approach, subsequent refinements, and a formal characterisation of their respective meaning. Of the 81 terms harvested from dictionaries, 13 were formalised and aligned to well-known part-whole relations. The results demonstrate both that distinctions are made that have been neglected in part-whole relation use, yet also exhibits some underspecification; e.g., a parthood for portions of specific solid materials, specific containers, such as the mouth, and a spatial parthood that entails relation composition. This is, to the best of our knowledge, the first systematic ontological assessment on part-whole relations outside the Western hemisphere. Further research on the remaining 18 part-whole relations is under way.

## References

- [1] M. G. Al Zamil and Q. Al-Radaideh. Automatic extraction of ontological relations from arabic text. *Journal of King Saud University - Computer and Information Sciences*, 26(4):462 – 472, 2014.
- [2] F. Ameke. Body parts in Ewe grammar. In *The Grammar of Inalienability: A Typological Perspective on Body Part Terms and the Part-Whole Relation*, pages 783–840. Mouton de Gruyter, 1996.
- [3] A. Artale, N. Guarino, and C. M. Keet. Formalising temporal constraints on part-whole relations. In G. Brewka and J. Lang, editors, *Proc. of KR'08*, pages 673–683. AAAI Press, 2008.
- [4] X. Cao, C. Cao, S. Wang, and H. Lu. Extracting part-whole relations from unstructured Chinese corpus. In *Proc. of FSKD'08*. IEEE Xplore, 2008.
- [5] H. Chappell and W. McGregor, editors. *The Grammar of Inalienability: A Typological Perspective on Body Part Terms and the Part-Whole Relation*. Mouton de Gruyter, 1996.
- [6] S. Climent. Individuation by partitive constructions in Spanish. In P. Bouillon and F. Busa, editors, *The language of word meaning*, pages 192–214. Cambridge University Press, 2001.
- [7] G. R. Dent and C. L. S. Nyembezi. *Scholar's Zulu Dictionary*. Shuter & Shooter Publishers, 4 edition, 2009.
- [8] C. M. Doke, D. M. Malcolm, J. M. A. Sikakana, and B. W. Vilakazi. *English-Zulu Zulu-English Dictionary*. Witwatersrand University Press, 1958. 918p.
- [9] M. Donnelly and T. Bittner. Summation relations and portions of stuff. *Philosophical Studies*, 143:167–185, 2009.
- [10] M. Fernández-López, A. Gómez-Pérez, and M. C. Suárez-Figueroa. Selecting and customizing a mereology ontology for its reuse in a pharmaceutical product ontology. In *Proc. of FOIS'08*, pages 181–194. IOS Press, 2008.

- [11] F. Frescura and J. Myeza. *Illustrated glossary of Southern African Architectural Terms*. Bilingual Glossary Series. UKZN Press, 2016.
- [12] G. Guizzardi. *Ontological Foundations for Structural Conceptual Models*. Phd thesis, University of Twente, The Netherlands. Telematica Instituut Fundamental Research Series No. 15, 2005.
- [13] G. Guizzardi. Modal aspects of object types and part-whole relations and the de re/de dicto distinction. In *proc. of CAISE'07*, volume 4495 of *LNCS*. Springer-Verlag, 2007. Trondheim, 2007.
- [14] L. M. Hayman. The syntax of body parts in Haya. In H. Chappell and W. McGregor, editors, *The Grammar of Inalienability: A Typological Perspective on Body Part Terms and the Part-Whole Relation*, pages 865–890. Mouton de Gruyter, 1996.
- [15] N. Hussey. The language barrier: The overlooked challenge to equitable health care. *SAHR*, pages 189–195, 2012/2013.
- [16] A. Ittoo and G. Bouma. Minimally-supervised extraction of domain-specific part-whole relations using wikipedia as knowledge-base. *Data & Knowledge Engineering*, 85:57–79, 2013.
- [17] C. M. Keet. A core ontology of macroscopic stuff. In K. Janowicz and S. Schlobach, editors, *Proc. of EKAW'14*, volume 8876 of *LNAI*, pages 209–224. Springer, 2014.
- [18] C. M. Keet. Relating some stuff to other stuff. In *Proc. of EKAW'16*, volume 10024 of *LNAI*, pages 368–383. Springer, 2016.
- [19] C. M. Keet. A note on the compatibility of part-whole relations with foundational ontologies. In *FOUST-II: 2nd Workshop on Foundational Ontology, JOWO 2017*, volume 2050 of *CEUR-WS*, page 10p, 2017.
- [20] C. M. Keet. Representing and aligning similar relations: parts and wholes in isiZulu vs English. In *Language, Data, and Knowledge 2017 (LDK'17)*, volume 10318 of *LNAI*, pages 58–73. Springer, 2017.
- [21] C. M. Keet and A. Artale. Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology*, 3(1-2):91–110, 2008.
- [22] C. M. Keet and L. Khumalo. On the verbalization patterns of part-whole relations in isizulu. In *Proc. of INLG'16*, pages 174–183. ACL, 2016. 5-8 September, 2016, Edinburgh, UK.
- [23] L. Khumalo. Advances in developing corpora in African languages. *Kuwala*, 1(2):21–30, 2015.
- [24] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Ontology library. WonderWeb Deliverable D18 (ver. 1.0, 31-12-2003), 2003. <http://wonderweb.semanticweb.org>.
- [25] R. Mizoguchi. YAMATO: Yet Another More Advanced Top-level Ontology. In *Proc. of AOW'10, CRPIT*, pages 1–16. CRPIT, 2010. Sydney : ACS.
- [26] R. Motschnig-Pitrik and J. Kaasboll. Part-whole relationship categories and their application in Object-Oriented Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 11(5):779–797, 1999.
- [27] OpenMRS. <https://www.transifex.com/openmrs/OpenMRS/>.
- [28] M. Poveda-Villalón, M. C. Suárez-Figueroa, and A. Gómez-Pérez. Validating ontologies with OOPS! In *Proc. of EKAW'12*, volume 7603 of *LNAI*, pages 267–281. Springer, 2012.
- [29] J. Rogers and A. Rector. GALEN's model of parts and wholes: experience and comparisons. In *Proceedings of the AMIA Symp 2000*, pages 714–718. AMIA, 2000.
- [30] C. Rosse and J. L. V. Mejino Jr. A reference ontology for biomedical informatics: the foundational model of anatomy. *J. of Biomedical Informatics*, 36(6):478–500, 2003.
- [31] SNOMED CT, last accessed: 27-1-2012. <http://www.ihtsdo.org/snomed-ct/>.
- [32] N. Tandon, C. Hariman, J. Urbani, A. Rohrbach, M. Rohrbach, and G. Weikum. Commonsense in parts: Mining part-whole relations from the web and image tags. In *Proc. of AAAI'16*, pages 243–250. AAAI Press, 2016.
- [33] A. Varzi. *Handbook of Spatial Logics*, chapter Spatial reasoning and ontology: parts, wholes, and locations, pages 945–1038. Berlin Heidelberg: Springer Verlag, 2007.
- [34] A. C. Varzi. Mereology. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Stanford, fall 2004 edition, 2004. <http://plato.stanford.edu/archives/fall2004/entries/mereology/>.
- [35] H. Velichko et al. Implicit personality conceptions of the nguni cultural-linguistic groups of south africa. *Cross-Cultural Research*, 45(3):235–266, 2011.
- [36] L. Vieu and M. Aurnague. Part-of relations, functionality and dependence. In *Categorization of Spatial Entities in Language and Cognition*. John Benjamins, Amsterdam, 2005.
- [37] M. Winston, R. Chaffin, and D. Herrmann. A taxonomy of part-whole relations. *Cognitive Science*, 11(4):417–444, 1987.
- [38] T. Yıldız, B. Diri, and S. Yıldırım. Acquisition of Turkish meronym based on classification of patterns. *Pattern Analysis and Applications*, 19(2):495–507, 2016.