Representing and aligning similar relations: parts and wholes in isiZulu vs English

C. Maria Keet

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

Abstract. Ontology-enabled medical information systems are used in Sub-Saharan Africa, which require localisation of Semantic Web technologies, such as ontology verbalisation, yet keeping a link with the English language-based systems. In realising this, we zoom in on the part-whole relations that are ubiquitous in medical ontologies, and the isiZulu language. The analysis of part-whole relations in isiZulu revealed both 'underspecification'—therewith also challenging the transitivity claim—and three refinements cf. the list of common part-whole relations. This was first implemented for the monolingual scenario so that it generates structured natural language from an ontology in isiZulu. Two new natural language-independent correspondence patterns are proposed to solve non-1:1 object property alignments, which are subsequently used to align the part-whole taxonomies informed by the two languages.

1 Introduction

With the more widespread uptake of ontologies, localisation and internationalisation of existing ontologies, as well as *de novo* ontology development in a language other than English is becoming more commonplace. This brings afore a new set of problems in general regardless the natural language, as well as language-specific issues. In this paper, we zoom in on object properties, and part-whole relations in particular, and as other language isiZulu. IsiZulu is a language in the Bantu language family that has about 12 million first language speakers and about 25 million people in South Africa can speak it. Tools with an isiZulu interface are being developed, such as the medical translation app mobilezulu[1] to assist doctors with the language barrier during consultations, and the Electronic Health Record system OpenMRS[2] is popular in Sub-Saharan Africa, which imports the medical ontology SNOMED CT [25]. Localisations of OpenMRS are under way[3], which, in turn, will assist with the automatic generation of patient summaries, so that they will adhere better to the treatment instructions [30]. Such 'intelligent' information systems require generation of natural language, with as minimum requirement to verbalise the ontology.

[1] `mobilezulu.org.za`
[2] `https://wiki.openmrs.org/display/projects/Home`
[3] `https://www.transifex.com/openmrs/OpenMRS/`

To be able to realise this, the structured knowledge has to be localised and verbalised. It is well-known that medical and healthcare terminologies, such as SNOMED CT and the Foundational Model of Anatomy [21] are replete with part-whole relations, such as 'each heart is part of one human', 'operating team has as member at least one doctor', and 'HIV test is involved in a pre-natal checkup'. While the nouns (OWL classes) are fairly straightforward to translate and standardise, the relations (OWL object properties) are a different matter. It is known how to represent part-whole relations, notably the relatively wide uptake of the taxonomy of part-whole relations of [13], which is also popular in NLP (e.g., [26]). Recent efforts in finding verbalisation patterns for those part-whole relations to generate natural language sentences in isiZulu [14] focussed on the patterns, but essentially revealed that there are no 1:1 mappings between the identified part-whole relations in (the conceptualisation by people who speak as first/home language) isiZulu and English. This is further confounded by the issue that the 'has part' reading direction does not have a single word for it. The former issue brings afore the question *how to deal with non-1:1 mappings among object properties*, which, to the best of our knowledge, current multilingual models and tools do not have a solution for [3, 6, 8, 12, 17], though separation of ontology and natural language and lexicalisation [4] is obviously a good principle to start from.

The aim of this paper is to solve these two natural language-motivated problems, being non-1:1 alignments for object properties and absence of single reusable labels. We investigate this in detail for the demarcated, and well-researched, part-whole relations, and take as a use case the language isiZulu. First, a brief ontological analysis is carried out on the part-whole relations that were proposed for isiZulu natural language generation (NLG) in [14]. This revealed that there are both generalisations up to requiring parthood to be non-transitive, but also three refinements compared to the typical list of part-whole relations. Second, the engineering issues for the monolingual case are addressed and implemented for the ontology verbalisation algorithms of [14]. with Owlready [15] as proof-of-concept. Third, this is extended to the multilingual case by proposing a refinement to the 'VAP' correspondence ontology design pattern, *HetOP*, and introducing a new pattern, *UnionOP*, so as to systematically handle non-1:1 object property alignments. This is then applied to aligning the taxonomy of part-whole relations to the part-whole relations informed by isiZulu.

In the remainder of the paper we describe related works in Sect. 2 and introduce the main contribution in Sect. 3 for the monolingual scenario and in Sect. 4 for the multilingual case. We discuss in Sect. 5 and conclude in Sect. 6.

2 Related works

The contributions presented in the following sections draw in particular from related works on part-whole relations and from so-called "correspondence patterns" ontology design patterns, and a few relevant aspects of linguistic annotation models.

2.1 Part-whole relations

Part-whole relations have been investigated especially in the areas of Ontology (analytic philosophy), conceptual modelling, linguistics and NLP, notably [10, 13, 20, 26, 29, 31]. Multiple types of part-whole relations have been proposed, which resulted in a fairly stable taxonomy of part-whole relations [13]. It distinguishes between 'real' parthood relations (mereology) and part-whole relations in natural language utterances only (meronymy). The primitive *part_of* relation in mereology is antisymmetric, reflexive, and transitive [27], whereas meronymic relations are not necessarily transitive and where 'part' is used loosely, such as in "a musician is part of [i.e., *member_of*] an orchestra". Similar to ideas discussed in [29], the part-whole relations taxonomy in [13] distinguishes relations also by the categories of the relata for their meaning and excludes certain undesirable inferences; e.g., involved-in is a parthood specifically among processes. This taxonomy (see Fig. 1) can be used with various different surface readings/labels, like using made of instead of constituted of or preferring has ingredient over stuff part. The set of relations themselves are not really contested. These relations have been proposed in research done by people from multiple countries and cultures who speak multiple natural languages, so one could assume a genericity or even universality of it.

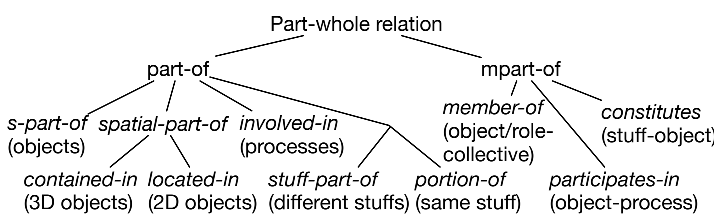

Fig. 1. Taxonomy of part-whole relations, based on [13], and informal description of their domain and range. The part-of branch is mereological parthood with transitivity; the mpart-of branch has relations that are non-transitive or intransitive.

Specific linguistically motivated analyses for languages other than English on their use of part-whole relations to confirm this are sparse. Vieu and Aurnague [29] focus on French, with an emphasis on parthood where the part has some particular function with respect to the whole, with "entities-as-a-lexical-type" for the Component Integral Whole parthood relation (s-part-of in Fig. 1). Thus, it remains within the same common set of recognised part-whole relations. To the best of our knowledge, no ontology research has been conducted on part-whole relations that is informed by a natural language in a family of languages other than Germanic or Italic. The works on mereology/meronymy in, notably, Arabic, Chinese, and Turkish focus on relation extraction, stating that they limit the extraction to the aforementioned typical set of part-whole relations or a subset thereof [1, 5, 32]. Upon further inspection, there are two noteworthily points. Cao at al. [5] did 'refine' constitution with an "Element-Object ... for convenient verification"; e.g., calcium as part of milk. It Is unclear whether the authors assert this is a part-whole relation semantically or linguistically distinguishable from the others. Yıldız et al [32] excluded the spatial part-whole relations, but 'constituted of' is distinguished from 'made of', with the former having a more built-type of flavour to it (examples of wholes given: system, program) and the latter more generic (examples of their wholes: questionnaire and public opinion). Finally, Keet and Khumalo [14] also started from the typical set of part-whole relations, but from a knowledge engineering and linguistic starting point and aimed at NLG, which resulted in some differences for isiZulu that we shall analyse in Sections 3 and 4.

2.2 Correspondence patterns and/or language models

There are two principal ways to deal with language-motivated mismatches of object properties (relations): either they are conceptually different, are represented as such in the ontology, and then a possibly heterogeneous alignment has to be asserted, or the underlying conceptualisation is the same (or similar enough), which then is represented in the ontology with one object property, and the language differences are dealt with in a language model by means of a separate language annotation file. The former requires *correspondence patterns* to assert, e.g., subsumption between elements in the different ontologies, and more complex mappings. Pattern alignment, rather than 1:1 mappings between ontology vocabulary elements, has been proposed in [22] with ontology design patterns (ODPs) that includes property equivalence and subsumption axioms, and which later also included data property to object property transformation [23]. The online ODP catalogue[4] contains 13 alignment patterns, but is sparse for complex object property mappings. It lists three patterns that constitute one relevant alignment case, called the "Vocabulary Alignment Pattern: Sub property of an external Property", which we shall extend and formalise in Section 4.1.

Several language models have been proposed in recent years. To assess them on their potential applicability for isiZulu, we first need to illustrate some pertinent aspects of rendering an axiom with a part-whole relation in isiZulu. The main point here is not the whole process of verbalising an axiom, but rather the constituents that will have to be dealt with in a language model for annotations. Let us take 'has part' in isiZulu and the common axiom type $C \sqsubseteq \exists R.D$, as presented in [14]; the verbalisation *pattern* is:

$\text{QCall}_{nc_{x,pl}}$ $\text{W}_{nc_{x,pl}}$ $\text{SC}_{nc_{x,pl}}$-CONJ-P_{nc_y} RC_{nc_y}-QC_{nc_y}-*dwa*

where W: entity playing whole; P: entity that plays the part; CONJ: Conjunction (for enumerative-and; *na-*); SC: Subject Concord for conjugation; PC: Possessive Concord; RC: Relative Concord; QCall: quantitative concord for universal quantification; and QC: quantitative concord for existential quantification. The equivalent of 'All humans have as part some heart' is *bonke abantu <u>banenhliziyo</u>*

[4] http://www.ontologydesignpatterns.org

eyodwa, with 'has part' underlined: the SC *ba-* from noun class 2 (nc2)'s *abantu* (the W) and the phonologically conditioned CONJ *na* + *inhliziyo* = *nenhliziyo*. Therefore it generates *bane-* for 'has part' in this sentence. With W='orchestra' (nc5, SC=*a-*) and P='musician' (*isazi somnyuziki*), the 'has part' is *anesazi somnyuziki* and with W='computer' (nc5) and P='CPU' (*umqondo womshini*), the 'has part' results in *anomqondo womshini*. There are six different SCs for the plural noun classes and one CONJ that has three phonologically conditioned variants, hence, there are 6*3=18 strings all having the same meaning of 'has part'. Ontologically, this ought not to be put in an ontology as 18 different object properties with equivalences, for it is one conceptualisation with context-dependent surface realisations [12], not 18 different types of relation. If only one label is used, then there needs to be some annotation and rules to govern generating, or selecting, the right form. The other main complication is best illustrated with 'contained in' that uses phonologically conditioned locative affixes to verbalise this notion. For instance, *imvilophu* 'envelope' becomes *emvilophini*, so both reading directions ('contains' and 'contained in') do not have a single name or label that is reusable for all sentences. (The list of the relevant sections of the verbalisation patterns is included in the third column of Table 1.)

Language models proposed for the Semantic Web include *lemon* [18] and its smaller ontolex-lemon W3C submission, OLiA [6], and the model in [12]. LIME [8] complements *lemon* for metadata, and is therefore not further considered. OLiA's permanent URL (purl) is offline despite trying over several days. Therefore, we consider in some detail only *lemon* and the positionalist model in [12]. *Lemon* [18] is orthogonal to the ontology, where an IRI of the class, property, or individual has a lexical entry from the lexicon and it must have a canonical form. One can define syntactic behaviour such as a property's subject and object as `Argument` and annotate, e.g., the case (e.g., genitive) in a `Frame`. This still requires a string for a lexical entry, not a stub for a system-generated identifier. For this to possibly work for multiple context-dependent labels for a relation, the `LexicalEntry` would need to be modified from `canonicalForm` and optional `altForm`s into relating an identifier as lexical form with a concept description (in natural language at least), and optionally a `Form`.

The model in [12] also offers additional annotation options, such as for, among others, case, tense, and prepositions. However, that model represents object properties differently from OWL, which is then mapped to OWL. Instead of two properties, like an ex:teaches and ex:taught-by, there is one relationship, say *ex:teaching* with two roles, one for each participating entity (e.g., [*lecturer*] and [*course*]), alike UML's association ends. This construction can have multiple *relational expressions* attached to it, such as e.g., teaches , taught by, lectures etc.. While this is ontologically preferable over *lemon*, it does not provide a solution for a property with no single label either, other than using an arbitrary string for the name of the relationship and linking it to a template or verbalisation pattern to generate the relational expressions.

Thus, no existing system or model can readily deal with unnamed or multiply named variants, and there are very limited object property mapping options.

3 Parts and wholes in isiZulu NLG: the monolingual case

In order to be able to address the multilingual setting, some issues have to be resolved for the monolingual setting first. While there are non-trivial aspects for the verbalisation in isiZulu, such as deep prepositions [12], the main issue here is that *containment* and all types of whole-part relations do not have a stable surface realisation as they do in English, which in OWL ontologies typically are merged in the property's naming or labelling (e.g., has part). Hence, there is no readily available string to name the object property with. The second issue is which part-whole relations exist, and the consequences that follow from it. To address this, we first subject to a brief ontological analysis those part-whole relations that have been investigated for isiZulu NLG (Section 3.1), then propose how to handle the unnamed properties (Section 3.2), and finally describe how it has been implemented (Section 3.3).

3.1 Ontological aspects of part-whole relations in isiZulu

Structuring the verbalisation patterns of [14] by their linguistic realisations for 'part' and analysis on relata as reported therein, then a taxonomy emerges that is substantially different from Fig. 1, which is depicted in Fig. 2. This taxonomy with eight relations is preliminary, in that not all terms denoting 'parts' in isiZulu have been investigated yet. There are two points of note already, however: there are more refined distinctions in part-whole relations—for portions, participation, and constitution—and more coarse-grained ones to the extent that the distinction between mereology and meronymy-only does not exist. *Ingxenye* 'part of' is used for parthood, involvement, membership, stuff parts, participation of individual objects (vs. collectives), and containment. This is a mix of mereological and meronymic part-whole relations and, according to the linguist (L. Khumalo), there is no difference. This means that it can result in erroneous deductions, for different things can be chained together that should not. For instance, *ingxenye* is used for 'hand is part of the musician' (structural parthood) and for 'musician is part of the orchestra' (membership), but a derivation from this is incorrect ontologically, as the hand is not part of the orchestra. That is, *ingxenye* (the generic parthood) is not transitive. One could try to contest the universality of transitivity of parthood in that, unlike the examples discussed in [28, 11, 2], really no distinctions are being made in this case. However, before conceding to non-transitivity, there are four points to consider. First, by making parthood non-transitive, one also loses desirable deductions and it will not assist in resolving distinguishing desirable from undesirable deductions—which was a reason to have multiple part-whole relations in the first place. Second, to the best of our knowledge, no empirical investigation has been carried out into ascertaining how many desirable and undesirable deductions one loses/gains in an ontology by asserting, or not, transitivity on parthood. That is, there is no estimate of its practical importance. Third, a partial order, which parthood is, *does* have the property of being transitive, both mathematically and ontologically. Fourth, there is no clear demarcation for pushing a tolerable little vs. too much, to shoe-horn the conceptualisation and language into the more widely used mereology vs non/in-transitive distinction; e.g., *ilunga* is a 'member of an organisation' and *ilungu* is a 'council member' that, albeit having narrower meanings cf. $member\_of$, might provide some wriggle room despite that it will not occur in text corpora like "member of" does in English.

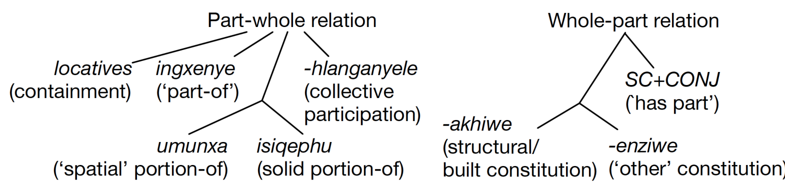

Fig. 2. Preliminary taxonomy based on the verbalisation patterns in [14].

The other aspect, refinement, has not been observed and investigated before, of which there are at least three cases. First, a difference is made between 'individual participation' and 'collective participation', as in a voter vs. the electorate participating in an election, and a doctor vs. an operating team participating in an operation. Depending on the foundational ontology (FO), this can be handled by the category of the participating objects—physical objects (e.g., a protein, human) or the roles they play (e.g., enzyme, voter) versus social agents (e.g., a company, electorate) as a minimum distinction. This would require all domain ontologies to adhere to some FO that contains this distinction. Whether this distinction can be pushed further ontologically, such as with Searle's collective intentions [24], remains to be investigated.

Second, there are special portions for space vs. for solid and solid-like objects, like between "the portion of the kitchen where the kitchen utensils are" vs "the sample of blood is a portion of the blood of the human". This can be solved also by taking into account the FO categories the participating objects belong to. If it concerns essentially space and a portion thereof only, then *umunxa* should be used (whereas when the focus is the physical object and space secondary, then it is containment), and if it is any amount of matter and a part that is of the same type of matter, then it is *isiqephu* for 'portion'.

Third, there is a distinction between *-akh-* and *-enz-*, which bears some resemblance between 'constituted of', as in "a vase is constituted of clay" and the more generic 'made of', as in "a pill is made of starch", although they are generally treated equivalently in other works (e.g., [13] and references therein) other than, perhaps, in Turkish [32]. The *-akh-* verb root is used for 'built' or composed things, whereas *-enz-* is used for all other cases. At the time of writing, it is not clear how to distinguish computationally between the two, i.e., beyond the general desideratum for constitution relating a physical object to an amount of matter (stuff). This means that for the time being, it is up to the modellers to choose one correctly. Note also that *akhiwe* and *enziwe* are only used in the whole[object]-to-part[stuff] reading direction, not from part to whole.

3.2 Processing unnamed object properties

We will now resolve handling relations that do not have a neat, single, label. Several options were explored, and we elaborate on two.

One can use an 'unnamed' object property in an axiom, provided the language supports declaring inverses and the property is named. OWL's `ObjectInverseOf` (Inv, for short) or `InverseObjectProperty` [19] can be used for that. Then, using Inv(partOf) amounts to hasPart; e.g., Human $\sqsubseteq \exists$Inv(partOf).Heart (or: Umuntu $\sqsubseteq \exists$Inv(ingxenye).Inhliziyo). While the axioms require more effort to understand, this is not an issue with a natural language layer on top of it. A downside is that it still requires at least one name, which the containment relation does not have. Also, OWL 2 EL is popular for large medical terminologies and SNOMED CT is represented in OWL 2 EL, but this profile does not have inverses.

The second option is to 'squeeze' it somehow into OWL's vocabulary element naming options: 1) use some arbitrary label (possibly a system-generated identifier) and describe the intention in the object property's annotation, as *lemon* does not have an attribute for this either; 2) use the English term in the ontology, ignoring the localisation; or 3) use some abbreviation of the English term. The linguist consulted (L. Khumalo, UKZN) preferred the arbitrary string option and the annotation field to describe the type of relation. This option can be realised with the positionalist model of [12] and with modification of *lemon* [18], as described in Section 2.2, or without either. Because currently all verbalisation knowledge is encoded in the verbaliser already, including the noun class information and processing of the deep prepositions, the simple annotation was chosen as proof-of-concept, for the complex alignments are the eventual target.

3.3 Implementation

The investigated part-whole relations of isiZulu have been represented in an OWL ontology, `PWzu.owl`, that imports `DOLCEmini.owl`—a module of the OWL-ized DOLCE foundational ontology [16] (i.e., of DLP3971.zip)—so as to constrain the domain and range of each relation with relatively well-known and defined entities. The context-dependent (multi-label) relations have been given arbitrary names, which are listed in the 3rd column of Table 1. These labels are then linked to the applicable verbalisation pattern in the verbaliser.

To enable testing as well as taking a step toward applicability in one of the use case scenarios (healthcare), the ontology verbaliser in isiZulu has been extended to be able to process OWL files. A Python script was already developed that implements the verbalisation algorithms, which needed only to be linked to OWL, which was achieved the OWL API for Python, Owlready [15].

To test it, we represented in OWL all test cases of [14] and those from earlier works on ontology verbalisation in isiZulu, totalling to 82 logical axioms of which 41 with part-whole relations; the others include named class subsumption, disjointness, and negated object properties in existential restrictions. All files are available from `http://www.meteck.org/files/geni/`, as well as those files mentioned in the next section. An annotated screenshot is shown in Fig. 3.

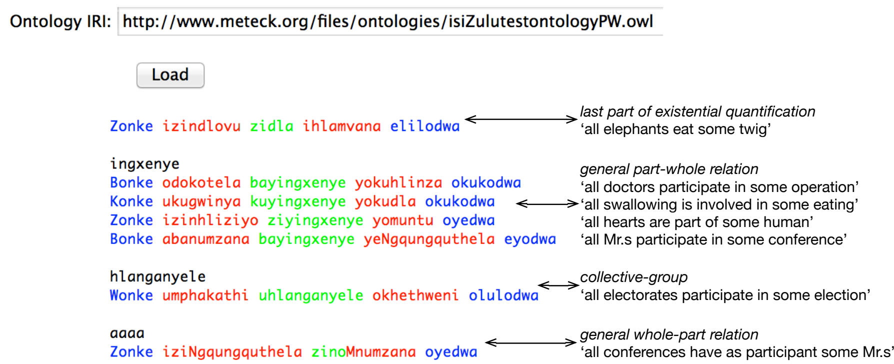

Fig. 3. Section of the GUI interface of the Semantic Web-enabled isiZulu verbaliser. Explanations were added on the right for clarification (not generated by the software). (Note: illustration was deemed more important than ontological precision.)

4 Parts and wholes in a multilingual setting

The principal problem to address is the alignment of the commonly structured part-whole relations with those in isiZulu. This will be needed in practice with broader adoption of, notably, SNOMED CT in Sub-Saharan Africa for EHR and patient discharge notes tailored to one's language. To assist with maintainability, one would want to keep a link between the source ontology and the localised one. IT may also be used in the other direction, to relate local knowledge to other knowledge in the world, such as African architecture [9]. As can be readily observed from Fig. 1 vs Fig. 2, this requires alignments that are not 1:1 mappings, hence, the need to resort to aligning *patterns*. This is described in Section 4.1 after which we apply it to the 'English↔isiZulu' part-whole relations in Section 4.2.

4.1 Non 1:1 mappings for object properties

We will first extend and refine the "Vocabulary Alignment Pattern: Sub property of an external Property" of the aforementioned ODP catalogue, so that it can then also be used for the more general case of property subsumption and its domain and range. The extended version is shown in Fig. 4 for two arbitrary ontologies O_1 and O_2, where A, B, C, D are classes and R_1, R_2 object properties. For precision, we formalise the pattern as follows, assuming a Semantic Web setting with OWL [19] and in a similar way as in [7]. Because it is a pattern, its elements refer not to vocabulary elements of a particular ontology, but any, we use calligraphic letters to distinguish them from OWL classes in the ontology to be aligned.

- *alignment pattern name*: *HetOP*
- *pattern elements*: $\mathcal{C}$, $\mathcal{D}$, $\mathcal{R}_1$ from O_1, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{R}_2$ from O_2 where $\mathcal{C}, \mathcal{D} \in V_C \cup$ `owl:Thing`.

– *alignment pattern contexts* (i.e., the fragment of interest):
  - *pattern* P_1 *in* O_1: $\exists\mathcal{R}_1.\mathcal{D} \sqsubseteq \mathcal{C}$, $\exists\mathcal{R}_1^-.\mathcal{C} \sqsubseteq \mathcal{D}$;
  - *pattern* P_2 *in* O_2: $\exists\mathcal{R}_2.\mathcal{B} \sqsubseteq \mathcal{A}$, $\exists\mathcal{R}_2^-.\mathcal{A} \sqsubseteq \mathcal{B}$.
– *Cross-ontology alignments*: $\mathcal{A} \sqsubseteq \mathcal{C}$ and $\mathcal{B} \sqsubseteq \mathcal{D}$ or $\mathcal{A} \equiv \mathcal{C}$ and $\mathcal{B} \sqsubseteq \mathcal{D}$ or $\mathcal{A} \sqsubseteq \mathcal{C}$ and $\mathcal{B} \equiv \mathcal{D}$, and $\mathcal{R}_2 \sqsubseteq \mathcal{R}_1$.
– *Global constraints* (to ensure the ontology does not become incoherent or inconsistent): $\mathcal{A} \sqcap \mathcal{C} \sqsubseteq \bot$ and $\mathcal{B} \sqcap \mathcal{D} \sqsubseteq \bot$ must NOT be asserted or derivable.

Note that `owl:Thing` as possible domain or range means that $\mathcal{R}_1$ thus may or may not have a user-defined domain and range declared. The three options for equivalence/subsumption (but not twice equivalence) follow the constraints for semantically correct role hierarchies as described in [13].

A second correspondence ODP for multilingual and heterogeneous alignments are where one ontology, and natural language, has one relation ($\mathcal{R}_1$) for which another language has two or more ($\mathcal{R}_i$ with $2 \leq i \leq n$). Aligning these differences can be accomplished systematically as follows.

– *alignment pattern name*: *UnionOP*
– *pattern elements*: $\mathcal{R}_1$ from O_1, $\mathcal{R}_i$ with $2 \leq i \leq n$ from O_2.
– *alignment pattern contexts*:
  - *pattern* P_1 *in* O_1: $\mathcal{R}_1$;
  - *pattern* P_2 *in* O_2: $\mathcal{R}_i$.
– *Cross-ontology alignments*: $\mathcal{R}_2 \sqsubseteq \mathcal{R}', \ldots, \mathcal{R}_n \sqsubseteq \mathcal{R}'$, and $\mathcal{R}' \equiv \mathcal{R}_1$ or $\mathcal{R}' \sqsubseteq \mathcal{R}_1$, with $\mathcal{R}'$ in O_1.
– *Global constraints*: the context and cross-ontology alignments do not violate *HetOP*.

The cross-ontology alignment ought to have an additional union axiom, $\mathcal{R}' \sqsubseteq \mathcal{R}_2 \sqcup \ldots \sqcup \mathcal{R}_n$, but this type of axiom is beyond OWL 2 DL. While one could eliminate $\mathcal{R}'$ and use $\mathcal{R}_1$ directly, the auxiliary property reduces the number of inter-ontology links and therewith simplifies maintenance. Whether one asserts equivalence or subsumption between $\mathcal{R}'$ and $\mathcal{R}_1$ depends on 1) the confidence one has on the exhaustiveness of the refined object properties, 2) any conflicting domain or range axioms.

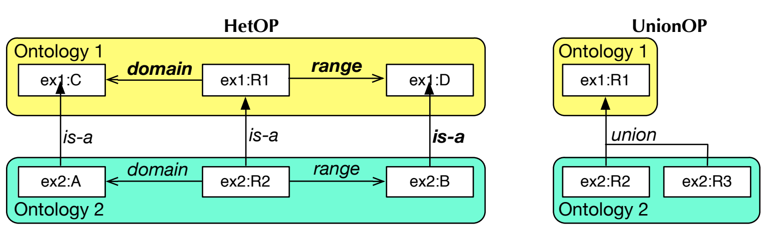

Fig. 4. Informal depiction of the *HetOP*, refining the VAP correspondence ODP (changes shown in bold face) and the new *UnionOP*.

The *HetOP* and *UnionOP* correspondence patterns can obviously be used for any set of different relations that have to be aligned, not just the part-whole relations we will be using them for. For instance, the Dutch *naspeuren* has both meanings of 'investigate' and 'trace' of the cause, which are two more refined notions for which *HetOP* may serve. The Dutch verb *tillen* (and *levantar* in Spanish) means both 'lift' or 'raise' that are similar but not synonyms, and 'swindle' (*estafar* in Spanish). So, one could assert either a $\mathsf{tillen}_{sense1} \equiv \mathsf{swindle} \equiv \mathsf{estafar}$ if each language has its own ontology, or add one object property with those labels in an annotation file for a multilingual ontology, and $\mathsf{tillen}_{sense2} \equiv \mathsf{levantar}$ (one OWL object property) with $\mathsf{lift} \sqsubseteq \mathsf{tillen}_{sense2}$ and $\mathsf{raise} \sqsubseteq \mathsf{tillen}_{sense2}$ as object sub property assertions. A quick check for verbs—candidates for object properties—in dictionaries suggest many more such cases.

4.2 Aligning the part-whole relations

The alignment between the two taxonomies was carried out manually in three stages. First, an informal alignment was carried out as a conceptualisation stage, so as to scope any issues, address naming and so on; the outcome of this informal alignment is included in Table 1. To ensure separation of concerns, the first column contains the name of the relationship, the second column lists a subset of the possible labels based on English that can be either syntactical differences or with terms that are generally used synonymously, and the third column contains the essential aspects of the isiZulu verbalisation patterns.

The second stage concerned the logical aspects and correspondence ODPs. For instance, *involvement* in the part-whole taxonomy has its domain and range restricted to DOLCE's Perdurant (roughly: processes), so then the *HetOP* pattern has to hold for pw:involvedIn ⊑ pwzu:ingxenye, for the latter does not have domain and range restrictions. Aligning isiZulu's *space-portion*—i.e., umunxa in the 'part of' reading direction—with an instantiation of *HetOP* would result in an inconsistency if it were to be tried with pw:portionOf, for umunxa's domain and range are DOLCE's Region whereas that of *portion* is DOLCE's Amount of Matter, which are located in disjoint branches in DOLCE. For this reason, one also cannot use *UnionOP*, though one can assert that pw:portionOf ⊑ pwzu:portion-of, with pwzu:portion-of the $\mathcal{R}'$ of the *UnionOP* specification. The bridge axioms are, principally, as follows:

- Equivalence mappings: pw:part-whole ≡ pwzu:part-whole, pw:whole-part ≡ pwzu:whole-part, pw:containedIn ≡ pwzu:ffff;
- Subsumption mappings: pwzu:hlanganyele ⊑ pw:participatesIn, pw:portionOf ⊑ pwzu:portion-of, pwzu:isiqephu ⊑ pw:portionOf;
- *HetOP* alignments (only property subsumption axioms listed): pw:memberOf ⊑ pwzu:ingxenye, pw:involvedIn ⊑ pwzu:ingxenye, pw:sPartOf ⊑ pwzu:ingxenye, pw:stuffpartOf ⊑ pwzu:ingxenye.
- *UnionOP* alignments (all axioms listed): pwzu:akhiwe ⊑ pwzu:constitution, pwzu:enziwe ⊑ pwzu:constitution, pwzu:constitution ≡ pw:constitutedOf.

Third, this was implemented in order to verify that an alignment is indeed feasible such that it does not lead to contradictions or undesirable deductions

Table 1. Summarised and informal version of the alignments. P=part, W=whole, SC=subject concord, CONJ=conjunction, LOC=locative prefix, LOCSUF=locative suffix, EP=epenthetic, COP=copula; the patterns in column 3 omit the parts about quantifiers and pluralisation.

Relationship name	**English surface realisations (notations and synonyms)**	**isiZulu part-whole patterns ("%": name in the OWL file)**
Reading direction: from Whole to Part		
partwhole	has_part, hasPart, part, ...	SC+CONJ+P % 'aaaa'
involvement	involves, sub-process, ...	SC+CONJ+P % 'aaaa'
membership	has_member, member, ...	SC+CONJ+P % 'aaaa'
stuffpart	hasStuffPart, hasSubstuff, has_ingredient, ...	SC+CONJ+P % 'cccc'
ind-participation	has_participant, participant, ...	SC+CONJ+P % 'aaaa'
col-participation	has_participant, participant, ...	SC+CONJ+P % 'bbbb'
containment	contains	SC+CONJ+P %'aaaa'
space-portion	has_portion, portion, piece, ...	SC+CONJ+P % 'dddd'
solid-portion	has_portion, portion, piece, ...	SC+CONJ+P % 'eeee'
built-constitution	constituted_of, madeOf, ...	SC+akhiwe nga+P
constitution	constituted_of, madeOf, ...	SC+enziwe nga+P
Reading direction: from Part to Whole		
partwhole	part_of, isPartOf, ...	SC+COP+ingxenye ya+W
involvement	involvedIn, ...	SC+COP+ingxenye ya+W
membership	member_of, isMemberOf, ...	SC+COP+ingxenye ya+W
stuffpart	stuff_part_of, ingredientOf	SC+COP+ingxenye ya+W
ind-participation	participates_in, ...	SC+COP+ingxenye ya+W
col-participation	participates_in, ...	SC+hlanganyele LOC+W+LOCSUF
containment	contained_in, ...	SC+EP+LOC+W+LOCSUF %'ffff'
space-portion	portion-of, ...	SC+COP+umunxa wa+W
solid-portion	portion-of, ...	SC+COP+isiqephu sa+W

in an ontology. To this end, we have taken the OWL file of the part-whole taxonomy, `PW.owl` and the new `PWzu.owl` with the isiZulu part-whole relations, and imported them into a new ontology that contains the aforementioned bridge axioms, `PWzuPWbridge.owl` (available from the aforementioned URL). There were no errors nor undesirable deductions.

5 Discussion

The solutions presented might appear specific to this scenario of part-whole relations. However, the solutions are generic for addressing 'unnamed' properties and, moreover, there are two new correspondence patterns for aligning object properties. It is also the first reported systematic assessment on, and, importantly, finding differences in, part-whole relations with respect to cultures with a language in a language family other than (Indo-)European.

The unintended 'byproduct' of attempting to verbalise part-whole relations in an ontology into isiZulu uncovered what may be considered different conceptualisations when one takes the category of domain and range as a desideratum to distinguish different part-whole relations. Given that there are a few papers that hint in a similar direction, such as the 'element-object' in Chinese [5], it suggests that there may be more generalisations and refinements in part-whole relations in other languages after all. A consequence of the differences in conceptualisation as perceived by amaZulu and their isiZulu language, was that transitivity does not hold for the main part-whole relation, *ingxenye*, when used as is. Transitivity is currently ignored in `PWzu.owl`, but that could be regained through the backdoor with the mappings to the part-whole relations taxonomy in `PW.owl`. It remains to be investigated how many 'interesting' deductions will be lost with the simple option and how much additional processing time the backdoor option would take.

While in hindsight the direction of the solution might be evident, it would have been useful to have had methodological guidance upfront. In particular, there were four possible cases that interfered with finding a solution: 1) same relationship with multilingual annotation vs. 2) different relationships, and 3) one 'base' ontology with possible alignments between annotations vs. 4) multiple localised ontologies with alignments between ontology vocabulary elements. For different relations, as in this case, one needs correspondence ODPs, as for multiple localised ontologies (though they are harder to maintain); for the same relation then annotations suffice, and for one base ontology with multiple language annotation models, perhaps annotation alignments will have to be devised.

Regarding the identified part-whole relations in isiZulu, we know there are more words used in prose to describe part-whole relations; e.g., *qukatha* 'containment' (roughly), *isididiyelo* 'ingredient' depending on how the whole came to be, and *ingqikithi* 'the essential part'. It remains to be seen whether they are synonyms or other refinements. A further point for investigation by linguists is that there may or may not be concept drift for the plurals of parts. A cursory check of *ingxenye* and its plural *izingxenye* in the 20 million-token isiZulu National Corpus (unpublished; pers. comm. L. Khumalo) indicated concept sameness in their use, but not *iminxa* (plural of *umunxa*) that relate abstract categories rather than spaces.

6 Conclusion

The analysis of part-whole relations in isiZulu revealed both 'underspecification' that complicates obtaining desirable deductions and three new refinements on part-whole relations for participation, spatial portions, and constitution. This was first implemented for the monolingual scenario that required a basic mechanism to deal with 'unnamed' object properties. It was shown to work to generate structured natural language from an ontology with isiZulu vocabulary to isiZulu sentences. The multilingual scenario required non-1:1 alignments between the common part-whole relations and those in isiZulu, for which two new correspondence patterns were introduced. These two patterns were used to align the part-whole relations. The patterns are specified such that they are natural language independent and thus can be used in aligning object properties in other ontologies as well.

Future research concerns a further analysis of part-whole relations in isiZulu and integrating the proof-of-concept tool with the model of [12].

Acknowledgments This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 93397). The author would like to thank M. Xakaza for assistance with tool development.

References

1. Al Zamil, M.G., Al-Radaideh, Q.: Automatic extraction of ontological relations from arabic text. Journal of King Saud University - Computer and Information Sciences 26(4), 462 – 472 (2014), special Issue on Arabic {NLP}
2. Barbier, F., Henderson-Sellers, B., Le Parc-Lacayrelle, A., Bruel, J.M.: Formalization of the whole-part relationship in the Unified Modelling Language. IEEE Transactions on Software Engineering 29(5), 459–470 (2003)
3. Bosca, A., Dragoni, M., Francescomarino, C.D., Ghidini, C.: Collaborative management of multilingual ontologies. In: Buitelaar, P., Cimiano, P. (eds.) Towards the Multilingual Semantic Web, pp. 175–192. Springer (2014)
4. Buitelaar, P., Cimiano, P., Haase, P., M., S.: Towards linguistically grounded ontologies. In: Proc. of ESWC'09. LNCS, Springer (2009)
5. Cao, X., Cao, C., Wang, S., Lu, H.: Extracting part-whole relations from unstructured chinese corpus. In: Proc. of FSKD'08. IEEE Xplore (2008)
6. Chiarcos, C., Sukhareva, M.: OLiA - ontologies of linguistic annotation. Semantic Web Journal 6(4), 379–386 (2015)
7. Fillottrani, P.R., Keet, C.M.: Patterns for heterogeneous tbox mappings to bridge different modelling decisions. In: Blomqvist, E., et al. (eds.) Proc. of ESWC'17. LNCS, vol. 10249, p. (in print). Springer (2017), 30 May - 1 June 2017, Portoroz, Slovenia
8. Fiorelli, M., Stellato, A., McCrae, J.P., Cimiano, P., Pazienza, M.T.: Lime: The metadata module for ontolex. In: Gandon, F., et al. (eds.) Proc. of ESWC'15. LNCS, vol. 9088, pp. 321–336 (2015)
9. Frescura, F., Myeza, J.: Illustrated glossary of Southern African Architectural Terms. Bilingual Glossary Series, UKZN Press (2016)
10. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. Phd thesis, University of Twente, The Netherlands. Telematica Instituut Fundamental Research Series No. 15 (2005)
11. Johansson, I.: Relations and predicates, chap. On the transitivity of the parthood relation, pp. 161–181. Ontos Verlag: Frankfurt (2004)
12. Keet, C.M., Chirema, T.: A model for verbalising relations with roles in multiple languages. In: Proc. of EKAW'16. LNAI, vol. 10024, pp. 384–399. Springer (2016), 19-23 November 2016, Bologna, Italy

13. Keet, C.M., Artale, A.: Representing and reasoning over a taxonomy of part-whole relations. Applied Ontology 3(1-2), 91–110 (2008)
14. Keet, C.M., Khumalo, L.: On the verbalization patterns of part-whole relations in isizulu. In: Proc. of INLG'16. pp. 174–183. ACL (2016), 5-8 September, 2016, Edinburgh, UK
15. Lamy, J.: Ontology-oriented programming for biomedical informatics. Studies in Health Technology and Informatics 221, 64–68 (2016)
16. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Ontology library. WonderWeb Deliverable D18 (ver. 1.0, 31-12-2003). (2003), http://wonderweb.semanticweb.org
17. McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: Interchanging lexical resources on the semantic web. Language Resources and Evaluation 46(4), 701–719 (2012)
18. McCrae, J., de Cea, G.A., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: The Lemon cookbook. Tech. rep., Monnet Project (2012)
19. Motik, B., Patel-Schneider, P.F., Parsia, B.: OWL 2 web ontology language structural specification and functional-style syntax. W3c recommendation, W3C (27 Oct 2009), http://www.w3.org/TR/owl2-syntax/
20. Motschnig-Pitrik, R., Kaasboll, J.: Part-whole relationship categories and their application in Object-Oriented Analysis. IEEE Transactions on Knowledge and Data Engineering 11(5), 779–797 (1999)
21. Rosse, C., Mejino Jr, J.L.V.: A reference ontology for biomedical informatics: the foundational model of anatomy. J. of Biomedical Informatics 36(6), 478–500 (2003)
22. Scharffe, F., Fensel, D.: Correspondence patterns for ontology alignment. In: Proc. of EKAW'08. LNAI, vol. 5268, pp. 83–92. Springer (2008)
23. Scharffe, F., Zamazal, O., Fensel, D.: Ontology alignment design patterns. Knowledge and Information Systems 40, 1–28 (2014)
24. Searle, J.R.: Collective intentions and actions. In: Cohen, P., Morgan, J., Pollak, M. (eds.) Intentions in Communication, pp. 401–415. MIT Press (1990), cambridge, MA, USA
25. SNOMED CT: (last accessed: 27-1-2012), `http://www.ihtsdo.org/snomed-ct/`
26. Tandon, N., Hariman, C., Urbani, J., Rohrbach, A., Rohrbach, M., Weikum, G.: Commonsense in parts: Mining part-whole relations from the web and image tags. In: Proc. of AAAI'16. pp. 243–250. AAAI Press (2016)
27. Varzi, A.C.: Mereology. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy. Stanford, fall 2004 edn. (2004), http://plato.stanford.edu/archives/fall2004/entries/mereology/.
28. Varzi, A.C.: A note on the transitivity of parthood. Applied Ontology 1, 141–146 (2006)
29. Vieu, L., Aurnague, M.: Part-of relations, functionality and dependence. In: Aurnague, M., Hickmann, M., Vieu, L. (eds.) Categorization of Spatial Entities in Language and Cognition. John Benjamins, Amsterdam (2005)
30. Wilcox, L., Morris, D., Tan, D., Gatewood, J., Horvitz, E.: Characterising patient-friendly micro-explanations of medical events. In: Proc. of CHI'11. pp. 29–32. ACM (2011)
31. Winston, M., Chaffin, R., Herrmann, D.: A taxonomy of partwhole relations. Cognitive Science 11(4), 417–444 (1987)
32. Yıldız, T., Diri, B., Yıldırım, S.: Acquisition of Turkish meronym based on classification of patterns. Pattern Analysis and Applications 19(2), 495–507 (2016)