July 2023

Towards an Ontology of Viewpoints

Frances GILLIS-WEBBER [a,1]
[a] *Computer Science Department, University of Cape Town, South Africa*
ORCiD ID: Frances Gillis-Webber https://orcid.org/0000-0002-3740-5904

Abstract. In a multilingual domain ontology developed using the labels approach, where each ontological entity is labelled with a language-tagged string, two scenarios result: (1) the ontology is 'language-independent', where there is an equal number of labels per natural language, or (2) the ontology is a 'primary-language' ontology, where one natural language takes precedence over the other languages used. In a multilingual ontology, it is assumed there is full equivalence between the different languages, however, each natural language, as an embodiment of a culture, differs in how it interprets and organises the world. The result is that although the viewpoint expressed by the multilingual domain ontology is thought to be universal, one natural language is very often privileged, typically English.

Using the culture-bound concepts of 'dowry' and 'bride price', we demonstrate the differences in perspective when considered for different languages and sub-domains. We propose an ontology, Model of Multiple Viewpoints (MULTI), where both language and culture are considered together, and language is classified as a social norm of a community. MULTI is formalised in OWL and aligned to DOLCE+DnS Ultralite, a foundational ontology suitable for modelling contexts. The evaluation of MULTI is done against the identified use cases. The expected result is that an ontology can be annotated with its viewpoint, thus making the viewpoint of the ontology explicit.

Keywords. viewpoints, multilingual ontologies, multilingualism

1. Introduction

When determining the requirements for the Web Ontology Language (OWL), internationalisation was identified as a goal, where the development of a multilingual ontology should be supported by OWL so that it can be usable by different countries and cultures [1]. Additionally, it was also specified that OWL should potentially allow for different views of the same ontology to be developed, with each view appropriate for a culture [1]. Whether this meant allowing for differing perspectives, for example, distinguishing between the perspectives of the colonised and the coloniser, or it meant providing a view of the ontology specific to a natural language from that culture, is unclear.

To realise the internationalisation goal, there was a requirement to support multiple language-specific labels within an ontology. This requirement for multilingual labels has indeed been met, however in doing so, the assumption was made that a (domain) ontology is a set of universal concepts, a nomenclature for which multiple language-specific labels can be set for a resource. With the exception of scientific terms, the reality is that there is

[1] Corresponding Author: F. Gillis-Webber, fgilliswebber@cs.uct.ac.za.

very often an anisomorphic relationship between the multilingual labels for a resource, even more so for those resources representative of concepts which are culture-bound, as each natural language, as an embodiment of a culture, differs in how it interprets and organises the world [2].

When developing a multilingual ontology using multilingual labels, an annotation is set using `rdfs:label`, and a language tag which conforms to BCP 47 [3] is used to identify the natural language of the annotation. There are two typical scenarios that result: (1) the ontology is natural language-independent, or (2) the ontology has a primary natural language. If the former, then the labels for each entity are mostly equivalent in number and the URI fragment identifiers are opaque. If the latter, then one natural language is privileged over the other natural languages used in the ontology, to the extent that the non-primary language labels are a translation of the primary language label. The consequence, particularly for the latter scenario, is that the primary language, as the viewpoint of the ontology by which the domain has been interpreted, is not made explicit.

Oxford English Dictionary defines a viewpoint as a "mental position or attitude from which a subject or question is considered; a point of view" [4], where 'perspective' is a synonym. In this paper, we propose an ontology to model different viewpoints, where the focus is on those viewpoints pertaining to natural language and culture. The ontology, Model of Multiple Viewpoints[2] (MULTI), is formalised in OWL and aligned to Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [5]. Using MULTI, the applicable viewpoints of an ontology or its entities are then specified by way of annotation.

The remainder of the paper is structured as follows. In Section 2, we introduce several language examples from which use cases are identified. MULTI is presented in Section 3, and an evaluation thereof is given in Section 4. This follows with a discussion in Section 5, and related works detailed in Section 6. The paper concludes with Section 7.

2. Motivation

If we consider two culture-bound concepts, 'dowry' and 'bride price', both are customs associated with the act of marriage. Dowry is defined in English, identified by the ISO 639-1 language code *en*, as goods, property or money provided by a daughter's family at time of marriage to ensure their daughter's economic security, and bride price is defined as money or property given by the bridegroom to the family of the bride [6]. In a hypothetical ontology, where the sub-domain is the act of marriage and its customs, both concepts are included, where each can be modelled as a sub-class of MarriageCustom. In the Nguni language, isiXhosa, there are equivalent terms: 'ikhazi' for dowry, and 'lobola' for bride price. However, if we consider the same concepts from the perspective of the AmaXhosa (the first-language speakers of isiXhosa), then the concept of 'dowry' as per the English definition no longer applies. Instead, within AmaXhosa culture, 'ikhazi' refers to the cattle (or money in lieu of cattle) paid as part of the bride price [6].

Staying with the same sub-domain, there is a variety of English called South African English (SAE) that is particular to South Africans. The custom of dowry in English is not practised by SAE first-language speakers, nor is the custom of lobola. South African Indian English (SAIE) is in turn a variety of SAE, spoken by local Indian communities. In

[2] `https://w3id.org/MULTI`

SAIE, the concept of dowry is relevant within this community, however the term 'thilak' is used instead of 'dowry' [7].

If we consider the representation of the concept of medical consent in an ontology but only within the context of Western medicine, then ConsentBySpouse can be defined by the following axioms:

$$\mathsf{InformedConsent} \sqcap \mathsf{UnconsciousPatient} \sqcap \mathsf{SpouseOfPatient} \sqcap \exists \mathsf{IsActive.LegalMarriage}, \quad (1)$$

$$\mathsf{CivilMarriage} \sqsubseteq \mathsf{LegalMarriage} \quad (2)$$

Customary marriage is recognised in South Africa as a legal marriage, of which the payment of 'lobola' is proof that the marriage was negotiated in accordance with local customs and thus fulfills in part the requirements for a valid customary marriage [8]. If medical consent is considered from the perspective of South Africa, then customary marriage is a legal marriage, in addition to civil marriage, if the requirements as per South African law are met.

In this paper, an ontology developed from the perspective of a language or region, is deemed a viewpoint. For language, this includes regional varieties and dialects. Similarly, if the ontology is intended to be more language- or region-neutral, with the aim to be 'language-independent' or 'universal', then this is also deemed a viewpoint. The viewpoints to be considered here, taken from the aforementioned examples, are shown in Table 1. Viewpoints 1–3 serve as use cases where a viewpoint can be specific to a natural language. For a viewpoint specific to a region, Viewpoint 4 for the country of South Africa is a use case. For a language-independent viewpoint, where it is not specific to any one language or region, Viewpoints 5 and 6 are use cases.

Table 1. The identified use cases, where VP is a viewpoint.

VP1	isiXhosa
VP2	South African English
VP3	South African Indian English
VP4	South Africa
VP5	Language-independent
VP6	Language-independent, with Western perspective

3. The Model of Multiple Viewpoints

Language and culture are inextricably linked, where the definition of culture used here is "the customary beliefs, social forms, and material traits of a racial, religious, or social group" [9]. We begin first with the representation of language in MULTI, shown in Figure 2, before presenting the cultural aspects. MULTI was aligned to the foundational ontology, DOLCE. The 'flavour' of DOLCE used was that of DOLCE+DnS Ultralite, a lightweight ontology suitable for modelling contexts [10]. The Model of Language Annotation (MOLA) [11], OntoLex-Lemon [12], and LexInfo [13] were also used, although by soft reuse only. Three distinct parts to representing a language in an ontology are shown for the English language in Figure 1. We will discuss each part in turn.

Borrowing the visualisation from [14], we adapt it to visualise MULTI. In the diagrams that describe MULTI (Figures 2, 3, and 4), each class is shown as a rectangle with a label,

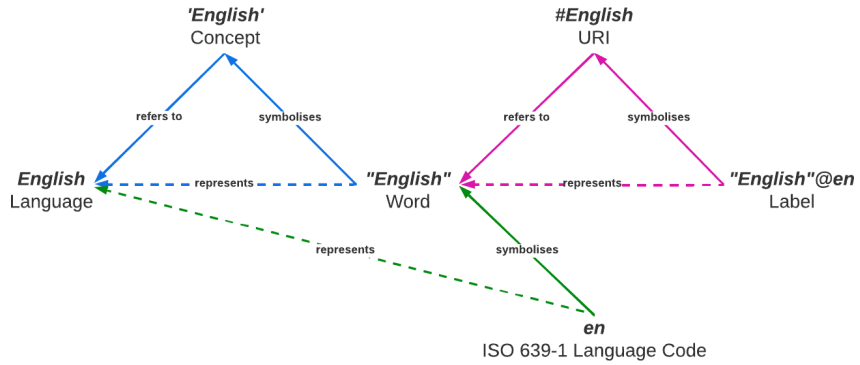

Figure 1. Using the semiotic triangle as a visualisation, there are three distinct parts when representing a natural language in MULTI, shown here for English: the blue triangle is the representation of the concept *'English'* (Part 1), the pink triangle is the representation of the same concept in an OWL ontology (Part 2), and the green triangle is the representation of English using a standardised language identifier, shown here for ISO 639 (Part 3).

and if a class is a sub-class, it is contained within its parent rectangle. Those classes with labels in **bold** are new classes (either newly created in MULTI or soft-reused from MOLA, Ontolex-Lemon, or LexInfo), where all other classes are from DOLCE. An arrow represents a relation between two classes, where the start of the arrow is the domain, and the end, the range. Each arrow is labelled with the property name, and with the exception of annotation axioms, are followed by some or only, both of which are natural language keywords from Manchester OWL Syntax for an existential or value restriction respectively. Unless explicitly specified otherwise, all object properties are from DOLCE. Description Logic is used for the axiomatisation of the ontology. Annotation assertions are shown using OWL, serialised in Functional Style Syntax (FSS).

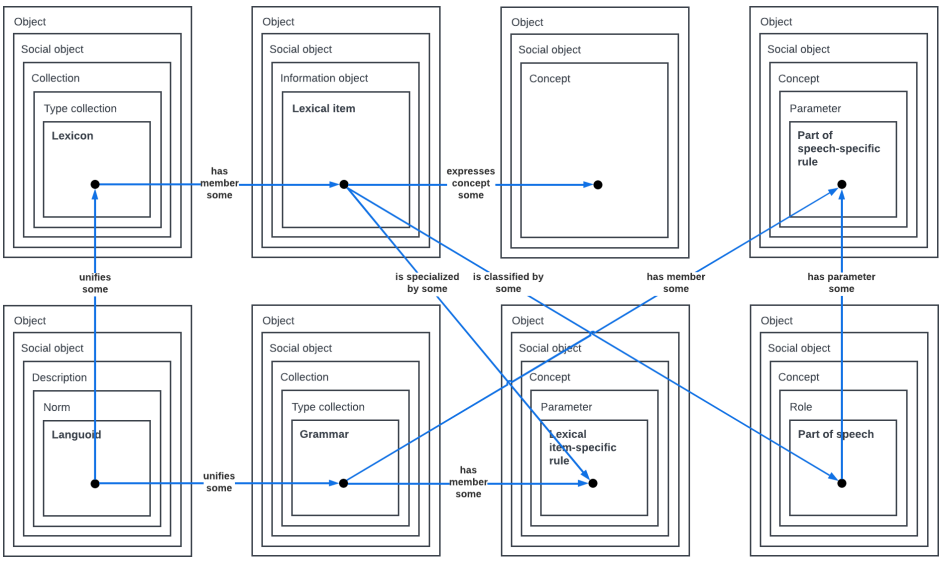

Figure 2. The visualisation of the classes and object properties for the representation of the concept 'language' (Part 1) in MULTI.

3.1. Language

A natural language comprises a lexicon and a grammar, where a lexicon is a mental dictionary which lists that language's words. Although a lexicon could have been treated as an aggregate of its members in MULTI, using, for example, DOLCE's ObjectAggregate, it was opted to treat a lexicon as a conceptual container of words. DOLCE's Collection was selected for this, using its sub-class TypeCollection (shown in *Ax1*), where the latter is a collection shared by a maximal set of individuals of the same type.

(Ax1) Lexicon $\sqsubseteq$ *TypeCollection*
(Ax2) LexicalItem $\sqsubseteq$ *InformationObject*
(Ax3) Lexicon $\sqsubseteq \exists$*hasMember.LexicalItem*
(Ax4) LexicalItem $\sqsubseteq \exists$*expressesConcept.Concept*

In *Ax2*, the term 'lexical item' is used instead of 'word', so as to account for those languages whose morphemes (smallest meaningful units of the language) rarely stand alone (unlike that of, for example, English).

A grammar pertains to the rules of a language, for which there is rules for the assignation of meaning, the combination of sounds into words, word formation, and syntax [15]. Within MULTI however, the representation of a grammar is simplified:

(Ax5) Grammar $\sqsubseteq$ *TypeCollection*
(Ax6) LexicalItemSpecificRule $\sqsubseteq$ *Parameter*
(Ax7) PartOfSpeechSpecificRule $\sqsubseteq$ *Parameter*
(Ax8) Grammar $\sqsubseteq \exists$*hasMember.LexicalItemSpecificRule*
(Ax9) Grammar $\sqsubseteq \exists$*hasMember.PartOfSpeechSpecificRule*
(Ax10) PartOfSpeech $\sqsubseteq$ *Role*
(Ax11) LexicalItem $\sqsubseteq \exists$*isClassifiedBy.PartOfSpeech*
(Ax12) PartOfSpeech $\sqsubseteq \exists$*hasParameter.PartOfSpeechSpecificRule*
(Ax13) LexicalItem $\sqsubseteq \exists$*isSpecializedBy.LexicalItemSpecificRule*

A grammar is similarly represented as that of Lexicon, using TypeCollection, however it has two types of members: LexicalItemSpecificRule and PartOfSpeechSpecificRule, both of which are sub-classes of DOLCE's Parameter class. In MULTI, constraints are encoded as parameters. A lexical item may be constrained by one or more grammar rules. Parts of speech (POS) is not made explicit in MULTI as not all languages share the same POS. Similar to that of lexical items, a POS may be constrained by one or more grammar rules. A POS is modelled as a Role, where a role is defined by DOLCE as a concept which classifies an object (in this case, a lexical item).

Axioms 1–13 pertain to the intension of language. Language is defined as "a system of conventional spoken, manual (signed), or written symbols by means of which human beings, as members of a social group and participants in its culture, express themselves" [16]. Social norms can be defined as those informal rules by which behaviour in societies is governed [17]. In a society, if a person chooses not to use the same language used by everybody else, then that person may not be able to communicate. Within the context of MULTI, language has thus been classified as a social norm. To do this, we use DOLCE's Description class, which is a SocialObject intended to either represent a conceptualisation or to be used as a 'descriptive context'. Description in turn has a sub-class Norm, to which Languoid is aligned (given in *Ax14*). The concept 'Language' is

modelled as a sub-class thereof, using the Languoid class from MOLA, where Languoid is either a language or a language family.

(Ax14) Languoid $\sqsubseteq$ *Norm*
(Ax15) Language $\sqsubseteq$ *Languoid*
(Ax16) LanguageFamily $\sqsubseteq$ *Languoid*
(Ax17) Languoid $\sqsubseteq \exists$*unifies.Lexicon*
(Ax18) Languoid $\sqsubseteq \exists$*unifies.Grammar*

To associate both a lexicon and a grammar with the conceptualisation of language, DOLCE's unifies property is used, which takes Description and Collection as domain and range respectively. *Axioms 1–18*, as well as their alignment to DOLCE, are visualised in Figure 2.

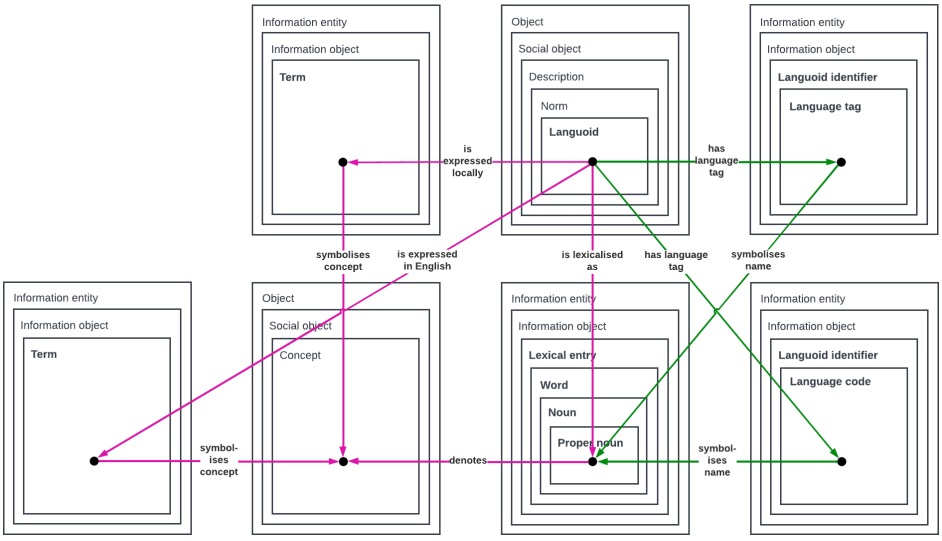

Figure 3. The visualisation of the classes and annotation properties for the representation of a language in an ontology (Parts 2 and 3) in MULTI.

Continuing with Languoid, we now focus on the annotation axioms of *Ax14*. These annotations, pertaining to Parts 2 and 3, are visualised in Figure 3.

(Ax19) Term $\sqsubseteq$ *InformationObject*

(AnnAx1) AnnotationAssertion(isExpressedLocally as av *)*
(AnnAx2) AnnotationAssertion(isExpressedInEnglish as av *)*

In MULTI, we distinguish between the term which expresses the concept Languoid, and the lexicalisation thereof. Two annotation assertions are given in *AnnAx1* and *AnnAx2*, where *as* is the annotation subject and *av* the annotation value. The annotation subject is a Language individual. The annotation value is a language-tagged string or a URI, and if the latter, then it is an individual of Term (from *Ax19*). The annotation value for isExpressedLocally is the term as used by first language speakers of their language, in contrast to isExpressedInEnglish. Both annotation properties are newly added to MULTI.

(Ax20) ProperNoun $\sqsubseteq$ *Noun* *

(Ax21) LexicalEntry $\sqsubseteq$ *InformationObject*

(AnnAx3) AnnotationAssertion(isLexicalisedAs as av *)*

The lexicalisation of a language name is classified as a proper noun, a name given for a specific thing. OntoLex-Lemon has a LexicalEntry class, for which there are several sub-classes, one of them Word. A lexical entry is defined by OntoLex-Lemon as the representation of "a unit of analysis of the lexicon" [12]. LexInfo in turn has the classes ProperNoun, a sub-class of Noun, which is then aligned to Word from OntoLex-Lemon. All four classes are used in MULTI for the lexicalisation, shown in *Axioms 20–21* (the asterisk indicates excluded axioms due to space constraints). The annotation assertion isLexicalisedAs is given in *AnnAx3*, and is newly added to MULTI.

As shown in Figure 3, denotes and symbolisesConcept are annotation properties which have Concept as an annotation value in an assertion, although this value can also be a URI from an external resource. denotes is an object property from OntoLex-Lemon and symbolisesConcept is newly added to MULTI.

Moving onto language codes, LanguageCode and LanguageTag are classes from MOLA, where the former represents ISO 639 language codes and the latter represents a language tag formed using the BCP 47 pattern. Both are included in MULTI, aligned to a new class LanguoidIdentifier:

(Ax22) LanguoidIdentifier $\sqsubseteq$ *InformationObject*

(AnnAx4) AnnotationAssertion(hasLangugeTag as av *)*
(AnnAx5) AnnotationAssertion(symbolisesName as av *)*

A language identifier is associated with a languoid using the hasLanguageTag object property from MOLA, created as an annotation property. This has been similarly done with the symbolisesName object property from DOLCE, shown in *AnnAx5*, where each language identifier is a reference of the proper noun in *Ax20*.

3.2. Community

We now turn our attention to the representation of viewpoints in MULTI. The word 'narrative' has several lexical senses, one of which is "a way of presenting or understanding a situation or series of events that reflects and promotes a particular point of view or set of values" [18]. DOLCE has a Narrative class, which is a sub-class of the previously-mentioned Description. Using this class to align to, we introduce the following axioms:

(Ax23) Viewpoint $\sqsubseteq$ *Narrative*
(Ax24) CommunitySpecificViewpoint $\sqsubseteq$ *Viewpoint*
(Ax25) CommunityAgnosticViewpoint $\sqsubseteq$ *Viewpoint*

As shown in *Axioms 24* and *25*, two types of viewpoints are distinguished, where our focus for now is on *Ax24*. A CommunitySpecificViewpoint unifies a Community, which has members of type NaturalPerson, and hasLocation some Place, where Place can be AbsoluteLocation or RelativeLocation. A Community has DemographicVariable as a quality, where each DemographicVariable is a value from CulturalExpression. See Figure 4 for a visualisation thereof.

Three types of communities are modelled in MULTI: linguistic community, community of identity, and a political community. LinguisticCommunity is defined here as a

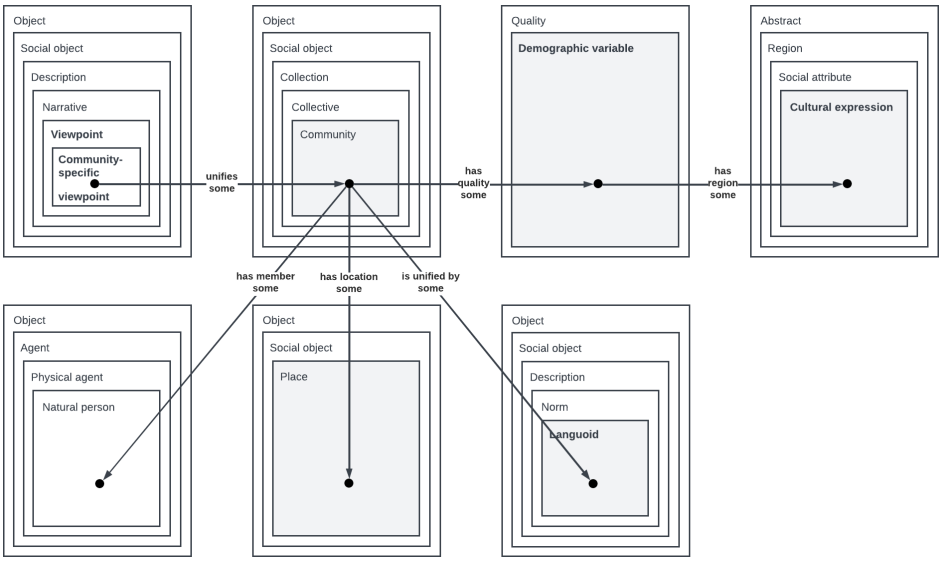

Figure 4. The visualisation of the classes and annotation properties for the representation of a community-specific viewpoint in MULTI. The grey areas indicate there are sub-classes which have not been included in the visualisation.

community where membership is predominantly based on language. If we recall the definitions of social norms and language, a LinguisticCommunity is a community for which a shared language is critical for communication. CommunityOfIdentity is a community whose membership is driven more by personal autonomy, and for which there is the following sub-classes:

(Ax26) LocationBasedCommunity $\sqsubseteq$ *CommunityOfIdentity*
(Ax27) ReligiousCommunity $\sqsubseteq$ *CommunityOfIdentity*
(Ax28) SubcultureCommunity $\sqsubseteq$ *CommunityOfIdentity*
(Ax29) VirtualCommunity $\sqsubseteq$ *CommunityOfIdentity*

Language is less critical here, although there may be a shared sociolect (a dialect of a social group). PoliticalCommunity is a community whose membership is less autonomous and is not driven primarily by language.

Demographics is the study of the characteristics of a social group. Each characteristic is specified as a demographic variable, where an example variable is 'age'. We employ the notion of demographic variables in MULTI, by which to specify the attributes of a community:

(Ax30) DemographicVariable $\sqsubseteq$ *Quality*
(Ax31) CollectiveDemographicVariable $\sqsubseteq$ *DemographicVariable*
(Ax32) IndividualDemographicVariable $\sqsubseteq$ *DemographicVariable*

We distinguish between two types of demographic variable: collective and individual (given in *Axioms 31–32*). CollectiveVariable are those characteristics which are prescribed to a group as a whole. Example variables are Income and Nationality, both of which are sub-classes to CollectiveDemographicVariable. IndividualDemographicVariable are

those characteristics which apply to the person, as member of said community. Examples of sub-classes of IndividualDemographicVariable include Age, and GenderIdentity.

DemographicVariable is aligned to DOLCE's Quality class, where a quality is "any aspect of an Entity ..., which cannot exist without that Entity" [10]. A Quality has a Region in Dolce, where a region is used as a value of a quality.

(Ax33) CulturalExpression $\sqsubseteq$ *SocialAttribute*

(Ax34) DemographicVariable $\sqsubseteq \exists$ *hasRegion.CulturalExpression*

In MULTI, CulturalExpression is aligned as a sub-class of SocialAttribute, which is in turn a sub-class of Region. Sub-classes of CulturalExpression are classified by collective and individual, of which an example sub-class for each respectively is IncomeScale, and GenderExpression. The values for a CulturalExpression are then specified using DOLCE's hasRegionDataValue data property.

A community is distinguished in MULTI by one or more cultural expressions. Each cultural expression consists of several values, for which a selection of those values is applied to a community as a demographic variable. From this, we can define our three types of Community:

(Def1) LinguisticCommunity $\equiv$ *Community* $\sqcap \exists$*isUnifiedBy.Language*

(Def2) CommunityOfIdentity $\equiv$ *Community* $\sqcap$ *(*$\exists$*hasQuality.DemographicVariable* $\sqcap$ *(*$\exists$*hasRegion.CulturalExpression))*

(Def3) PoliticalCommunity $\equiv$ *Community* $\sqcap$ *(*$\exists$*hasLocation.MetaphoricalLocation* $\sqcup$ $\exists$*hasLocation.AbsoluteLocation)*

Lastly, we can define our two viewpoints given in *Axioms 24–25*:

(Def4) CommunitySpecificViewpoint $\equiv$ *Viewpoint* $\sqcap$ *(*$\exists$*unifies.LinguisticCommunity* $\sqcup$ $\exists$*unifies.CommunityOfIdentity* $\sqcup$ $\exists$*unifies.PoliticalCommunity)*

(Def5) CommunityAgnosticViewpoint $\equiv$ *Viewpoint* $\sqcap$ *(*$\neg$ *CommunitySpecificViewpoint)*

To annotate an ontology or ontology entity with a viewpoint, *AnnAx6* can be used, where the range of *av* is a Viewpoint.

(AnnAx6) AnnotationAssertion(hasViewpoint as av *)*

4. Evaluation

Revisiting the use cases listed in Table 1, we evaluate MULTI for each use case. In Listings 2–7, OWL fragments are given for each use case, serialised in FSS. The namespaces of the ontologies used are defined in Listing 1 (the standard ontologies such as OWL and RDF are assumed to already be defined). MULTI is available to view: `https://w3id.org/MULTI`

```
1 Prefix(:=<http://example.com/ont/>)
2 Prefix(multi:=<https://w3id.org/MULTI#>)
3 Prefix(dul:=<http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#>)
4 Prefix(mola:=<https://w3id.org/MOLA#>)
5 Prefix(ontolex:=<http://www.w3.org/ns/lemon/ontolex#>)
6 Prefix(lexinfo:=<http://www.lexinfo.net/ontology/3.0/lexinfo#>)
```

```
7 Prefix(dbp:=<http://dbpedia.org/resource/>)
```

Listing 1: Namespaces for each of the ontologies used.

An example OWL fragment to model VP1 (isiXhosa) is shown in Listing 2. Due to space constraints, annotations such as `rdfs:label` have been excluded. The isiXhosa viewpoint pertains to a community of speakers, unified by the isiXhosa language. For this reason, we select both CommunitySpecificViewpoint (Line 1) and LinguisticCommunity (Line 2), where, if we recall our definition of a linguistic community, membership is predominantly based on language. In Line 3, Language from *Ax15* is used instead of Languoid, as Languoid supersumes both Language and LanguageFamily. If we were referring to the community of speakers unified by any of the Nguni languages (of which isiXhosa is one), then LanguageFamily would have been a better choice. MoLA has a variety of lects as a sub-class to Language, one of which is Dialect. If we were referring to the speakers of a community unified by one of the isiXhosa dialects, then Dialect would have been suitable.

```
1  ClassAssertion( multi:CommunitySpecificViewpoint :XhosaViewpoint )
2  ClassAssertion( multi:LinguisticCommunity :XhosaComm )
3  ClassAssertion( mola:Language :XhosaLang )
4  ClassAssertion( lexinfo:ProperNoun :XhosaLexEnt )
5  ClassAssertion( mola:LanguageCode :XhosaLangCode )
6  ObjectPropertyAssertion( dul:unifies :XhosaViewpoint :XhosaComm )
7  ObjectPropertyAssertion( dul:isUnifiedBy :XhosaComm :XhosaLang )
8  Annotation...ion( multi:isExpressedLocally :XhosaLang "isiXhosa"@xh )
9  Annotation...tion( multi:isExpressedInEnglish :XhosaLang "Xhosa"@en )
10 AnnotationAssertion( mola:hasLanguageTag :XhosaLang :XhosaLangCode )
11 AnnotationAssertion( multi:symbolisesName :XhosaLangCode :XhosaLexEnt )
12 AnnotationAssertion( multi:isLexicalisedAs :XhosaLang :XhosaLexEnt )
13 AnnotationAssertion( ont:denotes :XhosaLexEnt dbp:Xhosa_language )
```

Listing 2: OWL fragment for the isiXhosa Viewpoint (VP1)

An example OWL fragment to model VP2 (South African English) is shown in Listing 3. Due to space constraints, we exclude those annotation assertions for terms and lexical entries, as these are similar to that from VP1. The viewpoint of South African English pertains to a community of speakers who speak a variety of English spoken only in South Africa. Like that of VP1, this viewpoint is community-specific. The community is predominantly unified by language, however, we also include location.

In Line 3, we include a class assertion for `:SouthAfrica`, where `:SouthAfrica` is an IndependentCountry. As previously mentioned, in MULTI, we distinguish between absolute and relative locations. Country is a sub-class of AbsoluteLocation, however we also distinguish between those countries which are independently or dependently governed. Wales is an example of the latter, where it is governed (partly) by Britain, which is in turn an independent country, as it is self-governing. In Listing 2, a language code was associated with isiXhosa, however in Listing 3, a language tag was used as there is no ISO 639 language code available for South African English. The language tag, shown in Line 12, conforms to BCP 47, where "ZA" is the ISO 3166 code for South Africa.

```
1 ClassAssertion( multi:CommunitySpecificViewpoint :SAEngViewpoint )
2 ClassAssertion( multi:LinguisticCommunity :SAEngComm )
3 ClassAssertion( mola:Dialect :SAEngLang )
4 ClassAssertion( mola:LanguageTag :SAEngLangTag )
5 ClassAssertion( multi:IndependentCountry :SouthAfrica )
6 ObjectPropertyAssertion( dul:unifies :SAEngViewpoint :SAEngComm )
7 ObjectPropertyAssertion( dul:isUnifiedBy :SAEngComm :SAEngLang )
8 ObjectPropertyAssertion( dul:hasLocation :SAEngComm :SouthAfrica )
9 AnnotationAssertion( mola:hasLanguageTag :SAEngLang :SAEngLangTag )
10 AnnotationAssertion( rdfs:label :SAEngLangTag "en-ZA" )
```

Listing 3: OWL fragment for the South African English Viewpoint (VP2)

An example OWL fragment to model VP3 (South African Indian English) is shown in Listing 4, where, if we recall the definition, this is a variety of English spoken in South Africa by South Africans of Indian descent. This viewpoint follows similarly to Listings 2 and 3. In Line 3, Ethnolect is selected instead of VP2's Dialect, where "ethnolect" pertains to a variety of language associated with an ethnic group. In Line 5, there is a class assertion for Ethnicity, a DemographicVariable. This demographic variable is associated with the community asserted in Line 2 using the hasQuality object property in Line 11. The same demographic variable is associated with Population, a social attribute of type CollectiveCulturalExpression.

```
1 ClassAssertion( multi:CommunitySpecificViewpoint :SAIndEngViewpoint )
2 ClassAssertion( multi:LinguisticCommunity :SAIndEngComm )
3 ClassAssertion( mola:Ethnolect :SAIndEngLang )
4 ClassAssertion( multi:IndependentCountry :SouthAfrica )
5 ClassAssertion( multi:Ethnicity :SAIndDescent )
6 ClassAssertion( multi:Population :SAPopulation )
7 ObjectProper...tion( dul:unifies :SAIndEngViewpoint :SAIndEngComm )
8 ObjectProper...tion( dul:isUnifiedBy :SAIndEngComm :SAIndEngLang )
9 ObjectProper...tion( dul:hasLocation :SAIndEngComm :SouthAfrica )
10 ObjectProper...tion( dul:hasQuality :SAIndEngComm :SAIndDescent )
11 ObjectProperty...tion( dul:hasRegion :SAIndDescent :SAPopulation )
```

Listing 4: OWL fragment for the South African Indian English Viewpoint (VP3)

Moving onto VP4, we give an example OWL fragment to model a viewpoint by region, in this case, South Africa, shown in Listing 5. This viewpoint is still a CommunitySpecificViewpoint, however, membership of this community is no longer driven by language, and instead, membership is based on some political tie, in this instance, the laws applicable to a country. If a viewpoint pertained to a political party, then that would be a CommunityOfIdentity, as membership is autonomous.

```
1 ClassAssertion( multi:CommunitySpecificViewpoint :SAViewpoint )
2 ClassAssertion( multi:PoliticalCommunity :SAComm )
3 ClassAssertion( multi:IndependentCountry :SouthAfrica )
4 ObjectPropertyAssertion( dul:unifies :SAViewpoint :SAComm )
5 ObjectPropertyAssertion( dul:hasLocation :SAComm :SouthAfrica )
```

Listing 5: OWL fragment for the South Africa region Viewpoint (VP4)

An example OWL fragment to model VP5, a viewpoint deemed to be language-independent is shown in Listing 6. This viewpoint is represented simply in Line 1.

```
1 ClassAssertion( multi:CommunityAgnosticViewpoint :LangIndViewpoint )
```

Listing 6: OWL fragment for the language-independent Viewpoint (VP5)

VP6 and the notion of 'Western perspective' is harder to represent, in part due to there being no firm definition. A simple representation was opted for in Listing 7, specifying that the viewpoint is community-specific, and the community is political (not linguistic). The location of this community is specified using MetaphoricalLocation, a sub-class of Place. This community is characterised by the concept `:BioMedicine`, the name for the Western medicine system.

```
1 ClassAssertion( multi:CommunitySpecificViewpoint :WestViewpoint )
2 ClassAssertion( multi:PoliticalCommunity :WesternComm )
3 ClassAssertion( multi:MetaphoricalLocation :TheWest )
4 ClassAssertion( multi:MedicineSystem :BioMedicine )
5 ObjectPropertyAssertion( dul:unifies :WestViewpoint :WesternComm )
6 ObjectPropertyAssertion( dul:hasLocation :WesternComm :TheWest )
7 ObjectPro...tion( dul:isCharacterizedBy :WesternComm :BioMedicine )
```

Listing 7: OWL fragment for the language-independent Viewpoint with Western perspective (VP6)

5. Discussion and Future Work

In the preceding sections, we have detailed viewpoints specific to a community of speakers unified by language (VPs 1–3). We have also considered viewpoints from the perspective of a political community, where the notion of a political community is broad: a physical territory associated with a government for which there is rigid membership (VP4) to several countries across continents, unified by a political ideology (VP6). Each viewpoint is distinguished by a selection of attributes applicable to its type. For a viewpoint of a linguistic community, these attributes are qualities, each a demographic variable pertaining to the individual or the community. Each demographic variable is in turn within the 'region' of a cultural expression. An example of a demographic variable is Income which is a selection from a range of values in IncomeScale, and IncomeScale is a cultural expression of a social class. For a political community, although it was intended to use both demographic variables and cultural expressions as defining attributes, for VP6 this was not suitable. Instead, System was added as a sub-class of DOLCE's Concept, for which there are two sub-classes: LegalSystem, and MedicineSystem. Using the isCharacterizedBy object property, a community can be associated with a system.

Not discussed were those communities of identity, listed in *Axioms 26–29*. SubcultureCommunity is a culture within a culture, and an example thereof is the Goth sub-culture in South Africa, for which there is a shared identity of music, dress, and other creative expressions [19]. A VirtualCommunity pertains to online communities, for which traditional geographic and political boundaries may not apply, and instead

communities are bounded by topic of discussion or medium, where 'Tiktokers' is an example of the latter. LocationBasedCommunity refers to a community bounded by a specific location, such as 'New Yorkers'. Finally, a ReligiousCommunity pertains to a community for which membership is based on a shared religion. These communities are quite distinct. However, another community to consider, is that of immigrants to a country who have learned the local language, for which there is an overlap between community of identity and linguistic community. As future work, these and more community types will be identified and then modelled using MULTI, adapting the current model where necessary. The viewpoints considered in this paper have been more implicit. An example of explicit viewpoints is the viewpoint of the colonised, in contrast to the coloniser, both introduced briefly in Section 1. How to model these types of viewpoints will also have to be considered.

To contextualise MULTI, the broader aim is to localise an ontology to a viewpoint of a specific community, by first identifying conceptual differences, and then refactoring the axioms where necessary. The purpose of MULTI is to represent each of these viewpoints. Although the focus is on communities unified by language, MULTI has been developed so as to support other types of communities, as pertains to the adapted definition of ontology localisation in [20]. Regarding the ontology development aspects of MULTI, for methodology, the micro-level approach [21] was followed, with the author as the only ontologist. At time of writing, the F1, F2, A1, I1, and R1.1 principles of FAIR [22] are supported, with other outstanding principles (notably F4) to be addressed as future work.

6. Related Work

There are two main approaches to identifying a language in an ontology: use a language tag for annotations and other data property values, or represent a natural language in an ontology and use its URI when identifying the language of some thing. For the former, an ISO 639 language code [23] can be used, of which there are 7 917 codes at present. However, it is not possible to represent lesser-known dialects and other lects (such as sociolects) using a language code. Alternatively, a custom language tag that conforms to BCP 47 [3] can be used. BCP 47 provides a schema for parsing, but the identification of any custom sub-tags still requires human inspection to decode the string. The string also has length limitations, which makes it difficult to represent lects that are more nuanced in their differences, such as location identified by polygon region. There are other language registries which cover a wider range of languages, however each are problematic, where Gillis-Webber et al. has previously discussed the limitations of each [11].

To represent a natural language in an ontology, many ontologies have been developed for this purpose, of which we mention three. The Model of Language Annotation (MOLA) [11] models languages sufficiently against its identified competency questions but there is no ability to model aspects of a community and their culture related to a language. Lexvo.org is an ontology for language-related information [24], but its support for the description of a natural language is limited. Lingvoj Ontology [25] provides for the representation of language in relation to people, limited mostly to reading, speaking, writing, and understanding skills that a person or groups of persons may have.

Lastly, several ontologies pertaining to culture in the Linked Open Vocabularies [26] repository were also reviewed. Due to space constraints, none are listed here, but of those

that were able to load in Protégé, none were suitable for use as they were limited to the modelling of cultural events or artifacts. Of all the language, linguistic, and cultural ontologies that were reviewed, none met all the requirements that had been identified for the representation of viewpoints from the perspective of a community and their culture.

7. Conclusion

In conclusion, we have proposed an ontology to model multiple viewpoints. Each viewpoint is associated with a community, and the identity of that community may be specific to a language or political ideology, or some other cultural expression. A generalised viewpoint, agnostic to any one particular community, is also able to be modelled. An ontology is then annotated with its applicable viewpoint, thus making the view of the ontology explicit. When localising an ontology from the view of one community to another, this is the first step in the ontology transformation process.

Acknowledgments

This work was financially supported by Hasso Plattner Institute for Digital Engineering through the HPI Research School at UCT.

References

[1] W3C OWL Working Group. OWL Web Ontology Language Use Cases and Requirements: W3C Recommendation 10 February 2004. World Wide Web Consortium; 2004. Online; accessed: 2023, April 28. Available from: `https://www.w3.org/TR/webont-req/`.

[2] Hirst G. Overcoming linguistic barriers to the multilingual semantic web. In: Buitelaar P, Cimiano P, editors. Towards the Multilingual Semantic Web: Principles, Methods and Applications. Springer Berlin Heidelberg; 2014. p. 3-14.

[3] Phillips A, Davis M. BCP 47, RFC 5646: "Tags for Identifying Languages". IETF; 2009. Online; accessed: 2021, June 10. Available from: `https://www.rfc-editor.org/info/bcp47`.

[4] "viewpoint, n.". Oxford English Dictionary Third Edition; 2016. Online; accessed: 2023, January 15. Available from: `https://www.oed.com/view/Entry/223315?redirectedFrom=viewpoint`.

[5] Borgo S, Ferrario R, Gangemi A, Guarino N, Masolo C, Porello D, et al. DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering. vol. 17; 2022. p. 45-69.

[6] Mini BM, editor. The Greater Dictionary of isiXhosa: K to P. vol. 2. IsiXhosa National Lexicography Unit, University of Fort Hare; 2003.

[7] Mesthrie R. A Dictionary of South African Indian English. UCT Press; 2011.

[8] Getting married under customary law; 2011. Online; accessed: 2022, December 12. Available from: `https://www.justice.gov.za/services/getting-married-cusmar-law.html`.

[9] "culture, noun". Merriam-Webster Dictionary; 2023. Online; accessed: 2023, January 15. Available from: `https://www.merriam-webster.com/dictionary/culture`.

[10] Ontology:DOLCE+DnS Ultralite; 2010. Online; accessed: 2023, January 24. `http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite`.

[11] Gillis-Webber F, Tittel S, Keet M. A Model for Language Annotations on the Web. In: Villazón-Terrazas B, Hidalgo-Delgado Y, editors. Knowledge Graphs and Semantic Web. 1st Iberoamerican Conference, KGSWC 2019, Villa Clara, Cuba, June 23-30, 2019, Proceedings. vol. 1029 of Communications in Computer and Information Science. Springer, Cham; 2019. p. 1-16.

[12] Cimiano P, McCrae JP, Buitelaar P. Lexicon Model for Ontologies: Community Report. Ontology-Lexicon Community Group under the W3C Community Final Specification Agreement (FSA); 2016. Available from: `https://www.w3.org/2016/05/ontolex/`.

[13] Cimiano P, Buitelaar P, McCrae J, Sintek M. LexInfo: A declarative model for the lexicon-ontology interface. Journal of Web Semantics. 2011 03;9(1):29-51.

[14] Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L. Sweetening ontologies with DOLCE. In: Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. EKAW 2002. vol. 2473 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002. p. 166-81.

[15] "grammar". Encyclopedia Britannica; 2022. Online; accessed: 2023, January 15. Available from: `https://www.britannica.com/topic/grammar`.

[16] Crystal D, Robins R. "language". Encyclopedia Britannica; 2023. Online; accessed: 2023, January 15. Available from: `https://www.britannica.com/topic/language`.

[17] Social Norms. Stanford Encyclopedia of Philosophy; 2018. Online; accessed: 2023, January 12. Available from: `https://plato.stanford.edu/entries/social-norms`.

[18] "narrative, noun". Cambridge Dictionary; 2023. Online; accessed: 2023, January 15. Available from: `https://dictionary.cambridge.org/dictionary/english/narrative`.

[19] The dark, weird world of the Goths; 2004. Online; accessed: 2023, March 19. Available from: `https://www.iol.co.za/news/south-africa/the-dark-weird-world-of-the-goths-212417`.

[20] Cimiano P, Montiel-Ponsoda E, Buitelaar P, Espinoza Mejía M, Gomez-Perez A. A Note on Ontology Localization. Applied Ontology. 2010 04;5(2):127—137.

[21] Keet CM. An introduction to ontology engineering. vol. 20 of Computing. UK: College Publications; 2018.

[22] Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, et al. FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence. 2020 01;2(1-2):10-29.

[23] ISO. Language codes - ISO 639; n.d. Online; accessed: 2023, April 28. `https://www.iso.org/iso-639-language-codes.html`.

[24] de Melo G. Lexvo.org: Language-related Information for the Linguistic Linked Data Cloud. Semantic Web. 2015 01;6(4):393-400.

[25] Bernard Vatant. The Lingvoj Ontology (lingvo); n.d. Online; accessed: 2023, March 19. `https://lov.linkeddata.es/dataset/lov/vocabs/lingvo`.

[26] Vandenbussche PY, Atemezing G, Poveda-Villalón M, Vatant B. Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. Semantic Web. 2017 01;8:437-52.