Natural Language Engineering **17** (3): 283–309. © Cambridge University Press 2010
doi:10.1017/S1351324910000082

Enriching ontologies with multilingual information†

E. MONTIEL-PONSODA[1], G. AGUADO DE CEA[1], A. GÓMEZ-PÉREZ[1] and W. PETERS[2]

[1]*Ontology Engineering Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid, Spain*
e-mails: {emontiel, lupe, asun}@fi.upm.es
[2]*Sheffield Natural Language Processing Group, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK*
e-mail: w.peters@dcs.shef.ac.uk

(*Received 1 August 2009; revised 1 March 2010; accepted 1 May 2010; first published online 9 June 2010*)

Abstract

This paper presents a novel approach to ontology localization with the objective of obtaining multilingual ontologies. Within the ontology development process, ontology localization has been defined as the activity of adapting an ontology to a concrete linguistic and cultural community. Depending on the ontology layers – terminological and/or conceptual – involved in the ontology localization activity, three heterogeneous multilingual ontology metamodels have been identified, of which we propose one of them. Our proposal consists in associating the ontology metamodel to an external model for representing and structuring lexical and terminological data in different natural languages. Our model has been called Linguistic Information Repository (LIR). The main advantages of this modelling modality rely on its flexibility by allowing (1) the enrichment of any ontology element with as much linguistic information as needed by the final application, and (2) the establishment of links among linguistic elements within and across different natural languages. The LIR model has been designed as an ontology of linguistic elements and is currently available in Web Ontology Language (OWL). The set of lexical and terminological data that it provides to ontology elements enables the localization of any ontology to a certain linguistic and cultural universe. The LIR has been evaluated against the multilingual requirements of the Food and Agriculture Organization of the United Nations in the framework of the NeOn project. It has proven to solve multilingual representation problems related to the establishment of well-defined relations among lexicalizations within and across languages, as well as conceptualization mismatches among different languages. Finally, we present an extension to the Ontology Metadata Vocabulary, the so-called LexOMV, with the aim of reporting on multilinguality at the ontology metadata level. By adding this contribution to the LIR model, we account for multilinguality at the three levels of an ontology: data level, knowledge representation level and metadata level.

† This work is supported by two European Commission's projects, *Lyfecycle support for networked ontologies* (NeOn) (FP6-027595) and *Multilingual Ontologies for Networked Knowledge* (Monnet) (FP7-248458), the Spanish National Project *GeoBuddies* (TSI-2007-65677C02) and an I+D fellowship from the Universidad Politécnica de Madrid. We would also like to thank the anonymous journal reviewers.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

1 Introduction

Multilinguality in ontologies is nowadays demanded by institutions worldwide having a large number of resources in different natural languages. One of the institutions that has explicitly expressed the need for structuring the great amounts of information it has in different natural languages is the Food and Agriculture Organization (FAO) of the United Nations.[1] The FAO works with six official languages (English, French, Spanish, Arabic, Chinese and Russian), but, in fact, it manages information in more than fifteen languages. This impressive number gives just a rough idea of what an institution of this type has to do to provide solutions to vital issues such as nutrition, agriculture, forestry or fisheries, if it wants to customize the resulting information to users in places as far apart as Island or Thailand. If, for example, the Thai rice harvest is threatened by a severe plague, this may have to be reported in English and Italian at a meeting of the FAO at its headquarters in Rome. For this purpose, linguist experts (terminologists, translators and interpreters) at the FAO have to analyse documents in Thai, and together with domain experts define language equivalences for the type of rice and the plague agent involved in the disaster. This dynamic process has to ensure the creation of new concepts and terms in the other languages. Once this is achieved, the new information has to be updated in all resources dealing with agricultural issues. However, updating is not an easy task considering that the FAO manages a great variety of heterogeneous multilingual linguistic resources, such as

- glossaries: FAOTERM, the Fisheries Glossary,[2] the Aquaculture Glossary[3] and Globefish[4] (in English, French, Spanish, Arabic and Chinese);
- thesauri: AGROVOC (in English, French, Spanish, Arabic, Chinese, Czech, Japanese, Portuguese, Thai, Slovak, Lao, Hindi, German, Italian, Hungarian; and under development for Marati, Polish, Korean, Farsi, Malay, Amharic, Catalan and Russian), ASFA[5] (in English, French and Spanish);
- databases: AGRIS,[6] FIGIS.[7]

It should be noticed that resources have different levels of granularity regarding the type and quantity of information. For example, AGROVOC contains semantically related terms in more than ten languages, whereas FAOTERM contains the searched term and its definition in the six official languages of the FAO plus Italian and Latin. As for the quantity of information, some resources are more complete in the traditional languages of the FAO, but show important gaps in the others.

Last but not least, one of the crucial issues that multilingual organizations have to deal with is conceptualization mismatches. This means that some concepts or

[1] At the time of writing this paper, the FAO (http://www.fao.org) was participating as a Use Case in the NeOn project. For more information see http://www.neon-project.org
[2] http://www.fao.org/fi/glossary/default.asp
[3] http://www.fao.org/fi/glossary/aquaculture
[4] http://www.globefish.org
[5] http://www4.fao.org/asfa/asfa.htm
[6] http://www.fao.org/AGRIS
[7] http://www.fao.org/fishery/en

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

categorizations of reality are relevant in some cultures but not in others, i.e. some cultures give names to precise bits of reality, for which other cultures have no specific names. For example, the Thai language has different lexicalizations for rice according to its cooking stage: Khao dip (rice not cooked), khao suk (rice cooked), khao niew (sticky rice), khao chao (rice not sticky). These categorizations should be included in the FAO resources, as well as equivalents and definitions in the rest of languages.

The scenario here described aims at showing an illustrative example of the need for semantically organizing and customizing multilingual information within international organizations. As a consequence of this demand, the methodology designed within the NeOn project for a collaborative development of ontologies (Suárez-Figueroa and Gómez-Pérez 2008) has included a new scenario regarding the localization of ontologies. Ontology localization is understood as 'the adaptation of an ontology to a concrete linguistic and cultural community'.[8] However, the ontology localization activity can result in heterogeneous multilingual ontology metamodels depending on the ontology layers implied in the localizing activity and on the depth these layers undergo localization (cf. Section 2).

The novel approach to ontology localization we present in this paper intends to solve the issues of heterogeneity, distribution and cultural specificities by associating ontologies to a linguistic model that integrates the necessary multilingual information for ontology localization. In particular, our aim is to localize ontologies by associating them with a model called Linguistic Information Repository (LIR) (cf. Section 3), whose main features are (1) independence of the ontological model, (2) interoperability with existing standards representing lexical and terminological knowledge, (3) provision of a subset of linguistic descriptions to account for the linguistic realization of a domain ontology in different natural languages, (4) representation of term variants within one language and cultural specificities among different languages, (5) unified access to aggregated multilingual information related to ontology elements and (6) accessibility by committing to the OWL[9] as representational language, and the tool support it entails.

The rest of the paper is structured as follows: Section 4 shows the evaluation of the LIR against the FAO requirements. The representation modality adopted by the LIR is compared against the well-known representation modality offered by the OWL and Resource Description Framework Schema (RDF(S)) labelling functionality in Section 5. Section 6 is devoted to the LexOMV, an Ontology Metadata Vocabulary (OMV) extension that reports about multilinguality at the metadata level. And finally, Section 7 summarizes the main contributions of the paper.

[8] In this definition, language and culture are understood as an indivisible unit, since the language is a mirror of the cultural and historical heritage of a certain society. However, it is explicitly mentioned here to highlight the fact that the different categorizations of reality that cultures make are reflected in their linguistic realizations.

[9] OWL stands for Web Ontology Language, and embraces a family of knowledge representation languages for editing ontologies, endorsed by the World Wide Web Consortium or W3C (http://www.w3.org/)

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

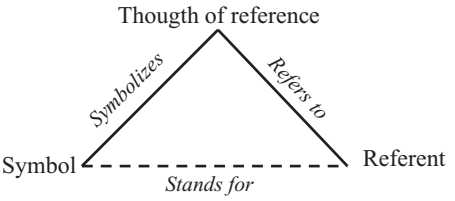


Fig. 1. Ogden and Richards' triangle.

2 Main trends in modelling multilinguality in ontologies

Ontologies, as any other system for representing knowledge, make use of 'thoughts of reference', also known as concepts, to refer to the real world. When comparing ontologies to linguistic systems, we may state that both have three main components: *signs* or symbols used to designate *concepts* or thoughts of mind, which refer to *phenomena in the real world*. In linguistic semantics this idea was represented as a triangle by Ogden and Richards (1923) (cf. Figure 1). A few years later, Morris (1938) in his approach to semiotics made a similar distinction by dividing the sign into sign vehicle (syntax), designatum (semantic) and interpreter (pragmatic), stating that the understanding of the world was dependant of the viewpoint of the interpreter. Coming back to the ontology field, multiple authors have tackled this topic, and we can basically distinguish six layers in any ontology, as summarized in (Barrasa 2007), namely

(I) lexical layer: characters and symbols that make up the syntax (ASCII encoding, UNICODE, etc.)
(II) syntactic layer: structure of characters and symbols, i.e. the grammar. It embraces different representation languages (e.g. RDF(S)[10], OWL, etc.)
(III) representation paradigm layer: paradigm followed in the representation of the ontology (frames, semantic networks, Description Logics, etc.) that allows for certain ways of expressing and structuring knowledge
(IV) terminological layer: terms or labels selected to name ontology elements
(V) conceptual layer: related to conceptualization decisions, such as granularity, expressiveness, perspective, etc.
(VI) pragmatic layer: final layout of the model according to the user's needs

According to this, we may state that only the terminological, conceptual and pragmatic layers are involved in the ontology localization activity. The terminological layer plays a decisive role in the localization activity since it is closely related to the names given to the different ontology elements. As a result of this activity, ontology labels will be expressed in more than one natural language. Regarding the conceptual layer, certain ontologies may require the adaptation of their conceptual structure in order to fit in the thoughts of reference of a specific linguistic and cultural community. As for the pragmatic layer, the needs of the final application

[10] RDF(S) stands for Resource Description Framework Schema, and it is a knowledge representation language for the authoring of ontologies, also endorsed by the W3C (see footnote 9)

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

will determine the type and quantity of linguistic information that is to be related to the ontology. The rest of the layers – lexical, syntactic and representation paradigm layers – should not be so strongly affected by the localization activity.

Up to now, the number of multilingual ontologies is still quite small compared to the total amount of ontologies available in the Web (cf. OntoSelect or Watson[11]). According to the state of the art, there are three main ways of obtaining a multilingual ontology-based system, depending on the layer(s) involved in the Localization Activity (each modality will be explained in more detail in the following sections):

- Including multilingual data in the ontology metamodel. This implies localization at the terminological layer since the ontology conceptualization remains unmodified.
- Combining the ontology metamodel with a mapping model. This allows localization at the conceptual layer since conceptualizations in different languages are mapped to each other.
- Associating the ontology metamodel to a multilingual linguistic model. Localization is performed at the terminological layer, although conceptual layer adaptations are also foreseen.

The appropriateness of the modelling modality will be principally determined by the requirements of the final task or application in which the ontology is to be used. It is also important to note at this stage that the research presented in this paper is the result of theoretical reflections, and has only one experimental application against the requirements of the FAO use case within the NeOn project, as already outlined in the introduction.

In the following subsections we offer a more exhaustive analysis of these three modalities for modelling multilinguality, as well as a brief overview of some of the most relevant applications that follow these approaches. We also provide a description of the main strengths and weaknesses of each option in order to support our decision of applying the third modelling modality in the approach followed in this research work (see Section 2.3).

2.1 Including multilingual data in the ontology metamodel

Including multilingual data in the ontology metamodel is currently the most widespread modelling modality within the ontological community. It consists of making use of the labelling facility of RDF(S) and OWL ontology representation languages.[12] This relies on two RDF(S) properties, `rdfs:label` and `rdfs:comment`, that can be used to define labels and descriptions in natural language for ontology elements. This system allows localization at the terminological layer, as labels for ontology classes can be expressed in various natural languages (see Figure 2). This modelling modality is described in more detail in Section 5.

[11] These are Semantic Web search engines to look for ontologies according to different criteria.
[12] http://www.isi.edu/in-notes/rfc3066.txt

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

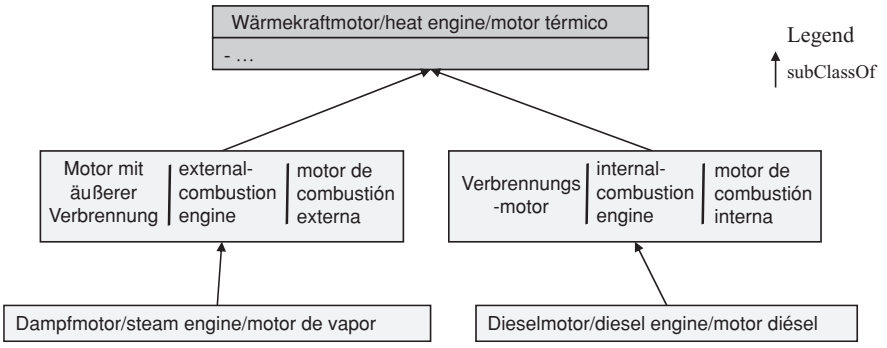


Fig. 2. Multilingual information included in Ontology Metamodel.

2.1.1 Disadvantages

The RDF(S) properties related to ontology elements make up a list of unconnected linguistic descriptions. It is not possible to establish semantic relations among labels and their corresponding definitions. Labels in different languages are understood as exact equivalents, but this is rarely the case (Edmonds and Hirst 2002). Moreover, this labelling system makes it hard to specify possible linguistic distinctions between labels in the same language (term variants) or in different languages. Besides, it is not possible to keep track of the provenance of the linguistic data although it may be relevant for evaluating the quality of the information. Neither is it possible the performance of complex operations with linguistic elements, since no semantic relations exist among them.

2.1.2 Advantages

Labels can be integrated in the ontology in as many languages as the user wishes. This model has proven to be more suitable for highly specialized domain ontologies, e.g. in engineering or technical domains, since domain-specific knowledge is more prone to be shared among different linguistic and cultural communities, and the equivalence relation among labels in different languages is deemed acceptable.

2.2 Combining the ontology metamodel with a mapping model

According to this approach, there are various modelling ways depending on the mapping arity and the graph form. The two main representation forms are:

- Binary mappings in an orthogonal graph. In this case, each monolingual ontology organizes knowledge of a certain culture, and is mapped to the rest of ontologies in a pairwise fashion.
- Binary mappings in a radial graph. In this option, monolingual ontologies are mapped to each other through an interlingua consisting of a set of common concepts for establishing equivalences (see Figure 3).

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

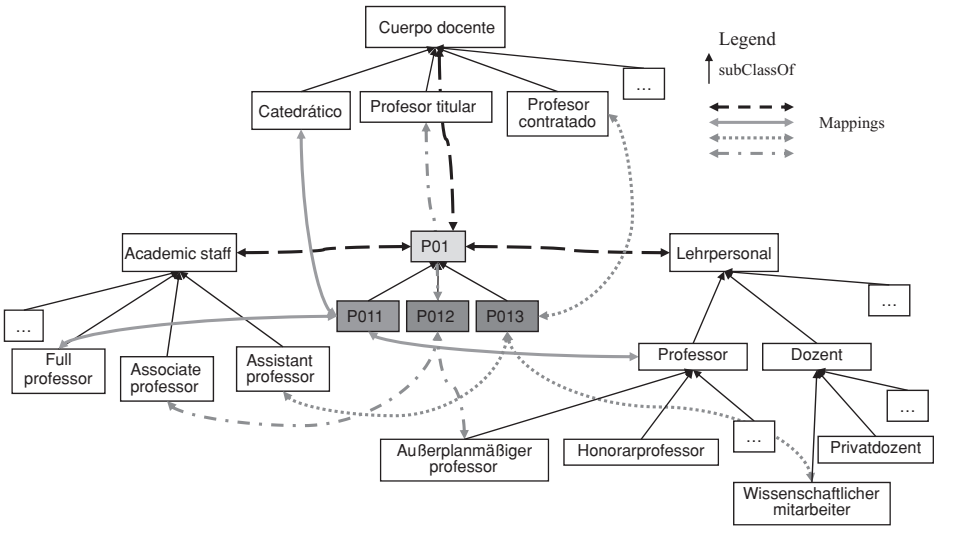


Fig. 3. Binary mappings in a radial graph.

The most representative application following this approach is EuroWordNet (EWN) (Vossen 2002).[13] This multilingual general lexicon consists of monolingual wordnets, each one reflecting the linguistic and cultural specificities of a certain language, linked to each other through an interlingual set of common concepts that caters for equivalences among ontologies. The crucial issue in the development of such multilingual models is the establishment of mappings among concepts in the different conceptualizations. Being aware of this problem, wordnet developers took as starting point either the set of interlingual common concepts (structured in a language-independent way by a Top Ontology and a Domain Ontology), or the English wordnet (WordNet1.5) (Fellbaum 1998), in order to guarantee a minimal level of compatibility between the independent wordnets. The risk of this second option, as the same authors anticipated, was that the resulting conceptualizations could be biased by the English one (Vossen 2004). EWN provides information about nouns, verbs, adjectives and adverbs organized in synsets (a set of words with the same part-of-speech that can be interchanged in a certain context) (Fellbaum 1998). Synsets are normally accompanied by glosses that describe their sense in a certain context.

A similar approach was followed by the multilingual general lexicon SIMPLE (Lenci *et al.* 2000). This application also consists of lexicon ontologies developed for each language and linked to each other. In fact, the common set of concepts identified in EWN was used as a core set of senses in order to provide a cross-language linkage. In this lexicon, however, the quantity and granularity of morphological and syntactic information (from the PAROLE lexicon Lenci *et al.* 2000) is much larger. For the

[13] Currently, the interest of mapping or aligning ontologies documented in different natural languages following this approach is increasing as reported in (Euzenat *et al.* 2009).

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

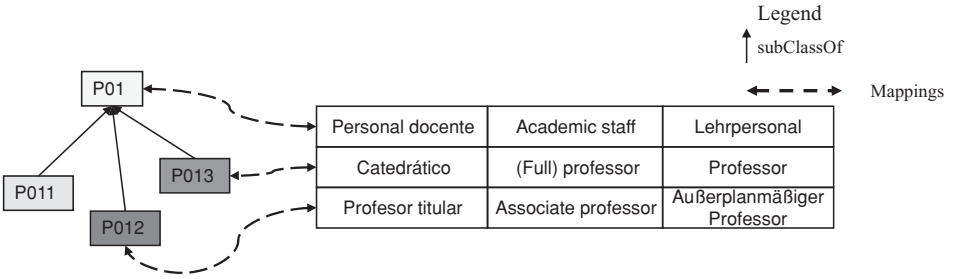


Fig. 4. Ontology metamodel associated with a multilingual linguistic model.

time being, these models are out of the scope of our research as they are not intended for actually providing multilinguality to domain ontologies already available.

2.2.1 Disadvantages

A great effort is needed to conceptualize the same domain in different natural languages. Three different types of expertise are required for that endeavour: domain expertise, linguistic expertise and ontology engineering expertise. In the case that domain ontologies already exist in different languages, the establishment of alignments among conceptualizations in different languages is by no means trivial, since each conceptualization reflects the cultural specificities of each language. This makes the linkage among conceptualizations very difficult, thus resulting in one conceptualization biasing the others. The quantity of linguistic information embedded in the ontology is often limited to labels and definitions associated with ontology classes making use of the RDF(S) properties.

2.2.2 Advantages

This option enables independent conceptualizations in each language, which may better capture the specificities of each culture. This approach may be more suitable for modelling ontologies in those knowledge domains highly dependent on the culture in which they have been conceived, such as the judiciary.

2.3 Associating the ontology metamodel with a multilingual linguistic model

In this modelling option (Figure 4), the elements of the ontology have links to linguistic data stored outside the ontology. The model for representing and organizing the linguistic information can be a data base, as in GENOMA-KB[14] or OncoTerm,[15] or an ontology, as in the case of LingInfo (Buitelaar, Sintek and Kiesel 2006), LexOnto (Cimiano *et al.* 2007) or a new model that merges both, LexInfo (Buitelaar *et al.* 2009).

[14] http://genoma.iula.upf.edu:8080/genoma
[15] http://www.ugr.es/ oncoterm/alpha-index.html

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

In this approach, conceptual and terminological layers are kept separate, and the localization activity is mainly carried out at the terminological layer. However, the ontology conceptualization layer can also undergo modifications, such as the creation of language specific ontology modules, in order to meet localization needs. The distinguishing aspect among the applications and approaches that follow this modelling modality is determined by the kind of linguistic classes that make up the different models. Depending on the linguistic needs of the final task or application, some models will be more suitable than others.

The LingInfo model focuses on the representation of the morphological and syntactic structures (segments, head and modifiers) of a term. LexOnto goes one step further in that it pursues to represent linguistic realizations of ontology elements. This model builds on the notion of subcategorization frames, i.e. linguistic predicate–argument structures that represent how an ontology label (noun, adjective or verb) is syntactically realized in a certain linguistic structure. These two models have been aligned because, according to authors, both pursue the same objective though putting the emphasis on complementary aspects, namely to provide 'more expressive lexicon models for ontologies' (Buitelaar *et al.* 2009). In addition, they have been made interoperable with the Lexical Markup Framework (LMF) ISO standard, a metamodel for describing computational lexicons (Francopoulo *et al.* 2006). These models have been designed with the aim of improving tasks such as ontology learning or ontology population from text, which has determined the set of linguistic information captured in the model.

The Human Genome Knowledge Base GENOMA-KB (Cabré *et al.* 2004) or the OncoTerm data base pursue rather terminological or translational objectives by linking a terminological multilingual database to highly specialized ontologies of the biology and oncology domains, respectively. Both approaches are built upon the OntoTerm[16] terminological management system. The linguistic information associated to ontology concepts is limited to terms and definitions in different languages accompanied by basic morphological information (part-of-speech, gender and number), and examples of sentences in which these terms appear.

The model we propose is a composite form of both types of approaches. On the one hand, the main objective of the LIR (Peters, Montiel-Ponsoda and Aguado de Cea 2007; Montiel-Ponsoda *et al.* 2008; Montiel-Ponsoda and Peters 2008) is to provide multilinguality to monolingual domain ontologies, in the line with GENOMA-KB and OncoTerm. On the other hand, the substantial quantity of linguistic information and its organization as an ontology[17] go more in line with the most recent proposals for linguistically grounding ontologies (Buitelaar *et al.* 2009). What differentiates the LIR from the rest of the approaches presented is the type of linguistic information associated to ontology elements. The set of linguistic classes that composes the LIR has been expressly designed to account for different lexicalizations within the same and across languages and conceptualization mismatches among different languages. However, morphosyntactic and other linguistics aspects are also captured but with

[16] http://ontoterm.com
[17] See (Montiel-Ponsoda and Peters 2008) for the OWL code.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

a lesser level of detail. Extensions of the model with further linguistic classes can be obtained from other models since the LIR also builds on the LMF reference ISO standard for computational lexicons.

2.3.1 Disadvantages

Since there is just one conceptualization, it is not as flexible as the model described in Section 2.2 above, which means that some language specificities can be lost, unless they are captured in an expressive linguistic model, i.e. at the terminological layer, or in specific ontology modules, i.e. at the conceptual layer, if so required by the final application.

2.3.2 Advantages

This type of representation allows the enrichment of domain ontologies with linguistically rich and complex models. Since these are external portable models, they can be associated to any domain ontology, and have been thought to be published with them. The different linguistic categories that compose the model are structured and semantically related. In this sense, it is possible to establish links between lexicalizations, definitions and sources of provenance. Regarding conceptualization mismatches between languages, these can be explicitly captured in the model. If additional linguistic information is required by the final application, these models can be extended thanks to the interoperability established with standard linguistic description models. Finally, linguists or domain experts without ontology development expertise can easily edit the terminological layer without dealing with the ontology.

3 The linguistic information repository

The LIR model is conceived as a hub to interconnect various standard descriptions for linguistic knowledge on the one hand, and ontological concepts on the other. In fact, it adopts a number of data categories[18] for linguistic description from standards in order to guarantee interoperability with existing and proposed standards for the representation and integration of terminological and linguistic knowledge. Its design is mainly based on the core package of the LMF (Francopoulo *et al.* 2006), similarly to the new LexInfo model (see Section 2.3). As already mentioned, LMF is an ISO standard specification to model computational lexicons. According to this standard, a Lexicon comprises lexical entries that are realized by word forms related to the different senses a word can have, as happens in WordNet.[19]

The rationale underlying the LIR is not to design a lexicon for different natural languages and then establish links to ontology concepts, but to associate multilingual linguistic knowledge to the conceptual knowledge represented in an arbitrary domain

[18] ISO 12620, see http://www.ttt.org/clsframe/datcats.html
[19] http://wordnet.princeton.edu/

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

ontology. In the LIR, each lexical entry can be realized by different word forms linked to the same word sense – constrained by the knowledge represented in the ontology concept – although word senses and concepts can not be said to overlap (Hirst 2004). The reason for this is that word senses are tightly related to the particular vision of a language, whereas ontology concepts try to capture objects of the real world in a formal way, and are defined according to expert criteria agreed by consensus. These criteria need not fully reflect the lexical meaning of the natural language label that lexicalizes the concept. In this sense, it could be stated that the LIR goes more in the line of what Pustejovsky (Pustejovsky 1995) defined as Sense Enumeration Lexicon, in which a unique sense is associated to a word string. This theory would not be adequate if our purposes were to design a lexicon for a language, in which all senses of a word should be accounted for. However, we argue that this is a suitable approach to enrich domain ontologies with multilingual information. According to the needs of the final application, LIR could be extended with further linguistic knowledge, such as morphological decomposition and syntactic complementation, as modelled in LMF or LexInfo. This knowledge could be obtained by navigating those models after establishing a connection between them.[20] The LIR also serves the objective of integrating and aggregating multilingual information contained in heterogeneous and distributed lexical sources by guaranteeing a homogeneous access to the information.

In the following, our purpose is to describe in more detail the classes that make up the LIR, as represented in Figure 5. The linguistic information captured in the LIR is organized around the `LexicalEntry` class. A lexical entry is considered a unit of form and meaning in a certain language. Therefore, it is associated to the `Language`, `Lexicalization` and `Sense` classes. A set of related lexicalizations or term variants shares the same meaning (represented by the sense) within the specific context of a certain cultural and linguistic universe. For example, *Food and Agriculture Organization* and *FAO* would be two lexicalizations in the same language linked to the same sense.

Thanks to the expressiveness of the `hasVariant` relation, it would be possible to say that the one is acronym of the other. The `Language` class at the `LexicalEntry` level allows launching searches in which just those lexical entries related to one natural language are shown to the user, thus displaying the ontology in the selected language. The `PartOfSpeech` class is also linked to the `LexicalEntry` class to avoid repetition in the various lexicalizations because all share the same part-of-speech. `Sense` is considered a language-specific unit of intensional lexical semantic description, which comes to fruition through the `Definition` class expressed in natural language. By keeping senses in the linguistic model independent from ontology concepts, we capture cultural and linguistic specificities that may slightly differ from the concept expressed in the ontology. `Sense` is kept as an empty class to interoperate with the above mentioned standards and linguistic representation

[20] For a detailed description of alignments between the LIR and LMF, and other lexical and terminological descriptions see (Peters, Gangemi and Villazón-Terrazas 2010).

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

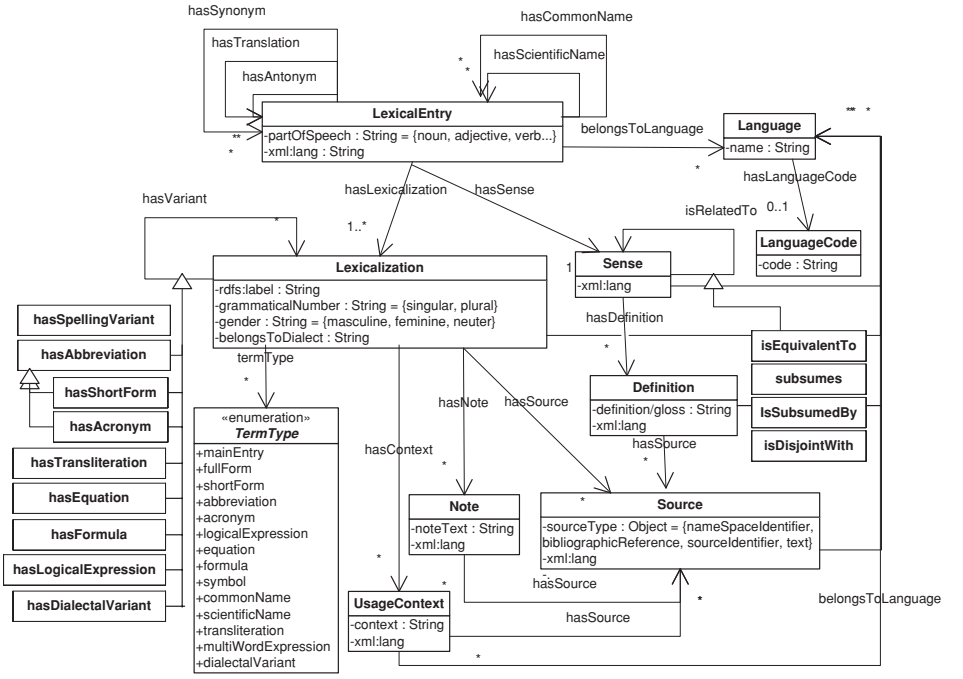


Fig. 5. The LIR model.

models. `Definition` has a pointer to the linguistic resource it has been obtained from. In this way reliability and authority of definitions are guaranteed.

Then, `Lexicalization` is related to its `Source` or provenance, to a `Note` class and to a `UsageContext` class. The `Source` class aims again at being a pointer to the resource where the information has been extracted from. `Note` is here linked to `Lexicalization`, but it could also be linked to any other class in the model. It may include supplemental information; for instance, usage specificities of a certain lexicalization within its language system. By linking `Note` to the `Sense` or `Definition` classes we can make explicit possible differences or nuances among senses within and across languages. Additionally, senses can be related by means of the relation `isRelatedTo`, or any of its specifications: `isEquivalentTo`, `subsumes`, `isSubsumedBy`, or `isDisjointWith`.

The `UsageContext` class provides examples of use (syntactic behaviour, collocational information) of a certain lexicalization in the language system to which it belongs. Finally, lexical semantic equivalences are established among lexical entries within the same language (`hasSynonym` or `hasAntonym`), or across languages (`hasTranslation`). Note that we use the `hasTranslation` label to establish equivalences between lexicalizations in different languages, although it is assumed that words identified as translation equivalents are rarely identical in sense. As Hirst (Hirst 2004) stated, 'more usually they are merely cross-lingual near-synonyms'. Nonetheless, for the practical reason of providing multilinguality, this approach is adopted.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

It remains to say that the LIR is linked to the `OntologyElement` class of the OWL metamodel, thus associating multilingual information with any element of the ontology. Finally, we must refer to the LabelTranslator NeOn plug-in, a translation supporting tool (Espinoza, Gómez-Pérez and Mena 2008) that provides semi-automatically translations for ontology lexicalizations. Currently, the languages supported by the plug-in are Spanish, English and German. Once translations are obtained for the labels of the original ontology, they are stored in the LIR. However, if the system does not support the language combination in which we are interested, we can still use this system to take advantage of the LIR *application programming interface* or API implemented in the NeOn Toolkit. In this sense, we can manually introduce the linguistic information we need. See Section 5 for some snapshots of the LIR API.

4 Evaluation of the LIR against FAO requirements

In this section our aim is to describe the requirements of the FAO as regards the representation of multilingual information in ontologies, which are the ones that guided us in the development of the LIR. As the FAO is a good exponent of an international organization with multilingual needs, we believe that its requirements can be representative of other international organizations. Since the research on models to linguistically enrich ontologies is incipient and the existing ones have been also created to cover particular needs or for specific tasks (see Buitelaar *et al.* 2006), it is difficult to perform a standard evaluation. In our case, we carried out an *ad hoc* evaluation against the multilinguality requirements of the FAO use case, as reported in Section 4.1 by means of some examples.

The FAO, in its commitment for improving information management and communication, is introducing semantic technologies in its information systems.[21] Regarding multilingual data, one of the most used and updated lexical resources within the FAO has traditionally been the AGROVOC thesaurus,[22] defined as a controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains. In 2003, the FAO initiated the development of the AGROVOC Concept Server (CS) (Liang *et al.* 2008), an ontology created *ad hoc* from the original thesaurus to overcome some of the main deficiencies of thesauri, which are summarized below. Although the CS solved some immediate needs, as reported in (Liang *et al.* 2008), the need for a portable model that would enrich FAO domain ontologies with multilingual information was still present.[23] For this reason the FAO asked for a multilingual representation system that could solve the following thesauri drawbacks:

[21] For example, in the framework of the NeOn project, ontologies have been created to manage information about fisheries for what is known as the Fish Stock Depletion Assessment System (FSDAS). See resulting ontologies in: http://www.neon-project.org/nw/Ontologies

[22] http://aims.fao.org/website/AGROVOC-Thesaurus/sub

[23] A description of the alignments between AGROVOC CS and LIR for an automatic population of LIR with AGROVOC CS data is included in (Peters *et al.* 2009).

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

- Thesaurus relationships (Broader Term (BT), Narrower Term (NT), Related Term (RT), Preferred Term (USE) and UsedFor) fall short of expressing semantic and lexical relations in a refined and precise way
- Thesaurus relationships do not cover all possible associations between terms in the sense that it is not possible to retrieve and distinguish an *acronym* from a *full form* description, a *synonym* from a *translation* or a *scientific name* from a *common name*.
- Thesauri do not allow lexical variants to be specified for dialects or local languages for a geographical region, such as the ones we could find between Spanish used in Spain and Spanish used in Latin America.
- Thesauri do not allow more than one translation per term to be set. According to this, for example, the English term *Field size* can be translated in French as *Taille des parcelles* or *Dimension des parcelles*. In the current AGROVOC thesaurus one of the translations is assigned as the translation of the descriptor, and the other as an associated non-descriptor.

Therefore, the LIR model was created with the purpose of overcoming the limitations of thesauri, on the one hand, and fulfilling the needs of portability and association of multilingual information to domain ontologies, on the other. In the next section we spell out with real examples how the LIR solves FAO multilingual representation need.

4.1 Benefits of the LIR to FAO's needs

The LIR model provides a very granular specification of relationships between elements of an ontology. In particular, it identifies well-defined relationships at the terminological layer used to represent ontological concepts. In FAO, not only could several resources such as AGROVOC or the Concept Server benefit from the LIR paradigm, but also recently developed domain-specialized ontologies have taken advantage of this model. The examples below show how some problems mentioned in Section 4 can be solved by modelling the multilingual information with the LIR model:

- establishment of well-defined relations within lexicalizations in one language,
- establishment of well-defined relations within lexicalizations across languages,
- conceptualization mismatches among different cultures and languages,
- representation of non-native language expressions.

In the following, we illustrate how LIR solves these problems.

4.1.1 Example 1: establishment of well-defined relations within lexicalizations in one language

The example in Figure 6 concerns the establishment of relations among term variants belonging to the same language. Specifically, this case exemplifies the use of various *acronyms* and *full forms* attached to one and the same concept.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

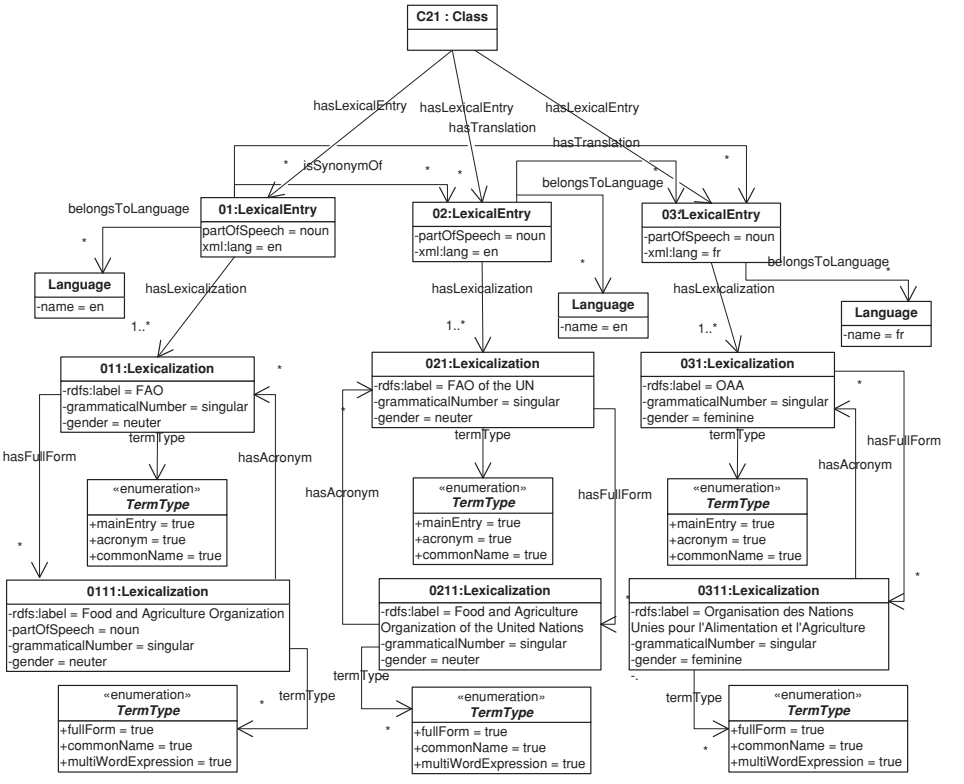


Fig. 6. Representation of acronyms and full forms within a language.

Three lexical entries (`01:LexicalEntry`, `02:LexicalEntry` and `03:LexicalEntry`) are associated with the same concept (`C21:Class`), which means that they are terms that identify one and the same concept. Two lexical entries (`01:LexicalEntry` and `02:LexicalEntry`) belong to English, whereas the third lexical entry (`03:LexicalEntry`) belongs to French. The two English lexical entries are considered synonyms, and both are translations of the French lexical entry. Each lexical entry contains two lexicalizations. For example, `01:LexicalEntry` includes `011:Lexicalization` and `0111:Lexicalization`, whose labels are *FAO* and *Food and Agriculture Organization*, respectively. FAO is the acronym for Food and Agriculture Organization, and, moreover, it is considered the main entry. *FAO of the UN* and *Food and Agriculture Organization of the United Nations* are deemed synonyms of FAO and Food and Agriculture Organization. Both lexical entries (`01:LexicalEntry` and `02:LexicalEntry`) are translations of *OAA* and *Organisation des Nations Unies pour l'Alimentation et l'Agriculture* in the French language.

Thanks to LIR it is possible to retrieve synonyms within the same language associated with the same concept, and distinguish different term types such as acronyms and full forms.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

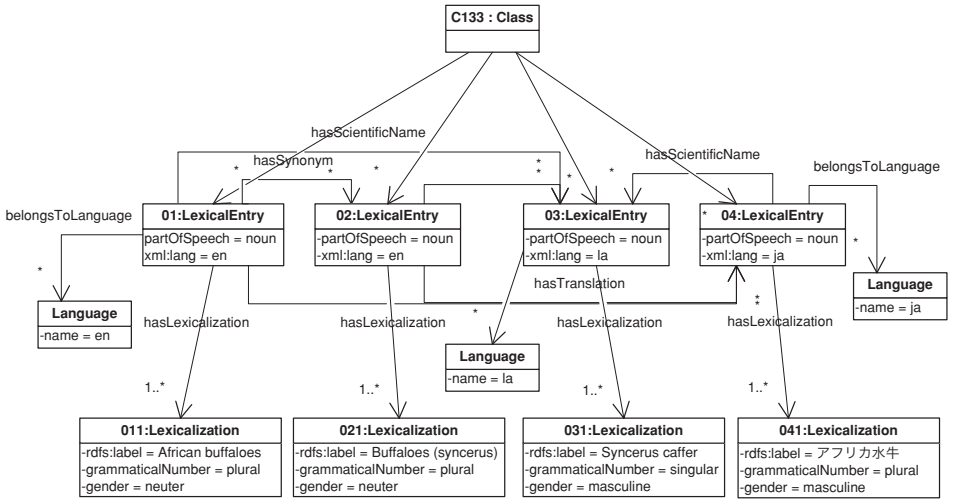


Fig. 7. Representation of scientific names and common names across languages.

4.1.2 Example 2: establishment of well-defined relations within lexicalizations across languages

The second example highlights the possibility given by the LIR model to represent scientific names and use them across languages (scientific names are in Latin and are internationally accepted over scientific communities).

Variants in the same language (e.g. *Buffaloes (syncerus)*) can therefore be connected to the same scientific term, such as the English and Japanese translations. We have illustrated in Figure 7 how the concept *buffaloes* (`C133:Class`) has four lexical entries associated (`01:LexicalEntry`, `02:LexicalEntry`, `03:LexicalEntry` and `04:LexicalEntry`). Two of them belong to the English language and contain synonymous lexicalizations (`011:Lexicalization` and `021:Lexicalization`). Then, we have a lexicalization in Latin that represents the scientific name, and it is accordingly related with the rest of lexical entries by means of the object property `hasScientificName`. Finally, `04:LexicalEntry` belongs to the Japanese language, which is also the common denomination in Japanese of the *Syncerus caffer* scientific name, and, at the same time, the translation of the two lexicalizations in English.

4.1.3 Example 3: conceptualization mismatches among different languages

More often than not, conceptualizations of the same domain coming from different communities show important discrepancies, because the granularity level with which some concepts are understood may not be the same. This results in a mismatch of terminological equivalents. The situation can be summarized in two cases: (a) one in which a culture makes a more fine-grained distinction of a certain reality parcel than the other, or (b) the opposite situation, in which a culture does not make so fine-grained distinctions but remains at a more underspecified level. An example of this has been presented at the introduction in Section 1 regarding the conception of rice in Thailand.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

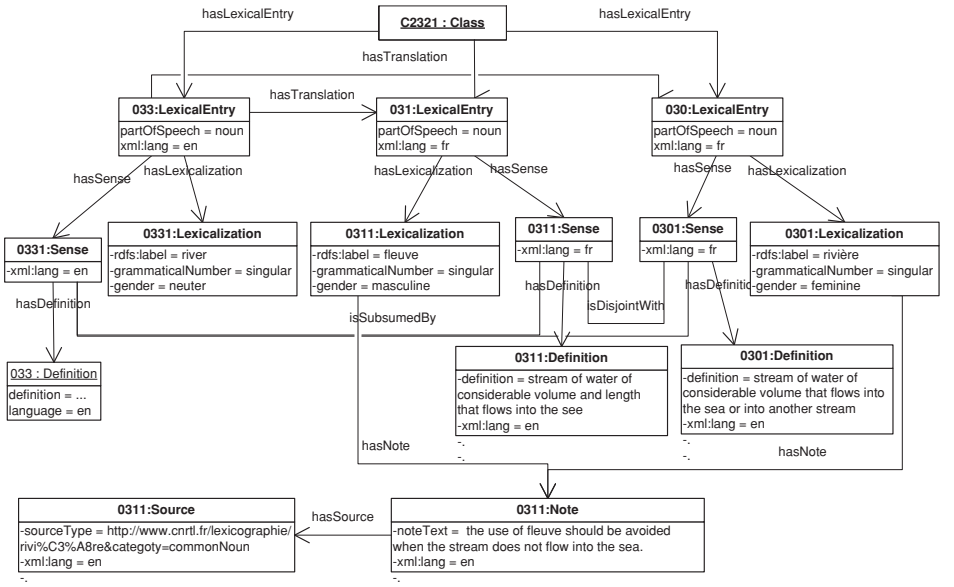


Fig. 8. Representation of conceptualization mismatches.

In order to explicitly express that kind of specificities among cultures, LIR has foreseen the classes `Sense`, `Definition` and `Note`, as well as the relations that specify the `isRelatedTo` relation among senses (`isEquivalentTo`, `subsumes`, `isSubsumedBy` and `isDisjointWith`). Let us imagine the case in which our ontology contains the class *river*. In English, river is defined as a *natural stream of water of usually considerable volume*. To the best of our knowledge, the French language has no exact equivalent, but a different granularity level represented by different terms. On the one hand, the term *course d'eau*, which is slightly more general, and could be considered a translation of *stream of water or watercourse*, and on the other hand, the terms *fleuve* and *rivière*, which are more specific. Broadly speaking, *fleuve* is a river that flows into the sea, whereas *rivière* is a river that can flow into the sea or into another stream.

We have tried to represent the following scenario in Figure 8. In this case, the ontology concept, river (`C2321:Class`), has three lexical entries associated to it (`033:LexicalEntry`, `031:LexicalEntry` and `030:LexicalEntry`). The lexicalization related to the English language is river, whereas there are two lexicalizations in French, *fleuve* and *rivière*. Basically, the three lexical entries correspond to the same object in the real world, as described in the ontology concept. However, LIR captures cultural specificities in the terminological layer by means of a more complex machinery of linguistic classes. In the first place, each lexical entry is assigned to a different `Sense` class, and a definition in natural language in the `Definition` class. At the linguistic level, these lexical entries are related by the `hasTranslation` relation, but at the semantic level the two French senses are related to the English sense by the `subsumes` relation. This means that the French lexical entries are more specific than the English one. Between them, the two lexical entries are related by

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

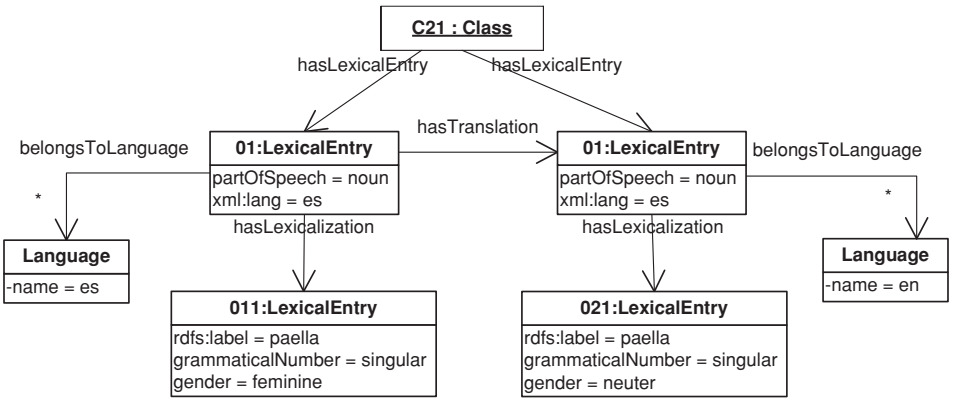


Fig. 9. Representation of non-native language expressions.

the `isDisjointWith` relation, which means that the individuals that are related to one cannot be related to the other. Finally, the `Note` class is used to make some comments about the use of the lexicalizations.

We should note here that our starting point is a given conceptualization that reflects how a certain community classifies reality. Then, by means of LIR we try to define translations or equivalences of those concepts in other languages. Considering our example of the concept *river*, it would be possible to modify the ontology on the basis of the linguistic information contained in LIR, if deemed necessary by the final application. In this case, two additional classes underlying *fleuve* and *rivière* would be added as subclasses of the concept river. Then, in the English language, we could describe those concepts as 'rivers that flow into the sea' or 'rivers that can flow into the sea or into other rivers', or we could simply associate the three concepts to the lexicalization *river*. The decision would depend on the needs of the final application.

4.1.4 Example 4: Representation of non-native language expressions

The last example we want to include here is related to the possibility offered by LIR of expressing that certain lexicalizations belonging to a specific language can be used in another language. This is the case of the Spanish word *paella*, a word also used in other languages such as English and Italian. By using the `belongToLanguage` link provided by the LIR model, we can express that a term is used in a specific country or a specific culture, and using the `xml:lang` attribute we can identify the real language of the term (see Figure 9).

5 Comparison of the LIR against the OWL and RDF(S) labelling functionality

In this section our aim is to compare the modelling modality presented in Section 2.1 against the LIR model. Nowadays, the most used modelling modality to document ontologies in natural language is the labelling functionality allowed by OWL and RDF(S) ontology representation languages. This consists in making use of

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

the annotations `rdfs:label` and `rdfs:comment`[24] to associate word forms and descriptions to ontology classes. Below we include an example of the ontology code for the class *Río*, in which two labels (*Río* and River) and one comment in Spanish are associated to the ontology class.

```
  <owl:Class rdf:about= '&Ontology1175677975;Rio'>
<rdfs:label>Río</rdfs:label>
<rdfs:label>River</rdfs:label>
<rdfs:comment>Masa de agua continental que fluye en su mayor parte
sobre la superficie del suelo</rdfs:comment>
```

The language of labels and definitions can be also specified using the 'language tagging' facility of RDF literals (e.g. River@en or Río@es). These RDF(S) properties can be complemented by the Dublin Core metadata[25] that have been created to describe resources of an information system. Examples of the Dublin Core Metadata elements are: title, creator, subject or description. Figure 10 shows how this is visualized in the ontology editor Protégé.

Taking into account that it is possible to attach as many annotations as wished to any ontology class, this functionality has been used to associate annotations in different natural languages to obtain a multilingual ontology. This is precisely one of the main advantages of this representation modality, namely, associating as much information in different languages as wished. However, we have identified several drawbacks for an appropriate exploitation of the resulting multilingual ontologies:

- All annotations are referred to the ontology element they are attached to, but it is not possible to define any semantic relations among the linguistic annotations themselves. This results in a bunch of semantically unrelated data whose motivation is difficult to understand even for a human user.
- When labels within the same language or in different languages are attached to the same ontology element, it is not possible to make explicit which is the relation existing among them.
- Finally, scalability issues will probably arise. If only a couple of languages are involved and not much linguistic information is needed, the RDF(S) properties can suffice. But if a higher number of languages is required, as seems to be the trend in the current demand, the linguistic information will become unmanageable.

In order to illustrate these issues, we have included a snapshot of an ontology of the hydrographical domain created in the ontology editor Protégé (see Figure 10). In this ontology the concept river in Spanish has information associated in Spanish and English about terms (by means of the 'label' annotation), definitions (introduced by the 'comment' annotation) and the source of provenance of synonyms and definitions

[24] Properties of the RDF Schema vocabulary, as recommended by the W3C consortium (http://www.w3.org/TR/rdf-schema/)
[25] http://dublincore.org

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

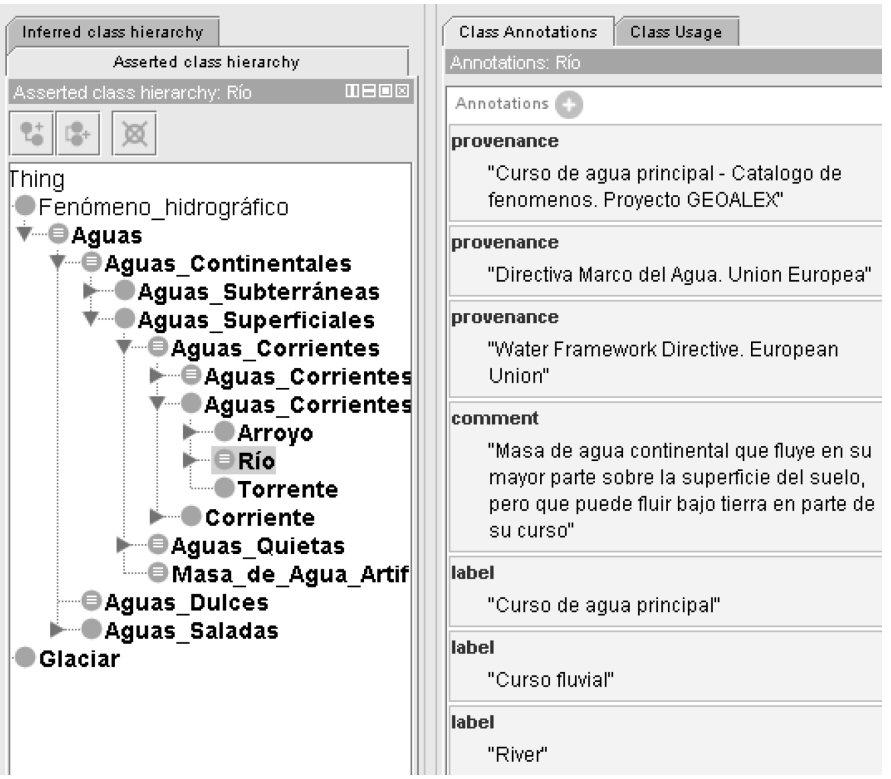


Fig. 10. Linguistic Information associated to the *Río* concept by means of the OWL and RDF(S) labelling functionality.

(by means of the 'provenance' annotation). However, no mechanisms are provided to establish relations between the labels and their corresponding definitions or sources of provenance. In the same sense, there are no possibilities for establishing a relation of synonymy among the term variants in the same language (*Curso de agua principal* – main watercourse, and *Curso fluvial* – watercourse), or a relation of translation to the labels in English (*River* in this example).

Fine-grained information such as the use of certain labels in certain discourse registers is equally missed. In this case, such an information would be useful to specify the difference between the use of *Río* in Spanish general documents, and the use of *Curso de agua principal* or *Curso fluvial* in technical documents or in the communication among experts.

As we have already illustrated in Section 4, those relations among labels in the same language or in a different language could be made explicit in LIR. To show this, we also include some snapshots of an ontology containing the river concept, making use of the LIR API implemented by the LabelTranslator plugin of the NeOn Toolkit.

Figure 11 illustrates several lexical entries associated to the same concept in different natural languages.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

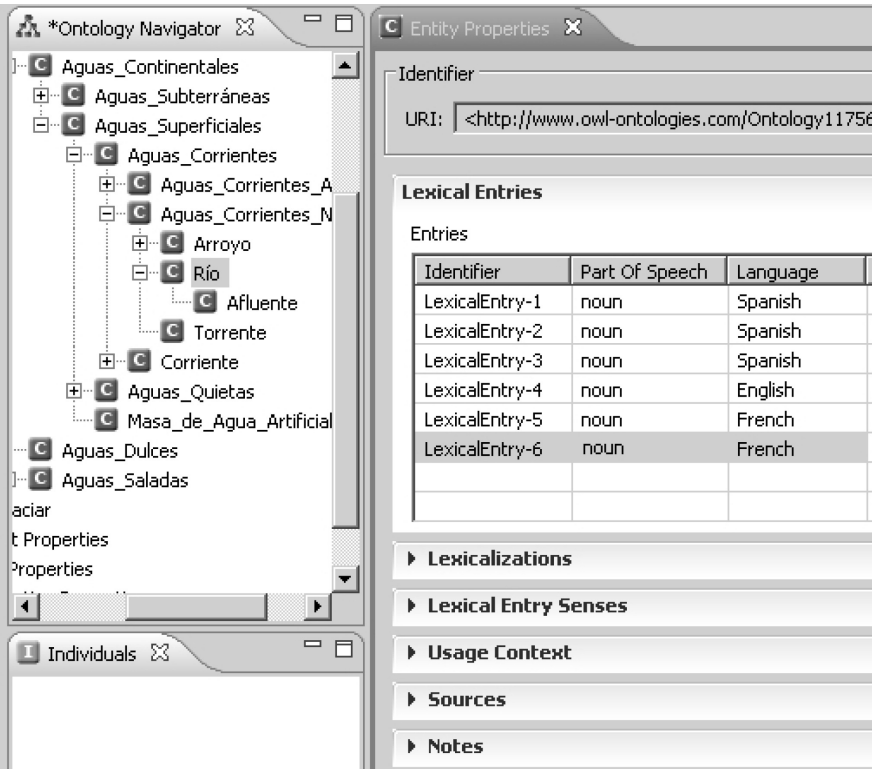


Fig. 11. Linguistic information associated to the *Río* concept by means of the LIR model.

Figure 12 shows some elements of the lexical information that can be related to each lexical entry. In this example, one lexical entry in French (LexicalEntry-5), whose lexicalization is *Rivière*, has one sense related to it (Sense-1), and its corresponding definition in French.

And, finally, Figure 13 shows how the relations of synonymy and translation are explicitly established among lexical entries within the same language and across languages.

6 Reporting about multilinguality at the ontology metadata level

The Ontology Metadata Vocabulary (OMV) is a metadata schema that captures reused relevant information about ontologies (Hartmann *et al.* 2006). OMV is designed as an ontology and it is implemented in OWL DL. It consists of the OMV core that provides fundamental information about an ontology and its life cycle (e.g. name, description, date of creation), and several OMV extensions. One of these extensions is LexOMV (Montiel-Ponsoda *et al.* 2007). LexOMV (Figure 14) is a metadata schema that captures linguistic information contained in ontologies (such as multilingual labels).

LexOMV consists of the following classes: `LinguisticData`, `OntologyElement`, `LinguisticElement`, and `NaturalLanguage`. By means of the `OntologyElement`

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

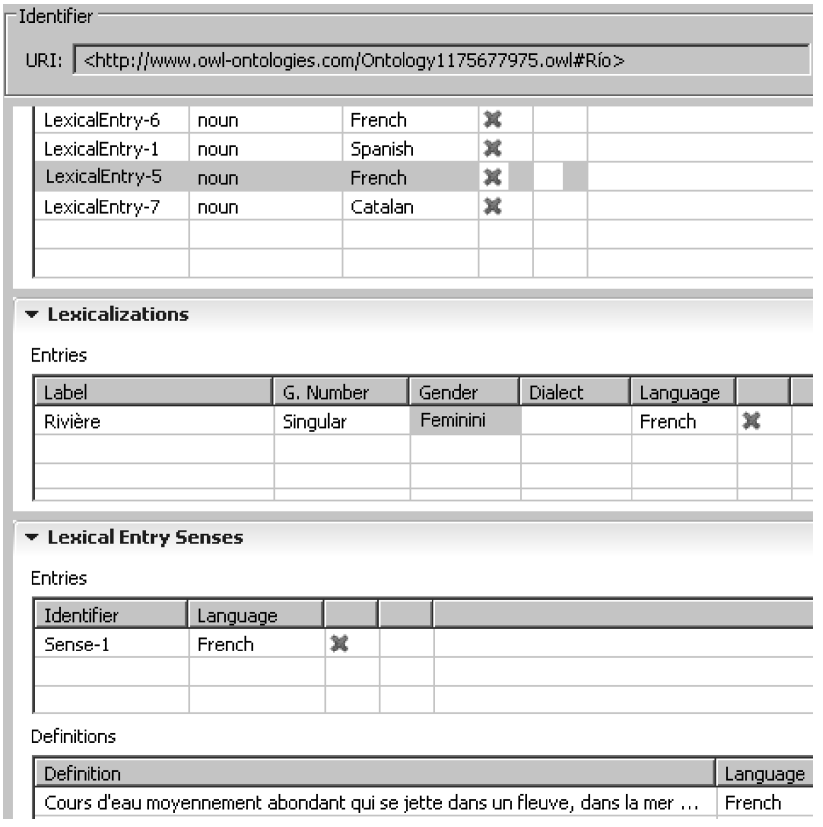


Fig. 12. Linguistic information associated to the lexical entry representing *Rivière*.

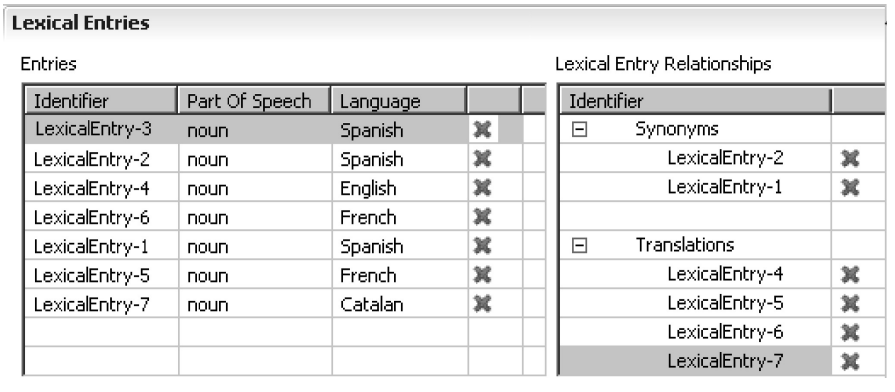


Fig. 13. Relations of synonymy and translation among labels associated to the *Río* concept.

class, we are able to make separate statements about the different elements in ontologies. Then, we define a class called `LinguisticElement`, in which we have included the attributes *name* referring to the name of the linguistic classes, e.g. definition, lexicalization, usage context, or part-of-speech, and *description* including an explanation of what is understood by these linguistic classes. As it is expected, we

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

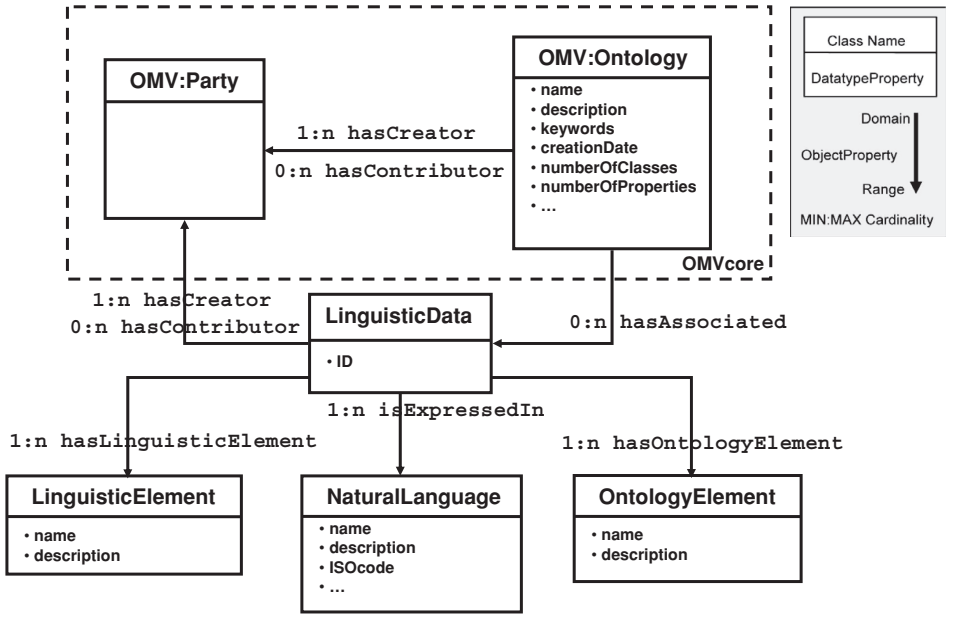


Fig. 14. LexOMV.

also define a class called `NaturalLanguage` with attributes such as *name*, *description* and *ISOcode* that allow us to refer to the different languages as defined by the ISO standard 639.[26] Finally, we define the class `LinguisticData` in order to associate the multilingual information with the rest of the ontology metadata.

Thus, to show that a certain linguistic element (let us say, Definition) is expressed in two languages (e.g. English and Spanish) for a certain type of ontology element (e.g. Class) in a given ontology, we link the ontology (described in the OMV Core) via the `hasAssociated` relation to the `LinguisticData` class where we integrate all the necessary information about which ontology elements have linguistic information associated to them, and in which natural languages. Thanks to the LexOMV, we inform the user, searching for ontologies with linguistic information of the various types of linguistic data included in the ontology in different languages. Furthermore, our extension allows us to describe who the authors and contributors of those linguistic data are by relating the `LinguisticData` class to the `Party` class of the OMV Core. According to this extension, we can now capture the author name or date of creation of the ontology next to information like 'this ontology includes lexicalizations and definitions of ontology classes in English and Spanish'.

6.1 Closing the circle: multilinguality at data, knowledge representation and metadata level

Figure 15 illustrates the different levels at which multilinguality can be present. In this figure we first identify the two levels at an ontology-based application affected

[26] http://www.loc.gov/standards/iso639-2/php/code list.php

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

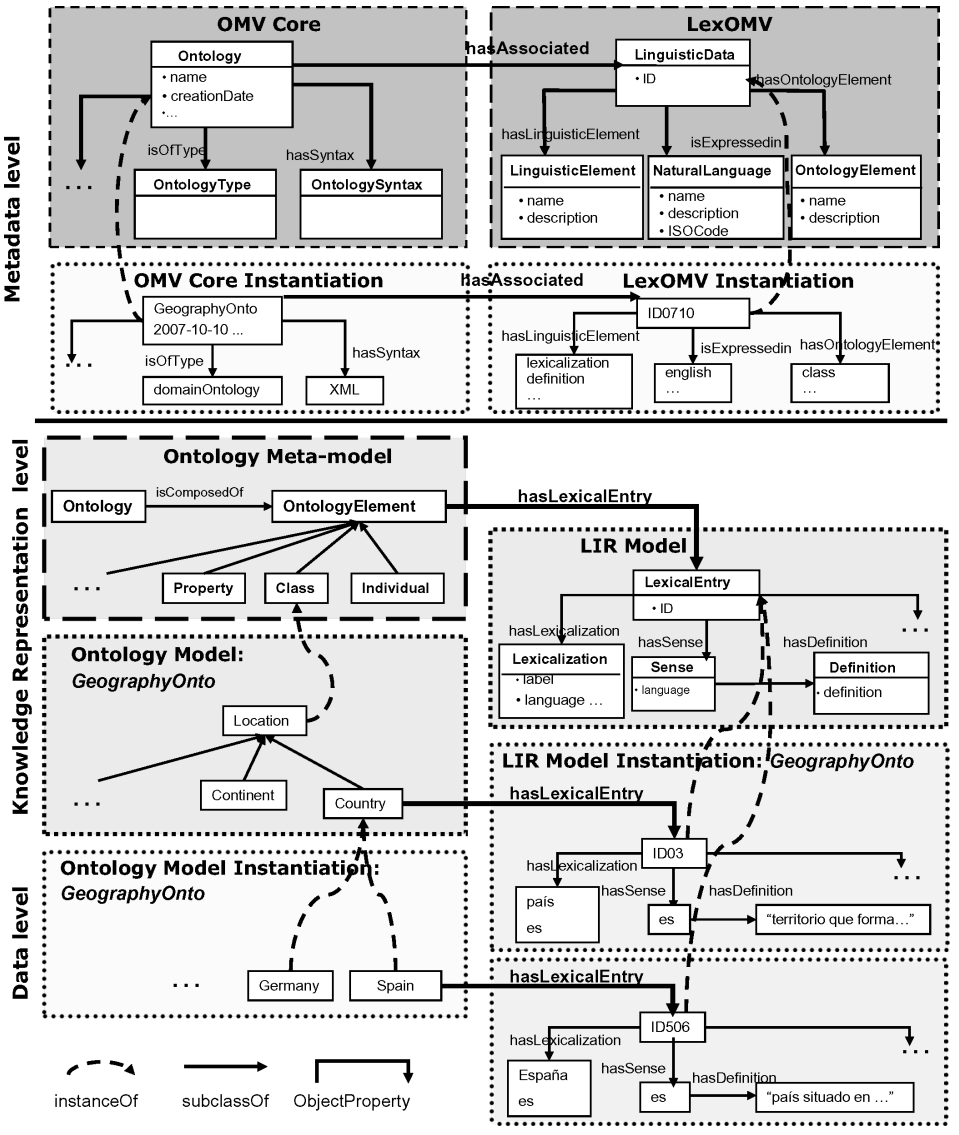


Fig. 15. Ontology structure levels affected by multilinguality.

by the inclusion of multilingual data: *knowledge representation* and *data* levels; and, second, at a higher level, the *metadata* level that reports about the data in the ontology. Depending on the heterogeneity layers implied in the localization activity – as identified in Section 2 – the knowledge representation level will be modelled in a different way. In our illustration, we have represented the modelling option 2.3, by including a sample of LIR associated to the ontology metamodel, and having as a result a multilingual ontology metamodel. The figure explains graphically how LIR is instantiated for a given domain ontology (GeographyOnto, in our example) and for its instances. The upper level of the figure represents how OMV Core and LexOMV are instantiated taking into account the information present in the lower

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

part of the figure. Therefore, LexOMV allows us to make the following assertions about the multilingual data included in the ontology: the GeographyOnto domain ontology has some linguistic elements (specifically lexicalizations and definitions) expressed in Spanish, associated to the ontology element class.

7 Conclusions

In this contribution we have raised the impending need of international organizations dealing with multilingual information for representing multilinguality in ontologies. In order to obtain multilingual ontologies, one of the main activities to be carried out during the ontology development process is the ontology localization activity, as explained in the paper. This activity may result in different options for modelling multilinguality, depending on the ontology layers implied in its development. We have discussed the three modelling options identified, paying attention to the suitability of associating the ontology metamodel to a multilingual linguistic model designed in the framework of the NeOn project, LIR. This model implies localization at the terminological layer and allows localization at the conceptual one. LIR has proven to have the following benefits, as showed in the instantiation examples: (1) establishment of relations between linguistic elements within the same language or across languages, and (2) solution to conceptualization mismatches among different cultures. Additionally, this model keeps the linguistic information associated to the ontology independent of the ontology metamodel, but with the possibility of establishing links to any ontology element. Within NeOn, the model has been provided accessibility to external resources by means of the LabelTranslator NeOn plug-in, which will facilitate the translation of the LIR elements in a semi-automatic way. Finally, and thanks to the LexOMV extension we are able to integrate the multilingual aspects of ontologies represented by LIR at the metadata level to enable search and reuse of multilingual ontologies.

References

Barrasa, J. 2007. *Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales*. PhD Thesis, Madrid, Spain: Universidad Politécnica de Madrid.

Buitelaar, P., Cimiano, P., Haase, P., and Sintek, M. 2009. Towards linguistically grounded ontologies. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvnen, R., Mizoguchi, E., Oren, M., Sabou, E., and Simperl (eds.), *The Semantic Web: Research and Applications, 6th European Semantic Web Conference*, ESWC 2009, pp. 111–125. Heraklion, Greece: Springer Verlag.

Buitelaar, P., Sintek, M., and Kiesel, M. 2006. A multilingual/multimedia lexicon model for ontologies. In Y. Sure and J. Domingue (eds.), *The Semantic Web: Research and Applications, 3rd European Semantic Web Conference*, ESWC 2006, pp. 502–513. Budva, Montenegro: Springer Verlag.

Cabré, M. T., Bach, C., Estopà, R., Feliu, J., Martínez, G., and Vivaldi, J. 2004. The GENOMA-KB project: towards the integration of concepts, terms, textual corpora and entities. In *Proceedings of 4th International Conference on Languages Resources and Evaluation*, LREC 2004, pp. 87–90. Lisboa, Portugal: ELRA/ELDA.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

Cimiano, P., Hasse, P., Herold, M., Mantel, M., and Buitelaar, P. 2007. LexOnto: a model for ontology lexicons for ontology-based NLP. In *Proceedings of OntoLex 2007, co-located at the 6th International Semantic Web Conference ISWC+ASWC 2007*, Busan, South Korea.

Edmonds, P., and Hirst, G. 2002. Near-synonymy and lexical choice. *Computational Linguistics* **28**(2): 105–144.

Espinoza, M., Gómez-Pérez, A., and Mena, E. 2008. Enriching an ontology with multilingual information. In S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis (eds.), *The Semantic Web: Research and Applications, 5th European Semantic Web Conference*, ESWC 2008, pp. 333–347. Tenerife, Spain: Springer Verlag.

Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Šváb-Zamazal, O., Svátek, V., Trojahn, C., Vouros, G., and Wang, S. 2009. Results of the ontology alignment evaluation initiative' 09. *ISWC workshop on Ontology Matching (OM 2009)*. Available at http://oaei.ontologymatching.org/2009/results/oaei2009.pdf

Fellbaum, C. (ed.) 1998. *Wordnet. An Electronic Lexical Database*. Cambridge: MIT Press.

Francopoulo, G., Monte, G., Calzolari, N., Monachini, M., Bel, N., Pet, M., and Soria, C. 2006. LMF for multilingual, specialized lexicons. In *Proceedings of 5th International Conference on Languages Resources and Evaluation*, LREC 2006, pp. 27–32, Genova, Italy.

Hartmann, J., Palma, R., and Paslaru Bontas, E. 2006. OMV-Ontology Metadata Vocabulary for the Semantic Web. OMV Report v.2.0. Available at http://ontoware.org/frs/download.php/336/OMV-ReportV2.1.pdf

Hirst, G. 2004. Ontology and the lexicon. In S. Staab and R. Studer (eds.), *Handbook on Ontologies and Information Systems*, pp. 1–21. Berlin: Springer Verlag.

Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowski, A., Peters, I., Peters, W., Ruimy, N., Villegas, M., and Zampolli, A. 2000. SIMPLE: a general framework for the development of multilingual lexicons. *International Journal of Lexicography* **3**: 249–263.

Liang, A. C., Lauser, B., Sini, M., Keizer, J., and Katz, S. 2008. From AGROVOC to the Agricultural Ontology Service/Concep Server. An OWL model for managing ontologies in the agricultural domain. In *Proceedings of the OWL: Experiences and Directions Workshop*, Manchester, UK. Available at ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf

Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., and Peters, W. 2008. Modelling multilinguality in ontologies. In *Proceedings of the 22nd International Conference on Computational Linguistics*, Coling 2008, Companion volume – Posters and Demonstrations, pp. 67–70, Manchester, UK.

Montiel-Ponsoda, E., Aguado de Cea, G., Suárez-Figueroa, M. C. Palma, R., and Peters, W. 2007. LexOMV: an OMV extension to capture multilinguality. In *Proceedings of OntoLex 2007*, co-located at the 6th International Semantic Web Conference ISWC+ASWC 2007, Busan, South Korea.

Montiel-Ponsoda, E., and Peters, W. (coordinators) 2008. Multilingual and localization support for ontologies. Technical report, NeOn Project Deliverable 2.4.2. Available at http://www.neon-project.org/web-content/images/Publications/neon_2008_d242.pdf

Morris, C. W. 1938. Foundations of the theory of signs. In *International Encyclopedia of Unified Science* **1**(2), Chicago: The University of Chicago Press.

Ogden, C. K., and Richards, I. A. 1923. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul.

Peters, W., Espinoza, M., Montiel-Ponsoda, E., and Sini, M. 2009. Multilingual and localization support for ontologies 3. Technical report, NeOn Project Deliverable 2.4.3. Available at http://www.neon-project.org/web-content/images/Publications/neon_2009_d243.pdf

Peters, W., Gangemi, A., and Villazón-Terrazas, B. 2010. Modelling and re-engineering linguistic/terminological resources. Technical report, D2.4.4 NeOn Project Deliverable. Available at http://www.neon-project.org/web-content/images/Publications/neon_2009_d244.pdf

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

Peters, W., Montiel-Ponsoda, E., and Aguado de Cea, G. 2007. Localizing ontologies in OWL. In *Proceedings of OntoLex 2007*, co-located at the 6th International Semantic Web Conference ISWC+ASWC 2007, Busan, South Korea.

Pustejovsky, J. 1995. *The Generative Lexicon*. MIT Press: Cambridge, MA.

Suárez-Figueroa, M. C., and Gómez-Pérez, A. 2008. A first attempt towards a standard glossary of ontology engineering terminology. In *Proceedings of the 8th International Conference on Terminology and Knowledge Engineering* TKE 2008, pp. 1–16, Copenhagen, Denmark.

Vossen, P. 2002. *EuroWordNet*. General document. Version 3, final. Available at http://www.vossen.info/docs/2002/EWNGeneral.pdf

Vossen, P. 2004. EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. *Semi-special issue on multilingual databases*, IJL **17** (2): 161–173.

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press

https://doi.org/10.1017/S1351324910000082 Published online by Cambridge University Press