A Fuzzy Ontology Extension of WordNet and EuroWordnet for Specialized Knowledge

Pilar León Araúz[1], Juan Gómez-Romero[2], and Fernando Bobillo[3]

[1] Department of Translation and Interpreting, University of Granada, Spain, pleon@ugr.es
[2] Department of Computer Science, Applied Artificial Intelligence Group, University Carlos III of Madrid, Spain, jgromero@inf.uc3m.es
[3] Department of Computer Science and Systems Engineering, University of Zaragoza, Spain, fbobillo@unizar.es

Abstract. The integration of linguistic and domain-oriented knowledge resources is a common task in many current projects with NLP objectives. WordNet and EuroWordNet offer large lexical resources for such purposes. Nevertheless, their exploitation in conjunction with specialized knowledge is still a non-trivial task. We propose to use fuzzy ontologies, which combine Fuzzy Logic theory and Description Logics, to represent the imprecise notions of prototypicality and representativeness with regards to synset membership, hyponymy and equivalence. We show that this approach is particularly appropriate to combine WordNet with other specific terminological resources, such as the environmental knowledge base EcoLexicon.

Keywords: WordNet, EuroWordNet, fuzzy ontologies

1 Introduction

The integration of linguistic and domain-oriented knowledge resources is a common task in many current projects with NLP objectives. WordNet [13] is the most extensive linguistic resource that has been most widely used for such purposes. As it is well-known, it groups nouns, verbs, adjectives and adverbs into synonym sets (*synsets*), where each one expresses a different concept. Synsets are related through semantic relations, composing a semantic network of concepts designated by various synonym words or, more accurately, *word senses*. WordNet is publicly available and has also been translated into RDF and OWL, the standard ontology languages in the Semantic Web, becoming one of the largest lexical ontologies available. This is one of the main reasons why it has been commonly used in the Semantic Web. However, WordNet was originally conceived as an English lexical resource and multilinguality has also become a challenge in the Semantic Web, which needs language-independent knowledge extraction procedures and representation formalisms. In this sense, the integration of different languages in EuroWordNet [20] emerged as the basis of tasks requiring multilingual information.

Aguado de Cea et al. (Eds.): Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012), pp.139-154. 19-22 June 2012, Madrid, Spain

Both WordNet and EuroWordNet have been successfully used in computational linguistics and NLP. They can be applied to improve information retrieval tasks from text sources: term extraction, word disambiguation, query expansion with synonym terms, automatic annotation of texts, etc. Nevertheless, their exploitation in conjunction with specialized knowledge is still a non-trivial task, partly because they only include general language knowledge [9]. This problem mostly affects the synsets defined in the leaf nodes of the taxonomy, since they group together words that might be generally considered synonyms by lay users, but have different meanings in specialized contexts. For this reason, specialized resources are not usually connected to either WordNet or EuroWordNet, thus missing their support for the analysis of common discourse.

Consequently, it seems natural to extend WordNet's and EuroWordNet's representations to set imprecise connections between some pieces of information rather than using their sharp equivalences and relations between terms. Despite the undisputed success of ontologies, they are not generally appropriate to deal with imprecise and vague knowledge. However, proposals for the creation of fuzzy ontologies based on fuzzy Description Logics have emerged [12]. In fuzzy ontologies, concepts denote fuzzy sets and relations denote fuzzy relations; therefore, the axioms are not in general either true or false, but they may hold to some degree of truth.

In this paper, we study the limitations of WordNet and EuroWordNet for specialized knowledge representation, and propose an extension based on fuzzy ontologies to overcome them. Previous research has focused on the use of WordNet for the automatic creation of fuzzy ontologies [10] or the automatic grading of fuzzy synsets [5,14]. Nonetheless we focus on the inherent imprecision of several contents from WordNet and EuroWordNet and, in particular, on their representation with fuzzy ontologies, which has not been explored to date. We show that the fuzzification of the hyponym, hypernym and synset membership relations, as well as the equivalence relationship in EuroWordNet is helpful to: 1) discriminate between the several word senses included in the same WordNet synset and their equivalents in EuroWordNet; and 2) provide a more fine-grained representation of the semantic distance between a concept and its hyponyms/hypernyms, as well as the equivalence degree between synsets in different languages. These features would allow the integration of specialized information into the WordNet/EuroWordNet schema without loss of knowledge.

2 Structure of WordNet and EuroWordNet

The basic elements in the WordNet knowledge base are words, word senses and synsets. Words correspond to terms; the lexical form is the written expression of the word. Since the same word can be used with different senses in different pragmatic contexts, a word has several associated senses –a sense is associated to a single word. According to WordNet specification, a synset is a collection of word senses that can be considered synonyms –hence, a word sense is included into one synset at most. Each synset is associated with a gloss, which is a brief textual description of the meaning of the synset. Synsets are interlinked by means of lexical and semantic relations, such as hyponymy, antonymy, etc. In this paper, we focus on the following WordNet's relations: the hyponymic relations hyponymOf and its inverse hypernymOf, which represent the notions of subtype-of and supertype-of, respectively, and on the relations inSynset and its inverse containsWordSense, which link words and synsets. Figure 1 represents these core elements of WordNet.

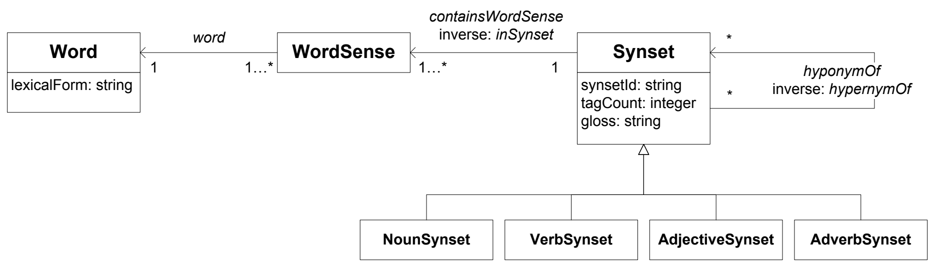

Fig. 1. UML schema of the core WordNet entities

The RDF/OWL translation proposed by the WordNet Task Force of the World Wide Web Consortium can be considered the official ontology-based version of WordNet [19]. The WordNet ontology defines three main classes: Word, WordSense and Synset. The elements of WordNet are included as instances of these three classes. Relations have also been instantiated to mirror WordNet associations. Note that this translation does not treat WordNet as a class hierarchy where the hyponym relationship is interpreted as a class inclusion axiom. While this is possible by adding a few axioms to the ontology, the semantics of the hyponym relation in WordNet discourages this interpretation.

EuroWordNet follows the same structure as the Princeton WordNet 1.5, but some specific changes were made motivated by multilinguality and the different conceptualizations lexicalized in every language. For this reason, each language module represents an autonomous and unique language-specific system of language-internal relations between synsets, which are connected through the Inter-Lingual-Index (ILI) [24] (Figure 2). This means that each language developed its own wordnet and linguistic equivalences were set through interlingual links stored in the ILI. In this way, language-specific synsets linked to the same ILI-record are equivalent across languages.

EuroWordNet introduced new relations and relational features. On the one hand, WordNet's original relations were extended, such as cross-part-of-speech synonymy and role relations such as agent, instrument, patient and location [25]. On the other hand, new multilingual relations had to be included between synsets and ILI-records, such as eq-Synonym, which is a direct match between a synset and an ILI-record –thus reflecting equivalence.

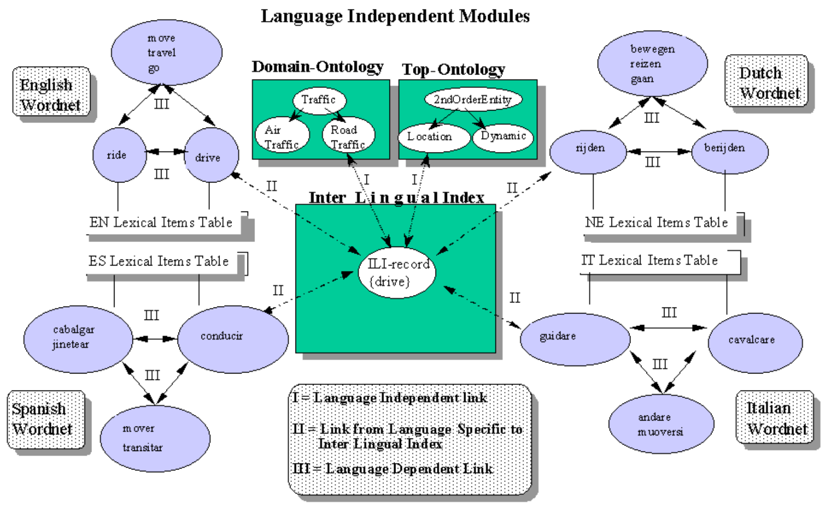

Fig. 2. Architecture of EuroWordNet (extracted from [24])

Moreover, other complex relations were added when there was a lexical gap in one language or when meanings did not exactly fit due to anisomorphism. eq-Near-Synonym is used when a meaning matches multiple ILI-record simultaneously, when multiple synsets match with the same ILI-record, or when there is some doubt about the precise mapping. The has-eq-Hyperonym relation is used when a meaning, in a particular language, is more specific than any available ILI-record. Conversely, has-eq-Hyponym relations are used when a meaning, in a particular language, can only be linked to more specific ILI-records [21,25]. For example, *dedo* in Spanish may refer to either *finger* or *toe*. Consequently, the synset {dedo} corresponds with the ILI-record of the synset {digit}, but its equivalents correspond with the ILI-records of the hyponym synsets {finger} and {toe}. In this paper we propose an extension of the equivalence relations eq-Synonym and eq-Near-Synonym by means of a new relation directly connecting synset pairs in two different languages, English and Spanish.

There have been several approaches for converting EuroWordNet into OWL [7,6], but none of these ontologies are publicly available. They were based on the need of extending EuroWordNet with other domain-specific resources for different purposes. In this sense, EuroWordNet also evolved into the creation of additional WordNets in multiple languages leading to a new proposal launched in 2006 called the Global WordNet Grid [17,22]. The Global WordNet Grid seeks to assemble all multilingual WordNets developed for both general and domain-specific applications.

A representative example of this interoperable grid can be found in the Kyoto project, whose main goal is to develop a content enabling system that provides semantic search and information access to large quantities of distributed multimedia data for both experts and the general public, and to apply this system to environmental information on a global scale [23]. EcoLexicon[1] [11] is also a multilingual specialized resource dealing with environment, which would undoubtedly benefit from its integration into the Global WordNet Grid. However, the use of WordNets in specialized contexts, either isolated or in combination with domain-specific terminological knowledge bases, is not a straightforward task.

3 Limitations of WordNet and EuroWordNet

As several researchers have claimed [15], synsets and their relations have not always been created according to the same criteria. For instance, concepts and individuals are rarely, if ever, part of the same synset, but they appear intermingled in the same hierarchical levels. This results in an expressivity lack due to the confusion between subsumption and instantiation [8]. The latter versions of WordNet have reviewed most cases by creating a different relation, has-Instance, but there are still some inconsistencies, as in the synset {Black Hills}, which is a hyponym of the synset {mountain} and is located as a sibling of other general concepts, such as {seamount} or {volcano}.

Gangemi et al. [8] also found heterogeneous levels of generality when comparing certain synsets and their too specific siblings. We have also found other problems related to the criteria on which synsets are built. The notion of synonymy used in the construction of WordNet synsets is described as follows: "two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value" [13]. This gives rise to different types of synsets, ranging from those whose members are fully synonyms to those where slight differentiating nuances may apply (see Section 3.1). Nevertheless, it appears striking that word senses such as those expressed by the words *animal, animal being, beast, brute, creature* and *fauna* may belong to the same synset whereas word senses like *typhoon* and *hurricane* are included in distinct synsets.

Some of these problems were partially solved in EuroWordNet. The ILI structure helped overcoming cross-linguistic differences as well as solving the limitations found in WordNet –avoiding sense duplication, enabling disjonction, tagging relations with different features, etc. [1]. However some of the latter drawbacks persist in EuroWordNet, and other problems related to the ILI and to the language-dependent coverage of word senses came along [6]. For instance, the linguistic differences between English and other languages gave rise to complex and obscure equivalence relations [22] (see Section 3.2).

3.1 Hyponymy and synset membership in WordNet

WordNet has been extensively criticized for having too fine-grained sense distinctions that are irrelevant for certain NLP applications. This might be true when dealing with word sense disambiguation tasks based on how polysemy is

[1] `http://ecolexicon.ugr.es`

addressed in WordNet. Nonetheless, we believe that the representation of specialized knowledge in WordNet needs a more fine-grained approach in terms of categorization. In the following examples, we will show how both synset members and hyponym synsets should not have the same status of representativeness. Theoretically, the notion of synset corresponds to that of meaning, and ultimately to that of a distinct concept, since all word senses included in a synset are close enough to represent the same concept in reality. Nevertheless, synset members show different degrees of similarity and their associated word forms cannot always be interchanged in a specialized context, especially those found in the leaf nodes.

S0: (n) **hill** (a local and well-defined elevation of the land) *"they loved to roam the hills of West Virginia"*

S1: (n) **knoll**, **mound**, **hillock**, **hummock**, **hammock** (a small natural hill)

S2: (n) **anthill** , **formicary** (a mound of earth made by ants as they dig their nest)

S3: (n) **kopje**, **koppie** (a small hill rising up from the African veld)

S4: (n) **molehill** (a mound of earth made by moles while burrowing)

Fig. 3. WordNet representation of synsets (synset identificators are ours)

In Figure 3, synsets *S2* {anthill, formicary} and *S3* {kopje, kopiie} are composed of two word senses designated by clear synonyms. Strictly speaking, both word senses in each synset could not be regarded as absolute synomyns of each other, since they are designated by words that reflect geographical or register variations that prevent them from being interchangeable in all pragmatic contexts. However, they represent the highest degree of similarity –or synonymy– found in WordNet synsets. Absolute synonymy has long been considered an extremely rare phenomenon, but they undoubtedly represent the same concept in reality.

In synset *S1* {knoll, mound, hillock, hummock, hammock}, word senses are still close concepts that share meaning identity, but their similarity degree differs slightly from that of synsets *S2* and *S3*. Even though all of them are low natural hills, as the synset gloss states, certain word senses add some other features to their shared nuclear meaning. For instance, any layman would know that *mound* represents a wider sense than the others. Conversely, *hummock* and *hammock* represent a slightly more specific sense, since they are usually related to the presence of ice or trees respectively. We could then say that some of these word senses (i.e. *knoll*) are closer –or more similar– than others (i.e. *hammock*) to the prototype of a small hill, the meaning represented in the synset. This means that all senses do not have equal degrees of synset membership.

Comparing sibling synsets, it is also evident that either *S2* or synset *S4* {molehill} are more accurate types of *mounds* than *knolls*, whereas synset *S3* is better categorized as a *knoll* than as a *hammock*. Consequently, there is also a clear relationship between the individual word senses of a synset and the word senses of its hyperonym synset. Nevertheless, WordNet does not represent it, since this is not necessarily important for general language purposes. Equally, from the three synsets that constitute the hyponyms of synset *S1*, anyone would easily point synset *S3* as the most prototypical exemplar of a small natural hill compared to synsets *S2* and *S4*. However, establishing the degree in which a whole synset is the most prototypical exemplar of another is not so obvious when dealing with longer sets of hyponymic synsets or multiple word senses in each of the synsets.

Prototype theory [18] provides an approach to account for such effects of prototypicality on categorization. Prototype theory is based on the fact that concepts are graded. They show different degrees of category membership according to the notions of family resemblance [26] and similarity. Initially they were widely accepted and applied by the psychology of concepts, but in the 80s cognitivists started to reject them in the realm of conceptual compositionality [16]. However, Zadeh showed that it was very unlikely that an adequate theory of prototypes could be constructed without an explicit use of fuzzy sets and related concepts [27].

Previous research has considered grading synset membership [14] and quantifying the degree of synonymy among word senses [5]. According to Oliveria and Gomes [14], the fuzzy membership of a word in a synset can be interpreted in two ways: 1) as the confidence level about using the word to indicate the meaning of the synset or 2) as the likelihood of a word conveying the meaning of each synset it belongs to. The first interpretation would be useful to distinguish among the different variants included in certain synsets. For instance, it seems obvious that in synsets such as {ocean floor, ocean bottom, seabed, sea bottom, Davy Jones's locker, Davy Jones}, the last two word senses are associated to words that would only be interchangeable in very specific contexts.

The second interpretation is closer to our approach. However, in our proposal, both intra-synsets and hyponymic relations (either among word senses or synsets) are regarded as fuzzy constructs that rely on different levels of prototypicality according to the notion of similarity. After all, similarity is the common cognitive feature involved in category membership, synonymy and hyponymy construction. Consequently, synsets should not be regarded as a mere set of more or less canonical synonyms. They are not even the even union of all word senses. Rather, they are prototypical categories with fuzzy boundaries whose members (senses) belong to it in different degrees. The hyponymic relation between synsets is also fuzzy and depends on the degree in which word senses are individually related to the word senses of a hyperonymic synset. This is especially important when a synset contains several word senses with different nuances, but even more for the integration of specialized knowledge in WordNet, since some of the concepts contained in EcoLexicon as clearly distinct notions are compiled in WordNet within the same synset.

In synset *S1* although word senses were not equally prototypical, they shared the same meaning identity. In synset *S5* (Figure 4), however, they do not point at all to the same concept in reality. In this case, the only true synonyms in the

S5: (n) **breakwater**, **groin**, **groyne**, **mole**, **bulwark**, **seawall**, **jetty** (a protective structure of stone or concrete; extends from shore into the water to prevent a beach from washing away)

S6: (n) **offshore breakwater**, **detached breakwater**

S7: (n) **convex seawall**, **rigid revetment**

S8: (n) **fish-tail groin**, **Y-shaped groin**

Fig. 4. WordNet (*S5*) and EcoLexicon (*S6*, *S7*, *S8*) combined synsets

synset are *groin* and *groyne*, which represent a geographical variant instead of a lexical relation. It is our guess that WordNet displays all of the concepts within the same synset due to its general language approach. Nonetheless, if we were to integrate all kinds of coastal protective structures from EcoLexicon, degrees would be vital for the development of its hyponym synsets.

In *S5 groin* (and *groyne*) can be regarded as the prototype because its definitional features match the gloss of the synset. Therefore, *groyne* and *groin* are related to their own synset with the maximum degree. Consequently, *bulwark*, which is the concept less related to the gloss (and thus, less similar to *groin*), has the lower degree in terms of synset membership. Following the general-language perspective of WordNet's synsets, we have included what could be three different hyponymic synsets extracted from EcoLexicon. Synset *S8* is composed of two word senses that are fully synonymic and thus should have the same synset membership degree. Synsets *S6* and *S7* are not strict synomyms but, according to metalinguistic information in lexicographic resources, even if they do not exactly point to the same real-world concept from a specialized perspective, there is often no distinction between *offshore* and *detached breakwaters* and *convex seawalls* and *rigid revetments*. This is why they belong to the same synset but should not have the same degree of prototypicality or synset membership. By grading synset membership and the hyponymic relation between word senses, we can infer the degree in which a whole synset is a hyponym of its superordinate synset (see Section 4.2).

3.2 Equivalence in EuroWordNet

Synsets pairs in different languages can also be assigned a degree in terms of equivalence, which is also linked to the notion of similarity. As stated above, in EuroWordNet they found it necessary to add new relations which clarify particular similarities or differences in the semantic constituency of words which have a close meaning across different language [1]. This is how they considered a new construct called *fuzzynym* [20], which is used when the relation among word senses is undertermined. In the same way, the eq-Near-Synonym relation is said to be typically used for fuzzy concepts in fuzzy cross-lingual matching scenarios [25]. However, these notions do not seem to be related at all to fuzzy logic.

S9: **breakwater**, **groin**, **groyne, mole**, **bulwark**, **seawall**, **jetty** (a protective structure of stone or concrete; extends from shore into the water to prevent a beach from washing away)

S9b: **dique**, **rompeolas**, **malecón**, **espigón**

Fig. 5. EuroWordNet synsets in English and Spanish

S10: **marsh**, **marshland**, **fen** (low-lying wet land with grassy vegetation; usually is a transition zone between land and water)

S10b: **cenagal**, **tolla**, **paúl**, **paular**, **pantano**, **marjal**, **ciénaga**

Fig. 6. EuroWordNet synsets in English and Spanish

Cross-linguistic differences are based on the fact that each language has its own set of concepts depending on their lexicalization patterns. Therefore, EuroWordNet theoretically allows for any structuring of the set of concepts (ILI) that is used to interlink the meanings across languages [21]. Nonetheless, they followed a minimalistic (conservative) approach with regards to restructuring the ILI, which led to a rather English-centered conceptualization.

Among other methods, WordNet synsets were translated into equivalent synsets in other languages by exploiting several bilingual resources. They were also identified on the basis of the possibility of a word being replaced by another in a specific context [25]. This does not seem to be the case for those found in Figures 5 and 6, at least from a specialized knowledge perspective.

Synset *S9b* (in Spanish) is composed again of word senses that are not synomyms either in Spanish. Moreover, it also seems quite evident that the equivalence relation among individual word senses in English and Spanish are not the same. *Espigón* is now the prototype, because its sense corresponds with the gloss, and would thus be the direct equivalent of *groyne* and *groin*, but not of *bulwark*. Although in a different degree, *breakwater* could be the equivalent of both *dique* and *rompeolas* but definitely not of *malecón*.

Nevertheless, in Figure 6, none of the senses in Spanish match exactly the prototype described in the gloss. At first sight, these senses may seem closer than those described in *S9*. In fact, these concepts are often confused in non-specialized literature, but their associated words point to different concepts in the real world. In a specialized knowledge scenario, the definitional feature *grassy vegetation* is the key to distinguish all of them. Accordingly, *marsh* is the prototype of the English synset but in the Spanish synset the prototype, and thus the direct equivalente of *marsh* (*marisma*), is missing.

Consequently, the equivalence relation of word senses within a synset can also be graded. What can now be inferred is the degree in which a synset in Spanish is the equivalent of a synset in English and viceversa (see Section 4.2).

Table 1. Syntax and semantics of the fuzzy DL $\mathcal{SR}^{-}\mathcal{I}$.

Concept	**Syntax** (C)	**Semantics** ($C^{\mathcal{I}}(x)$)
(C1)	A	$A^{\mathcal{I}}(x)$
(C2)	$\top$	γ_0
(C3)	$\bot$	γ_p
(C4)	$C \sqcap D$	$C^{\mathcal{I}}(x) \otimes D^{\mathcal{I}}(x)$
(C5)	$C \sqcup D$	$C^{\mathcal{I}}(x) \oplus D^{\mathcal{I}}(x)$
(C6)	$\neg C$	$\ominus C^{\mathcal{I}}(x)$
(C7)	$\forall R.C$	$\inf_{y \in \Delta^{\mathcal{I}}}\{R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)\}$
(C8)	$\exists R.C$	$\sup_{y \in \Delta^{\mathcal{I}}}\{R^{\mathcal{I}}(x,y) \otimes C^{\mathcal{I}}(y)\}$
Role	**Syntax** (R)	**Semantics of** $R^{\mathcal{I}}(x,y)$
(R1)	R	$R^{\mathcal{I}}(x,y)$
(R2)	R^{-}	$R^{\mathcal{I}}(y,x)$
Axiom	**Syntax** (τ)	**Semantics** ($\mathcal{I}$ satisfies τ if ...)
(A1)	$\langle a\!:\!C \bowtie \alpha\rangle$	$C^{\mathcal{I}}(a^{\mathcal{I}}) \bowtie \alpha$
(A2)	$\langle (a,b)\!:\!R \geq \alpha\rangle$	$R^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}}) \geq \alpha$
(A3)	$\langle C_1 \sqsubseteq C_2 \geq \alpha\rangle$	$\inf_{x \in \Delta^{\mathcal{I}}}\{C_1^{\mathcal{I}}(x) \Rightarrow C_2^{\mathcal{I}}(x)\} \geq \alpha$
(A4)	$\langle R_1 \ \ldots \ R_m \sqsubseteq R \geq \alpha\rangle$	$\inf_{x_1, x_{n+1} \in \Delta^{\mathcal{I}}}\{\sup_{x_2 \ldots x_n \in \Delta^{\mathcal{I}}}\{$ $R_1^{\mathcal{I}}(x_1,x_2) \otimes \cdots \otimes R_n^{\mathcal{I}}(x_n, x_{n+1})\}\} \Rightarrow R^{\mathcal{I}}(x_1, x_{n+1})) \geq \alpha$

4 Fuzzy extension of WordNet and EuroWordNet

4.1 Fuzzy ontologies

The main formalism behind ontology languages are Description Logics (DLs), a family of logics for representing structured knowledge. Each logic is denoted by using a string of capital letters which identify the expressivity of the logic and therefore its complexity. For instance, the standard language for ontology representation OWL 2 is equivalent to $\mathcal{SROIQ}(\mathbf{D})$. Fuzzy DLs are the extension of DLs inthe fuzzy case. Essentially, in fuzzy ontologies fuzzy concepts denote fuzzy sets of individuals, whereas fuzzy roles (or properties) denote fuzzy binary relations among individuals. For the sake of clarity, in the rest of this paper we will consider a relatively simple fuzzy DL, fuzzy $\mathcal{SR}^{-}\mathcal{I}$, which is a fragment of fuzzy $\mathcal{SROIQ}(\mathbf{D})$ [2].

A *fuzzy Knowledge Base* (KB) or *fuzzy ontology* contains a finite number of axioms stating information about *fuzzy concepts* (or classes) (denoted C, D), *fuzzy roles* (or relations, denoted R) and *individuals* (denoted a, b). Fuzzy concepts can be atomic or complex –i.e., inductively formed from other concepts. Roles can be atomic or the inverse of a role. Table 1 shows the concepts, roles and axioms of fuzzy $\mathcal{SR}^{-}\mathcal{I}$, where $\bowtie \in \{\geq, \leq\}$, α is a degree of truth, $\otimes$ is a t-norm, $\ominus$ is a negation, $\oplus$ is a t-conorm, and $\Rightarrow$ is an implication. Usually, the set of degrees of truth $\mathcal{N}$ is assumed to be $[0,1]$, but we will assume that the set of degrees of truth $\mathcal{N}$ is a finite ordered set $\{\gamma_0, \gamma_1, \ldots, \gamma_p\}$ with the minimum value γ_0 representing absolute falsity, and the maximum value γ_p representing absolute truth. We assume that the operators $\otimes, \oplus, \ominus, \Rightarrow$ are those of Zadeh fuzzy logic except in axioms (A3) and (A4) which use Gödel implication [2].

The semantics of the logic is given by a fuzzy interpretation. A fuzzy interpretation $\mathcal{I}$ is a pair $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a non empty set $\Delta^{\mathcal{I}}$ (the interpretation domain) and a fuzzy interpretation function $\cdot^{\mathcal{I}}$ mapping: 1) an *individual* a to an element $a^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, 2) a fuzzy *concept* C to a function $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightarrow \mathcal{N}$, and 3) a fuzzy *role* R to a function $R^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \rightarrow \mathcal{N}$. $C^{\mathcal{I}}$ (resp. $R^{\mathcal{I}}$) denotes the membership function of the fuzzy concept C (resp. fuzzy role R) w.r.t. $\mathcal{I}$. $C^{\mathcal{I}}(a^{\mathcal{I}})$ (resp. $R^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}})$) gives us to what extent the individual a can be considered an element of the fuzzy concept C (resp. to what extent (a, b) can be considered an element of the fuzzy role R) under the fuzzy interpretation $\mathcal{I}$.

There are several reasoning algorithms for fuzzy ontologies. We will focus on a family of algorithms called *crisp representation algorithms*. Given a fuzzy ontology, crisp representation algorithms solve the reasoning tasks by computing an equivalent OWL 2 ontology. This makes it possible to reuse existing reasoners, editors, etc. The interested reader can find the details of the process in [2]. DeLorean[2] is an implementation of a crisp representation algorithm for Fuzzy OWL 2. For implementation details, we refer the reader to [3].

4.2 Fuzzy ontology of WordNet and EuroWordNet

The fuzzy extension of WordNet implies the following: 1) both containsWordSense and its inverse relation inSynset are fuzzy, as the membership of a sense to a synset is a matter of degree; 2) similarly, hyponymOf and its inverse relation hypernimOf are fuzzy; 3) we define a new fuzzy relation R1 between senses, representing the degree to which a sense is a subtype of another one –this relation is transitive. As for EuroWordNet's equivalent synsets, we need to create: 4) a new fuzzy relation equivalentTo between synsets in different languages; and 5) another fuzzy relation R2 between the senses of each equivalent synset –this relation is symmetrical.

Finally, we add the following axioms:

$$\langle \mathsf{containsWordSense}\ \mathsf{R1}\ \mathsf{inSynset} \sqsubseteq \mathsf{hyponymOf} \geq 1 \rangle \tag{1}$$

$$\langle \mathsf{containsWordSense}\ \mathsf{R2}\ \mathsf{inSynset} \sqsubseteq \mathsf{equivalentTo} \geq 1 \rangle \tag{2}$$

Axiom 1 is equivalent to axiom 3, which specifies how to compute the minimal degree of a synset being an hypernym of another synset:

$$\langle \mathsf{containsWordSense}\ \mathsf{R1}^{-}\ \mathsf{inSynset} \sqsubseteq \mathsf{hypernymOf} \geq 1 \rangle \tag{3}$$

According to the semantics of these axioms, the fuzzy reasoner automatically computes lower bounds for the degrees of: 1) synset *s1* being an hyponym of a synset *s2*; 2) synset *sy1* being an equivalent of synset *s2*. Note that if there

[2] `http://webdiis.unizar.es/~fbobillo/delorean`

				S5						
				0.9	*1*	*1*	*0.8*	*0.2*	*0.5*	*0.9*
				1	2	3	4	5	6	7
				breakwater	**groin**	**groyne**	**mole**	**bulwark**	**seawall**	**jetty**
S6	*1*	8	**offshore breakwater**	1	0.7	0.7	0.6	0.2	0.2	0.6
	0.9	9	**detached breakwater**	1	0.8	0.8	0.6	0.3	0.4	0.7
S7	*1*	10	**convex seawall**	0.6	0.5	0.5	0.4	0.8	1	0.4
	0.9	11	**rigid revetment**	0.5	0.5	0.5	0.3	0.8	0.9	0.4
S8	*1*	12	**fish-tail groin**	0.9	1	1	0.8	0.5	0.5	0.9
	1	13	**Y-shaped groin**	0.9	1	1	0.8	0.5	0.5	0.9

Fig. 7. Similarity degrees between senses (shown in plain text) and prototype resemblance degrees between senses and synsets (shown in italics). The white numbers in shaded cells denote the sense identification number. The labels *S5*, *S6*, etc. denote the synset identification number. (Identification numbers are ours)

are several senses related through the property $\mathsf{R1}$, the maximum degree of truth is considered for the $\mathsf{hyponymOf}$ relation –resp. for property $\mathsf{R2}$ and the $\mathsf{equivalentTo}$ relation.

In fact, the semantics implies that:

$$(\mathsf{hyponymOf})^{\mathcal{I}}(\mathsf{sy1},\mathsf{sy2}) \geq \sup_{\mathsf{sense1},\mathsf{sense2}\in\Delta^{\mathcal{I}}}\{\mathsf{containsWordSense}^{\mathcal{I}}(\mathsf{sy1},\mathsf{sense1}) \otimes \mathsf{R1}^{\mathcal{I}}(\mathsf{sense1},\mathsf{sense2}) \otimes \mathsf{inSynset}^{\mathcal{I}}(\mathsf{sense2},\mathsf{sy2})\} \quad (4)$$

$$(\mathsf{equivalentTo})^{\mathcal{I}}(\mathsf{sy1},\mathsf{sy2}) \geq \sup_{\mathsf{sense1},\mathsf{sense2}\in\Delta^{\mathcal{I}}}\{\mathsf{containsWordSense}^{\mathcal{I}}(\mathsf{sy1},\mathsf{sense1}) \otimes \mathsf{R2}^{\mathcal{I}}(\mathsf{sense1},\mathsf{sense2}) \otimes \mathsf{inSynset}^{\mathcal{I}}(\mathsf{sense2},\mathsf{sy2})\} \quad (5)$$

Succinctly put, this means that the degree of hyponymy between two synsets is obtained as follows:

1. For each pair of senses, the degrees of $\mathsf{containsWordSense}$, $\mathsf{R1}$ and $\mathsf{inSynset}$ are combined by using the $\otimes$ operator –the minimum, in this case.
2. The degree of the $\mathsf{hyponymOf}$ relation is the supremum of the values resulting from the previous combination.

Figure 7 provides an example on degree assignment with synset membership and hyponymic relations. For instance, for $S6$ (with respect to $S5$) the supremum happens in the following case:

				S9						
				0.9	*1*	*1*	*0.8*	*0.2*	*0.5*	*0.9*
				1	2	3	4	5	6	7
				breakwater	**groin**	**groyne**	**mole**	**bulwark**	**seawall**	**jetty**
S9b	*0.9*	14	**dique**	0.9	0.9	0.9	0.5	0.2	0.3	0.8
	0.9	15	**rompeolas**	1	0.6	0.6	0.6	0.3	0.6	0.5
	0.8	16	**malecón**	0	0	0	0.9	0.3	0.2	0.3
	1	17	**espigón**	0.6	1	1	0.8	0.2	0.5	0.9

Fig. 8. Equivalence degrees between senses (shown in plain text) and prototype resemblance degrees between senses and synsets (shown in italics)

$$\begin{aligned}
&\mathsf{containsWordSense}^{\mathcal{I}}(\mathsf{S6}, \mathsf{OffshoreBreakwater}) \otimes \\
&\mathsf{R1}^{\mathcal{I}}(\mathsf{Offshore\ Breakwater}, \mathsf{Breakwater}) \otimes \\
&\mathsf{inSynset}^{\mathcal{I}}(\mathsf{Breakwater}, \mathsf{S5}) = \\
&\min\{1, 1, 0.9\} = 0.9
\end{aligned}$$

Therefore, it holds that $(\mathsf{hyponymOf})^{\mathcal{I}}(\mathsf{S6}, \mathsf{S5}) \geq 0.9$. Similarly, $(\mathsf{hyponymOf})^{\mathcal{I}}$ $(\mathsf{S7}, \mathsf{S5}) \geq 0.6$ and $(\mathsf{hyponymOf})^{\mathcal{I}}(\mathsf{S8}, \mathsf{S5}) = 1$. Without our fuzzy extension, $S5$ is as hyponym of $S6$ as it is of $S7$, but now we are able to quantify the degree of hyponymy. The application of the supremum, established by the semantics of the axiom, has the effect of discarding those senses that are less representative within their own synsets or with regards to their hypernyms.

The degree of the $\mathsf{equivalentTo}$ relation is computed in the same way, but replacing $\mathsf{R1}$ by $\mathsf{R2}$. Figure 8 shows the degree assigment to equivalent word senses within a synset in English *S9* and its equivalent synset in Spanish *S9b*. According to these values, it holds that $(\mathsf{equivalentTo})^{\mathcal{I}}(\mathsf{S9}, \mathsf{S9b}) = 1$. In this case, this degree is obtained because both synsets include at least one sense which is completely equivalent to the prototype of the other. These senses are *groin* and *groyne* in *S9* and *espigón* in *S9b*.

In cases of linguistic anisomorphism, or when none of the equivalent senses match the prototype, the degree is no longer equals to 1. This happens in the example shown in Figure 9, where $(\mathsf{equivalentTo})^{\mathcal{I}}(\mathsf{S10}, \mathsf{S10b}) \geq 0.8$. Notice that the pair formed by *marsh* and *marjal* is determinant to obtain this degree. *Marsh* is the prototype of *S10*, whereas *marjal* is not that of *S10b*, but it is the closest sense of the synset and nearly the equivalent of the English sense. Interestingly enough, *ciénaga* and *fen* are totally equivalent, but none of them have the maximum degree with respect to their synset. Therefore, their combined

				S10		
				1	*0.9*	*0.7*
				1	2	3
				marsh	**marshland**	**fen**
S10b	*0.7*	18	**cenagal**	0.7	0.6	1
	0.4	19	**tolla**	0.4	0.3	0.5
	0.5	20	**paúl**	0.5	0.4	0.9
	0.5	21	**paular**	0.5	0.4	0.9
	0.7	22	**pantano**	0.7	0.6	0.9
	0.8	23	**marjal**	0.8	0.7	0.5
	0.7	24	**ciénaga**	0.7	0.6	1

Fig. 9. Equivalence degrees between senses (shown in plain text) and prototype resemblance degrees between senses and synsets (shown in italics)

value is less than the supremum, and they do not affect the final equivalence degree.

Given a fuzzy ontology $\mathcal{K}$ representing WordNet, EuroWordnet and EcoLexicon knowledge, we would be interested in solving three *reasoning tasks*:

- Computing the minimal degree for a synset $\mathsf{sy1}$ being a (possibly non-direct) hyponym of a synset $\mathsf{sy2}$. This can be solved by computing the Best Degree Bound (BDB) $\mathsf{bdb}(\mathcal{K}, (\mathsf{sy1}, \mathsf{sy2}) : \mathsf{hyponymOf})$[3].
- Computing the minimal degree for a synset $\mathsf{sy1}$ being a (possibly non-direct) hypernym of a synset $\mathsf{sy2}$ using $\mathsf{bdb}(\mathcal{K}, (\mathsf{sy1}, \mathsf{sy2}) : \mathsf{hypernymOf})$.
- Computing the minimal degree for a synset $\mathsf{sy1}$ being an equivalent of a synset $\mathsf{sy2}$ using $\mathsf{bdb}(\mathcal{K}, (\mathsf{sy1}, \mathsf{sy2}) : \mathsf{equivalentTo})$.

There are two possibilities to reason under this approach. The first one is to use an appropriate fuzzy ontology reasoner such as `DeLorean`, which is, to the best of our knowledge, the only existing reasoner that supports the expressivity of fuzzy $\mathcal{SR}^{-}\mathcal{I}$. In this paper we have assumed this option. A second option is to create an *ad hoc* crisp representation with the same semantics of the fuzzy ontology (following the same approach as for the reduction procedure internally performed by `DeLorean`) that can be processed with any classical ontology reasoner, such as `Pellet` or `HermiT`. This alternative is discussed in [4].

[3] $\mathsf{bdb}(\mathcal{K}, \tau)$ is the maximal degree α such that $\mathcal{K}$ entails the axiom $\langle \tau \geq \alpha \rangle$ [2].

5 Conclusions

The examples in the previous section shows that fuzzy ontologies allow us to extend WordNet and EuroWordNet with a more fine-grained description of synsets, hyponymic relations and equivalence, which is necessary in specialized knowledge. Fuzzy ontologies facilitate the addition of new synsets and new general axioms, such as the ones that are used to automatically infer the degree of hyponymic and equivalence relations. It would also be possible to add other axioms in order to calculate the hyponymy and equivalence degrees between arbitrary senses, but this remains as a matter for future research. Another direction could be the application of different semantics to calculate the inferred degrees. For example, instead of using axioms and the supremum, we could use a weighted average to avoid overriding less relevant senses. Moreover, we could even consider the fuzzification of other relations in WordNet and EuroWordNet to better represent antonymy, linguistic variations, etc. Last but not least, extending WordNet and EuroWordNet to include imprecise knowledge requires a considerable effort to define synset membership, similarity and equivalence degrees. We assume that experts can assign these values, but it may not be desirable or even possible in several cases, even though they would only be necessary in the synsets representing specialized knowledge. It would be interesting to study and compare different techniques to automatically obtain these degrees from the compared analysis of terminological resources (matching between WordNet and EcoLexicon semantic networks, lexical comparison of WordNet glosses and EcoLexicon definitions), and the development of corpus-based text processing procedures.

References

1. Alonge, A., Bloksma, L., Calzolari, N., Castelln, I., Mart, M., Peter, W., Vossen, P.: The linguistic design of the eurowordnet database. Computers and Humanities 32, 19–43 (1998)
2. Bobillo, F., Delgado, M., Gómez-Romero, J.: Crisp representations and reasoning for fuzzy ontologies. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 17(4), 501–530 (2009)
3. Bobillo, F., Delgado, M., Gómez-Romero, J.: DeLorean: A reasoner for fuzzy OWL 2. Expert Systems with Applications 39(1), 258–272 (2012)
4. Bobillo, F., Gómez-Romero, J., León Araúz, P.: Fuzzy Ontologies for Specialized Knowledge Representation in WordNet. In: Proceedings of the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2012) (2012)
5. Borin, L., Forsberg, M.: From the peoples synonym dictionary to fuzzy synsets - first steps. In: Proceedings of the LREC 2010 workshop Semantic relations. Theory and Applications (2010)
6. De Luca, E.W., Numberger, A.: LexiRes RDF/OWL Editor: Maintaining Multilingual Resources. Journal for Language Technology and Computational Linguistics 24(2), 56–76 (2009)
7. De Luca, E.W., Eul, M., Nürnberger, A.: Converting EuroWordNet in OWL and Extending it with Domain Ontologies. In: Kunze, C., Lemnitzer, L., Osswald, R.

(eds.) Proceedings of the GLDV-2007 Workshop on Lexical-Semantic and Ontological Resources. vol. 336, p. 3948. Informatik. Berichte (2007)
8. Gangemi, A., Guarino, N., Oltramari, A.: Conceptual analysis of lexical taxonomies: the case of wordnet top-level. In: FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems. ACM Press (2001)
9. Gómez-Pérez, A., Corcho, O., Fernández-López, M.: Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer (2004)
10. Lai, L.F., Wu, C.C., Lin, P.Y., Huang, L.T.: Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search. In: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011). pp. 2684–2689 (2011)
11. León Araúz, P., Reimerink, A., Faber, P.: Environmental knowledge in EcoLexicon. In: Jassem, K., Fuglewicz, P., Piasecki, M. (eds.) Proceedings of the Computational Linguistics Applications conference (CLA'11). pp. 9–16 (2011)
12. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. Journal of Web Semantics 6(4), 291–308 (2008)
13. Miller, G.A.: WordNet: A lexical database for English. Communications of the ACM 38(11), 39–41 (1995)
14. Oliveira, G., Gomes, P.: Automatic discovery of fuzzy synsets from dictionary definitions. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI) (2011)
15. Oltramari, A., Gangemi, A., Guarino, N., Masolo, C.: Restructuring WordNet's top-level: The OntoClean approach. In: Simov, K. (ed.) Proceedings of the Ontologies and Lexical Knowledge Bases Workshop (OntoLex'02). pp. 17–26 (2002)
16. Osherson, D., Smith, E.: On the adequacy of prototype theory as a theory of concepts. Cognition 9, 35–58 (1981)
17. Pease, A., Fellbaum, C., Vossen, P.: Building the global wordnet grid. In: Proceedings of the CIL-18 Workshop on Linguistic Studies of Ontology (2008)
18. Rosch, E.: Cognition and Categorization, chap. Principles of categorization. Lawrence Erlbaum Associates (1978)
19. van Assem, M., Gangemi, A., Schreiber, G.: RDF/OWL representation of WordNet (W3C Working Draft) (Jun 2006), `http://www.w3.org/TR/wordnet-rdf/`
20. Vossen, P. (ed.): EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers (1998)
21. Vossen, P.: EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an interlingual index. International Journal of Lexicography 17(2), 161–173 (2004)
22. Vossen, P.: From WordNet, EuroWordNet to the Global Grid. In: eLex2009 conference (2009)
23. Vossen, P., Aguirre, E., Calzolari, N., Fellbaum, C., Hsieh, S., Huang, C., Isahara, H., Kanzaki, K., Marchetti, A., Monachini, M., Neri, F., Raffaelli, R., Rigau, G., M., T.: KYOTO: A system for Mining, Structuring and Distributing Knowledge Across Languages and Cultures. In: Association, E.L.R. (ed.) Proceedings of LREC08 (2008)
24. Vossen, P., Bloksma, L., Rodriguez, H., Climent, S., Calzolari, N., Roventini, A., Bertagna, F., Alonge, A., Peters, W.: The EuroWordnet base concepts and top ontology. Tech. rep., Centre National de la Recherche Scientifique (1998)
25. Vossen, P.: WordNet, EuroWordNet and Global WordNet. Revue française de linguistique appliquée 7, 27–38 (2002/1)
26. Wittgenstein, L.: Philosophical Investigations. Basil Blackwell (1953)
27. Zadeh, L.: A note on prototype theory and fuzzy sets. Cognition 12, 291–297 (1982)