Towards Effective Correction Methods Using WordNet Meronymy Relations

Javier Álvez
Lorea Group

Itziar Gonzalez-Dios
Ixa Group, HiTZ Center

German Rigau
Ixa Group, HiTZ Center

University of the Basque Country (UPV/EHU)
{javier.alvez,itziar.gonzalezd,german.rigau}@ehu.eus

Abstract

In this paper, we analyse and compare several correction methods of knowledge resources with the purpose of improving the abilities of systems that require commonsense reasoning with the least possible human-effort. To this end, we cross-check the WordNet meronymy relation *member* against the knowledge encoded in a SUMO-based first-order logic ontology on the basis of the mapping between WordNet and SUMO. In particular, we focus on the knowledge in WordNet regarding the taxonomy of animals and plants. Despite being created manually, these knowledge resources —WordNet, SUMO and their mapping— are not free of errors and discrepancies. Thus, we propose three correction methods by semi-automatically improving the alignment between WordNet and SUMO, by performing some few corrections in SUMO and by combining the above two strategies. The evaluation of each method includes the required human-effort and the achieved improvement on unseen data from the WebChild project, that is tested using first-order logic automated theorem provers.

1 Introduction

The areas of commonsense knowledge representation and commonsense reasoning are of great interest for their application in many tasks related to Natural Language Processing (NLP) e.g. Recognizing Textual Entailment (RTE) (Bos and Markert, 2006; Dagan et al., 2013; Abzianidze, 2017), Natural Language Inference (NLI) (Bowman et al., 2015) or Interpretable Semantic Textual Similarity (ISTS) (Lopez-Gazpio et al., 2017). In the literature, among the knowledge resources, WordNet (Fellbaum, 1998) is one of the most frequently used semantic resources that is applied to NLP tasks. Furthermore, WordNet interlinks many other semantic resources e.g. the EuroWordNet Top Ontology (Rodríguez et al., 1998), or SUMO[1] (Niles and Pease, 2001).

When linking lexical resources such as WordNet (Fellbaum, 1998) and ontologies such as SUMO (Niles and Pease, 2001), DOLCE (Gangemi et al., 2002) or OpenCYC (Reed and Lenat, 2002), Prevot et al. (2005) generalised these three methodological options: restructuring, populating and aligning. But, moreover, ontologies and lexical resources can also be used to cross-check them and validate the knowledge content encoded.

In order to automatically cross-check the knowledge in WordNet and SUMO, Álvez et al. (2015, 2019) introduced a general framework that enables evaluating the competency of SUMO-based ontologies like Adimen-SUMO (Álvez et al., 2012) and proposed a method for the automatic creation of *competency questions* (CQs) (Grüninger and Fox, 1995). Their proposal is based on several predefined question patterns (QPs) that are instantiated using information from WordNet and its mapping into SUMO (Niles and Pease, 2003). In addition, the authors described an application of first-order logic (FOL) automated theorem provers (ATPs) for the automatic evaluation of the proposed CQs. However, a low percentage of the meronymy pairs from WordNet can be validated against SUMO using the proposed framework, as reported by Álvez and Rigau (2018); Álvez et al. (2018). Overall, three possible causes for this low validation ratio have been identified:

- Incorrect mappings between WordNet and SUMO: two cases are presented in Table 1. The first one is valid because the knowledge from WordNet, SUMO and its mapping is correctly aligned: individuals ($parent_n^1$) with an instance of *BiologicalAttribute* as property can be members of instances of *FamilyGroup* ($family_n^2$). However, the second case is invalid:

[1] http://www.ontologyportal.org

	Part		Whole	
Valid	*parent*$_n^1$: a father or mother; (...)	Subsumed by **BiologicalAttribute**	*family*$_n^2$: primary social group; (...)	Subsumed by **FamilyGroup**
Invalid	*hyaena*$_n^1$: doglike nocturnal mammal (...)	Subsumed by **Canine**	*family_Hyaenidae*$_n^1$: hyenas	Subsumed by **Canine**

Table 1: Valid and invalid examples of the relation *member*

Canine (whole) is characterised as an individual (i.e. not a group); therefore, it cannot have members. In order to be able to validate the pair, *family_Hyaenidae*$_n^1$ should be corrected to be subsumed by *GroupOfAnimals*.

- Discrepancies in the knowledge encoded in WordNet and SUMO: the groups (species, genus, family, order, ...) in the taxonomy of animals and plants are connected by the relation *member* of WordNet, while the relation *member* of SUMO connects individuals (which cannot be groups) to their groups.

- Limitations of ATPs.

In this paper, our aim is shedding light on the sources of difficulty when correcting knowledge resources, which is a mainly manual and never ending task. Exactly, we want to discover which correction methods and strategies lead to maximising the improvement with the least possible human-effort. To this end, we consider three correction approaches: i) the correction of the mapping between WordNet and SUMO on the basis of the WordNet hierarchy and our manual error analysis of the results reported in Álvez et al. (2018); ii) the correction of the knowledge in SUMO in order to its alignment to WordNet; iii) the combination of the previous two approaches. We report on a practical evaluation of the impact of each correction method on unseen data provided by the WebChild project (Tandon et al., 2014, 2017), which is a large collection of commonsense knowledge that has been automatically extracted and disambiguated from Web contents. To the best of our knowledge, this is the first work dealing with the problem of correcting FOL commonsense resources.

Outline. First, we present the related work in the next section and review the knowledge resources and evaluation framework in Section 3. Then, we describe the proposed correction methods in Section 4 and provide the evaluation results in Section 5. Finally, we conclude and outline the future work in Section 6.

2 Related Work

In this section, we present the works related to meronymy knowledge and its acquisition, cross-checking resources, mapping error detection and ontology debugging and repairing.

Meronymy is a semantic relation that *connects* the parts and the whole. This connection can be functional, homeomeric/homeomerous (consisting of similar parts), separable or simultaneous (Campenhoudt, 1996). In the typology of meronymy relations, the most important subrelations are constituent-object, member-collection and material-object. The importance of meronymy is pointed out by vor der Brück and Helbig (2010), which extract meronymy relations from Wikipedia by means of a logic-oriented approach. According to them, meronymy is necessary for many NLP tasks such as question answering. Following their example, if someone asks about the earthquakes in Europe, then the question could be answered thanks to the meronymy relation if we had the data of each European country.

Both manual and automated attempts have been made to acquire meronymy knowledge. Among the first ones, there are more than 22,000 meronymic pairs in WordNet (Fellbaum, 1998), that have been manually constructed and reviewed. WordNet is a large lexical database of English where nouns, verbs, adjectives and adverbs are grouped into sets of synonyms called *synsets*,[2] each one denoting a distinct concept. Moreover, synsets are interlinked

[2]In this paper, we will refer to the synsets using the format *word*$_p^s$, where s is the sense number and p is the part-of-speech: n for nouns and v for verbs e.g. *plant*$_n^2$ means that the word *plant* is a noun and that we are referring to its second sense in WordNet.

by means of lexical-semantic relations. WordNet encodes three main meronymy relations that relate noun synsets: i) *part*, the general meronymy relation; ii) *member*, which relates particulars and groups; and iii) *substance*, which relates physical matters and things. In total, WordNet v3.0 includes 22,187 (ordered) meronymy relations (around 10 % of the relations between synset pairs in WordNet): 9,097 pairs using *part*, 12,293 pairs using *member* and 797 pairs using *substance*. For example, the synsets $tongue_n^1$ and $mouth_n^1$ are related by *part*, $lamb_n^1$ and $genus\_Ovis_n^1$ are related by *member*, and $neuroglia_n^1$ and $glioma_n^1$ are related by *substance*.

Furthermore, additional relations were manually added to WordNet in Lebani and Pianta (2012) on the basis of featural descriptions. However, the coverage of the collected meronymy knowledge is quite restricted. This limitation is also present in some automated proposals like ConceptNet (Speer et al., 2017), which has been obtained by crowdsourcing and contains around 20,000 meronymy relation pairs between non-disambiguated words.

The coverage of the automatically acquired meronymy knowledge is larger in other works. For example, PWKB (the part-whole KB) (Tandon et al., 2016), which has been integrated into WebChild v2.0 (Tandon et al., 2017), consists of almost 6 millions of disambiguated meronymy pairs that have been obtained from Web contents and image tags by combining pattern-based information extraction methods and logical reasoning. However, this KB suffers from low salience since more pairs were obtained by expanding a small set of relations. A complementary resource is hasPartKB (Bhakthavatsalam et al., 2020), which contains more salient and accurate hasPart relations (around 50,000) extracted from a large corpus of generic statements. Finally, Quasimodo (Romero et al., 2019) and Aristo Tuple KB (Mishra et al., 2017) contain several thousands of non-disambiguated meronymy pairs, but their coverage is rather limited.

The knowledge in all the above cited resources is restricted to relation pairs. Regarding general knowledge, SUMO (Niles and Pease, 2001) contains both facts and axioms that encode more abstract information and properties about meronymy.

In relation with cross-checking knowledge resources, Álvez et al. (2008) exploit the EuroWordNet Top Ontology (Rodríguez et al., 1998) and its mapping to WordNet for detecting many ontological conflicts and inconsistencies in the WordNet nominal hierarchy.

Most of the works presented for error correction both in the mapping and in the ontologies have been proposed for OWL ontologies. Relating mapping error detection and correction, many methods have been proposed to detect mapping errors or invalid mappings between ontologies, knowledge resources, dictionaries and thesauri (Reis et al., 2015). Similar to us, Pathak and Chute (2009) reasoning strategies for the biomedical domain. Exactly, they use description logics to detect inconsistencies since they consider that ontologies are consistent, and therefore, errors come from the mappings. Wang and Xu (2012) divided the mapping errors in four categories (from now on, Wang&Xu classification): redundant, imprecise, inconsistent and abnormal mappings. Correction strategies are presented in Abacha et al. (2016) for the biomedical domain, where questions are proposed to experts in order to validate the mapping and the ontology. Surveys on mapping maintenance and ontology matching are respectively presented in Reis et al. (2015) and Ochieng and S. Kyanda (2018). Relating ontology error detection, recent work on ontology debugging involves detecting hidden modelling errors: Teymourlouie et al. (2018) use DBpedia during the ontology debugging process to detect contradictions in ontologies that seem coherent. Unfortunately, as far as we know, no correction approach has been proposed.

3 Knowledge Resources and Evaluation Framework for WordNet Meronymy

In this section, we describe the knowledge resources and framework that enable the automatic evaluation of the meronymy relation *member* of WordNet by using Automated Theorem Provers (ATPs).

Adimen-SUMO (Álvez et al., 2012) is a first-order logic (FOL) ontology obtained by means of a suitable transformation of most of the knowledge (around 88 % of the axioms) in the *top* and *middle* levels of SUMO (Niles and Pease, 2001). Adimen-SUMO enables the application of state-of-the-art FOL ATPs such as Vampire (Kovács and Voronkov, 2013) and E (Schulz, 2002) in order to automatically reason on the basis of the knowledge in SUMO (Niles and Pease, 2001). SUMO is organised around the notions of *particulars* (also called *instances* or *objects*) and *classes* by means of the meta-predicates *instance* and *subclass*. Amongst them, SUMO also differentiates *relations* and *attributes*, and provides specific predicates for their use that are inherited by Adimen-SUMO e.g. *subrelation* and *attribute*. We denote the nature of SUMO concepts by adding as subscript the following symbols: o for SUMO objects, c for SUMO classes, r for SUMO relations, a for SUMO attributes and A for classes of SUMO attributes. For example: $Waist_o$, $GroupOfAnimals_c$, $material_r$, $Solid_a$ and $BiologicalAttribute_A$.

WordNet and SUMO are linked by means of a semantic mapping that connects WordNet synsets to SUMO concepts using three relations: *equivalence*, *subsumption* and *instance* (Niles and Pease, 2003). The mapping relation *equivalence* connects WordNet synsets and SUMO concepts that are semantically equivalent. *Subsumption* (or *instance*) is used when the semantics of the WordNet synsets is less general than (or instance of) the semantics of the SUMO concepts to which the synsets are connected. For example, the synset $lamb^1_n$ is connected to $Lamb_c$ by *equivalence* and $neuroglia^1_n$ is connected to $Tissue_c$ by *subsumption*. From now on, we denote the semantic mapping relations by concatenating the symbols '$=$ ' (*equivalence*), '$+$' (*subsumption*) and '@' (*instance*) to the corresponding SUMO concept e.g. $lamb^1_n$ is connected to $Lamb_c$= and $neuroglia^1_n$ is connected to $Tissue_c+$.

For the automatic evaluation of the WordNet meronymy relations, we apply the framework introduced in Álvez et al. (2019), which is an adaptation of the method proposed in Grüninger and Fox (1995) for the formal design and evaluation of ontologies on the basis of *Competency Questions* (CQs). This framework enables the use of ATPs in order to automatically classify CQs as follows: CQs are decided to be *passing* (if proved to be entailed by the ontology), *non-passing* (their negations are proved to be entailed by the ontology) and *unresolved* (neither the CQs nor their negations are proved to be entailed by the ontology).

Furthermore, we adapt the *Question Patterns* (QPs) for the meronymy relation *member* introduced in Álvez and Rigau (2018); Álvez et al. (2018), which enable the translation of its semantics into a suitable CQ or yield a semantically incorrect conjecture according to the restrictions for relations provided by SUMO. Those QPs employs the translation of the mapping information of synsets into Adimen-SUMO statements that is described in Álvez et al. (2019), which characterises the semantics of WordNet synsets in terms of SUMO instances and requires the use of a new variable for each synset. There is a different QP for each possible combination of mapping relations, which states the quantification of the introduced variables and the logical connectives that enable the construction of the final CQ. For example, the synsets $sheep^1_n$ and $flock^5_n$ are respectively connected to $Sheep_c$= and $Group_c+$. Thus, we use the second QP proposed in Álvez and Rigau (2018) because of the use of the mapping relations *equivalence* and *subsumption*, and obtain the following conjecture:

```
(forall (?X)
    (=>
        ($instance ?X Sheep)
        (exists (?Y)
            (and
                ($instance ?Y Group)
                (member ?X ?Y) ) ) ) )
```

Finally, the WordNet meronymy pairs on *member* can be classified according to the following categories depending of: a) if the *member* pair is translated into a CQ, then it is decided to be *validated*, *unvalidated* or *unknown* if the CQ is passing, non-passing or unresolved respectively; b) the *member* pairs that yield to semantically incorrect conjectures are classified as *unvalidated*.

By using the above described framework and regarding the original versions of SUMO and its mapping from WordNet, from the 12,293 *member* pairs provided by WordNet only 19 are validated, while 11,963 pairs are unvalidated and 311 remains unknown. Moreover, from the 11,963 unvalidated pairs, only 24 yield a correct CQ. That is, the direct application of the introduced evaluation framework just allows to validate a mere 1.5% of the member pairs encoded in WordNet and, apparently, most of the unvalidated pairs yield semantically incorrect SUMO conjectures. This may be an indication of both misalignment in the knowledge encoded in WordNet and SUMO and the existence of a large number of discrepancies in their mapping.

4 Knowledge Correction Methods

In this section, we introduce the proposed correction strategies for knowledge resources. For our analysis and interventions, we have used the information contained in the Multilingual Central Repository (Gonzalez-Agirre et al., 2012). Exactly, we have consulted: the Basic Level Concepts (BLCs) (Izquierdo et al., 2007), which are frequent and salient concepts in WordNet that try to represent as many concepts as possible (abstract concepts) and as many distinctive features as possible (concrete concepts); the Top Concept Ontology (TCO) (Rodríguez et al., 1998); the Semantic Files (SF) from WordNet and the WordNet Domains (WND) (Bentivogli et al., 2004). Moreover, we have also consulted SUMO and its documentation.

4.1 Correction of the mapping

We have performed two kinds of interventions in order to realign the mapping between WordNet and SUMO: 1) structural corrections in the BLCs; 2) opportunistic corrections based on an error analysis. In both phases, we have performed a manual analysis that has served as the basis for proposing some criteria in order to automatically propagate or expand the corrections.

For performing structural corrections, from the 800 BLCs in WordNet we have manually inspected the topmost 200 ones. To that end, we have used information from WordNet, TCO and SUMO and for each BLC, we have decided whether the mapping was correct or not. If we have not considered it as correct, we have proposed a new mapping for it. During this intervention we have tried to make as few changes as possible; so, if the original mapping was acceptable, then it has not been changed. It is important to note that at this correction phase we have considered all the synsets from WordNet without restricting to those that are related to meronymy, that is, what we correct can appear or not in our benchmark.

This way, we have manually corrected the mapping of 50 BLCs (25 %). This manual correction can be classified in two types: a) groups that are characterised as individual classes (38 synsets), most of them related to plants and animals; b) punctual mapping errors (12 synsets). Following Wang&Xu classification, these errors are imprecise or inconsistent mappings: exactly, 10 are imprecise mappings and 40 are inconsistent. For example:

- $\textit{dicot\_genus}^1_n$ ("*genus of flowering plants having two cotyledons (embryonic leaves) in the seed which usually appear at germination*") and $\textit{fish\_genus}^1_n$ ("*any of various genus of fish*") belong to the first type of corrections because they were incorrectly connected to $\textit{FloweringPlant}_c+$ and $\textit{Fish}_c+$ and have now been linked to $\textit{Group}_c+$ and $\textit{GroupOfAnimals}_c+$ respectively.
- $\textit{agency}^1_n$ ("*an administrative unit of government:*") and $\textit{substance}^1_n$ ("*the real physical matter of which a person or thing consists:*") belong to the second type of corrections because $\textit{agency}^1_n$ was imprecisely connected to $\textit{PoliticalOrganization}_c=$ (updated to $\textit{GovernmentOrganization}_c+$) and $\textit{substance}^1_n$ was incorrectly connected to $\textit{Object}_c=$ (corrected to $\textit{Substance}_c=$).

During this intervention, we have been able to revise and correct when necessary around 20 BLCs per hour and, in total, we have spent 10 hours.

After the manual correction of the BLCs, we have automatically propagated the corrected BLC mappings to their hyponyms based on the following criterion:

> Propagate the corrected as long as the hyponym and its BLC are equally mapped in the original mapping.

By proceeding in this way, we have corrected a total of 3,883 mappings.

For the opportunistic correction of the mapping based on an error analysis, we have inspected the unclassified pairs in the experimentation introduced at the end of Section 3. More concretely, we have grouped the synset pairs according to their mapping to SUMO and ordered them by frequency. Apparently, most of the detected errors are due to the fact that species, genera, families, orders, etc. (taxonomic biological classification) and galaxies, constellations, etc. (collections of planets, stars, asteroids, etc.) are connected to SUMO classes representing individuals and not groups (group errors as presented before and inconsistent according to Wang&Xu classification). In order to correct this type of errors, we have designed four very simple heuristics:

1. If the synset is an hyponym of $\textit{group}^1_n$ in WordNet, is connected to both *Animal*+ and *Group*+ in the TCO, is connected to a subclass of $\textit{Animal}_c$ in SUMO and some of the words *family*, *genus*, *order*, *suborder*, *class*, *phylum*, *subphylum*, *kingdom*, *subkingdom*, *division*, *subdivision*, *algae*, *superfamily*, *subfamily*, *superorder*, *group*, *subclass* or *superclass* occurs in its gloss, then map the synset to $\textit{GroupOfAnimals}_c+$.

2. If the synset is a hyponym of $group_n^1$ in WordNet, is connected to both *Plant*+ and *Group*+ in the TCO, is connected to a subclass of $Plant_c$ in SUMO and some of the words *family*, *genus*, *order*, *suborder*, *class*, *phylum*, *subphylum*, *kingdom*, *subkingdom*, *division*, *subdivision*, *algae*, *superfamily*, *subfamily*, *superorder*, *group* occurs in its gloss, then map the synset to $Group_c$+.

3. If the synset is a hyponym of $group_n^1$ in WordNet, is connected to *Group*+ in the TCO, is connected to a subclass of either $Microorganism_c$, $Virus_c$, $Bacterium_c$ or $Fungus_c$ in SUMO and some of the words *family*, *genus*, *order*, *suborder*, *class*, *phylum*, *subphylum*, *kingdom*, *subkingdom*, *division*, *subdivision*, *algae*, *superfamily*, *subfamily*, *superorder*, *group* occurs in its gloss, then map the synset to $Group_c$+.

4. If the synset is connected to a subclass of $AstronomicalBody_c$ in SUMO and the word *constellation* occurs in its gloss, then map the synset to $Group_c$+.

It is worth noting that there is no concept for representing groups of either plants, microorganisms, viruses, bacteria, fungi or astronomical bodies in SUMO. For example, the synset $animal\_kingdom_n^1$ ("*taxonomic kingdom comprising all living or extinct animals*") was incorrectly connected to $Animal_c$= and its mapping has been corrected to $GroupOfAnimals_c$+.

Furthermore, corrections have been also propagated as described for structural corrections. This way, the mapping of 1,961 synsets has been corrected with an human-effort of 2 hours.

4.2 Matching Knowledge Discrepancies

The objective of this intervention is detecting and solving the knowledge discrepancies between WordNet and SUMO that prevent the validation of many pairs where the mapping information is correct. For this purpose, we have augmented the manual error analysis described in the above subsection by also considering unvalidated pairs.

Overall, most of the detected conflicts are related to organisms. With respect to unvalidated pairs, the main problem is that the relation between taxonomic groups cannot be expressed in terms of SUMO due to the domain restrictions of the SUMO predicate $member_r$. In particular, the first argument of $member_r$ is restricted to be an instance of $SelfConnectedObject_c$, which is disjoint with the SUMO class $Collection_c$ and hence disjoint with the SUMO class $Group_c$. Consequently, we cannot construct a SUMO statement that expresses that an instance of $Group_c$ is a member of another instance of $Group_c$, as required for the validation of the examples in Table 1. In order to overcome this problem, we have proposed to replace the domain restriction of the first argument of the SUMO predicate $member_r$: instead of being instance of $SelfConnectedObject_c$, our proposal is restricting the first argument of $member_r$ to be instance of $Object_c$, which is superclass of $Group_c$ (1 axiom corrected). In addition, the characterization of $GroupOfPeople_c$ and $GroupOfAnimals_c$ has to be accordingly updated: in the new proposed axiomatization, the members of $GroupOfPeople_c$ can be instances of either $Human_c$ or $GroupOfPeople_c$, and the members of an instance of $GroupOfAnimals_c$ can be either instances $Animal_c$ that are not instance of $Human_c$ or instances of $GroupOfAnimals_c$ (2 axioms corrected).

Regarding unclassified *member* pairs, by a manual inspection of SUMO we have detected that the characterization of concepts representing groups is too weak. More concretely, there is no concept for the representation of groups of plants and the existing concepts for the representation of groups —$Group_c$ for general groups; $GroupOfPeople_c$, $AgeGroup_c$, $FamilyGroup_c$, $SocialUnit_c$, $EthnicGroup_c$ and $BeliefGroup_c$ for groups of people; GroupOfAnimals and $Brood_c$ for groups of animals— are only partially characterised. More concretely, the nature of the members of each kind of group is properly restricted, but individuals (including the instances of $Agent_c$) are not restricted to belong to some groups. In order to solve these issues, we have created and characterised a new concept for groups of plants ($GroupOfPlants_c$, 3 new axioms) and introduced another 9 new axioms for the characterization of groups).

In total, our interventions have required a human-effort of 2 hours.

4.3 Joining Mapping and Ontology Corrections

In order to integrate both interventions, we have made some changes in the mapping.

On one hand, we have updated the mapping of

9 synsets from the top 200 BLCs from $Group_c+$ to $GroupOfPlants_c+$, and this change has been propagated to 1,961 synsets. On the other hand, we have redefined the second heuristic presented in Subsection 4.1 in order to map the synset to $GroupOfPlants_c+$. The updated heuristic is the following:

- If the synset is an hyponym of the synset $group_n^1$ in WordNet and is connected to a subclass of $Plant_c$ in SUMO, then map the synset to $GroupOfPlants_c+$

This heuristic is directly applied to 356 synsets and propagated to another 85 synsets. In total, we have updated 2,411 mappings that were previously mapped to $Group_c+$.

All these interventions have been performed with almost no human-effort.

5 Evaluation

In this section, we evaluate the proposed knowledge correction methods on both seen and unseen data, which is extracted from the WebChild project. In Table 2, we report on the results obtained by applying the evaluation framework described in Section 3 for the different intervention phases in WordNet (initial, correction of the mapping, matching knowledge discrepancies and joint intervention) and in WebChild data (initial and joint intervention). For each phase, we provide the number of pairs that are validated/unvalidated/unknown (Validated, Unvalidated and Unknown columns respectively) and three metrics that measure the performance of the evaluation: recall (calculated as the ratio between validated pairs and total pairs); precision (calculated as the ratio between validated pairs and validated+unvalidated pairs); and $F1$ (calculated as the harmonic mean of precision and recall) values. In the case of unvalidated pairs, we provide both the total number of pairs (T column) and the number of pairs which yield a correct CQ (C column).

Regarding seen data, it is easy to see that matching knowledge discrepancies outperforms mapping correction, although the improvement is low in both cases: correcting the mapping turns almost a half of the previously unvalidated pairs into unknown while matching knowledge discrepancies increases a bit the number of validated pairs. However, by combining both interventions the improvement is much higher: the amount of validated pairs is 500 times bigger and the amount of unvalidated pairs is almost 15 times smaller.

With respect to the data extracted from the WebChild project, the combined intervention heavily improves the results again, although the impact is a bit lower: many pairs still remain unknown and the ratio between validated and unvalidated pairs is lower than in the case of WordNet. For a better understanding of these results, we have manually analysed a sample of WebChild pairs consisting of five randomly selected cases from each output (validated, unvalidated and unknown).

Considering the validated pairs, 4/5 have been classified as validated for good reasons, e.g. $Acrocomia_n^1$ is member of $Palmae_n^1$. The only error is a wrong pair in the knowledge base: the synset $genus_n^2$ (“*(biology) taxonomic group containing one or more species*”) is incorrectly asserted to be member of $Carapidae_n^1$ (“*pearlfishes: related to the Brotulidae*”).

From the unvalidated pairs, 2/5 pairs are wrong so they have been correctly classified as unvalidated e.g. $superphylum_n^1$ is not a member of $locative\_role_n^1$. However, 3/5 pairs are correct and should have been validated, but there are mapping errors e.g. $Auriculariaceae_n^1$ is member of $Tremellales_n^1$, although the pair is classified as unvalidated because $Tremellales_n^1$ is still mapped to $Fungus_c$.

Finally, in relation to unknown pairs, one pair is correct —$rice\_weevil_n^1$ is member of $Sitophylus_n^1$— and 4/5 pairs are wrong, e.g. $relative_n^1$ is not a member of $Ming\_dynasty_n^1$. However, these pairs cannot be resolved by ATPs because the required information is missed in the ontology or, as in the case of the correct pair, due to resource (specially time) restrictions.

6 Conclusions and Future Work

In this paper we have reported on several correction methods for the knowledge about meronymy in WordNet, SUMO and their mapping with the aim of improving the abilities of systems that require commonsense reasoning. To this end, we have applied FOL ATPs on a large set of CQs automatically constructed on the basis of several predefined QPs and the knowledge of the involved resources. Since finding and correcting errors in knowledge resources has always been time-consuming and required quite a lot of manual work, we have focused on the human-effort required for each cor-

Data	Phase	Validated	Unvalidated T	Unvalidated C	Unknown	Recall	Precision	$F1$
WordNet	Initial	19	11,963	24	311	0.002	0.002	0.002
	Mapping	29	6,561	5,811	5,703	0.002	0.004	0.003
	Knowledge	132	11,603	30	558	0.011	0.011	0.011
	Joint	10,071	808	58	1,414	0.819	0.926	0.869
WebChild	Initial	82	35,377	102	3,368	0.002	0.002	0.002
	Joint	18,569	3,526	136	17,032	0.475	0.840	0.607

Table 2: Evaluation of the knowledge correction methods

rection strategy. As a result, we have been able to increase the number of WordNet pairs that can be validated against the knowledge in SUMO with a total human-effort of 14 hours. All the resources —the corrected mapping, the augmented ontology and the experimental reports— are available at the Adimen-SUMO webpage.[3]

By analysing our evaluation results on WordNet, it seems at first glance to be worth investing effort correcting and matching the knowledge of the involved resources, since the improvement is slightly higher (see Table 2) and has required less human-effort (2 hours against 12 hours), although the combined strategy leads to the better results with almost no additional human-effort. More concretely, at the initial stage only a 0.15 % of the *member* pairs in WordNet could be validated and our interventions have enabled the validation of almost 82 % of the pairs.

Regarding the evaluation on unseen data, we have confirmed that our interventions are correct, although there is still a lot of work to do. Furthermore, our detailed analysis revealed some aspects for future work. For example, the capture of metonymy, solving additional misalignments (e.g. classifying humans as animals) and the need of analysing the inheritance of relations.

Moreover, we plan to test if the improved knowledge resources also obtain better results in other benchmarks based on antonymy and semantic roles (Álvez et al., 2017), and we would like to carry out similar experiments in other datasets e.g. BLESS[4] (Baroni and Lenci, 2010). Additionally, we also plan to consider additional WordNet relations: for example, the remaining relations about meronymy *part* and *substance*, cause or the semantic roles described in the Morphosemantic links (Fellbaum et al., 2009).

Longer term research includes a new mapping between WordNet and SUMO on the basis of formulae instead of labels, with the aim of providing a more precise definition of the semantics of synsets in terms of the SUMO language.

[3] http://adimen.si.ehu.es/web/AdimenSUMO

[4] https://sites.google.com/site/geometricalmodels/shared-evaluation

Acknowledgements

This work has been partly supported by: (i) the project DeepKnowledge (PID2021-127777OB-C21), funded by MCIN/AEI/10.13039/501100011033 and FEDER *Una manera de hacer Europa*; (ii) the European Union (ERDF funds) under grant PID2020-112581GB-C22; (iii) the project AWARE (TED2021-131617B-I00), funded by MCIN/AEI/10.13039/501100011033 and the European Union (NextGenerationEU/PRTR); (iv) the Basque Government (IXA excellence research group IT1570-22); (v) the University of the Basque Country (UPV/EHU) under project LoRea (GIU21/044).

References

A. B. Abacha, J. C. Dos Reis, Y. Mrabet, C. Pruski, and M. Da Silveira. 2016. Towards natural language question generation for the validation of ontologies and mappings. *Journal of Biomedical Semantics*, 7(1):48.

L. Abzianidze. 2017. LangPro: Natural language theorem prover. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP 2017)*, pages 115–120. Association for Computational Linguistics.

J. Álvez, J. Atserias, J. Carrera, S. Climent, E. Laparra, A. Oliver, and G. Rigau. 2008. Complete and consistent annotation of WordNet using the Top Concept Ontology. In *Proc. of the* 6^{th} *Int. Conf. on Language Resources and Evaluation (LREC 2008)*, pages 1529–1534. European Language Resources Association (ELRA).

J. Álvez, I. Gonzalez-Dios, and G. Rigau. 2018. Cross-checking WordNet and SUMO using meronymy. In *Proc. of the 11th Int. Conf. on Language Resources and Evaluation (LREC 2018)*, pages 4570–4577. European Language Resources Association (ELRA).

J. Álvez, P. Lucio, and G. Rigau. 2012. Adimen-SUMO: Reengineering an ontology for first-order reasoning. *Int. J. Semantic Web Inf. Syst.*, 8(4):80–116.

J. Álvez, P. Lucio, and G. Rigau. 2015. Improving the competency of first-order ontologies. In *Proc. of the 8th Int. Conf. on Knowledge Capture (K-CAP 2015)*, pages 15:1–15:8. ACM.

J. Álvez, P. Lucio, and G. Rigau. 2017. Black-box testing of first-order logic ontologies using WordNet. *CoRR*, abs/1705.10217.

J. Álvez, P. Lucio, and G. Rigau. 2019. A framework for the evaluation of SUMO-based ontologies using WordNet. *IEEE Access*, 7:36075–36093.

J. Álvez and G. Rigau. 2018. Towards cross-checking WordNet and SUMO using meronymy. In *Proc. of the 9th Global WordNet Conference (GWC 2018)*, pages 25–33.

M. Baroni and A. Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

L. Bentivogli, P. Forner, B. Magnini, and E. Pianta. 2004. Revising the Wordnet domains hierarchy: semantics, coverage and balancing. In *Proc. of the Workshop on Multilingual Linguistic Ressources*, pages 101–108. Association for Computational Linguistics, COLING.

S. Bhakthavatsalam, K. Richardson, N. Tandon, and P. Clark. 2020. Do dogs have whiskers? A new knowledge base of haspart relations. *CoRR*, abs/2006.07510.

J. Bos and K. Markert. 2006. Recognising textual entailment with robust logical inference. In *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, pages 404–426, Berlin, Heidelberg. Springer.

S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 632–642. Association for Computational Linguistics.

M. Van Campenhoudt. 1996. Recherche d'équivalences et structuration des réseaux notionnels: le cas des relations méronymiques. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(1):53–83.

I. Dagan, D. Roth, M. Sammons, and F. M. Zanzotto. 2013. Recognizing textual entailment: Models and applications. *Synthesis Lectures on Human Language Technologies*, 6(4):1–220.

C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

C. Fellbaum, A. Osherson, and P. E. Clark. 2009. Putting semantics into WordNet's "morphosemantic" links. In *Human Language Technology. Challenges of the Information Society*, pages 350–358, Berlin, Heidelberg. Springer.

A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. 2002. Sweetening ontologies with DOLCE. In A. Gómez-Pérez et al., editor, *Knowledge Engin. and Knowledge Manag.: Ontologies and the Semantic Web*, LNCS 2473, pages 166–181. Springer.

A. Gonzalez-Agirre, E. Laparra, and G. Rigau. 2012. Multilingual Central Repository version 3.0. In *Proc. of the 8th Int. Conf. on Language Resources and Evaluation (LREC 2012)*, pages 2525–2529. European Language Resources Association (ELRA).

M. Grüninger and M. S. Fox. 1995. Methodology for the design and evaluation of ontologies. In *Proc. of the Workshop on Basic Ontological Issues in Knowledge Sharing (IJCAI 1995)*.

R. Izquierdo, A. Suárez, and G. Rigau. 2007. Exploring the automatic selection of basic level concepts. In *Proc. of the Int. Conf. on Recent Advances on Natural Language Processing (RANLP'07)*, volume 7.

L. Kovács and A. Voronkov. 2013. First-order theorem proving and Vampire. In N. Sharygina and H. Veith, editors, *Computer Aided Verification*, LNCS 8044, pages 1–35. Springer.

G. Lebani and E. Pianta. 2012. Encoding commonsense lexical knowledge into WordNet. In *Proc. of the 6th Global WordNet Conference (GWC 2012)*, pages 159–166.

I. Lopez-Gazpio, M. Maritxalar, A. Gonzalez-Agirre, G. Rigau, L. Uria, and E. Agirre. 2017. Interpretable semantic textual similarity: Finding and explaining differences between sentences. *Knowledge-Based Systems*, 119:186 – 199.

B. D. Mishra, N. Tandon, and P. Clark. 2017. Domain-targeted, high precision knowledge extraction. *Transactions of the Association for Computational Linguistics*, 5(0):233–246.

I. Niles and A. Pease. 2001. Towards a standard upper ontology. In *Proc. of the 2nd Int. Conf. on Formal Ontology in Information Systems (FOIS 2001)*, pages 2–9. ACM.

I. Niles and A. Pease. 2003. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proc. of the IEEE Int. Conf.*

on Inf. and Knowledge Engin. (IKE 2003), volume 2, pages 412–416. CSREA Press.

P. Ochieng and Swaib S. Kyanda. 2018. Large-scale ontology matching: State-of-the-art analysis. *ACM Comput. Surv.*, 51(4):75:1–75:35.

J. Pathak and C. Chute. 2009. Debugging mappings between biomedical ontologies: Preliminary results from the NCBO bioportal mapping repository. *Nature Proceedings*, 2009:95–98.

L. Prevot, S. Borgo, and A.Oltramari. 2005. Interfacing ontologies and lexical resources. In *Proc. of OntoLex 2005 - Ontologies and Lexical Resources*, pages 91–102.

S. L. Reed and D. B. Lenat. 2002. Mapping ontologies into Cyc. In *Papers from Workshop on Ontologies For The Semantic Web (AAAI 2002)*, pages 1–6. AAAI Press.

J. C.r Dos Reis, C. Pruski, and C. Reynaud-Delaître. 2015. State-of-the-art on mapping maintenance and challenges towards a fully automatic approach. *Expert Systems with Applications*, 42(3):1465 – 1478.

H. Rodríguez, S. Climent, P. Vossen, L. Bloksma, W. Peters, A. Alonge, F. Bertagna, and A. Roventini. 1998. The top-down strategy for building EuroWordNet: Vocabulary coverage, base concepts and top ontology. In *EuroWordNet: A multilingual database with lexical semantic networks*, pages 45–80. Springer.

J. Romero, S. Razniewski, K. Pal, J. Z. Pan, A. Sakhadeo, and G. Weikum. 2019. Commonsense properties from query logs and question answering forums. In *Proc. of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, pages 1411—1420, New York, NY, USA. ACM.

S. Schulz. 2002. E - A brainiac theorem prover. *AI Communications*, 15(2-3):111–126.

R. Speer, J. Chin, and C. Havasi. 2017. ConceptNet 5.5: An open multilingual graph of general knowledge. In *Proc. of the 31st AAAI Conference on Artificial Intelligence (AAAI'17)*, pages 4444—4451. AAAI Press.

N. Tandon, G. de Melo, F. Suchanek, and G. Weikum. 2014. WebChild: Harvesting and organizing commonsense knowledge from the Web. In *Proc. of the 7th ACM Int. Conference on Web Search and Data Mining (WSDM '14)*, pages 523–532. Association for Computing Machinery.

N. Tandon, G. de Melo, and G. Weikum. 2017. WebChild 2.0 : Fine-grained commonsense knowledge distillation. In *Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, pages 115–120.

N. Tandon, C. Hariman, J. Urbani, A. Rohrbach, M. Rohrbach, and G. Weikum. 2016. Commonsense in parts: Mining part-whole relations from the web and image tags. In *Proc. of the 30th AAAI Conf. on Artif. Intell. (AAAI 2016)*, pages 243–250. AAAI Press.

M. Teymourlouie, A. Zaeri, M. Nematbakhsh, M. Thimm, and S. Staab. 2018. Detecting hidden errors in an ontology using contextual knowledge. *Expert Systems with Applications*, 95:312–323.

T. vor der Brück and H. Helbig. 2010. Meronymy extraction using an automated theorem prover. *Journal for Language Technology and Computational Linguistics*, 25(1):57–81.

Peng Wang and Baowen Xu. 2012. Debugging ontology mappings: a static approach. *Computing and Informatics*, 27(1):21–36.