Commonsense Reasoning
Using WordNet and SUMO: a Detailed Analysis

Javier Álvez
LoRea Group

Itziar Gonzalez-Dios
Ixa Group

German Rigau
Ixa Group

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

Abstract

We describe a detailed analysis of a sample of large benchmark of commonsense reasoning problems that has been automatically obtained from WordNet, SUMO and their mapping. The objective is to provide a better assessment of the quality of both the benchmark and the involved knowledge resources for advanced commonsense reasoning tasks. By means of this analysis, we are able to detect some knowledge misalignments, mapping errors and lack of knowledge and resources. Our final objective is the extraction of some guidelines towards a better exploitation of this commonsense knowledge framework by the improvement of the included resources.

1 Introduction

Any ontology tries to provide an explicit formal semantic specification of the concepts and relations in a domain (Noy and McGuinness, 2001; Guarino and Welty, 2002; Guarino and Welty, 2004; Gruber, 2009; Staab and Studer, 2009; Álvez et al., 2012). As with other software artefacts, ontologies typically have to fulfil some previously specified requirements. Usually both the creation of ontologies and the verification of its requirements are manual tasks that require a significant amount of human effort. In the literature, some methodologies collect the experience in ontology development (Gómez-Pérez et al., 2004; Guarino and Welty, 2004) and in ontology verification (Gangemi et al., 2006).

In order to evaluate the *competency* of SUMO-based ontologies in the sense proposed by Grüninger and Fox (1995), Álvez et al. (2019) propose a method for the semi-automatic creation of *competency questions* (CQs). Concretely, they adapt the methodology to evaluate the ontologies so that it can be automatically applied using automated theorem provers (ATPs). The construction of CQs is based on several predefined *question patterns* (QPs) that yield a large set of problems (dual conjectures) by using information from WordNet and its mapping into SUMO. For example, the synsets $machine^1_v$ and $machine^1_n$ are related by the morphosemantic relation *instrument* in the *Morphosemantic Links* (Fellbaum et al., 2009) of WordNet, as depicted in Figure 1. In the same figure, the mappings of the synsets are also provided: $machine^1_n$ and $machine^1_v$ are connected to $Machine_c{=}$ and $Making_c{+}$, where the symbol '=' means that $machine^1_n$ is semantically equivalent to the $Machine_c$, while '+' means that the semantics of $Making_c$ is more general than the semantics of $machine^1_v$. Hence, it is possible to state the the relationship of $machine^1_n$ and $machine^1_v$ in terms of SUMO as follows:

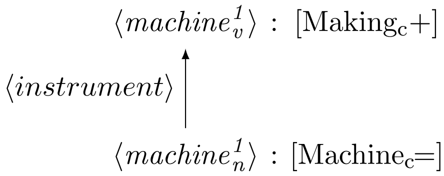

Figure 1: An example of WordNet and its mapping to SUMO

```
(forall (?Y)                                    (1)
    (=>
        (instance ?Y Machine)
        (exists (?X)
            (and
                (instance ?X Making)
                (instrument ?X ?Y)))))
```

The problem that results from Figure 1 consists of the above conjecture, which is considered to be true according to our commonsense knowledge, and its negation, which is therefore assumed to be false.

State-of-the-art ATPs for first-order logic (FOL) such as Vampire (Kovács and Voronkov, 2013) or E (Schulz, 2002) have been proved to provide advanced reasoning support to large FOL conversions of expressive ontologies (Ramachandran et al., 2005; Horrocks and Voronkov, 2006; Pease and Sutcliffe, 2007; Álvez et al., 2012). However, the semi-decidability of FOL and the poor scalability of the known decision procedures have been usually identified as the main drawbacks for the practical use of FOL ontologies. In particular, given an unsolved problem (i.e. a problem such that ATPs do not find any proof for its pair of conjectures) it is not easy to know if (a) the conjectures are not entailed by the ontology or (b) although some of the conjecture is entailed, ATPs have not been able to find the proof within the provided execution-time and memory limits. On the contrary, given a solved problem, it is hard to know whether the solution is obtained for a good reason, because an expected result does not always indicate a correct ontological modelling.

In this paper, we provide a detailed analysis of the large commonsense reasoning benchmarks created semi-automatically by (Álvez et al., 2017; Álvez et al., 2019). The aim of this analysis is to shed light on the commonsense reasoning capabilities of both the benchmark and the involved knowledge resources. To that end, we have randomly selected a sample of 169 problems (1% of the total) following a uniform distribution and manually inspected their source knowledge and results. By means of this detailed analysis, we are able to evaluate the quality of automatically created benchmarks of problems and to detect hidden problems and misalignments between the knowledge of WordNet, SUMO and their mapping.

Outline of the paper. In order to make the paper self-contained, we first introduce the ontology, the mapping to WordNet and the evaluation framework in Section 2. Next, we provide a full summary and the main conclusions obtained from our manual analysis in Section 3. Then, we individually examine some of the problems in Section 4. Finally, we provide some conclusions and discuss future work in Section 5.

2 Commonsense Reasoning Framework

In this section we describe briefly the whole commonsense reasoning reasoning framework. First, we present the knowledge resources needed and then the reasoning framework.

2.1 Resources

The resources we present in this section are FOL-SUMO, WordNet and the semantic mapping between them.

SUMO[1] (Niles and Pease, 2001) is an upper level ontology proposed as a starter document by the IEEE Standard Upper Ontology Working Group. SUMO is expressed in SUO-KIF (Standard Upper Ontology Knowledge Interchange Format (Pease, 2009)), which is a dialect of KIF (Knowledge Interchange Format (Genesereth et al., 1992)). The syntax of both KIF and SUO-KIF goes beyond FOL and, therefore, SUMO axioms cannot be directly used by FOL ATPs without a suitable transformation (Álvez et al., 2012).

To the best of our knowledge, there are two main proposals for the translation of the two upper levels of SUMO into a FOL formulae that are described in Pease and Sutcliffe, (2007), Pease et al. (2010) and Álvez et al. (2012) respectively. Both proposals have been developed under the *Open World Assumption* (OWA) (Reiter, 1978) and are currently included in the *Thousands of Problems for Theorem Provers* (TPTP) problem library[2] (Sutcliffe, 2009). In this paper, we use Adimen-SUMO v2.6, which is freely available at `https://adimen.si.ehu.es/web/AdimenSUMO`. From now on, we refer to Adimen-SUMO v2.6 as FOL-SUMO.

The knowledge in SUMO, and therefore in FOL-SUMO, is organised around the notions of *instance* and *class*. These concepts are respectively defined in SUMO by means of the predicates *instance* and *subclass*.[3] Additionally, SUMO also differentiates between *relations* and *attributes*, which are organized using the predicates *subrelation* and *subAttribute* respectively. For simplicity, from now on we denote the nature of SUMO concepts by adding as subscript the symbols o (SUMO instances that are neither relations nor attributes), c (SUMO classes that are nei-

[1] `http://www.ontologyportal.org`
[2] `http://www.tptp.org`
[3] It is worth noting that term *instance* is overloaded since it denotes both the SUMO predicate and the SUMO concepts that are defined by using that predicate.

ther classes of relations nor classes of attributes), r (SUMO relations) and a (SUMO attributes). For example: $Waist_o$, $Artifact_c$, $customer_r$ and $Female_a$.

WordNet (Fellbaum, 1998) is linked with SUMO by means of the mapping described in Niles and Pease (2003). This mapping connects WordNet synsets to terms in SUMO using three relations: *equivalence*, *subsumption* and *instantiation*. We denote the mapping relations by concatenating the symbols '=' (*equivalence*), '+' (*subsumption*) and '@' (*instantiation*) to the corresponding SUMO concept. For example, the synsets $horse^1_n$, $education^4_n$ and $zero^1_a$ are connected to $Horse_c$=, $EducationalProcess_c$+ and $Integer_c$@ respectively. *equivalence* denotes that the related WordNet synset and SUMO concept are equivalent in meaning, whereas *subsumption* and *instantiation* indicate that the semantics of the WordNet synset is less general than the semantics of the SUMO concept. In particular, *instantiation* is used when the semantics of the WordNet synsets refers to a particular member of the class to which the semantics of the SUMO concept is referred.

The mapping between WordNet and SUMO can be translated into the language of SUMO by means of the proposal introduced in Álvez et al. (2017). This translation characterises the mapping information of a synset in terms of SUMO instances by using equality (for SUMO instances) and the SUMO predicates $instance_r$ and $attribute_r$ (for SUMO classes and attributes respectively). For example, the noun synsets $smoking^1_n$ and $breathing^1_n$ are respectively connected to $Smoking_c$= and $Breathing_c$=. Thus, the SUMO statements that result by following the proposal described in Álvez et al. (2017) is:

$$\text{(instance ?X Smoking)} \quad (2)$$

$$\text{(instance ?X Breathing)} \quad (3)$$

2.2 Evaluation Framework

The competency of SUMO-based ontologies can be automatically evaluated by using the framework described in Álvez et al. (2019) and the resources mentioned above. This framework is based on the use of *competency questions* (CQs) or *problems* (Grüninger and Fox, 1995) derived from the knowledge in WordNet and its mapping to SUMO by means of several predefined *question patterns*. In this paper, we have considered the following QPs:

WordNet Relation	QP	Problems
Hyponymy	Noun #1	7,539
	Noun #2	1,944
	Verb #1	1,765
	Verb #2	304
Antonymy	#1	91
	#2	574
	#3	2,780
Morphosemantic Links	Agent	829
	Instrument	348
	Result	788
Total	**–**	**16,972**

Table 1: Creation of problems on the basis of QPs

- The four QPs based on *hyponymy* —2 QPs for nouns and 2 QPs for verbs— and the three QPs based on *antonymy* introduced in Álvez et al. (2019).
- The three QPs based on the *Morphosemantic Links agent*, *instrument* and *result* introduced in Álvez et al. (2017).

In Table 1, we report on the number of CQs/problems that results from each QP.

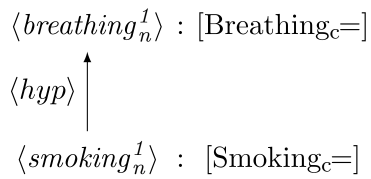

Figure 2: Example for Noun #2: $smoking^1_n$ and $breathing^1_n$

For example, the second QP based on *hyponymy* focuses on pairs of hyponym synsets *(hypo,hyper)* such that the hyponym *hypo* is connected to SUMO using *equivalence*. In those cases, the semantics of *hypo* is equivalent to the semantics of the SUMO statement that results from its mapping information. Further, the semantics of *hyper* is more general than the semantics of *hypo*. Consequently, we can state that the set of SUMO instances related to *hyper* is a superset of the set of SUMO instances connected to *hypo*. In particular, the noun synset $smoking^1_n$ ("*the act of smoking tobacco or other substances*") is hyponym of $breathing^1_n$ ("*the bodily process of inhalation and exhalation; the process of taking in oxygen from inhaled air and releasing carbon dioxide by exhalation*") (see Figure 2). By the instantiation of the second QP based on hyponymy using statements (2-3) which result from their mapping information, the following CQ that states that every instance of *Smoking*$_c$ is also instance of *Breathing*$_c$ can be obtained:

$$(\textbf{forall}\ \ (?\texttt{X})\quad (=> \quad (\text{instance } ?\texttt{X} \text{ Smoking}) \quad (\text{instance } ?\texttt{X} \text{ Breathing}))) \tag{4}$$

Given a set of CQs and an ontology, the evaluation framework proposes to perform two dual tests using FOL ATPs for each CQ: the first test is to check whether, as expected, the conjecture stated by the CQ is entailed by the ontology (*truth-test*); the second one is to check its complementary (*falsity-test*). If ATPs find a proof for either the truth- or the falsity-test, then the CQ is classified as *solved* (or *resolved*). In particular, the CQ is *passing*/*non-passing* if ATPs find a proof for the truth-test/falsity-test. Otherwise (that is, if no proof is found), the CQ is classified as *unresolved* or *unknown*. In this last case, we do not know whether (a) the conjectures are not entailed by the ontology or (b) although (some of) the conjectures are entailed, ATPs have not been able to find the proof within the provided execution-time and memory limits.

3 Detailed Analysis of the Experimental Results

In this section, we report on a detailed and manual analysis of the experimental results obtained from a small number of the CQs described in Section 2.

From this experimentation, we have randomly selected a sample of 169 problems (1% of the total) following a uniform distribution and analysed the results obtained for those problems by focusing on two questions: 1) we analyse the quality of mapping of the involved synsets and 2) we analyse the knowledge required for solving the problems.

Regarding the quality of the mapping (first question), we classify the mapping of synsets as either *correct* or *incorrect* according to the following criteria: a mapping is classified as *correct* if the semantics associated with the SUMO concept and with the synset are compatible, and it is classified as *incorrect* otherwise. For example, both the verb synset *machine*^1_v and the adjective synset *homemade*^1_a are connected to *Making*$_c$+, where the semantics of the SUMO class *Making*$_c$ is *"The subclass of Creation*$_c$ *in which an individual Artifact*$_c$ *or a type of Artifact*$_c$ *is made"*. Since the semantics of the verb synset *machine*^1_v is *"Turn, shape, mold, or otherwise finish by machinery"*, we classify the mapping of *machine*^1_v as *correct*. On the contrary, the semantics of the adjective synset *homemade*^1_a is *"made or produced in the home or by yourself"*. Thus, we classify the mapping of *homemade*^1_a as *incorrect*.

In addition, synsets with a correct mapping are classified as either *correct and precise* or *only correct*: a correct mapping is also considered as *correct and precise* if the semantics of the synset and the SUMO concept are equivalent, and it is classified as *only correct* (that is, correct but not precise) if the semantics of the SUMO concept is more general than the semantics of the synset. For example, the mapping of *machine*^1_v to *Making*$_c$ is classified as *only correct* since the semantics of *Making*$_c$ is more general than the semantics of *machine*^1_v. By contrast, the mapping of the noun synset *machine*^1_n to *Machine*$_c$= is classified as *correct and precise* since the semantics of *machine*^1_n is *"Any mechanical or electrical device that transmits or modifies energy to perform or assist in the performance of human tasks"* and the semantics of *Machine*$_c$ is *"Machine*$_c$*'s are Device*$_c$*'s that that have a well-defined resource and result and that automatically convert the resource into the result"*.

Regarding the required knowledge (second questions), we distinguish three cases:

- If the problem is solved, then we classify the knowledge in the proof provided by ATPs as either *correct* or *incorrect* depending on whether it matches our world knowledge or not.
- If the problem is unsolved and the mapping of the two involved synsets is correct, then we manually check whether the problem can be entailed by the knowledge in the ontology.
- If the problem is unsolved and the mapping of some of the involved synsets is incorrect, then the knowledge in the problem does not match our world knowledge and, consequently, it is not subject of classification.

It is worth noting that, in the case of unsolved problems such that the required knowledge is classified as existing, ATPs cannot find a proof for its

Question Pattern	#	Entailed					Incompatible					Unsolved			Total				
																	S		
		S	CM	IM	CK	IK	S	CM	IM	CK	IK	U	CM	IM	CM	IM	CK	IK	U
Noun #1 (7,539)	80	39	33 (5)	6	39	0	15	7 (0)	8	15	0	26	19 (0)	7	59 (5)	21	54	0	26
Noun #2 (1,944)	15	9	9 (5)	0	9	0	2	2 (2)	0	2	0	4	3 (2)	1	14 (9)	1	11	0	4
Verb #1 (1,765)	13	5	3 (1)	2	5	0	0	0 (0)	0	0	0	8	6 (0)	2	9 (1)	4	5	0	8
Verb #2 (304)	2	0	0 (0)	0	0	0	0	0 (0)	0	0	0	2	2 (1)	0	2 (1)	0	0	0	2
Antonym #1 (91)	1	0	0 (0)	0	0	0	0	0 (0)	0	0	0	1	0 (0)	1	0 (0)	1	0	0	1
Antonym #2 (584)	6	1	0 (0)	1	1	0	0	0 (0)	0	0	0	5	3 (1)	2	3 (1)	3	1	0	5
Antonym #3 (2,780)	33	9	4 (0)	5	9	0	0	0 (0)	0	0	0	24	7 (0)	17	11 (0)	22	9	0	24
Agent (829)	5	1	1 (0)	0	1	0	0	0 (0)	0	0	0	4	3 (1)	1	4 (1)	1	1	0	4
Instrument (348)	2	0	0 (0)	0	0	0	0	0 (0)	0	0	0	2	2 (2)	0	2 (2)	0	0	0	2
Result (788)	12	1	1 (0)	0	1	0	0	0 (0)	0	0	0	11	6 (4)	5	7 (4)	5	1	0	11
Total problems (16,972)	**169**	**65**	**51 (11)**	**14**	**65**	**0**	**17**	**9 (2)**	**8**	**17**	**0**	**87**	**51 (11)**	**36**	**111 (24)**	**58**	**82**	**0**	**87**

Table 2: Detailed analysis of problems

truth- or falsity-test because of the lack of time or memory resources.

In Table 2 we summarise some figures of our detailed analysis, where problems are organised according to their QP. The name of the QP and the number of resulting CQs is given in the first column (Question Pattern column) and the remaining columns are grouped into five main parts. In the first part (#, one column), we provide the number of problems of each category that have been randomly chosen. In the second and third parts (Entailed and Incompatible, five columns each), we provide the result of our quality analysis for the solved problems that have been classified as entailed (its truth-test has been proved) and incompatible (its falsity-test has been proved) respectively. Concretely we show:

- The number of solved problems (S column).
- The number of solved problems with a correct (CM column) and incorrect mapping (IM column). Additionally, in the CM column we provide the number of solved problems with a correct and precise mapping between brackets.
- The number of solved problems that have been proved on the basis of correct (CK column) and incorrect knowledge (IK column).

In the fourth part (Unsolved, three columns), we provide the result of our analysis for the unsolved problems:

- The number of unsolved problems (U column).
- The number of solved problems with a correct (CM column) and incorrect mapping (IM column). As before, in the CM column we provide the number of solved problems with a correct and precise mapping between brackets.

Finally, in the last part (Total, five columns) we summarise the result of our analysis:

- The number of problems with a correct (correct and precise between brackets) and incorrect mapping (CM and IM columns).
- The number of solved problems (S columns) that have been proved on the basis of correct (CK column) and incorrect knowledge (IK column).
- The number of unsolved problems (U column).

In total, the synsets in 111 problems (66 %) are decided to be correctly connected to SUMO and, among them, the synsets in 24 problems (14 %) are decided to be precisely connected. Thus, some of the synsets are not correctly connected to SUMO in 58 problems (34 %). Further, 82 problems (49 %) are solved and the knowledge of the ontology that is used in the proofs reported by ATPs is decided to be correct (100 %) according to our world knowledge. Among solved problems, 65 problems (79 %) are classified as entailed and 17 problems (21 %) are classified as incompatible. By manually analysing incompatible problems, we have discovered that the knowledge of WordNet and SUMO related to all the problems with a correct mapping is not well-aligned. Thus, we can conclude that this reasoning framework also enables the correction of the alignment between WordNet and SUMO. For example, $cloud_n^1$ ("*any collection of particles (e.g., smoke or dust) or gases that is visible*") is hyponym of $physical\_phenomenon_n^1$ ("*a natural phe-*

nomenon involving the physical properties of matter and energy") in WordNet. However, $cloud_n^1$ and $physical\_phenomenon_n^1$ are respectively connected to $Cloud_c$= and $NaturalProcess_c$=, which are inferred to be disjoint classes in FOL-SUMO.

Further, the mapping of the involved synsets is classified as correct in 51 of 65 entailed problems (78 %), while only 14 problems (22 %) are classified as entailed with an incorrect mapping. By contrast, the percentage of problems with an incorrect mapping is much higher among incompatible and unsolved problems: 42 % (8 of 17 entailed problems) and 41 % (36 of 87 unsolved problems) respectively. This is especially the case of the problems from the antonym categories: 26 of 40 antonym problems (65 %) have an incorrect mapping. This fact reveals the poor quality of the mapping of SUMO to WordNet adjectives. Finally, we have manually checked that 45 of the 51 unsolved problems with a correct mapping (88 %) cannot be entailed by the knowledge in SUMO, which sets an upper bound on the number of problems that can be classified as solved although augmenting the knowledge of the ontology and correcting the mapping and the alignment between WordNet and SUMO.

Next, we summarise the main conclusions drawn from our detailed analysis:

- The solutions of all the solved problems (with either correct or incorrect mapping) are based on correct knowledge of the ontology (CK columns). This means that we have not discovered incorrect knowledge in the ontology by inspecting the proofs provided by ATPs.

- The mapping of a half third of the problems is classified as incorrect (58 of 169 problems) and, among them, almost a half of the problems belong to the antonym categories (26 of 58 problems). This is mainly due to the poor quality of the mapping of SUMO to WordNet adjectives because many of them are connected to SUMO processes instead of SUMO attributes. Further, the number of problems with a precise mapping among the problems with a correct mapping is very low (24 of 111 problems). However, this is not surprising due to the large difference between the number of concepts defined in the core of SUMO (around 3,500 concepts) and WordNet (117,659 synsets).

- Among incompatible problems, the ones with a correct mapping (9 of 17 problems) enable the detection of misalignments between the knowledge of WordNet and SUMO.

- The number of solved problems among the *Morphosemantic Links* problems with a correct mapping is very low (only 2 of 13 problems), which reveals that FOL-SUMO lacks the required information about processes in SUMO.

- Most of the unsolved problems with a correct mapping —45 of 51 problems (88 %)— are due to the lack of information in the core of SUMO. However, we have also discovered 6 problems for which either its truth- or falsity-test is entailed by knowledge in the core of SUMO although it cannot be proved by ATPs within the given resources of time and memory. Thus, ATPs are able to solve 82 of 88 the problems (93 %) that are entailed by the current knowledge of the ontology.

4 Exhaustive Analysis of some Problems

In this section, we present a detailed analysis of some of the examples that have been reported in Table 2.

4.1 Examples of Entailed Problems

Next, we present two examples among the 65 problems that are classified as entailed. The mapping information is correct in the first example, while it is incorrect in the second one.

4.1.1 Case 1: Correct mapping

The first example we present involves the synsets $army_n^1$ ("*a permanent organization of the military land forces of a nation or state*") and $armed\_service_n^1$ ("*a force that is a branch of the armed forces*"). These synsets are respectively mapped to the SUMO classes $Army_c$ and $MilitaryService_c$ by equivalence.

In WordNet $army_n^1$ is hyponym of $armed\_service_n^1$ and in SUMO $Army_c$ is subclass of $MilitaryService_c$. In this case, the knowledge in both resources and in the mapping is correctly aligned, so we get an entailed problem. In Table 2, we report 51 entailed problems with a correct mapping.

4.1.2 Case 2: Incorrect mapping

The second example of entailed problem involves the synsets *atmospheric_electricity*^1_n ("*electrical discharges in the atmosphere*") and *electrical_discharge*^1_n ("*a discharge of electricity*"). These synsets are respectively mapped to the SUMO classes *Lightning*$_c$ and *Radiating*$_c$ by subsumption.

These synsets are related by hyponymy-hyperonymy in WordNet and by subclass in SUMO, as in the previous case. But, the mapping seems misleading for *electrical_discharge*^t_o *Radiating*$_c$: ("*Processes in which some form of electromagnetic radiation, e.g. radio waves, light waves, electrical energy, etc., is given off or absorbed by something else.*"). However, this case is resolved because by chance the knowledge in WordNet and in the incorrect mapping to SUMO is aligned.

We have discovered 14 entailed problems with an incorrect mapping.

4.2 Examples of Incompatible Problems

Next, we present three examples of problems that are classified as incompatible due to several reasons.

4.2.1 Case 1: Knowledge misalignment

The first example we present involves the SUMO classes *Smoking*$_c$ and *Breathing*$_c$ and the synsets *smoking*^1_n ("*the act of smoking tobacco or other substances*") and *breathing*^1_n ("*the bodily process of inhalation and exhalation; the process of taking in oxygen from inhaled air and releasing carbon dioxide by exhalation*") .

The synset *smoking*^1_n is hyponym of *breathing*^1_n in WordNet, which are respectively connected to *Smoking*$_c$= and *Breathing*$_c$=. These classes are disjoint in SUMO. That is, instances of *Smoking*$_c$ cannot be instances of *Breathing*$_c$. So, according to the knowledge in SUMO, it is not possible to breath and smoke at the same time, but, according to WordNet smoking is a subtype of breathing. In this case we have, therefore, a knowledge misalignment problem: the knowledge in one of the resources contradicts the knowledge in the other one.

Another example of this type of cases involves the SUMO classes *Cloud*$_c$ and *NaturalProcess*$_c$ and the synsets *cloud*^1_n ("*any collection of particles (e.g., smoke or dust) or gases that is visible*") and *physical_phenomenon*^1_n ("*a natural phenomenon involving the physical properties of matter and energy* "), which are mapped to SUMO respectively by equivalence and subsumption.

In WordNet *cloud*^1_n is hypomym of *physical_phenomenon*^1_n, but in SUMO they belong to different hierarchies: *Cloud*$_c$ is subclass of *Substance*$_c$ and *NaturalProcess*$_c$ is subclass of *Process*$_c$, and these classes are disjoint as in the previous example.

From the incompatible problems reported in Table 2, the knowledge is misaligned in 5 problems.

4.2.2 Case 2: Imprecise mappings

The next example involves the SUMO classes *Transfer*$_c$ ("*Any instance of Translocation where the agent and the patient are not the same thing.*") and *Removing*$_c$ ("*The Class of Processes where something is taken away from a location. Note that the thing removed and the location are specified with the CaseRoles patient and origin, respectively.*"). The involved synsets are *fetch*^1_v ("*go or come after and bring or take back*") and *carry_away*^1_v ("*remove from a certain place, environment, or mental or emotional state; transport into a new location or state*"). *fetch*^1_n is mapped to *Transfer*$_c$ via equivalence while *carry_away*^1_n is mapped to *Removing*$_c$ via subsumption.

fetch^1_v and *carry_away*^1_v are antonyms in WordNet, but their corresponding SUMO classes are related via subclass in SUMO: *Removing*$_c$ is subclass of *Transfer*$_c$. In our opinion this is a case of imprecise mapping, although correct, the class *Transfer*$_c$ is too general for the synset *fetch*^1_v.

In Table 2, we report two incompatible problems with a correct but imprecise mapping.

4.3 Examples of Unresolved Problems

Next, we present two examples of problems that are unresolved due to different causes.

4.3.1 Case 1: Lack of knowledge

The first example corresponds to the problems that are not solved due to the lack of knowledge in the ontology and involves the synsets *machine*^1_v (*Making*$_c$+) and *machine*^1_n (*Machine*$_c$=). These synsets are related via morphosemantic relation *instrument*. However, there is no similar knowledge encoded in SUMO, so this example remains unresolved.

In Table 2, we report 45 problems with a correct mapping that are unresolved due to the lack of knowledge in the ontology.

4.3.2 Case 2: Lack of resources

The second example corresponds to resolvable problems that remain unresolved due to the lack of resources (mainly time) of ATPs. This example involves the synset $male_a^3$ linked to $Male_c+$ and the sysnset $female_a^1$ linked to $Female_c=$ as antomyms. In this case, although all the knowledge is correct the ATPs cannot find the prove for it.

Among the problems reported in Table 2, we have found 6 problems with a correct mapping that can be solved but that remain unresolved due to the lack of resources of ATPs.

5 Conclusion and Future Works

In this paper we have presented a detailed analysis of a sample of large benchmark of commonsense reasoning problems that has been automatically obtained from WordNet, SUMO and their mapping.

Based on this analysis, we can detect that although the framework enables the resolution of around 49 % of the total problems, only 36 % of the total are resolved for the good reasons: 60 problems resolved with a correct mapping. We have also detected that the mapping requires a general revision and correction: in particular, in the case of adjectives. On the contrary, the knowledge in SUMO involved in the revised proofs seems to be correct according to our commonsense knowledge. Further, the problems classified as incompatible enable the detection of misalignments between WordNet and SUMO, while the problems classified as unknown can be taken as a source of knowledge for the augmentation of SUMO. Actually, we are planning to develop an automatic procedure for the augmentation of SUMO on the basis of the knowledge in WordNet. Finally, we have detected some problems that can be solved on the basis of the knowledge of SUMO but that are not solved due to the limitation of resources of ATPs.

Acknowledgments

This work has been partially funded by the the project DeepReading (RTI2018-096846-B-C21) supported by the Ministry of Science, Innovation and Universities of the Spanish Government, the projects CROSSTEXT (TIN2015-72646-EXP) and GRAMM (TIN2017-86727-C2-2-R) supported by the Ministry of Economy, Industry and Competitiveness of the Spanish Government, the Basque Project LoRea (GIU18/182) and BigKnowledge – *Ayudas Fundación BBVA a Equipos de Investigación Científica 2018*.

References

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. 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.

C. Fellbaum, A. Osherson, and P.E. Clark. 2009. Putting semantics into WordNet's "Morphosemantic" Links. In Z. Vetulani and H. Uszkoreit, editors, *Human Language Technology. Challenges of the Information Society*, LNCS 5603, pages 350–358. Springer.

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

A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann. 2006. Modelling ontology evaluation and validation. In Y. Sure and J. Domingue, editors, *The Semantic Web: Research and Applications*, LNCS 4011, pages 140–154. Springer.

M. R. Genesereth, R. E. Fikes, D. Brobow, R. Brachman, T. Gruber, P. Hayes, R. Letsinger, V. Lifschitz, R. Macgregor, J. Mccarthy, P. Norvig, R. Patil, and L. Schubert. 1992. Knowledge Interchange Format version 3.0 reference manual. Technical Report Logic-92-1, Stanford University, Computer Science Department, Logic Group.

A. Gómez-Pérez, M. Fernández-López, and O. Corcho-García. 2004. Ontological engineering. *Computing Reviews*, 45(8):478–479.

T. Gruber. 2009. Ontology. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 1963–1965. Springer.

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)*.

N. Guarino and C. A. Welty. 2002. Evaluating ontological decisions with OntoClean. *Commun. ACM*, 45(2):61–65, feb.

N. Guarino and C. A. Welty. 2004. An overview of OntoClean. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, pages 151–171. Springer.

I. Horrocks and A. Voronkov. 2006. Reasoning support for expressive ontology languages using a theorem prover. In J. Dix et al., editor, *Foundations of Information and Knowledge Systems*, LNCS 3861, pages 201–218. Springer.

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.

I. Niles and A. Pease. 2001. Towards a standard upper ontology. In Guarino N. et al., editor, *Proc. of the* 2^{nd} *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 H. R. Arabnia, editor, *Proc. of the IEEE Int. Conf. on Inf. and Knowledge Engin. (IKE 2003)*, volume 2, pages 412–416. CSREA Press.

N. F. Noy and D. L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05 and SMI-2001-0880, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics.

A. Pease and G. Sutcliffe. 2007. First-order reasoning on a large ontology. In Sutcliffe G. et al., editor, *Proc. of the Workshop on Empirically Successful Automated Reasoning in Large Theories (CADE-21)*, CEUR Workshop Proceedings 257. CEUR-WS.org.

A. Pease, G. Sutcliffe, N. Siegel, and S. Trac. 2010. Large theory reasoning with SUMO at CASC. *AI Communications*, 23(2-3):137–144.

A. Pease. 2009. Standard Upper Ontology Knowledge Interchange Format. Retrieved June 18, 2009, from `http://sigmakee.cvs.sourceforge.net/sigmakee/sigma/suo-kif.pdf`.

D. Ramachandran, R. P. Reagan, and K. Goolsbey. 2005. First-orderized ResearchCyc: Expressivity and efficiency in a common-sense ontology. In P. Shvaiko et al., editor, *Papers from the Workshop on Contexts and Ontologies: Theory, Practice and Applications (AAAI 2005)*, pages 33–40. AAAI Press.

R. Reiter, 1978. *On Closed World Data Bases*, pages 55–76. Springer, Boston, MA.

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

S. Staab and R. Studer. 2009. *Handbook on Ontologies*. Springer, 2^{nd} edition.

G. Sutcliffe. 2009. The TPTP problem library and associated infrastructure. *J. Automated Reasoning*, 43(4):337–362.