Semi-Automatic Mapping of WordNet to Basic Formal Ontology

Selja Seppälä,[1] Amanda Hicks,[2] Alan Ruttenberg[1]

[1]University at Buffalo
Buffalo, NY, USA
[2]University of Florida
Gainesville, FL, USA

seljamar@buffalo.edu, aehicks@ufl.edu, alanruttenberg@gmail.com

Abstract

We present preliminary work on the mapping of WordNet 3.0 to the Basic Formal Ontology (BFO 2.0). WordNet is a large, widely used semantic network. BFO is a domain-neutral upper-level ontology that represents the types of things that exist in the world and relations between them. BFO serves as an integration hub for more specific ontologies, such as the Ontology for Biomedical Investigations (OBI) and Ontology for Biobanking (OBIB). This work aims at creating a lexico-semantic resource that can be used in NLP tools to perform ontology-related text manipulation tasks. Our semi-automatic mapping method consists in using existing mappings between WordNet and the KYOTO Ontology. The latter allows machines to reason over texts by providing interpretations of the words in ontological terms. Our working hypothesis is that a large portion of WordNet synsets can be semi-automatically mapped to BFO using simple mapping rules from KYOTO to BFO. We evaluate the method on a randomized subset of synsets, examine preliminary results, and discuss challenges related to the method. We conclude with suggestions for future work.

1 Introduction

Ontologies are often used in combination with natural language processing (NLP) tools to carry out ontology-related text manipulation tasks such as automatic annotation of biomedical texts with ontology terms. These tasks involve categorizing relevant terms from texts under the appropriate categories. This requires coupling ontologies with lexical resources. Several projects have realized these kinds of mappings with upper-level ontologies that are extended by domain-specific ontologies (**?**; Gangemi et al., 2010; Niles and Pease, 2003; Pease and Fellbaum, 2010). However, no such resource is available for the Basic Formal Ontology (BFO), which is widely used in the biomedical domain.

We describe and evaluate a semi-automatic method for mapping the large lexical network WordNet 3.0 (WN) to BFO 2.0 exploiting an existing mapping between WN and the KYOTO Ontology (hereafter ‘KYOTO’). Our hypothesis is that a large portion of WN, primarily nouns and verbs, can be semi-automatically mapped to BFO 2.0 types by means of simple mapping rules exploiting the KYOTO Ontology.

In section 2, we give a brief overview of the ontological and lexical resources involved in the task: BFO, WN, and KYOTO. In section 3, we motivate and describe our methodology. In section 4, we evaluate the method and present preliminary results. In section 5, we discuss the major challenges related to this task. We conclude with suggestions for future work.

2 Ontological and Lexical Resources

The mapping methodology described below, in section 3, takes as input WordNet 3.0, which is mapped to the KYOTO 3 Top Ontology by way of KYOTO 3 Middle. KYOTO 3 Top is an extension of the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE-Lite-Plus, version 3.9.7). The KYOTO 3 Top Ontology was extended to a middle level ontology KYOTO 3 Middle by manually mapping Base Concepts automatically generated from WN (Herold et al., 2009). We use those existing mappings to create mapping rules from KYOTO to BFO 2.0. Hereafter, we briefly describe each of these ontological and lexical resources. We briefly present the aspects of BFO, WN, and the KYOTO Ontology that are relevant to this work. For more details, see the cited references.

The **Basic Formal Ontology (BFO)** is a domain-neutral upper-level ontology (Arp et al., 2015; Smith et al., 2012; Spear, 2006). It represents the types of things that exist in the world and relations between them. BFO serves as an integration hub for mid-level and domain-specific ontologies, such as the Ontology for Biomedical Investigations (OBI) and Ontology for Biobanking (OBIB). It is widely used in biomedical and other domain-specific ontologies,[1] which thus become interoperable (Smith and Ceusters, 2010). BFO is subdivided into CONTINUANTS (e.g., OBJECTS and FUNCTIONS) and OCCURRENTS (e.g., PROCESSES and EVENTS). Continuants can be either independent (e.g., physical OBJECTS like persons and hearts) or dependent (e.g., the ROLE of a person as a physician and the FUNCTION of a heart to pump blood). The most recent version, BFO 2.0, represents 35 types to which previous versions (BFO 1.0 and BFO 1.1) have been mapped in Seppälä et al. (2014).

WordNet 3.0 is a large lexical network linking over 117000 sets of synonymous English words (synsets) by means of semantic relations; it is widely used in NLP tasks (Fellbaum, 1998; Miller, 1995). Noun and verb synsets are linked via the hypernym relation.[2] WN 3.0 distinguishes between types and instances, meaning named entities. It also links a subset of synsets to topic domains (e.g., 'medicine') and semantic labels (e.g., 'noun.artifact').

The **KYOTO Ontology** is part of a project aimed at representing domain-specific terms in a computer-tractable axiomatized formalism to allow machines to reason over texts in natural language (Vossen et al., 2010). It links WordNets of different languages to ontology classes, on the basis of a mapping of the English WN to KYOTO. The approximately 2000 classes of KYOTO are subdivided into three layers: (1) The top-most layer is based on DOLCE-Lite-Plus. DOCLE shares a number of relevant characteristics with BFO: domain neutrality; bi-partition into 'endurants' (CONTINUANTS) and 'perdurants' (OCCURRENTS); strict hierarchical *is_a* taxonomy; distinction between independent and dependent entities. (2) The second layer is composed of noun and verb synsets constituting a set of Base Concepts (BCs) as well as some adjectives or qualities. (3) The third layer contains domain-specific classes (e.g., from the environmental domain) and some corresponding synsets.

3 Mapping Method

Our semi-automatic mapping method involves three main steps:

1. Manually creating mappings:
   - from KYOTO to BFO on the basis of existing mappings of DOLCE to BFO 1.0 and BFO 1.1 (Grenon, 2003; Khan and Keet, 2013; Seyed, 2009; Temal et al., 2010), ignoring the axiomatization incompatibilities;
   - from BFO 1.0 and BFO 1.1 to BFO 2.0 on the basis of work in Seppälä et al. (2014);
   - from WN semantic labels to BFO 2.0.

2. Manually creating mapping rules using the above mappings and extending them with more specific rules from other KYOTO types. The rules map to BFO 2.0 leaf types or, when BFO has no leaf-level type to represent the referent of a synset, to intermediary types (e.g., MATERIAL ENTITY, the direct parent of three leaf types).

3. Implementing the resulting mapping rules in a Python pipeline using the natural language toolkit for Python that integrates WN 3.0 (NLTK 3.0).[3]

The rules are of the form `'KYOTO/WN > BFO 2.0'`, for example:

```
'#non-agentive-social-object
     > disposition'
'accomplishment > process'
'noun.act > process'
```

The implementation first lists all KYOTO types that subsume or otherwise characterize a WN synset using the WN-KYOTO mapping data files.[4] We only retained types related to the synsets through equivalence and subclass

[1]See `http://ifomis.uni-saarland.de/bfo/users`.

[2]Adjectives and adverbs are linked by way of other semantic relations.

[3]Natural Language Toolkit for Python (NLTK), version 3.0, `http://www.nltk.org`.

[4]`http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index9c60.html?option=\\com\_content&view=article&id=429&Itemid=156`.

relations, plus the following ones deemed useful for creating appropriate mapping rules:

```
'DOLCE-Lite.owl#generically-dependent-on'
'DOLCE-Lite.owl#specifically-constantly
    -dependent-on'
'ExtendedDnS.owl#realized-by'
'ExtendedDnS.owl#realizes'
```

For example, the synset `immunity.n.02` is linked to the following types:

```
'Kyoto#condition__status
    -eng-3.0-13920835-n',
'Kyoto#state-eng-3.0-00024720-n',
'ExtendedDnS.owl#situation',
'ExtendedDnS.owl#non-agentive
    -social-object',
'ExtendedDnS.owl#social-object',
'DOLCE-Lite.owl#non-physical-object',
'DOLCE-Lite.owl#non-physical-endurant',
'DOLCE-Lite.owl#endurant',
'DOLCE-Lite.owl#spatio-temporal
    -particular',
'DOLCE-Lite.owl#particular'
```

Second, the mapping rules are applied starting from the more specific type in the types list: the program tests whether a given string on the left-hand side of the rule (e.g., `'#non-agentive-social-object'`) matches a string in the types list; if the strings match, the program assigns to that synset the corresponding BFO 2.0 type (e.g., `'disposition'`). Thus, the synset `immunity.n.02` is categorized as referring to a subtype of the BFO type DISPOSITION.

4 Evaluation and Results

In a first step, we evaluated the method on the 106 synsets in KYOTO marked with a 'medicine' topic domain (Seppälä, 2015a). The aim in this first step was to get a rough idea of the feasibility of the method, the results that could be expected, and the possible challenges. The medicine gold standard was created collectively by a BFO developer and a BFO expert. The experts followed an intuitive categorization criterion: assign the most specific BFO type of which the referent of the synset is a subtype. Following this principle, for each synset we may obtain a statement of the form "the WN synset X refers to a subtype of the (leaf) BFO type Y". For example, "the WN synset `immunity.n.02` refers to a subtype of the BFO type DISPOSITION". This task revealed difficult interpretation issues related to adjectives. 71.7% of the assigned BFO types were correct (63.2% of the synsets were assigned the expected BFO type; 8.5% a superclass). As hypothesized, all the correctly categorized synsets were nominal and verbal. 27.4% of the assigned BFO types were incorrect (mostly adjectives). One synset was not matched by any rule.

In a second step, we focused on nouns and verbs, and left adjectives for future work. After examining the erroneous cases in the first evaluation, we created a new ruleset, which we tested on a new randomly extracted sample of 100 nouns and 100 verbs (hereafter the 'POS-sample').

To create the corresponding gold standard, two of the authors, experts of BFO, first pre-annotated the POS-sample independently. They followed the same intuitive annotation criteria as with the medicine gold standard. The annotations were compared and the synsets separated into 'easy cases' (where both annotators agreed) and 'difficult cases' (in case of disagreement), respectively 101/200 and 99/200 synsets. Second, two BFO developers (annotators A & B) independently (i) reviewed the easy cases for validation and (ii) annotated the difficult cases. They were asked to apply the same intuitive annotation criteria as in previous steps. The annotators agreed on 2/3 of the latter sample. Finally, annotator B examined the cases on which they disagreed and decided on the final BFO type to assign considering the comments left by annotator A. Some difficult cases were collectively discussed to reach consensus. We discuss the challenging cases in section 5.

The baseline was created by a BFO developer and discussed with a BFO expert to resolve a few problematic cases (see, for example, section 5.5). The baseline mapping rules map, whenever possible, WN's top-level nouns to lowest level BFO 2.0 types and all verbs to BFO PROCESS.

In the following, we review the main results of our mappings. We limit the evaluation of the medicine sample to nouns and verbs to allow for comparisons with the performance of the rules on the POS-sample.

Figure 1 shows the overall performance of the first and the new rulesets compared to the baseline when applied to the medicine nouns and verbs sample. A total of 85% of the sample's synsets were correctly mapped with the new ruleset, which is considerably better than with the baseline rules (76%) and the first ruleset (77%). In the baseline

% of WN-BFO mappings	medicine n-v sample								
	baseline			first ruleset			new ruleset		
	n	v	total	n	v	total	n	v	total
correct	55	100	76	70	85	77	72	100	85
partial	17	0	9	0	12	6	0	0	0
incorrect	28	0	15	28	2	16	26	0	14
no mapping	0	0	0	2	0	1	2	0	1
total	100	100	100	100	100	100	100	100	100

Figure 1: Performance of the rulesets on the medicine nouns and verbs sample.

and the first experiment, respectively 9% and 6% of the synsets were mapped to a parent BFO type (see 'partial' row). There were no occurrences of partial mapping with the new ruleset. We did indeed correct some mapping rules with lower-level BFO types. However, assigning lower-level BFO types cannot be done with all of WN's top level categories. This means that with the baseline mapping method there will always be synsets that are not assigned an adequate (lowest level possible) BFO type. The percentage of incorrect mappings is steady across the applied mapping rules, around 15%. The proportion of synsets for which no rule was able to output a mapping is 1% for the KYOTO-based rules. As the baseline mapping rules were propagated all the way down the WN hierarchy, there are no such cases.

These results show that the performance of the KYOTO-based rulesets applied to the medicine nouns and verbs sample is (i) comparable to that of the baseline with the first ruleset and (ii) better with the new ruleset. While rather unsurprising, since the new ruleset was tuned on the medicine sample, this result nevertheless suggests that developing a more sophisticated mapping method, the KYOTO-based method, has advantages over a simple mapping of WN's top levels to BFO types.

% of WN-BFO mappings	pos sample					
	baseline			new ruleset		
	n	v	total	n	v	total
correct	41	99	70	42	86	64
partial	25	0	12.5	0	0	0
incorrect	34	1	17.5	53	7	30
no mapping	0	0	0	5	7	6
total	100	100	100	100	100	100

Figure 2: Performance of the new ruleset on the POS-sample.

Figure 2 shows the results of our second evaluation on a randomly extracted sample of 100 noun and 100 verb synsets. The overall performance of the new ruleset on the POS-sample is lower than that of the baseline ruleset, respectively 64% and 70%. However, the baseline ruleset itself performs lower here than in the medicine sample. Moreover, a closer look at the mappings reveals that while the new ruleset introduced some errors and non-matches, it also has the advantage of avoiding partial matches (when a synset is tagged with a superordinate BFO type instead of the lowest possible type). In 16 cases, the partial matches in the baseline correspond to the BFO type ENTITY, which is the uppermost level of the ontology and not relevant for a resource mapping WN to BFO.[5] With the baseline method, these cases could only be manually resolved; with the KYOTO-based method, we can test new rules to capture the incorrectly mapped cases (11/15 synsets) and the cases that were not mapped at all (3/15 synsets).

Moreover, the analysis of these new results reveals useful information for improving the WN-BFO mapping method. A notable example is the case of verb synsets: the baseline rules systematically mapped them to the BFO type PROCESS. This yielded only one error due to the fact that the erroneously mapped synset does actually not refer to any BFO type (see the discussion in section 5.3). This suggests that the KYOTO-based rules can be improved with this verb mapping rule. To test this hypothesis, we performed a supplementary test. Figure 3 shows the prospective performance of the new ruleset complemented with the verb mapping rule on the POS-sample. The overall performance rose from 64% to 70.5%, slightly higher than that of the baseline. However, the performance for verbs reached 99% (from 86% with the ruleset).

WN-BFO mappings	n	v	total	%
correct	42	99	141	70.5
partial	0	0	0	0
incorrect	53	1	54	27
no match	5	0	5	2.5
total	100	100	200	100

Figure 3: Prospective performance of the new ruleset combined with a systematic mapping of verbs to BFO PROCESS on the POS-sample.

The results further show that the performance for verbs was higher than that for nouns. Indeed, nouns refer to a large array of entity types (to

[5]These mappings could as well be considered incorrect.

10 BFO types in the POS-sample). Although the new ruleset did include rules for all the expected BFO types in the gold standard (as checked after the experiment), it did not capture the correct types in the KYOTO-types lists associated to the synsets. We suspect that the issue might be related to the ordering of the rules.

Finally, the results show that a total of 12 synsets (6%) were not mapped to BFO at all. This means that none of the rules was matched to the KYOTO-types lists associated with these synsets. These cases covered mostly deverbal nouns (8/12 synsets) that should be mapped to PROCESS. Further work is needed on the unmapped cases.

To summarize, we make the following general observations.

- Verbs are better covered than nouns, which can be explained by the fact that nouns refer to a wider array of BFO categories.

- Nouns are best covered by the rulesets, since the KYOTO-mappings allow for creating more specific rules that map to BFO leaf- or lower-intermediate types. This is not possible with a mapping of WN's top level nouns to BFO.

- Verbs are best covered by the baseline ruleset, which can be explained by the fact that most verbs refer to sub-types of BFO process. The results thus show that the KYOTO-based rules can be successfully complemented with a baseline rule that consists in systematically mapping verb synsets to BFO PROCESSES.

5 Discussion

5.1 Usefulness of a Lexico-semantic Resource Linked to BFO

While BFO may be seen as too small and high level for ontology-related text manipulation tasks, we believe that a lexico-semantic resource linked to BFO is still useful. Such a resource can be used, for example, to semi-automatically create BFO-compliant domain ontologies from existing domain-specific terminological resources.[6]

BFO could also be used as an alternative upper level ontology in the KYOTO project, in which case it needs to be mapped to WN. Substituting DOLCE with BFO would allow for comparisons of the performances of these upper level ontologies in applied tasks.

The fact that BFO types are very general is not a problem either. BFO reveals fundamental distinctions between the types, which involve different kinds of relations to different types of entities. When associated to WN synsets, such information may, for example, help definition authors write better definitions, as proposed in Seppälä (2015b).

5.2 Non-trivial Mappings Between Ontologies

Mapping DOLCE to BFO is not trivial. The former is meant to capture our use of language and conceptualization of the world; the latter is a realist ontology and excludes from its scope unicorns and other putative non-real entities. Consequently, their categories do not align in every case and are in some cases governed by different axioms.

In the following, we describe the problems that arise from these mapping issues and our solutions to them.

Axiomatic divergence: The number of DOLCE and BFO types that are axiomatically mappable is relatively reduced compared to the total number of types (Khan and Keet, 2013). Our solution to the axiomatization issue is to ignore axioms. Indeed, unlike work carried out, e.g., in Gangemi et al. (2003), this work is neither aimed at mapping DOLCE to BFO, nor at axiomatizing WN. Instead, we attempt to answer the question: to what types of entities do WN synsets refer? As mentioned in section 4, the resulting mappings are to be read as "a WN synset X refers to something that is a subtype of BFO type Y" — we exclude instances for now.

Incomplete rule coverage: Incomplete mappings between DOLCE and BFO lead to an incomplete rule coverage. Our solution to this issue was to extend the coverage of the rules by mapping other types and relations included in KYOTO as well as relevant semantic labels in WN to BFO types. As we saw in section 4, the KYOTO-based rules can also be complemented with rules from the baseline mapping ruleset. In addition to the verb mappings, we can test other WN top-level mappings, for example, when no mapping rule applies. Additional mapping rules could further be tested using mappings of WN to other upper level ontologies, such as the Suggested Upper

[6] See work in this direction in Seppälä (2015b).

Merged Ontology (SUMO).[7] The potential issue with SUMO is that it allows for multiple inheritance, which might result in categorization problems. Using other upper level ontologies also involves creating mappings between their categories and BFO categories.

One-to-many mapping types: In some cases, a single DOLCE type maps to several BFO types. For example, DOLCE 'feature' ambiguously refers to a BFO INDEPENDENT CONTINUANT, FIAT OBJECT PART, or SITE. Our solution was to map, whenever possible, to BFO leaf types, i.e., in this example to FIAT OBJECT PART and SITE. For the DOLCE type 'non-physical object' which is mapped to both FUNCTION and ROLE, we chose the rule to output the lowest common genus REALIZABLE ENTITY. A further disambiguation step is required to chose between the two. This might be done using additional mapping rules based on KYOTO and new sources.

Non-mapping types: Several DOLCE types have no matching type in BFO, as is the case for the DOLCE type 'abstract'. Conversely, some BFO types have no corresponding type in DOLCE, such as OBJECT BOUNDARY. A number of these cases might be captured by adding new rules using KYOTO and other sources.

While some of the above issues might not have a straightforward solution or no solution at all, even a partial mapping should be sufficient to cover a large portion of WN, leaving a smaller subset of problematic cases.

5.3 Heterogeneous Semantic Networks

One of the difficulties mapping WordNet to any ontology is that this task involves aligning semantic networks that were constructed with different aims and criteria. WN represents linguistic usage; BFO, entities in the world. While there are enough similarities between wordnets and ontologies to make this task possible, there are enough discrepancies to pose specific challenges. In this section, we discuss some of the challenges we encountered in our work.

Systematic polysemy: Systematic polysemy was one source of disagreement for generating the gold standard. For example, the synset `red-green_dichromacy.n.01` has the definition "confusion of red and green". Should this be mapped to BFO's DISPOSITION or PROCESS? On the one hand, a person with red-green dichromacy would not distinguish red from green if they were looking at a red and green object. This suggests that it is a disposition that inheres in the person and is realized by confusing red and green. On the other hand, 'confusion' is polysemous between a process and a result of a process. Similarly, we had nine cases of synsets such as `carpet_beetle.n.01` that can be used to describe a single organism, as in "I saw a carpet beetle in my bedroom", or an entire population, as in "The carpet beetle is not endangered." When we speak of a species being endangered, we are not speaking of a threat to individual organisms, but instead a threat to the population as a whole. For the gold standard, we tagged these as BFO OBJECTS (as individual organisms) rather than OBJECT AGGREGATES (as populations). Further investigation is needed to determine more nuanced ways of handling systematic polysemy.

Hierarchical discrepancies: Some of the mapping errors are the result of the discrepancy between WordNet's hypernym relation and `rdfs:subClassOf` used by BFO. For example, in WN `symptom.n.01` and `sign.n.06` are both descended from the Base Concept (BC) `cognition.n.01`. However, this is not ontologically precise. Symptoms such as a fever are not literally cognitions. However, WordNet's hypernym relations, in contrast to `rdfs:subClassOf`, are not intended to express relations among types of things but pyscho-linguistic intuitions of native English speakers. In the future, strategies for dealing with these cases need to be developed. Such strategies could include semi-automatic methods for ontologically evaluating WordNet's hierarchies, as in RUDIFY (Hicks and Herold, 2009). Another approach we are considering is to iteratively refine the mapping rules as errors are found.

Ontological distinctions: Other errors relate to ontological distinctions that WN is not intended to capture, e.g., `carrier.n.09`'s BC is correctly 'person', but this does not capture the distinction between rigid and non-rigid properties (Guarino and Welty, 2002). In the current version of the rules, these synsets are annotated with the BFO type ROLE, a non-rigid property, using WN's semantic type 'noun.person'.

Non-existent entities: Finally, we found that mapping a linguistic resource to a realist ontology raises the question of what to

[7] See `http://www.adampease.org/OP/`.

do with synsets that describe entities that are not real. For example, how can we map WN's `mythical_creature.n.01` and `metempsychosis.n.01` to a realist ontology such as BFO? This issue is particularly challenging for automating synset annotation since the system and the rules it uses have no way of telling apart existent entities from non-existent ones.[8]

While we have not resolved all questions around WN synsets that don't map to BFO, they raise interesting issues. A stimulating challenge will be to provide BFO-compliant interpretations of unmatched WN synsets.

5.4 Nature of Some Entity Types

Achieving consensus on the gold standard can also be challenging when the BFO community has ongoing discussions about the nature of some entity types such as language, measurements, and quantities. For example, measurements of temporal intervals are not modeled as such in BFO. So `track_record.n.01` would be mapped to ONE-DIMENSIONAL TEMPORAL REGION instead of the measurement of the time interval.

However, in the Information Artifact Ontology (IAO) and the Ontology for Biomedical Investigations (OBI), ontologies that extend BFO, measurements are categorized under GENERICALLY DEPENDENT CONTINUANTS.[9] Thus, for the time being and unless BFO proposes another way to represent these types of entities, we mostly consider these cases to refer to subtypes of GENERICALLY DEPENDENT CONTINUANTs.

5.5 Other Challenges

Issues arising from WN definitions: In a few cases, deciding what BFO type to assign was difficult due to vague, ambiguous, or unclear definitions of synsets. For example, the definition of `attribute.n.02` ("an abstraction belonging to or characteristic of an entity") is rather vague; it is thus difficult to determine which BFO type to assign to the corresponding synset and, consequently, to its hyponyms. When that happend in the elaboration of the baseline mapping rules, we examined the direct hyponyms of the category and assigned, whenever possible, a BFO leaf- or lower-intermediate type that seemed to cover most of the hyponyms.

Annotation mistakes: In two cases, our annotators made obvious mistakes, such as tagging the verb `die.v.02` with MATERIAL ENTITY instead of PROCESS. We confirmed with the annotators that these were in fact errors on their part and corrected the annotations with the appropriate BFO type.

Unknown issues: In some cases, we did not find any obvious explanation for the disagreement. For example, annotator A assigned the type REALIZABLE ENTITY to `federal_tax_lien.n.01` and annotator B, GENERICALLY DEPENDENT CONTINUANT. These cases were resolved by annotator B during the final examination of the annotations, as described in section 4.

6 Conclusion and Future Work

We presented a method to semi-automatically map WordNet 3.0 synsets to BFO 2.0 types via the KYOTO Ontology. Our preliminary results are encouraging, but reveal a number of challenges as addressed in the discussion section. More work is thus needed to see if the method scales to the full WN.

Future work will include:

- adding the verb-mapping baseline rule to improve verb mappings;
- examining our results and future development-sample results in more detail to investigate which parts of the rules are most productive, which ones cause errors, etc., and refine and reorder the rules;
- testing if complementing the KYOTO-based rules with other baseline rules improves the mapping results;
- testing if mapping all or part of the Base Concepts to BFO and propagating the mappings downwards would perform better or could be used in combination with the current method;
- resolving issues related to systematic polysemy by determining specific principles on their processing with BFO developers;
- studying the case of adjectives and their processing in terms of BFO types.

[8] The question of non-existent entities is itself an issue in ontology in general. For an overview, see: http://plato.stanford.edu/entries/nonexistent-objects/.

[9] See OBI 'value specification' (OBI:0001933) and IAO 'measurement datum' (IAO:0000109).

Acknowledgements

Work on this paper was supported in part by the Swiss National Science Foundation (SNSF) and by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida UL1 TR000064. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Thanks also to the anonymous reviewers, Christopher Crowner, and Barry Smith.

References

Robert Arp, Barry Smith, and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Aldo Gangemi, Roberto Navigli, and Paola Velardi. 2003. The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 820–838.

Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. 2010. Interfacing WordNet with DOLCE: towards OntoWordNet. In Chu-ren Huang, Nicoletta Calzolari, and Aldo Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*, pages 36–52. Cambridge University Press.

Pierre Grenon. 2003. BFO in a Nutshell: A Bi-categorial Axiomatization of BFO and Comparison with DOLCE. IFOMIS Report 06/2003. Technical report, Institute for Formal Ontology and Medical Information Science (IFOMIS), University of Leipzig, Leipzig, Germany.

Nicola Guarino and Christopher Welty. 2002. Evaluating ontological decisions with ontoclean. *Commun. ACM*, 45(2):61–65, February.

Axel Herold, Amanda Hicks, German Rigau, and Egoze Laparra. 2009. Central Ontology Version - 1 Deliverable 6.2. Technical report.

Amanda Hicks and Axel Herold. 2009. Evaluating ontologies with rudify. In Jan L. G. Dietz, editor, *Proceedings of the 2nd International Conference on Knowledge Engineering and Ontology Development (KEOD'09)*, pages 5–12. INSTICC Press.

Zubeida Casmod Khan and C. Maria Keet. 2013. Addressing issues in foundational ontology mediation. In *Proceedings of KEOD'13*, pages 5–16, Vilamoura, Portugal, September 19–22. SCITEPRESS.

Egoitz Laparra, German Rigau, and Piek Vossen. 2012. Mapping WordNet to the Kyoto ontology. In *LREC*, pages 2584–2589.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

I. Niles and A. Pease. 2003. Linking Lexicons and Ontologies: Mapping Wordnet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416. Citeseer.

Adam Pease and Christiane Fellbaum. 2010. Formal ontology as interlingua: The SUMO and WordNet linking project and global WordNet. In Chu-ren Huang, Nicoletta Calzolari, and Aldo Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge University Press.

Selja Seppälä, Barry Smith, and Werner Ceusters. 2014. Applying the realism-based ontology-versioning method for tracking changes in the basic formal ontology. In *8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*, Rio de Janeiro, Brazil.

Selja Seppälä. 2015a. Mapping WordNet to the Basic Formal Ontology using the KYOTO ontology. In *Proceedings of ICBO 2015*.

Selja Seppälä. 2015b. An ontological framework for modeling the contents of definitions. *Terminology*, 21(1):23–50.

A. Patrice Seyed. 2009. Bfo/dolce primitive relation comparison. In *Nature Precedings*.

Barry Smith and Werner Ceusters. 2010. Ontological Realism: A Methodology for Coordinated Evolution of Scientific Ontologies. *Applied Ontology*, 5:139–188.

Barry Smith, Mauricio Almeida, Jonathan Bona, Mathias Brochhausen, Werner Ceusters, Melanie Courtot, Randall Dipert, Albert Goldfain, Pierre Grenon, Janna Hastings, William Hogan, Leonard Jacuzzo, Ingvar Johansson, Chris Mungall, Darren Natale, Fabian Neuhaus, Anthony Petosa Robert Rovetto, Alan Ruttenberg, Mark Ressler, and Stefan Schulz, 2012. *Basic Formal Ontology 2.0: DRAFT SPECIFICATION AND USER'S GUIDE*, July.

Andrew D. Spear, 2006. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany.

Lynda Temal, Arnaud Rosier, Olivier Dameron, and Anita Burgun. 2010. Mapping BFO and DOLCE. *Studies In Health Technology And Informatics*, 160(Pt 2):1065–1069.

Piek Vossen, German Rigau, Eneko Agirre, Aitor Soroa, Monica Monachini, and Roberto Bartolini. 2010. KYOTO: an open platform for mining facts. In *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, pages 1–10.