Mapping WordNet to the Basic Formal Ontology using the KYOTO ontology

Selja Seppälä [1]*

[1]Department of Philosophy, University at Buffalo, USA

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; Laparra *et al.*, 2012; 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.[1]

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, which includes an upper-level ontology similar to BFO. 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 another ontology already linked to WN.

2 ONTOLOGICAL AND LEXICAL RESOURCES

The **Basic Formal Ontology (BFO)** is a domain-neutral upper-level ontology (Smith *et al.*, 2012). 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 the Cell Line Ontology (CLO), 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). 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., the 'noun.artifact' lexicographer file contains "nouns denoting man-made objects"[3]).

The **KYOTO ontology** (hereafter KYOTO) 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 the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE-Lite-Plus, version 3.9.7) and OntoWordNet (Gangemi *et al.*, 2003). **DOLCE** 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). (3) The third layer contains domain-specific classes (e.g. from the environmental domain).

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.
3. Implementing the 33 resulting mapping rules in a Python pipeline using the natural language toolkit for Python that integrates WN 3.0[4] (NLTK 3.0).

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 a WN synset using the WN-KYOTO mapping data files.[5] For example, the synset `immunity.n.02` is linked to:

*To whom correspondence should be addressed: seljamar@buffalo.edu

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

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

[3] See http://wordnet.princeton.edu/man2.1/lexnames.5WN.html.

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

[5] http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index9c60.html?option=com_contentview=articleid=429Itemid=156

```
'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 ones (BFO leaf nodes): the program tests if a given string (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

We manually evaluated the method on the 106 synsets in KYOTO marked with a 'medicine' topic domain. 72% of the assigned BFO types were correct (63% of the synsets were assigned the expected BFO type; 8% a superclass). As hypothesized, all the correctly categorized synsets were nominal and verbal. 27% of the assigned BFO types were incorrect (mostly adjectives). One synset was not matched by any rule.

5 DISCUSSION

WN is too large to be manually mapped to BFO. Using the properties of the hypernym hierarchy, we could have approached the problem by mapping the top levels of WN to the relevant BFO types, and propagating the mapped BFO types downwards. However, WN's organization fails to comply with basic ontological principles (Gangemi *et al.*, 2010). Moreover, that method would only cover nouns and verbs, while KYOTO also includes adjectives.

Mapping DOLCE to BFO is not trivial: their categories do not align in every case and are in some cases governed by different axioms. 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. However, these differences will not matter for our purposes here. Mapping WN to BFO is not trivial: WN represents linguistic usage; BFO, entities in the world. WN thus includes synsets that, in BFO terms, do not refer (at all or to a BFO type, e.g. `positive.a.04`). 10 synsets in the evaluation set posed categorization issues.

Our solutions to these issues are: (1) to extend the coverage of the rules by adding other types included in KYOTO and WN's semantic labels; (2) to ignore the axiomatizations. Indeed, 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? The resulting mappings are to be read as 'a WN synset X refers to something that is a subtype of BFO type Y', as in 'the synset `immunity.n.02` refers to a subtype of the BFO type DISPOSITION' — we exclude instances for now. Even a partial mapping should be sufficient to cover a large portion of WN, leaving a smaller subset of problematic cases. An interesting challenge will be to provide BFO-compliant interpretations of unmatched WN synsets.

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 more work is needed to see if the method scales to the full WN. Future work will include: extending the evaluation set of medical synsets using hyponymy relations and other domain resources; carrying out more thorough evaluations, e.g., by randomly extracting samples of synsets grouped by part of speech; augmenting the mapping rules by exploiting other resources, e.g., WN-SUMO mappings and ontologies extending BFO.

ACKNOWLEDGEMENTS

Work on this paper was supported by the Swiss National Science Foundation (SNSF). Thanks also to Christopher Crowner, Barry Smith, and Alan Ruttenberg.

REFERENCES

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

Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. (2003). Sweetening WordNet with DOLCE. *AI magazine*, **24**(3), 13–24.

Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. (2010). Interfacing WordNet with DOLCE: towards OntoWordNet. In C.-r. Huang, N. Calzolari, and A. Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*, pages 36–52. Cambridge University Press.

Grenon, P. (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.

Khan, Z. C. and Keet, C. M. (2013). Addressing issues in foundational ontology mediation. In *Proceedings of KEOD'13*, pages 5–16, Vilamoura, Portugal. SCITEPRESS.

Laparra, E., Rigau, G., and Vossen, P. (2012). Mapping WordNet to the Kyoto ontology. In *LREC*, pages 2584–2589.

Niles, I. and Pease, A. (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.

Pease, A. and Fellbaum, C. (2010). Formal ontology as interlingua: The SUMO and WordNet linking project and global WordNet. In C.-r. Huang, N. Calzolari, and A. Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge University Press.

Seppälä, S., Smith, B., and Ceusters, W. (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.

Seyed, A. P. (2009). BFO/DOLCE Primitive Relation Comparison. In *Nature Precedings*.

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

Smith, B., Almeida, M., Bona, J., Brochhausen, M., Ceusters, W., Courtot, M., Dipert, R., Goldfain, A., Grenon, P., Hastings, J., Hogan, W., Jacuzzo, L., Johansson, I., Mungall, C., Natale, D., Neuhaus, F., Rovetto, A. P. R., Ruttenberg, A., Ressler, M., and Schulz, S. (2012). *Basic Formal Ontology 2.0: DRAFT SPECIFICATION AND USER'S GUIDE*.

Temal, L., Rosier, A., Dameron, O., and Burgun, A. (2010). Mapping BFO and DOLCE. *Studies In Health Technology And Informatics*, **160**(Pt 2), 1065–1069.

Vossen, P., Rigau, G., Agirre, E., Soroa, A., Monachini, M., and Bartolini, R. (2010). KYOTO: an open platform for mining facts. In *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, pages 1–10.