An Evaluation of the Brandeis Semantic Ontology

Catherine Havasi, Anna Rumshisky, James Pustejovsky
Computer Science Dept., Brandeis University
415 South St.
Waltham, MA 02454
{havasi, arum, jamesp}@cs.brandeis.edu

Abstract

The Brandeis Semantic Ontology (BSO) is a new English resource in the generative lexicon tradition. Although still in development, the BSO is a sizable resource containing both a type system and a network of qualia relations. In this paper, we demonstrate that the BSO tends to contain correct qualia relationships by matching the current progress on the BSO with two different sources for qualia relation pairs. We compare with pairs semi-automatically selected from the British National Corpus and pairs automatically extracted from ConceptNet. We find that the current BSO data matches well with the extracted qualia pairs, but that the BSO is missing many of the qualia relations that could be extracted from corpora. We briefly propose a course of action to correct this problem.

1 Introduction

Generative Lexicon (GL) is a theory of lexical semantics that provides a different perspective on many of NLP's most important questions, from lexical creativity to the nature of word meaning. To this end, GL is meant to be included in a wide variety of approaches to NLP tasks [8][9]. One major hurdle to the inclusion of GL data and principles in NLP systems is the lack of any viable way for most researchers to create the network of qualia relations required to fully utilize GL [4]. Creating such a network of qualia relations would be a large investment of time and resources, and current GL implementations place burdensome constraints on any lexical development project.

The Lab for Linguistics and Computation (LLC) has been developing a large GL ontology and lexicon, called the Brandeis Semantic Ontology (BSO), for use by the general research community [10][11] to help alleviate this problem and to encourage the widespread use of GL by the NLP community.

We wish to use an evaluative comparison to show how well our current qualia reflect standard word uses found in corpora and elicited from humans. We would also like to evaluate the completeness of our collected qualia. In this paper we compare the current BSO information with pairs semi-automatically selected from the British National Corpus (BNC) and automatically extracted from ConceptNet. This information would allow us to further direct our development of the BSO and to develop ways to quickly aid in the addition of other qualia pairs to our resource.

1.1 What is the BSO?

The Brandeis Semantic Ontology is a large GL-based English-language lexical resource comprised of an ontology and a lexicon. At the heart of the BSO is the network of qualia structures found in the ontology. There are four types of qualia in the BSO: formal, agentive, telic and constitutive. These types are defined in 4.

A qualia structure is intended to express the componential aspect of a word's meaning [2]. Thus, these are an especially important part of the BSO, as they represent the connections and relations between words. For example, if one were to look up the phrase BEER in the BSO, one

1. FORMAL: the basic type distinguishing the meaning of a word;
2. CONSTITUTIVE: the relation between an object and its constituent parts;
3. TELIC: the purpose or function of the object, if there is one;
4. AGENTIVE: the factors involved in the object's origins or "coming into being".

Figure 1: Types of qualia in the BSO

would find its type is ALCOHOLIC BEVERAGE and its type's type is BEVERAGE. We give an example of this in 1.1.

The BSO's lexicon contains information about specific words and links from each word to an ontological category and thus to the ontology's qualia structures. The BSO will make available four pieces of GL information for each lexical item: type structure, argument structure, event structure, and qualia structure. Some of these are provided through access to information inside the entry of a lexical item's corresponding ontological type. We can only hope to capture at most a few of the prototypical values of these fields in each ontological or lexical item. At the time of this paper's publication, the BSO has been created mainly through introspection.

The specification of the ontology largely matches that proposed by the SIMPLE specification [1]. The ontology consists of three major types: entity, event, and property. Most of the current work has been done in the entity and event hierarchies, and each of these is divided into three smaller hierarchies: natural, artifactual, and complex.

Currently, the BSO has over 40,000 lexical entries and 3,700 ontological nodes as of the time of this study. Of these ontological nodes, only 1,800 have qualia relations. These statistics are subject to change since the BSO is currently under development.

2 Evaluating the BSO

First and foremost, we wanted to test the correctness of the qualia relations in the BSO and to test the possibility of using other resources to bootstrap the creation of the BSO database. We also choose to test the coverage of the BSO's qualia relations even though the BSO is unfinished in this respect and thus we expect the number to be fairly low.

We have extracted qualia information using a naïve process from two lexical resources to use in this evaluation: the British National Corpus (BNC) and ConceptNet [7]. From these resources, we extract word pairs which match basic criteria for a specific qualia type. We selected telic relationships from both ConceptNet and the BNC, while agentive relationships were selected from just the BNC. We then measured how many of these pairs were present in the BSO as of January 2007. We also tested the matching BSO qualia connections to detect how similar these connections were to those in the test sets.

2.1 The BNC and ConceptNet

Our test sets were selected based on several criteria. Corpora were selected from which we could easily extract the noun-verb pairs for use in the evaluation. We also needed variety in the types and compositions of lexical resources. Although work [4] has been done in recognizing qualia relationships in text, much of this work requires extensive semantic tagging and doesn't differentiate between the different types of qualia relations.

Type: ALCOHOLIC BEVERAGE
Indirect Agentive: CREATE MATERIAL ENTITY
Constitutive: ALCOHOL,...
Telic: DRINK ACTIVITY

Figure 2: A BSO Qualia Structure for Beer

2.2 The BNC and GL

To use data from the BNC, noun-verb pairs which could potentially form a qualia relation were extracted. We chose to only extract noun-verb pairs from this resource since we felt we would recover an adequate number of pairs using this method.

We used the Bonito[12] concordance tool and query patterns which were developed from patterns explored in [6], such as *[lempos="begin-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]*. Some of these patterns can be seen in 3.

[lempos="begin-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="finish-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="enjoy-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="start-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="like-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="dislike-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="want-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

[lempos="attempt-v"] [tag="D.*"]* [tag="AJ.*"]* [tag="N.*"]

Figure 3: Sample Patterns for extracting data from the BNC

The concordance produced 17,090 results, the vast majority of which were either coercions or non-qualia word relationships. In our evaluation, we did not examine words used in nonstandard contexts. An annotator trained in GL reviewed these pairs and determined that 292 of them contained a telic or agentive relationship, and we used these pairs in our study. We realize that it is unlikely that both the patterns and the annotator correctly identified all the qualia patterns in the BNC. However, we were only interested in a sample of the qualia present in the BNC rather than an exhaustive search.

2.3 What is ConceptNet?

Open Mind Common Sense (OMCS)[13] is a large-scale project to collect, process and use "common sense" data collected from volunteers who use its internet sites. This data is elicited from visitors using fill-in-the-blank frames, confirmation of computer-inferred statements, and free text entry. Since 2000, the project has collected over 700,000 English language common sense statements. Although solicited from volunteers, the corpus largely contains correct information. A sample of the corpus was evaluated by human judges yielding positive results: three quarters of the sample assertions were found to be "largely true," over four fifths were found to be largely objective or sensible, and 84% of the sample assertions were deemed common enough to be known by someone by high school [13].

ConceptNet[7] is a representation and wrapping of the Open Mind Common Sense corpus that is easy for a variety of applications to process. It is a semantic network of over 20 relation types and two hundred thousand relations. These relations are created by a pattern-matching process [7]. Each sentence in the OMCS database is compared with an ordered list of patterns, which are regular expressions that can also include additional constraints on phrase types based on the output of a natural language tagger and chunker.

2.4 ConceptNet and GL

Some of the relation types in ConceptNet line up well with qualia relations. If we revisit the example about beer from earlier, we can see that the OMCS database contains assertions whose information matches the information currently contained in the BSO. Some examples can be found in 4.

Since the relation types in ConceptNet matched those in the BSO so well, we took matching ConceptNet relations and extracted the word pairs in the relation for use in the evaluation. In

OMCS Statement	Corresponding BSO Qualia Role
"Beer is a type of alcoholic beverage".	Type
"beer is for drinking."	Telic
"Something you find in beer is alcohol.	Constitutive
"Beer is made from hops and barely".	Constitutive
"all beer has the property of being brewed."	Agentive

Figure 4: OMCS database entries and their corresponding qualia roles.

general, **IsA** maps well to the formal qualia, **PartOf** to the constitutive, **UsedFor** the telic. There are 46,742 **UsedFor** relationships in ConceptNet, which should be sufficient for our evaluation and extracted pairs of words which were present in such relationships.

Although the data in ConceptNet is inherently somewhat noisy since it was collected from the internet, we felt that the level of noise was small enough that human review of relevant qualia pairs were not necessary for this test. This causes the ConceptNet test to have more noise than the BNC test, and the results should be viewed accordingly.

3 Experiment One: Comparison with the BNC

3.1 Qualia Node Coverage

We began by measuring the percentage of annotated BNC pairs that are found in the BSO in order to obtain an estimate of what percentage of the nodes had been populated. Of the 292 pairs selected, 43 matches were found, showing that of the candidate nodes, only 15% of the nodes were populated.

Each potential pair found in the BNC was verified as a telic qualia candidate by checking if the first word in the pair was a noun and the second a verb. We then check to ensure that the BSO has a telic or agentive qualia record for the first item in the pair. These two steps were fully automated.

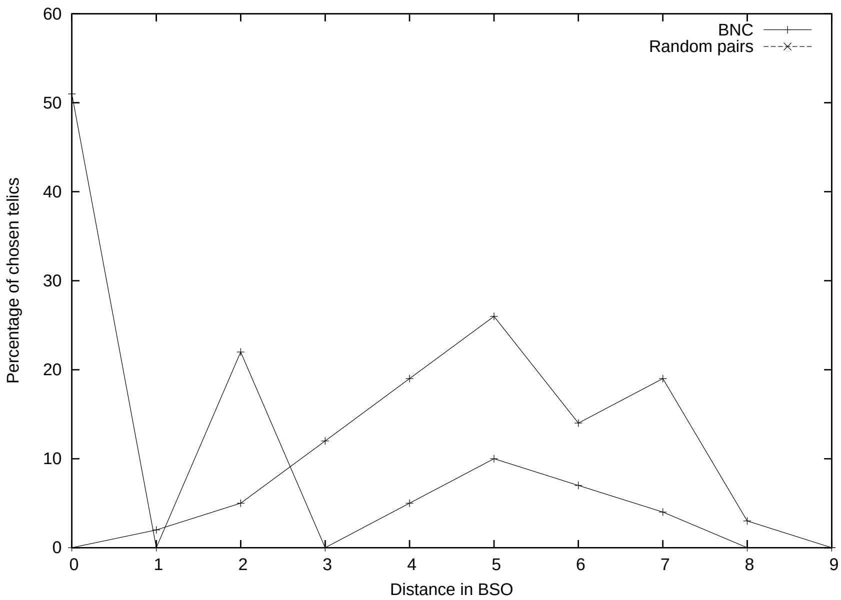

Figure 5: Distance between target qualia and BSO qualia for BNC pairs

3.2 Accuracy

In our experiments, two metrics were used to test the accuracy of the BSO against the sample pairs from both the BNC and ConceptNet: distance and score. Once a pair has passed the above coverage test, we know that the initial word in the pair is recorded as having an appropriate qualia entry in the BSO. The type, or formal, structure of the BSO makes up an ontological tree, and this tree can be used to calculate the similarity between two lexical items by measuring the distance between them, as in [15]. As one of our accuracy metrics, we measured the distance between the second item of an extracted qualia pair and the BSO's qualia for the first pair[1]. We then compared these distances with similarly computed distances for 43 randomly chosen pairs of lexical items from the BSO. As can be seen in 3.1, the BNC-BSO pairs performed significantly better than the random pairs. In fact, 51% of the BSO-BNC pairs had distance zero, signifying that the BSO correctly had the same qualia pair as the sample. The average distance between the the BSO and target qualia was 1.84 steps as opposed to 5.00 in the random sample. This indicates that the qualia pairs which were not exact matches were close to each other in the tree structure.

As the BSO is still under construction, the density of nodes in different areas of the type tree are very different. To account for the density variation appropriately, we improve on the distance metric by developing a score function which uses the branching factors of the nodes along the path between the BSO and target qualia items. If the two items referenced the same node, the score was one. Otherwise, the score was the product of the inverses of the branching factors for each node, where the branching factor is defined to be the number of ontological children of a node. Thus, a higher score indicates a better match. The average score of a pair in the BNC was 0.56, while in the random sampling the average score was 0.11. A full comparison of scores can be seen in 4.2. This shows that mismatches between the BSO and target qualia obtained from the BNC tend to be found in dense area areas of the graph which have many sparsely-populated, specific ontological nodes. We are actively working to reduce the occurrence of these areas in the BSO.

4 Experiment Two: Comparison with ConceptNet

We used the observed connections between ConceptNet relation types and BSO qualia to compare ConceptNet's data to the telic, formal and constitutive relations in the expert-created BSO. At the time of the study, ConceptNet did not have enough data collected in an agentive-related predicate to complete an evaluation of the agentive qualia.

4.1 Qualia Node Coverage

The same test metrics from Experiment One were run on the ConceptNet pairs. In order to compare these relations, we needed a mapping between nodes in the BSO and nodes in ConceptNet. The BSO focuses on representing single words, so we need to restrict the predicates we examine in ConceptNet to those connecting two concepts that normalize to a single word. Such predicates make up 11.1% of the UsedFor relations, 21.0% of IsA, and 31.2% of the PartOf relations in ConceptNet. This is 2,011 IsA relations, 384 UsedFor relations, and 57 PartOf relations.

4.2 Accuracy

We again ran our distance metric with the target qualia from ConceptNet and the qualia present in the BSO. The results can be seen in figure 4.1. We found that 22.0% of UsedFor relations, 17.5% of PartOf relations and 21.2% of IsA relations were direct matches between the representations. We then ran the same study with 538 random lexical pairs and in the random study only 1.2% were direct matches.

The results show that ConceptNet predicates appear as qualia relationships in the BSO many times more often than random chance; see 4.2. A Pearson's chi-square test of independence showed that the difference was very statistically significant, with $p < 0.0001$ for each relation type. 500 random predicates had 6 hits and 494 misses; in comparison, PartOf predicates had 10 hits and

[1]Since we used both telic and agentive in this case, we disambiguated qualia type when the first item had multiple qualia by checking to see which path was shortest and using only this path.

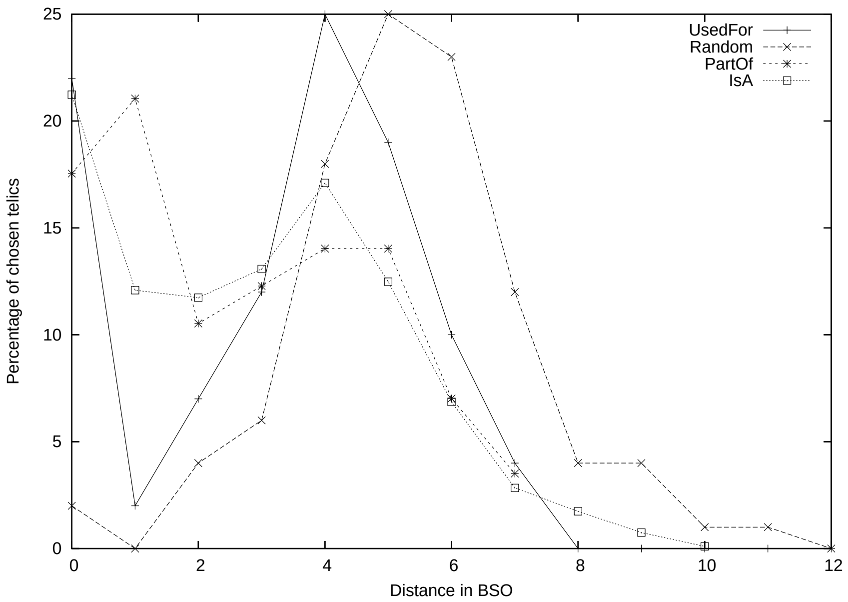

Figure 6: Distance between target qualia and BSO qualia for ConceptNet pairs.

47 misses ($\chi^2 = 49.0, df = 1$), UsedFor predicates had 84 hits and 300 misses ($\chi^2 = 101.5, df = 1$), and IsA predicates had 427 hits and 1586 misses ($\chi^2 = 112.6, df = 1$).

The branching factor score was then calculated. We found that the branching factor score for ConceptNet pairs was 0.32, which exceeds the score of 0.06 for a random sampling of pairs. These results are very similar to those found with the BNC pairs.

5 Discussion of Results and Future Directions

The telic, constitutive, and agentive qualia relations present in the BSO match well with qualia extracted from two very different English language resources. This is very promising, as it demonstrates that the developing BSO tends to contain correct qualia relationships. It also shows that the BSO's qualia relations in the BSO line up well with standard word use in corpora, as well as more colloquial use.

The BSO performs much better than random chance, and most "near misses" are located in dense areas of the BSO tree and are thus more likely to fall prey to minute distinctions between types, a known problem the BSO developers are trying to address.

However, we found that the BSO does not contain the vast majority of qualia relations that we have found in the BNC and ConceptNet. This could be because our pairs do not adequately represent the qualia structures used in everyday language, but it is more likely that the BSO is simply not complete enough to contain these qualia relations.

In the future, we could use resources such as ConceptNet to quickly and automatically acquire potential qualia relation pairs to be reviewed by a human. These pairs would then help us more rapidly populate the BSO. Preliminary work has been done in this direction, and we hope that such techniques will make the creation of a resource such as the BSO easier and less time-consuming. We hope that such techniques will assist expert annotators by providing them with a large body of appropriate empirical data to assist in generating qualia relations.

6 Acknowledgments

We would like to thank the entire Open Mind group, especially Rob Speer and Jason Alonso for the use of ConceptNet 3. Catherine Havasi is funded by an NSF Fellowship.

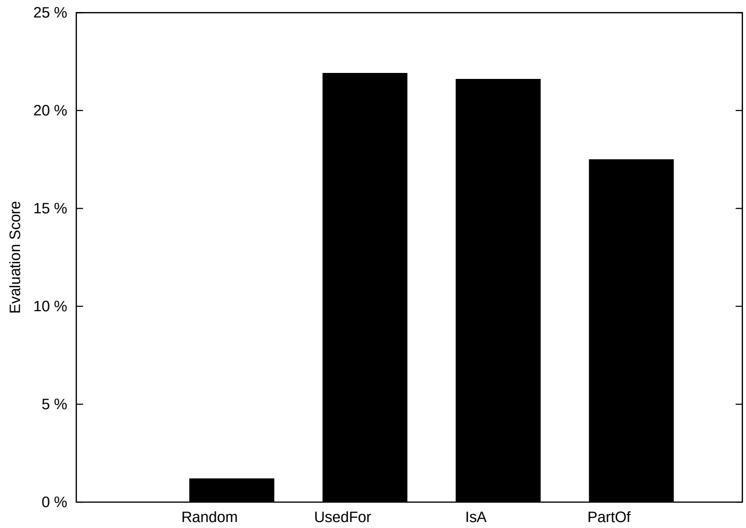

Figure 7: A very significant percentage of applicable relationships in ConceptNet match with relationships in the BSO

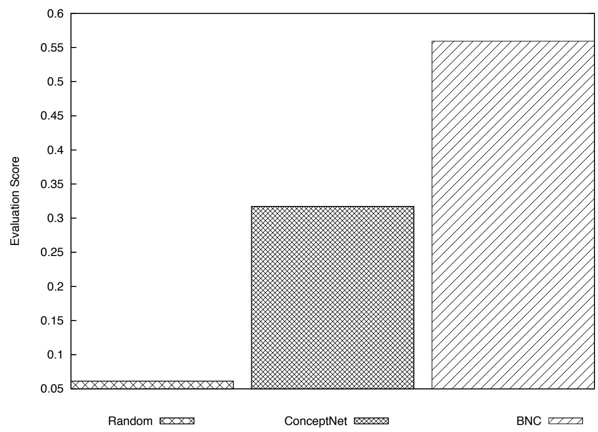

Figure 8: A Comparison between scores

References

[1] F. Busa, N. Calzolari, and A. Lenci. *Generative Lexicon and the SIMPLE Mode: Developing Semantic Resources for NLP.* Cambridge University Press, Cambridge, 2001.

[2] N Calzolari. *Acquiring and Representing Semantic Information in a Lexical Knowledge Base.* Springer Verlag, New York, 1992.

[3] V. Claveau, P. Sébillot, C. Fabre, and P. Bouillon. Learning semantic lexicons from a part-of-speech and semantically tagged corpus using inductive logic programming. *Journal of Machine Learning Research*, 4:493–525, 2003.

[4] Vincent Claveau, Pascale Sébeillot, Céile Fabre, and Pierrette Bouillon. Learning semantic lexicons from a part-of-speech and semantically tagged corpus using inductive logic programming. *Journal of Machine Learning Research*, 2003.

[5] Catherine Havasi, Robert Speer, and Jason B. Alonso. ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. Submitted to the 22nd Conference on Artificial Intelligence, Vancouver, Canada, 2007.

[6] Maria Lapata and Alex Lascarides. A probabilistic account of logical metonymy. *Computational Linguistics*, 2003.

[7] Hugo Liu and Push Singh. ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 2004.

[8] James Pustejovsky. Lexical knowledge representation and natural language processing. *Artificial Intelligence*, 63:193–223, 1993.

[9] James Pustejovsky. *The Generative Lexicon.* MIT Press, Cambridge, MA, 1995.

[10] James Pustejovsky, Catherine Havasi, Roser Saurí, Patrick Hanks, and Anna Rumshisky. Towards a generative lexical resource: The Brandeis Semantic Ontology. *Proceedings of the Fifth Language Resource and Evaluation Conference*, 2006.

[11] A. Rumshisky, P. Hanks, C. Havasi, and J. Pustejovsky. Constructing a corpus-based ontology using model bias. In *The 19th International FLAIRS Conference, FLAIRS 2006*, Melbourne Beach, Florida, USA, 2006.

[12] Pavel Rychly and Pavel Smrz. Manatee, Bonito and word sketches for Czech. In *Second International Conference on Corpus Linguisitcs*, 2004.

[13] Push Singh, Thomas Lin, Erik T. Mueller, Grace Lim, Travell Perkins, and Wan Li Zhu. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proceedings of First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Irvine, California, 2002.

[14] Robert Speer. Open Mind Commons: An inquisitive approach to learning common sense. *Proceedings of the Workshop on Common Sense and Interactive Applications*, 2007.

[15] Ben Wellner, James Pustejovsky, Catherine Havasi, Roser Sauri, and Anna Rumshisky. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of SIGDIAL 2006*, Sydney, Australia, 2006.