

Handling Subtle Sense Distinctions through Wordnet Semantic Types

Sofia Stamou^{*+} Dimitris Christodoulakis^{*+}

^{*}Computer Engineering and Informatics Department
Patras University, 26500, and

⁺Research Academic Computer Technology Institute
61 Riga Feraiou, 26221, Patras, Greece
{stamou, dxri}@cti.gr

Abstract

In this paper we challenge the question of whether there is value in having multiple layers of semantic information associated with corpus semantic annotation. In this context we introduce a semantic annotation experiment in which novice annotators were asked to assign sense tags to a set of polysemous corpus nouns, using Wordnet as their referential sense repository. Wordnet is a rich sense inventory that provides explicit information of the semantic types associated with every word sense. To measure the effect semantic types' knowledge has on the sense assignment process, we carried out two annotation sessions. In the first session, annotators relied exclusively on Wordnet synsets to annotate corpus nouns, whereas in the second session the same pool of annotators examined Wordnet synsets in conjunction with their semantic types, prior assigning a sense tag. Comparing annotators' performance in both sessions shows that when consulting semantic types, annotators assigned more salient senses to highly polysemous nouns, whereas for the same set of terms, when relying exclusively on Wordnet synsets, annotators tended to assign narrower senses, which whatsoever were more error-prone. Results indicate that semantic types have a potential in dealing with subtle sense distinctions in the course of corpus annotation.

Introduction

Recently there has been a major breakthrough in the construction of linguistically annotated corpora (Hinrichs and Simov, 2002). Currently available corpora display various levels of annotations, ranging from morpho-syntactic (Marcus et al., 1993) to semantic (Kingsbury et al., 2002), and dependency-based annotations (Bohmová et al., 2000). A critical element in corpus annotation is ambiguity resolution. While morphosyntactic ambiguity can be effectively tackled via available Part-of-Speech (POS) taggers and shallow syntactic parsers, Word Sense Disambiguation (WSD) is a long-standing burden in corpus annotation (Resnik and Yarowsky, 1997). Many efforts in lexical ambiguity resolution suggest to use the context of the word to be disambiguated together with information about each of its word senses to solve the problem.

In this paper we investigate how multiple layers of semantic information contribute in coping with polysemy, in the course of annotating a small set of nouns of the Hellenic National Corpus (HNC) (<http://corpus.ilsp.gr>). To that end we evaluate the quality of the annotations appended to a set of corpus polysemous nouns by a group of novice annotators, who employed Greek Wordnet (Stamou et al., 2002) as their referential sense repository. Wordnet is a rich sense inventory that provides explicit information of the semantic types associated with each word sense. Semantic types are represented via Wordnet's Base Concepts (BCs), under which word senses are organized. In the framework of HNC annotation, BCs serve in anchoring an additional layer of semantic information to the tagged terms. This layer conveys information about the semantic types (i.e., human entity, abstraction, artefact, etc.) of corpus terms and deals with Wordnet's fine-grained sense distinctions.

The challenge in our evaluation was not only to assess the quality of the sense tags assigned to the corpus terms, but also to examine whether combining Wordnet senses

together with information of their semantic types results in better sense distinctions. To evaluate the impact semantic types' information has on corpus annotation, we carried out two experimental sessions in which annotators were asked to match a set of polysemous corpus nouns to a sense from Wordnet, with and without employing information of the senses semantic types. Annotations were cross-checked and evaluated on the basis of the semantic overlap among the tags appended to the same set of terms in each session.

Our findings show that when employing semantic types' information, annotators assigned more accurate tags to nouns referring to concrete entities, such as objects and persons. Essentially, the same was true for nouns having abstract referents, whose sense discrimination was more error-prone. To verify that improved tags were due to semantic types' knowledge, we manually examined tagging inconsistencies between the two annotation sessions. Evaluation of the results indicates that for highly polysemous nouns, annotators preferred to assign more salient senses when semantic types' information was available, as opposed to the narrow senses selected for the same set of terms when relying solely on Wordnet's senses. Following the observations of (Gale et al., 1992) and (Miller et al., 1994) that salient senses are in most cases appropriate sense tags, we infer that semantic types' knowledge helps annotators to deal with polysemy.

The remainder of the paper is organized as follows: in section (2) we report on how Wordnet was employed in the annotation process and we describe the experiments carried out. In section (3) we define the objectives of our evaluation and we present our experimental results. In section (4) we provide a detailed discussion of our results in an attempt to shed light on the effect semantic types' knowledge has in the annotation process. To that end we also propose some improvements that could be made to Wordnet, in order for the latter to be a powerful resource for corpus annotation. In section (5) we conclude the

paper and we address the usefulness of semantic types in other applications beyond corpus annotation.

The Annotation Framework

To examine the effect that semantic types' information has on corpus annotation, we performed an experiment, in which a group of three novice annotators was asked to match a set of corpus nouns to a Wordnet sense. Although our experiment exhibits many similarities to a hand-tagging task reported in (Fellbaum et al., 1997), nevertheless the objective of our approach is significantly different. In particular, we wanted to examine whether multiple layers of semantic information can produce better quality corpus annotations. To set up our experiment we selected a set of 350 polysemous corpus nouns and we extracted a set of corpus sentences in which these nouns occurred.¹ We removed HTML tags from these sentences, converted them in plain text format and imported them in a corpus annotation software (Stamou et al., 2003), which provides three levels of annotations i.e., morphological, semantic and semantic type annotations. The first annotation layer involves using a morphosyntactic tagger (Orphanos and Christodoulakis, 1999) to automatically assign morpho-syntactic tags to corpus terms. Morphosyntactic tagging involves markup of corpus tokens with information on their POS, lemma and morphosyntactic features, such as number, person, gender. The output of the morpho-syntactic tagger was manually revised and corrected by a team of linguists.

The second layer involves assigning sense tags to the 350 selected nouns. Word senses are obtained from Greek Wordnet's synsets and are manually appended by the annotators. Sense assignment involves examining all possible Wordnet senses of a given term, in order to select the one that best reflects the term's meaning in the given context. The number of alternative word senses for each of the 350 nouns ranged from two to eight. Finally, the third annotation layer involves employing semantic types' information in conjunction with Wordnet's senses in order to assign sense tags to the corpus nouns. Semantic types' information is conveyed through Wordnet's BCs. Every Wordnet sense is linked via hierarchical relations (i.e., hypernymy, hyponymy) to broader senses, the so-called BCs. These are lexical elements that represent universal concepts and serve to the clustering of word senses that share common semantic attributes and/or features.

Experimental setup

At the beginning of the experiment we carried out a training session, in which annotators were given specific guidelines on how to use the annotation software and what their task would be. The functionality of the annotation module is pretty simple: it displays one corpus sentence at a time, in which the term to be tagged is highlighted. In a separate field on the screen, the target sentence appears morphosyntactically tagged. Annotators could not modify the morphosyntactic tags but they were encouraged to consult the morphosyntactically tagged sentences, whenever they thought it would be helpful in their task. In a separate field on the screen all Wordnet senses matching to the target noun are displayed in a random order. Word senses are provided as synonym sets along with defining

glosses. Annotators were asked to examine all displayed senses prior selecting the one that best matches a term's usage in the given sentence. Selection is performed simply by clicking on the most suitable Wordnet sense. Sentences having their target noun annotated are stored in a database. Annotators were asked to study carefully all senses prior deciding on the most appropriate, since they could not alter their selection afterwards. Finally, while using the third annotation level, annotators were given additional semantic information for the nouns to be tagged, besides their Wordnet senses. This information involves the Wordnet's BCs that correspond to each of the displayed senses. Annotators could simply view the BC associated with each sense and acquire information of the sense's semantic type. Semantic types are employed as a guide towards deciding which of the senses is more appropriate. Again, sense assignment is performed by clicking on the most suitable sense. At the end of the training session annotators felt confident in using the software and they had a clear understanding of the objectives of their task.

The Experiment

Our experiment was split into two sessions. The same pool of annotators participated in both sessions and their task was essentially to match a set of 350 polysemous corpus nouns to the most appropriate Wordnet sense. What differentiated the two sessions was the semantic information that was available to the annotators. In particular, during the first session, annotators had to decide on a sense tag, simply by examining all possible Wordnet senses associated with a given term. Based on this knowledge, annotators had to select only one sense for each term occurrence and could not change their selection, once it was made. Wordnet senses were displayed as a list of glosses in a random order. In the second session, annotators besides using Wordnet senses, they had also to employ each sense's BC as a reference and then to decide which of the displayed senses is more suitable. We speculated that Wordnet's BC would inform annotators about the terms' semantic types and would help them discriminate between subtle senses. Thus, annotators were advised to take into consideration semantic types' information when they would not be able to discriminate between fine-grained senses. A time period of a week elapsed between the two sessions to reassure that the annotators would not be influenced by their previous judgements. There was not a time limitation in either session but it took the annotators approximately 3-5 hours to finish their task in each session. Annotators worked independently from each other.

Experimental Results

Annotations that were assigned to corpus nouns in each session were cross-checked and evaluated on the basis of: (i) their accuracy, (ii) their overlap percentage, and (iii) the degree of polysemy that the selected corpus nouns displayed. Annotations' accuracy is determined as:

$$\% \text{ correct} = 100 \times \frac{\# \text{correct_sense_tags}}{\# \text{assigned_sense_tags}}$$

Correct tags were those that were appended to the 350 corpus nouns by experienced lexicographers at the end of

¹ One sentence per word occurrence.

the experiment. The degree of the annotations' overlap was calculated by comparing annotations' accuracy in the two sessions. To evaluate obtained results we cross-checked the sense tags assigned in each session and we manually examined their inconsistencies.

An analysis of our results indicates that when annotators employed semantic types' information to distinguish between fine-grained Wordnet senses, they preferred the most *central* or *general* meaning among candidates. This annotation strategy displays a pattern similar to the frequency tagging condition, reported in (Fellbaum et al., 1997), in which taggers preferred the most frequent sense, being at the top of the list of senses. Even though in our experiment Wordnet senses were randomly ordered, nevertheless their semantic types' knowledge gave the annotators a global understanding of their semantics and somehow motivated them to prefer the most general meaning, as the safest choice. This was mainly the case for highly polysemous nouns, whose Wordnet senses shared all the same BC. Assigning the most inclusive meaning resulted in 76% tagging accuracy. This means that annotators assigned to 266 out of the 350 nouns a salient Wordnet sense, when they employed semantic types' information. Conversely, for the same set of terms, annotators when relying exclusively on Wordnet senses, assigned a correct sense tag to 224 nouns, which corresponds to 64% tagging accuracy.

To ensure that selection of the most central meaning was not biased by the annotators' knowledge that all Wordnet senses were of the same semantic type, we cross-checked and compared annotations between the two sessions, i.e., with and without employing semantic types. Our results show that without utilizing semantic types information, annotators assigned a different sense tag to 42 nouns, which accounts to 12% tagging discrepancies between the two sessions. 36 out of these 42 nouns were assigned a less central, but nevertheless inaccurate sense, when BCs were not consulted, reducing subsequently tagging accuracy by 10,2%. An interesting finding is that tagging errors were produced for those terms that displayed the greatest degree of polysemy among all examined terms (i.e. they had more than five Wordnet senses). Annotation disagreements increased with increasing polysemy, whereas agreement rates were highest for nouns having only two Wordnet senses. Disagreements due to polysemy are compatible with the findings reported by Jorgenson (1990), whose subjects discriminated only about three senses of highly polysemous nouns. Conversely, for nouns whose different senses had distinct semantic types (cf. Example 1), taggers succeeded in assigning the correct sense tag, when BCs were taken into consideration, in 95% of the cases, which accounts to 332 correct tags out of the 350.

```
[result]
sense 1:
  something that follows and is caused by something
  else
  has BC (PHENOMENON)
sense 2:
  a set of elements that follows an action
  has BC (EVENT)
sense 3:
  a statement that solves a problem
has BC (ABSTRACTION)
```

Example 1: Senses with distinct semantic types

Moreover, semantic types' information assisted annotators to discriminate between senses of nouns with abstract referents, as for example the noun *democracy*. Semantic types' knowledge, encoded in Wordnet's BCs, helped annotators discriminate between the usage of the term democracy in the sense of *political orientation* and the use of democracy in the sense of *political system*. Democracy as a political orientation has *psychological feature* as its BC, whereas in the sense of political system it has *group* as BC. This tagging improvement implies that semantic types' information contributes into a better understanding of the distinctions between word senses. On the other hand, the same set of terms when tagged on the basis of Wordnet senses alone, resulted in 88% tagging accuracy,² confirming our intuition that semantic types' knowledge can accommodate sufficiently difficulties associated with subtle sense distinctions.

In sum, we found that Wordnet is a valuable resource for semantic annotation, but does not always help in dealing with polysemy problems. This is essentially the case for highly polysemous terms, whose senses are all of the same semantic type, i.e. they all share the same BC. In such cases annotators cannot successfully discriminate between the many fine-grained senses and as such they are subjected to erroneous annotations. In our annotation framework, sense assignment is augmented through the use of BCs, which explicitly append semantic types' tags to Wordnet senses, in an attempt to improve the annotators performance. Our evaluation reveals that semantic types have a potential in helping annotators discriminate between multiple senses, by providing a fairly clear representation of the terms' semantics.

Discussion

Annotations' accuracy in both sessions implies that the inexperienced annotators found their task duable and the annotation software easy to use. The improved tagging accuracy, when BCs were employed, demonstrates that Wordnet's semantic types are a useful semantic resource for distinguishing between multiple word senses. However, this is true only if different senses belong to distinct semantic types. This is mainly because BCs give annotators a clear idea on the terms' semantics and make them capable of understanding how senses are split in Wordnet.

Conversely, for Wordnet senses that they are all of the same semantic type, it was found that when BCs were consulted, the assigned tags represented more general meanings of the terms. Even though general meanings were appropriate in most cases, due to the fact that our corpus contains general-language vocabulary, however we assume that semantic types' information being identical across all senses, might have confused annotators. Thus, they preferred the most central meaning of a term over the other senses, as the safest choice. Another possible explanation for the improved annotations, when employing semantic types, might be the fact that in the second annotation session, taggers were somehow familiar with the corpus and the sense tagging process. Even though we tried to eliminate this possibility by allowing a time interval between the two sessions, nevertheless the

² 308 out of the 350 nouns were assigned a correct sense tag, when relying on Wordnet senses alone.

annotators' familiarity with their task might have slightly effected the quality of the annotations produced.

In general, we found that BCs have a potential in facilitating corpus annotation tasks, however we argue that their potential is limited to nouns' annotations. For other syntactic categories (e.g., verbs, adjectives and adverbs), BCs are less informative of the terms' semantic types. To accommodate this problem, we propose the exploitation of Wordnet's Top-Ontology, instead of the BCs, as a guide towards discriminating between multiple word senses. Another approach to this direction might be similar to the one proposed in Ng and Lee (1996) in which a combination of multiple knowledge sources was employed against annotation. We also found that when annotators relied solely on Wordnet senses to assign a sense tag, they tended to prefer narrower meanings, which however were not always correct. This might be attributed either to Wordnet's fine-grained sense distinctions, or to the annotators' inability to distinguish between multiple senses, given the fact that they were all novice taggers.

Summarizing, we found that Wordnet is a rich sense inventory that can substitute the dictionary look-up process towards semantic annotation tasks, but it still needs further improvements to that end. The most crucial enhancement that Wordnet needs to go through is a preliminary clustering of its multiple word senses so that redundancies are significantly reduced. It would also be useful that the annotators employ Wordnet's semantic relations towards distinguishing between multiple senses. Intuitively, we believe that semantic links would provide the annotators with a better understanding of the terms' semantic properties, resulting hence into more accurate annotations. An overall analysis of our results indicates that locating the correct usage of words in a repository of word senses is a hard task, whose difficulty increases as the number of the senses grows. However, Wordnet due to its structure can provide multiple layers of semantic information, which, when utilized in semantic annotation, can improve the annotations produced.

Conclusions

We reported on a corpus annotation approach, in which a group of novice annotators employed multiple layers of semantic information encoded in Wordnet to semantically annotate a set of 350 polysemous corpus nouns. The main challenge in this attempt was to investigate if semantic types' information represented in Wordnet's BCs can contribute in discriminating between subtle senses in the course of corpus annotation. To that end we carried out two experimental annotation sessions, with and without employing semantic types' information respectively. Obtained annotations were cross-checked and evaluated on the basis of their tagging accuracy overlap. An analysis of the results shows that Wordnet's BCs have a potential in helping annotators distinguish between multiple senses, but nevertheless their potential is limited to those senses that belong to different semantic types. For all other cases, semantic types' knowledge does not add value to the annotation process and thus it cannot contribute in helping annotators pick-up the best matching sense. Irrespectively of corpus annotation, we argue that semantic types are a useful source of semantic information that can contribute in other applications, such as text summarization, text classification, information extraction and so forth.

References

- Bohmova A., Hajic J., Hajicova E. & Hladka B. (2000) The Prague Dependency Treebank: A Three-Level Annotation Scenario. In *Treebanks: Building and Using Syntactically Annotated Corpora*. Kluwer Academic Publishers, Dordrecht
- Fellbaum Ch., Grabowski J. & Landes Sh. (1997) Analysis of a Hand-Tagging Task. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What and How?* Washington D.C., USA
- Gale W., Church K. & Yarowsky D. (1992) Estimating Upper and Lower Bounds on the Performance of Word-Sense Disambiguation Programs. In *Proceedings of the 30th ACL Meeting*, pp: 249-256
- Hellenic National Corpus: <http://corpus.ilsp.gr>
- Hinrichs E. & Simov K. (Eds.) (2002) *First Workshop on Treebanks and Linguistic Theories*, accessible via (<http://www.BulTreeBank.org/Proceedings.html>)
- Jorhenson J. (1990) The Psycholinguistic Reality of Word Senses. In *Journal of Psycholinguistic Research*, vol.19, pp.167-190
- Kingsbury P., Palmer M. & Marcus M. (2002) Adding Semantic Annotation to the Penn Treebank. In *Proceedings of the Human Language Technology Conference*, San Diego, CA, USA
- Marcus M., Marcinkiewicz M.A. & Santorini B. (1993) Building a Large Annotated Corpus of English: The Penn Treebank. In *Computational Linguistics*, vol.19, no.2, pp.313-330
- Miller G., Chodorow M., Landes S., Leacock C. & Thomas R. (1994) Using a Semantic Concordance for Sense Identification. In *Proceedings of the ARPA Human Language Technology Workshop*, San Francisco, USA
- Ng H. & Lee H. (1996) Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach. In *Proceedings of the 34th ACL Meeting*, Santa Cruz, CA, USA, pp. 40-47
- Orphanos G. & Christodoulakis D. (1999) Part-of-Speech Tagging and Unknown Word Guessing with Decision Trees. In *Proceedings of the EACL Conference*, Bergen, Norway
- Resnik P. & Yarowsky D. (1997) A Perspective on Word Sense Disambiguation Methods and their Evaluation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What and How?* Washington D.C., USA
- Stamou S., Oflazer K., Pala K., Christodoulakis D., Cristea D., Tufis D., Koeva S., Totkov G., Dutoit D. & Grigoriadou M. (2002) BALKANET: A Multilingual Semantic Network for Balkan Languages. In *Proceedings the 1st Global Wordnet Conference (GWC)*, Mysore, India
- Stamou S., Andrikopoulos V. & Christodoulakis D. (2003) Towards Developing a Semantically Annotated Treebank Corpus for Greek. In *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories (TLT)*, Vaxjo, Sweden