Reducing the Granularity of a Computational Lexicon via an Automatic Mapping to a Coarse-Grained Sense Inventory

Roberto Navigli

Dipartimento di Informatica
Università di Roma "La Sapienza"
Roma, Italy
navigli@di.uniroma1.it

Abstract

WordNet is the reference sense inventory of most of the current Word Sense Disambiguation systems. Unfortunately, it encodes too fine-grained distinctions, making it difficult even for humans to solve the ambiguity of words in context. In this paper, we present a method for reducing the granularity of the WordNet sense inventory based on the mapping to a manually crafted dictionary encoding sense groups, namely the Oxford Dictionary of English. We assess the quality of the mapping and discuss the potential of the method.

1. Introduction

Word Sense Disambiguation (WSD) is the task of choosing the appropriate senses of words in context. WSD obtained disappointing results in the last years, even in presence of large dictionaries and fast computational resources. Most of the disambiguation approaches developed so far adopt the WordNet dictionary (Fellbaum, 1998). This choice is mostly due to the free availability of this resource, its wide coverage of English, and the existence of a number of standard test sets based on it. Unfortunately, WordNet encodes sense distinctions that are difficult to recognize even for human annotators (Edmonds and Kilgariff, 1998).

The inter-annotator agreement between human annotators using WordNet as a sense inventory has been recently estimated around 70% (Snyder and Palmer, 2004; Chklovski and Mihalcea, 2002). This is a figure that state-of-the-art automatic systems find it difficult to outperform. Furthermore, even if a system were able to exceed such an upper bound, it would be unclear how to interpret such a result.

The major obstacle to effective WSD is therefore the fine granularity of the WordNet sense inventory, rather than the performance of the best WSD systems. Ng et al. (1999) show that the adoption of a coarse-grained sense inventory leads to an increase in inter-annotator agreement which is much higher than the reduction of the polysemy degree.

In this paper we describe a large-scale method for clustering WordNet senses via a mapping to a coarse-grained sense inventory (Section 2.), namely the Oxford Dictionary of English (Soanes and Stevenson, 2003). We show that this method is well-founded and accurate with respect to a manually-made mapping (Section 3.). We conclude the paper with an account of related work (Section 4.), and final remarks (Section 5.).

2. The Method

In this section, we provide an approach to the automatic construction of a coarse-grained sense inventory based on the mapping of WordNet senses to coarse senses in the Oxford Dictionary of English.

2.1. The Resources

WordNet (Fellbaum, 1998) is a computational lexicon of English which encodes concepts as synonym sets (*synsets*), according to psycholinguistic principles. For each word sense, WordNet provides a gloss (i.e. a textual definition) and relations such as hypernymy (e.g. an apple *kind-of* edible fruit), meronymy (e.g. a computer *has-part* CPU), etc. The Oxford Dictionary of English (ODE) (Soanes and Stevenson, 2003)[1] provides a hierarchical structure of senses, distinguishing between homonymy (i.e. completely distinct senses, like story as a narration and story as a structure) and polysemy (e.g. story as a narration and as a news report). Each polysemous sense is further divided into a *core sense* and a set of *subsenses*. For each sense (both core and subsenses), the ODE provides a textual definition, and possibly hypernyms and domain labels.

In Table 1 we show an excerpt of the sense inventories of the noun *story* as provided by both dictionaries[2].

The structure of the ODE senses is clearly hierarchical: if we were able to map with a high accuracy WordNet senses to ODE entries, then a sense clustering could be trivially induced from the mapping. As a result, the granularity of the WordNet inventory would be drastically reduced. Furthermore, disregarding errors, the clustering would be well-founded, as the ODE sense groupings were manually crafted by expert lexicographers. In the next section we illustrate a general way of constructing sense descriptions that we use for determining a complete mapping between the two dictionaries.

2.2. Constructing Sense Descriptions

For each word w, and for each sense S of w in a given dictionary $D \in \{\text{WORDNET}, \text{ODE}\}$, we construct a sense description $d_D(S)$ as a bag of words:

---

[1] The ODE was kindly made available by Ken Litkowski (CL Research) in the context of a license agreement.

[2] In the following, we denote a WordNet sense with the convention $w\#p\#i$ where w is a word, p a part of speech and i is a sense number; analogously, we denote an ODE sense with the convention $w\#p\#h.k$ where h is the homonym number and k is the k-th polysemous entry under homonym h.

Table 1: The sense inventory of *story#n* in WordNet and ODE (definitions are abridged, bullets (●) indicate a subsense in the ODE, arrows (→) indicate hypernymy, DOMAIN LABELS are in small caps).

story#n (WordNet)	
#1	A message that tells the particulars of an act or occurrence or course of events (→ message).
#2	A piece of fiction that narrates a chain of related events (→ fiction).
#3	Structure consisting of a room or set of rooms comprising a single level of a multilevel building (→ structure).
#4	A record or narrative description of past events (→ record).
#5	A short account of the news (→ news).
#6	A trivial lie (→ lie).

story#n (ODE)	
#1.1	**Core:** An account of imaginary or real people and events told for entertainment (→ account) ● A plot or storyline ● A piece of gossip (→ gossip) ● A false statement.
#1.2	**Core:** JOURNALISM A report in a newspaper, magazine, or broadcast (→ article).
#1.3	**Core:** An account of past events (→ account) ● The facts about the present situation (→ situation).
#1.4	**Core:** COMMERCE The commercial prospects or circumstances of a particular company (→ situation).
#2.1	**Core:** Storey.

$$d_{\text{D}}(S) = def_{\text{D}}(S) \cup hyper_{\text{D}}(S) \cup domains_{\text{D}}(S)$$

where:

- $def_{\text{D}}(S)$ is the set of words in the textual definition of S (excluding usage examples), automatically lemmatized and part-of-speech tagged with the RASP statistical parser (Briscoe and Carroll, 2002);
- $hyper_{\text{D}}(S)$ is the set of direct hypernyms of S in the taxonomy hierarchy of D ($\emptyset$ if hypernymy is not available);
- $domains_{\text{D}}(S)$ includes the set of domain labels possibly assigned to sense S ($\emptyset$ when no domain is assigned).

Specifically, in the case of WordNet, we generate $def_{\text{WN}}(S)$ from the gloss of S, $hyper_{\text{WN}}(S)$ from the noun and verb taxonomy, and $domains_{\text{WN}}(S)$ from the subject field codes, i.e. domain labels produced semi-automatically by Magnini and Cavaglià (2000) for each WordNet synset (we exclude the general-purpose label, called FACTOTUM).

For example, for the second WordNet sense of *story#n* we obtain the following description:

$$d_{\text{WN}}(story\#n\#2) = \{piece\#n, fiction\#n, narrate\#v, chain\#n, related\#a, event\#n\} \cup \{fiction\#n\} \cup \{\text{LITERATURE}\#\text{N}\}$$

In the case of the ODE, $def_{\text{ODE}}(S)$ is generated from the definitions of the core sense and the subsenses of the entry S. Hypernymy (for nouns only) and domain labels, when available, are included in the respective sets $hyper_{\text{ODE}}(S)$ and $domains_{\text{ODE}}(S)$. For example, the first ODE sense of *story#n* is described as follows:

$$d_{\text{ODE}}(story\#n\#1.1) = \{account\#n, imaginary\#a, real\#a, people\#n, \ldots, statement\#n\} \cup \{account\#n, gossip\#n\}$$

Notice that, for every S, $d_{\text{D}}(S)$ is non-empty as a definition is always provided by both dictionaries. This approach to sense descriptions is general enough to be applicable to any other dictionary with similar characteristics (e.g. the Longman Dictionary of Contemporary English in place of ODE).

2.3. Mapping Word Senses

In order to produce a coarse-grained version of the Word-Net inventory, we aim at defining an automatic mapping between WordNet and ODE, i.e. a function $\mu : Senses_{\text{WN}} \to Senses_{\text{ODE}} \cup \{\epsilon\}$, where $Senses_{\text{D}}$ is the set of senses in the dictionary D and ϵ is a special element assigned when no plausible option is available for mapping (e.g. when the ODE encodes no entry corresponding to a WordNet sense). Given a WordNet sense $S \in Senses_{\text{WN}}(w)$ we define $\hat{m}(S)$, the best matching sense in the ODE, as:

$$\hat{m}(S) = \arg\max_{S' \in Senses_{\text{ODE}}(w)} match(S, S')$$

where $match : Senses_{\text{WN}} \times Senses_{\text{ODE}} \to [0, 1]$ is a function that measures the degree of matching between the sense descriptions of S and S'. We define the mapping μ as:

$$\mu(S) = \begin{cases} \hat{m}(S) & if\, match(S, \hat{m}(S)) \geq \theta \\ \epsilon & otherwise \end{cases}$$

where θ is a threshold below which a matching between sense descriptions is considered unreliable. Finally, we define the clustering of senses $c(w)$ of a word[3] w as:

$$c(w) = \{\mu^{-1}(S') : S' \in Senses_{\text{ODE}}(w), \mu^{-1}(S') \neq \emptyset\} \cup \{\{S\} : S \in Senses_{\text{WN}}(w), \mu(S) = \epsilon\}$$

where $\mu^{-1}(S')$ is the group of WordNet senses mapped to the same sense S' of the ODE, while the second set includes singletons of WordNet senses for which no mapping can be provided according to the definition of μ.

For example, an ideal mapping between entries in Table 1 would be as follows:

$$\mu(story\#n\#1) = story\#n\#1.1, \mu(story\#n\#2) = story\#n\#1.1, \\ \mu(story\#n\#3) = story\#n\#2.1, \mu(story\#n\#5) = story\#n\#1.2, \\ \mu(story\#n\#4) = story\#n\#1.3, \mu(story\#n\#6) = story\#n\#1.1,$$

resulting in the following clustering:

$$c(story\#n) = \\ \{\{story\#n\#1, story\#n\#2, story\#n\#6\}, \\ \{story\#n\#3\}, \{story\#n\#4\}, \{story\#n\#5\}\}$$

In the next Section we describe an implementation of the *match* function based on the use of semantic information.

---

[3]We assume that words are always part-of-speech tagged.

2.3.1. Semantic matching

To match definitions in a semantic manner we adopted a knowledge-based Word Sense Disambiguation algorithm, Structural Semantic Interconnections (SSI, Navigli and Velardi (2004))[4].

SSI exploits an extensive lexical knowledge base, built upon the WordNet lexicon and enriched with collocation information representing semantic relatedness between sense pairs. Collocations are acquired from existing resources (like the Oxford Collocations, the Longman Language Activator, collocation web sites, etc.). Each collocation is mapped to the WordNet sense inventory in a semi-automatic manner and transformed into a *relatedness* edge (Navigli and Velardi, 2005).

Given a word context $C = \{w_1, ..., w_n\}$, SSI builds a graph $G = (V, E)$ such that $V = \bigcup_{i=1}^{n} Senses_{\text{WN}}(w_i)$ and $(S, S') \in E$ if there is at least one semantic interconnection between S and S' in the lexical knowledge base. A *semantic interconnection pattern* is a relevant sequence of edges selected according to a manually-created context-free grammar, i.e. a path connecting a pair of word senses, possibly including a number of intermediate concepts. The grammar consists of a small number of rules, inspired by the notion of lexical chains (Morris and Hirst, 1991).

SSI performs disambiguation in an iterative fashion, by maintaining a set C of senses as a semantic context. Initially, $C = V$ (the entire set of senses of words in C). At each step, for each sense S in C, the algorithm calculates a score of the degree of connectivity between S and the other senses in C:

$$Score_{SSI}(S, C) = \frac{\sum_{S' \in C \setminus \{S\}} \sum_{i \in IC(S, S')} \frac{1}{length(i)}}{\sum_{S' \in C \setminus \{S\}} |IC(S, S')|}$$

where $IC(S, S')$ is the set of interconnections between senses S and S'. The contribution of a single interconnection is given by the reciprocal of its length, calculated as the number of edges connecting its ends. The overall degree of connectivity is then normalized by the number of contributing interconnections. The highest ranking sense S of word w is chosen and the senses of w are removed from the semantic context C. The algorithm terminates when either $C = \emptyset$ or there is no sense such that its score exceeds a fixed threshold.

Given a word w, semantic matching is performed in two steps. First, for each dictionary $D \in \{\text{WORDNET, ODE}\}$, and for each sense $S \in Senses_{\text{D}}(w)$, the sense description of S is disambiguated by applying SSI to $d_D(S)$. As a result, we obtain a semantic description as a bag of concepts $d_D^{sem}(S)$. Notice that sense descriptions from both dictionaries are disambiguated with respect to the WordNet sense inventory.

Second, given a WordNet sense $S \in Senses_{\text{WN}}(w)$ and an ODE sense $S' \in Senses_{\text{ODE}}(w)$, we define $match_{SSI}(S, S')$ as a function of the direct relations connecting senses in $d_{\text{WN}}^{sem}(S)$ and $d_{\text{ODE}}^{sem}(S')$:

$$match_{SSI}(S, S') = \frac{|c \rightarrow c' : c \in d_{\text{WN}}(S), c' \in d_{\text{ODE}}(S')|}{|d_{\text{WN}}(S)| \cdot |d_{\text{ODE}}(S')|}$$

---

[4] SSI is available online at: http://lcl.di.uniroma1.it/ssi

where $c \rightarrow c'$ denotes the existence of a relation edge in the lexical knowledge base between a concept c in the description of S and a concept c' in the description of S'. Edges include the WordNet relation set (synonymy, hypernymy, meronymy, antonymy, similarity, nominalization, etc.) and the *relatedness* edge mentioned above (we adopt only direct relations to maintain a high precision).

For example, some of the relations found between concepts in $d_{\text{WN}}^{sem}(story\#n\#3)$ and $d_{\text{ODE}}^{sem}(story\#n\#2.1)$ are:

$story\#n\#3$	$relation$	$story\#n\#2.1$
building#n	$\xrightarrow{has-part}$	storey#n
single#a	$\xrightarrow{related-to}$	storey#v
structure#n	$\xrightarrow{has-kind}$	storey#v
multilevel#a	$\xrightarrow{related}$	storey#n
level#n	$\xrightarrow{synonym}$	storey#n

contributing to the final value of the function on the two senses:

$$match_{SSI}(story\#n\#3, story\#n\#2.1) = \frac{6}{10 \cdot 1} = 0.6$$

3. Evaluation

We evaluated the accuracy of the mapping produced with the semantic method described in the previous Section.

We produced a gold-standard data set by manually mapping 5,077 WordNet senses of 763 randomly-selected words to the respective ODE entries. The words in the data set were distributed as follows: 466 nouns, 231 verbs, 50 adjectives, 16 adverbs. The data set was created by two annotators and included only polysemous words. These words had 2,600 senses in the ODE.

Overall, 4,599 out of 5,077 WordNet senses had a corresponding sense in ODE (i.e. the ODE covered 90.58% of the WordNet senses in the data set), while 2,053 out of the 2,600 ODE senses had an analogous entry in WordNet (i.e. WordNet covered 78.69% of the ODE senses). The clustering of WordNet senses induced from the manual mapping was 49.85% of the original size and the average degree of polysemy decreased from 6.65 to 3.32.

One could argue that, given the inter-annotator agreement figures discussed in the introduction, our data set would be only partially reliable as a gold standard. We claim that this is not the case: intuitively, the agreement on our mapping task should be much higher, due to the coarse granularity of the target sense inventory. This intuition is also substantiated by a quantitative assessment: 548 WordNet senses of 60 words were mapped to ODE entries by both annotators, with an agreement of 92.7% (κ agreement: 85.4%).

In Table 2 we report the precision and recall of the semantic function in providing the appropriate association for the set of senses having a corresponding entry in ODE (i.e. excluding the cases where a sense ϵ was assigned by the manual annotators, cf. Section 2.3.). We compare its performance with a simple lexical function calculating the word overlap between sense descriptions from WordNet and the ODE. We also report in the Table the accuracy of the two functions when we view the problem as a classification task: an automatic association is correct if it corresponds to the

Table 2: Performance of the lexical and semantic mapping functions.

Function	Prec.	Recall	F1	Accuracy
Lexical	84.74%	65.43%	73.84%	66.08%
Semantic	86.87%	79.67%	83.11%	77.94%

manual association provided by the annotators or if both assign no answer (equivalently, if both provide an ϵ label). All the differences between the lexical and the semantic function are statistically significant ($p < 0.01$).

4. Related Work

Dolan (1994) describes a method for mapping word senses with the use of lexical information provided in the electronic version of LDOCE (textual definitions, semantic relations, domain labels, etc.). Unfortunately, no evaluation is provided.

Most of the approaches in the literature make use of the WordNet structure to cluster its senses. Peters et al. (1998) exploit specific patterns in the WordNet hierarchy (e.g. sisters, autohyponymy, twins, etc.) to group word senses. They study semantic regularities or generalizations obtained and analyze the effect of clustering on the compatibility of language-specific wordnets. Mihalcea and Moldovan (2001) study the structure of WordNet for the identification of sense regularities: to this end, they provide a set of semantic and probabilistic rules. Another approach exploits the (dis)agreements of human annotators to derive coarse-grained sense clusters (Chklovski and Mihalcea, 2002), where sense similarity is computed from confusion matrices.

Finally, Agirre and Lopez (2003) analyze a set of methods to cluster WordNet senses based on the use of confusion matrices of the results of WSD systems, translation equivalences, and topic signatures (word co-occurrences extracted from the web). They assess the acquired clusterings against 20 words from the Senseval-2 sense groupings.

Compared to our approach, most of these methods do not evaluate the clustering produced with respect to a gold-standard clustering, nor is an assessment of WSD attempted. Indeed, such an evaluation would be difficult and time-consuming without a coarse sense inventory like that of ODE.

5. Conclusions

In this paper, we presented a semantic approach to the construction of a coarse-grained sense inventory for the Word-Net dictionary via a mapping to the Oxford Dictionary of English. We showed that the resulting mapping is precise and accurate. As we mentioned in Section 3., the mapping can be easily validated with a high inter-annotator agreement. Lexicographers can also use our method to discover missing senses in the source dictionary (WordNet in our case).

As a future work, we plan to acquire a complete clustering of the WordNet inventory and show that Word Sense Disambiguation can benefit from such coarse distinctions.

We believe that this is a crucial research topic to be investigated.

6. References

Eneko Agirre and Oier Lopez. 2003. Clustering wordnet word senses. In *Proc. of Conf. on Recent Advances on Natural Language (RANLP)*. Borovets, Bulgaria.

Ted Briscoe and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proc. of 3rd Conference on Language Resources and Evaluation*. Las Palmas, Gran Canaria.

Tim Chklovski and Rada Mihalcea. 2002. Building a sense tagged corpus with open mind word expert. In *Proc. of ACL 2002 Workshop on WSD: Recent Successes and Future Directions*. Philadelphia, PA.

William B. Dolan. 1994. Word sense ambiguation: Clustering related senses. In *Proc. of 15th conference on Computational linguistics (COLING)*. Morristown, N.J.

Philip Edmonds and Adam Kilgariff. 1998. Introduction to the special issue on evaluating word sense disambiguation systems. *Journal of Natural Language Engineering*, 8(4).

Christiane Fellbaum, editor. 1998. *WordNet: an Electronic Lexical Database*. MIT Press.

Bernardo Magnini and Gabriela Cavaglià. 2000. Integrating subject field codes into wordnet. In *Proc. of the 2nd Conference on Language Resources and Evaluation (LREC)*. Athens, Greece.

Rada Mihalcea and Dan Moldovan. 2001. Automatic generation of a coarse grained wordnet. In *Proc. of NAACL Workshop on WordNet and Other Lexical Resources*. Pittsburgh, PA.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1).

Roberto Navigli and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated websites. *Computational Linguistics*, 30(2).

Roberto Navigli and Paola Velardi. 2005. Structural semantic interconnections: a knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(7).

Hwee T. Ng, Chung Y. Lim, and Shou K. Foo. 1999. A case study on the inter-annotator agreement for word sense disambiguation. In *Proc. of ACL Workshop: Standardizing Lexical Resources*. College Park, Maryland.

Wim Peters, Ivonne Peters, and Piek Vossen. 1998. Automatic sense clustering in eurowordnet. In *Proc. of the 1st Conference on Language Resources and Evaluation (LREC)*. Granada, Spain.

Benjamin Snyder and Martha Palmer. 2004. The english all-words task. In *Proc. of ACL 2004 SENSEVAL-3 Workshop*. Barcelona, Spain.

Catherine Soanes and Angus Stevenson, editors. 2003. *Oxford Dictionary of English*. Oxford University Press.