Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems

Simone Paolo Ponzetto
Department of Computational Linguistics
Heidelberg University
ponzetto@cl.uni-heidelberg.de

Roberto Navigli
Dipartimento di Informatica
Sapienza Università di Roma
navigli@di.uniroma1.it

Abstract

One of the main obstacles to high-performance Word Sense Disambiguation (WSD) is the knowledge acquisition bottleneck. In this paper, we present a methodology to automatically extend WordNet with large amounts of semantic relations from an encyclopedic resource, namely Wikipedia. We show that, when provided with a vast amount of high-quality semantic relations, simple knowledge-lean disambiguation algorithms compete with state-of-the-art supervised WSD systems in a coarse-grained all-words setting and outperform them on gold-standard domain-specific datasets.

1 Introduction

Knowledge lies at the core of Word Sense Disambiguation (WSD), the task of computationally identifying the meanings of words in context (Navigli, 2009b). In the recent years, two main approaches have been studied that rely on a fixed sense inventory, i.e., supervised and knowledge-based methods. In order to achieve high performance, supervised approaches require large training sets where instances (target words in context) are hand-annotated with the most appropriate word senses. Producing this kind of knowledge is extremely costly: at a throughput of one sense annotation per minute (Edmonds, 2000) and tagging one thousand examples per word, dozens of person-years would be required for enabling a supervised classifier to disambiguate all the words in the English lexicon with high accuracy. In contrast, knowledge-based approaches exploit the information contained in wide-coverage lexical resources, such as WordNet (Fellbaum, 1998). However, it has been demonstrated that the amount of lexical and semantic information

contained in such resources is typically insufficient for high-performance WSD (Cuadros and Rigau, 2006). Several methods have been proposed to automatically extend existing resources (cf. Section 2) and it has been shown that highly-interconnected semantic networks have a great impact on WSD (Navigli and Lapata, 2010). However, to date, the real potential of knowledge-rich WSD systems has been shown only in the presence of either a large manually-developed extension of WordNet (Navigli and Velardi, 2005) or sophisticated WSD algorithms (Agirre et al., 2009).

The contributions of this paper are two-fold. First, we relieve the knowledge acquisition bottleneck by developing a methodology to extend WordNet with millions of semantic relations. The relations are harvested from an encyclopedic resource, namely Wikipedia. Wikipedia pages are automatically associated with WordNet senses, and topical, semantic associative relations from Wikipedia are transferred to WordNet, thus producing a much richer lexical resource. Second, two simple knowledge-based algorithms that exploit our extended WordNet are applied to standard WSD datasets. The results show that the integration of vast amounts of semantic relations in knowledge-based systems yields performance competitive with state-of-the-art supervised approaches on open-text WSD. In addition, we support previous findings from Agirre et al. (2009) that in a domain-specific WSD scenario knowledge-based systems perform better than supervised ones, and we show that, given enough knowledge, simple algorithms perform better than more sophisticated ones.

2 Related Work

In the last three decades, a large body of work has been presented that concerns the development of automatic methods for the enrichment of existing resources such as WordNet. These in-

clude proposals to extract semantic information from dictionaries (e.g. Chodorow et al. (1985) and Rigau et al. (1998)), approaches using lexico-syntactic patterns (Hearst, 1992; Cimiano et al., 2004; Girju et al., 2006), heuristic methods based on lexical and semantic regularities (Harabagiu et al., 1999), taxonomy-based ontologization (Pennacchiotti and Pantel, 2006; Snow et al., 2006). Other approaches include the extraction of semantic preferences from sense-annotated (Agirre and Martinez, 2001) and raw corpora (McCarthy and Carroll, 2003), as well as the disambiguation of dictionary glosses based on cyclic graph patterns (Navigli, 2009a). Other works rely on the disambiguation of collocations, either obtained from specialized learner's dictionaries (Navigli and Velardi, 2005) or extracted by means of statistical techniques (Cuadros and Rigau, 2008), e.g. based on the method proposed by Agirre and de Lacalle (2004). But while most of these methods represent state-of-the-art proposals for enriching lexical and taxonomic resources, none concentrates on augmenting WordNet with associative semantic relations for many domains on a very large scale. To overcome this limitation, we exploit Wikipedia, a collaboratively generated Web encyclopedia.

The use of collaborative contributions from volunteers has been previously shown to be beneficial in the Open Mind Word Expert project (Chklovski and Mihalcea, 2002). However, its current status indicates that the project remains a mainly academic attempt. In contrast, due to its low entrance barrier and vast user base, Wikipedia provides large amounts of information at practically no cost. Previous work aimed at transforming its content into a knowledge base includes open-domain relation extraction (Wu and Weld, 2007), the acquisition of taxonomic (Ponzetto and Strube, 2007a; Suchanek et al., 2008; Wu and Weld, 2008) and other semantic relations (Nastase and Strube, 2008), as well as lexical reference rules (Shnarch et al., 2009). Applications using the knowledge contained in Wikipedia include, among others, text categorization (Gabrilovich and Markovitch, 2006), computing semantic similarity of texts (Gabrilovich and Markovitch, 2007; Ponzetto and Strube, 2007b; Milne and Witten, 2008a), coreference resolution (Ponzetto and Strube, 2007b), multi-document summarization (Nastase, 2008), and text generation (Sauper and Barzilay, 2009).

In our work we follow this line of research and show that knowledge harvested from Wikipedia can be used effectively to improve the performance of a WSD system. Our proposal builds on previous insights from Bunescu and Paşca (2006) and Mihalcea (2007) that pages in Wikipedia can be taken as word senses. Mihalcea (2007) manually maps Wikipedia pages to WordNet senses to perform lexical-sample WSD. We extend her proposal in three important ways: (1) we *fully automatize* the mapping between Wikipedia pages and WordNet senses; (2) we use the mappings to *enrich an existing resource*, i.e. WordNet, rather than annotating text with sense labels; (3) we deploy the knowledge encoded by this mapping to perform *unrestricted* WSD, rather than apply it to a lexical sample setting.

Knowledge from Wikipedia is injected into a WSD system by means of a mapping to WordNet. Previous efforts aimed at automatically linking Wikipedia to WordNet include full use of the first WordNet sense heuristic (Suchanek et al., 2008), a graph-based mapping of Wikipedia categories to WordNet synsets (Ponzetto and Navigli, 2009), a model based on vector spaces (Ruiz-Casado et al., 2005) and a supervised approach using keyword extraction (Reiter et al., 2008). These latter methods rely only on text overlap techniques and neither they take advantage of the input from Wikipedia being semi-structured, e.g. hyperlinked, nor they propose a high-performing probabilistic formulation of the mapping problem, a task to which we turn in the next section.

3 Extending WordNet

Our approach consists of two main phases: first, a mapping is automatically established between Wikipedia pages and WordNet senses; second, the relations connecting Wikipedia pages are transferred to WordNet. As a result, an extended version of WordNet is produced, that we call Word-Net++. We present the two resources used in our methodology in Section 3.1. Sections 3.2 and 3.3 illustrate the two phases of our approach.

3.1 Knowledge Resources

WordNet. Being the most widely used computational lexicon of English in Natural Language Processing, WordNet is an essential resource for WSD. A concept in WordNet is represented as a synonym set, or *synset*, i.e. the set of words which share a common meaning. For instance, the con-

cept of soda drink is expressed as:

$$\{ \text{pop}_n^2, \text{soda}_n^2, \text{soda pop}_n^1, \text{soda water}_n^2, \text{tonic}_n^2 \}$$

where each word's subscripts and superscripts indicate their parts of speech (e.g. n stands for noun) and sense number[1], respectively. For each synset, WordNet provides a textual definition, or gloss. For example, the gloss of the above synset is: "a sweet drink containing carbonated water and flavoring".

Wikipedia. Our second resource, Wikipedia, is a collaborative Web encyclopedia composed of pages[2]. A Wikipedia page (henceforth, Wikipage) presents the knowledge about a specific concept (e.g. SODA (SOFT DRINK)) or named entity (e.g. FOOD STANDARDS AGENCY). The page typically contains hypertext linked to other relevant Wikipages. For instance, SODA (SOFT DRINK) is linked to COLA, FLAVORED WATER, LEMONADE, and many others. The title of a Wikipage (e.g. SODA (SOFT DRINK)) is composed of the lemma of the concept defined (e.g. soda) plus an optional label in parentheses which specifies its meaning in case the lemma is ambiguous (e.g. SOFT DRINK vs. SODIUM CARBONATE). Finally, some Wikipages are redirections to other pages, e.g. SODA (SODIUM CARBONATE) redirects to SODIUM CARBONATE.

3.2 Mapping Wikipedia to WordNet

During the first phase of our methodology we aim to establish links between Wikipages and WordNet senses. Formally, given the entire set of pages $Senses_{\text{Wiki}}$ and WordNet senses $Senses_{\text{WN}}$, we aim to acquire a mapping:

$$\mu : Senses_{\text{Wiki}} \rightarrow Senses_{\text{WN}},$$

such that, for each Wikipage $w \in Senses_{\text{Wiki}}$:

$$\mu(w) = \begin{cases} s \in Senses_{\text{WN}}(w) & \text{if a link can be established,} \\ \epsilon & \text{otherwise,} \end{cases}$$

where $Senses_{\text{WN}}(w)$ is the set of senses of the *lemma* of w in WordNet. For example, if our

---

[1]We use WordNet version 3.0. We use word senses to unambiguously denote the corresponding synsets (e.g. plane_n^1 for { airplane_n^1, aeroplane_n^1, plane_n^1 }).

[2]http://download.wikipedia.org. We use the English Wikipedia database dump from November 3, 2009, which includes 3,083,466 articles. Throughout this paper, we use Sans Serif for words, SMALL CAPS for Wikipedia pages and CAPITALS for Wikipedia categories.

mapping methodology linked SODA (SOFT DRINK) to the corresponding WordNet sense soda_n^2, we would have $\mu(\text{SODA (SOFT DRINK)}) = \text{soda}_n^2$.

In order to establish a mapping between the two resources, we first identify different kinds of disambiguation contexts for Wikipages (Section 3.2.1) and WordNet senses (Section 3.2.2). Next, we intersect these contexts to perform the mapping (see Section 3.2.3).

3.2.1 Disambiguation Context of a Wikipage

Given a target Wikipage w which we aim to map to a WordNet sense of w, we use the following information as a disambiguation context:

- **Sense labels**: e.g. given the page SODA (SOFT DRINK), the words soft and drink are added to the disambiguation context.

- **Links**: the titles' lemmas of the pages linked from the Wikipage w (outgoing links). For instance, the links in the Wikipage SODA (SOFT DRINK) include soda, lemonade, sugar, etc.

- **Categories**: Wikipages are classified according to one or more categories, which represent meta-information used to categorize them. For instance, the Wikipage SODA (SOFT DRINK) is categorized as SOFT DRINKS. Since many categories are very specific and do not appear in WordNet (e.g., SWEDISH WRITERS or SCIENTISTS WHO COMMITTED SUICIDE), we use the lemmas of their syntactic heads as disambiguation context (i.e. writer and scientist). To this end, we use the category heads provided by Ponzetto and Navigli (2009).

Given a Wikipage w, we define its disambiguation context $Ctx(w)$ as the set of words obtained from some or all of the three sources above.

3.2.2 Disambiguation Context of a WordNet Sense

Given a WordNet sense s and its synset S, we use the following information as disambiguation context to provide evidence for a potential link in our mapping μ:

- **Synonymy**: all synonyms of s in synset S. For instance, given the synset of soda_n^2, all its synonyms are included in the context (that is, tonic, soda pop, pop, etc.).

- **Hypernymy/Hyponymy**: all synonyms in the synsets H such that H is either a hypernym (i.e., a generalization) or a hyponym (i.e., a specialization) of S. For example, given soda_n^2, we include the words from its hypernym { $\mathsf{soft\ drink}_n^1$ }.

- **Sisterhood**: words from the sisters of S. A sister synset S' is such that S and S' have a common direct hypernym. For example, given soda_n^2, it can be found that $\mathsf{bitter\ lemon}_n^1$ and soda_n^2 are sisters. Thus the words bitter and lemon are included in the disambiguation context of s.

- **Gloss**: the set of lemmas of the content words occurring within the gloss of s. For instance, given $s = \mathsf{soda}_n^2$, defined as "a sweet drink containing carbonated water and flavoring", we add to the disambiguation context of s the following lemmas: sweet, drink, $\mathsf{contain}$, $\mathsf{carbonated}$, water, $\mathsf{flavoring}$.

Given a WordNet sense s, we define its disambiguation context $Ctx(s)$ as the set of words obtained from some or all of the four sources above.

3.2.3 Mapping Algorithm

In order to link each Wikipedia page to a WordNet sense, we developed a novel algorithm, whose pseudocode is shown in Algorithm 1. The following steps are performed:

- Initially (lines 1-2), our mapping μ is empty, i.e. it links each Wikipage w to ϵ.

- For each Wikipage w whose lemma is monosemous both in Wikipedia and WordNet (i.e. $|Senses_{\text{Wiki}}(w)| = |Senses_{\text{WN}}(w)| = 1$) we map w to its only WordNet sense w_n^1 (lines 3-5).

- Finally, for each remaining Wikipage w for which no mapping was previously found (i.e., $\mu(w) = \epsilon$, line 7), we do the following:

  - lines 8-10: for each Wikipage d which is a redirection to w, for which a mapping was previously found (i.e. $\mu(d) \neq \epsilon$, that is, d is monosemous in both Wikipedia and WordNet) and such that it maps to a sense $\mu(d)$ in a synset S that also contains a sense of w, we map w to the corresponding sense in S.
  - lines 11-14: if a Wikipage w has not been linked yet, we assign the most likely sense to w based on the maximization of the conditional probabilities $p(s|w)$ over the senses

---

Algorithm 1 The mapping algorithm

Input: $Senses_{\text{Wiki}}$, $Senses_{\text{WN}}$
Output: a mapping $\mu : Senses_{\text{Wiki}} \rightarrow Senses_{\text{WN}}$

1: **for each** $w \in Senses_{\text{Wiki}}$
2:      $\mu(w) := \epsilon$
3: **for each** $w \in Senses_{\text{Wiki}}$
4:      **if** $|Senses_{\text{Wiki}}(w)| = |Senses_{\text{WN}}(w)| = 1$ **then**
5:          $\mu(w) := w_n^1$
6: **for each** $w \in Senses_{\text{Wiki}}$
7:      **if** $\mu(w) = \epsilon$ **then**
8:          **for each** $d \in Senses_{\text{Wiki}}$ s.t. d redirects to w
9:              **if** $\mu(d) \neq \epsilon$ and $\mu(d)$ is in a synset of w **then**
10:                 $\mu(w) :=$ sense of w in synset of $\mu(d)$; **break**
11: **for each** $w \in Senses_{\text{Wiki}}$
12:      **if** $\mu(w) = \epsilon$ **then**
13:          **if** no tie occurs **then**
14:              $\mu(w) := \underset{s \in Senses_{\text{WN}}(w)}{\operatorname{argmax}} p(s|w)$
15: **return** μ

$s \in Senses_{\text{WN}}(w)$ (no mapping is established if a tie occurs, line 13).

As a result of the execution of the algorithm, the mapping μ is returned (line 15). At the heart of the mapping algorithm lies the calculation of the conditional probability $p(s|w)$ of selecting the WordNet sense s given the Wikipage w. The sense s which maximizes this probability can be obtained as follows:

$$\mu(w) = \underset{s \in Senses_{\text{WN}}(w)}{\operatorname{argmax}} p(s|w) = \underset{s}{\operatorname{argmax}} \frac{p(s,w)}{p(w)}$$
$$= \underset{s}{\operatorname{argmax}} p(s,w)$$

The latter formula is obtained by observing that $p(w)$ does not influence our maximization, as it is a constant independent of s. As a result, the most appropriate sense s is determined by maximizing the joint probability $p(s,w)$ of sense s and page w. We estimate $p(s,w)$ as:

$$p(s,w) = \frac{score(s,w)}{\displaystyle\sum_{\substack{s' \in Senses_{\text{WN}}(w),\\ w' \in Senses_{\text{Wiki}}(w)}} score(s',w')},$$

where $score(s,w) = |Ctx(s) \cap Ctx(w)| + 1$ (we add 1 as a smoothing factor). Thus, in our algorithm we determine the best sense s by computing the intersection of the disambiguation contexts of s and w, and normalizing by the scores summed over all senses of w in Wikipedia and WordNet.

3.2.4 Example

We illustrate the execution of our mapping algorithm by way of an example. Let us focus on the

Wikipage SODA (SOFT DRINK). The word soda is polysemous both in Wikipedia and WordNet, thus lines 3–5 of the algorithm do not concern this Wikipage. Lines 6–14 aim to find a mapping μ(SODA (SOFT DRINK)) to an appropriate WordNet sense of the word. First, we check whether a redirection exists to SODA (SOFT DRINK) that was previously disambiguated (lines 8–10). Next, we construct the disambiguation context for the Wikipage by including words from its label, links and categories (cf. Section 3.2.1). The context includes, among others, the following words: soft, drink, cola, sugar. We now construct the disambiguation context for the two WordNet senses of soda (cf. Section 3.2.2), namely the sodium carbonate (#1) and the drink (#2) senses. To do so, we include words from their synsets, hypernyms, hyponyms, sisters, and glosses. The context for $soda_n^1$ includes: salt, acetate, chlorate, benzoate. The context for $soda_n^2$ contains instead: soft, drink, cola, bitter, etc. The sense with the largest intersection is #2, so the following mapping is established: μ(SODA (SOFT DRINK)) = $soda_n^2$.

3.3 Transferring Semantic Relations

The output of the algorithm presented in the previous section is a mapping between Wikipages and WordNet senses (that is, implicitly, synsets). Our insight is to use this alignment to enable the *transfer of semantic relations from Wikipedia to WordNet*. In fact, given a Wikipage w we can collect all Wikipedia links occurring in that page. For any such link from w to w', if the two Wikipages are mapped to WordNet senses (i.e., $\mu(w) \neq \epsilon$ and $\mu(w') \neq \epsilon$), we can transfer the corresponding edge $(\mu(w), \mu(w'))$ to WordNet. Note that $\mu(w)$ and $\mu(w')$ are noun senses, as Wikipages describe nominal concepts or named entities. We refer to this extended resource as WordNet++.

For instance, consider the Wikipage SODA (SOFT DRINK). This page contains, among others, a link to the Wikipage SYRUP. Assuming μ(SODA (SODA DRINK)) = $soda_n^2$ and μ(SYRUP) = $syrup_n^1$, we can add the corresponding semantic relation ($soda_n^2$, $syrup_n^1$) to WordNet[3].

Thus, WordNet++ represents an extension of WordNet which includes semantic associative relations between synsets. These are originally

---

[3]Note that such relations are *unlabeled*. However, for our purposes this has no impact, since our algorithms do not distinguish between *is-a* and other kinds of relations in the lexical knowledge base (cf. Section 4.2).

found in Wikipedia and then integrated into Word-Net by means of our mapping. In turn, Word-Net++ represents the English-only subset of a larger multilingual resource, BabelNet (Navigli and Ponzetto, 2010), where lexicalizations of the synsets are harvested for many languages using the so-called Wikipedia inter-language links and applying a machine translation system.

4 Experiments

We perform two sets of experiments: we first evaluate the intrinsic quality of our mapping (Section 4.1) and then quantify the impact of WordNet++ for coarse-grained (Section 4.2) and domain-specific WSD (Section 4.3).

4.1 Evaluation of the Mapping

Experimental setting. We first conducted an evaluation of the mapping quality. To create a gold standard for evaluation, we started from the set of all lemmas contained both in Word-Net and Wikipedia: the intersection between the two resources includes 80,295 lemmas which correspond to 105,797 WordNet senses and 199,735 Wikipedia pages. The average polysemy is 1.3 and 2.5 for WordNet senses and Wikipages, respectively (2.8 and 4.7 when excluding monosemous words). We selected a random sample of 1,000 Wikipages and asked an annotator with previous experience in lexicographic annotation to provide the correct WordNet sense for each page title (an empty sense label was given if no correct mapping was possible). 505 non-empty mappings were found, i.e. Wikipedia pages with a corresponding WordNet sense. In order to quantify the quality of the annotations and the difficulty of the task, a second annotator sense tagged a subset of 200 pages from the original sample. We computed the inter-annotator agreement using the kappa coefficient (Carletta, 1996) and found out that our annotators achieved an agreement coefficient κ of 0.9, indicating almost perfect agreement.

Table 1 summarizes the performance of our disambiguation algorithm against the manually annotated dataset. Evaluation is performed in terms of standard measures of precision (the ratio of correct sense labels to the non-empty labels output by the mapping algorithm), recall (the ratio of correct sense labels to the total of non-empty labels in the gold standard) and F_1-measure ($\frac{2PR}{P+R}$). We also calculate accuracy, which accounts for

	P	R	F_1	A
Structure	**82.2**	68.1	74.5	81.1
Gloss	81.1	64.2	71.7	78.8
Structure + Gloss	81.9	**77.5**	**79.6**	**84.4**
MFS BL	24.3	47.8	32.2	24.3
Random BL	23.8	46.8	31.6	23.9

Table 1: Performance of the mapping algorithm.

empty sense labels (that is, calculated on all 1,000 test instances). As baseline we use the most frequent WordNet sense (MFS), as well as a random sense assignment. We evaluate the mapping methodology described in Section 3.2 against different disambiguation contexts for the WordNet senses (cf. Section 3.2.2), i.e. structure-based (including synonymy, hypernymy/hyponymy and sisterhood), gloss-derived evidence, and a combination of the two. As disambiguation context of a Wikipage (Section 3.2.1) we use all information available, i.e. sense labels, links and categories[4].

Results and discussion. The results show that our method improves on the baseline by a large margin and that higher performance can be achieved by using more disambiguation information. That is, using a richer disambiguation context helps to better choose the most appropriate WordNet sense for a Wikipedia page. The combination of structural and gloss information attains a slight variation in terms of precision (−0.3% and +0.8% compared to Structure and Gloss respectively), but a significantly high increase in recall (+9.4% and +13.3%). This implies that the different disambiguation contexts only partially overlap and, when used separately, each produces different mappings with a similar level of precision. In the joint approach, the harmonic mean of precision and recall, i.e. F_1, is in fact 5 and 8 points higher than when separately using structural and gloss information, respectively.

As for the baselines, the most frequent sense is just 0.6% and 0.4% above the random baseline in terms of F_1 and accuracy, respectively. A χ^2 test reveals in fact no statistically significant difference at $p < 0.05$. This is related to the random distribution of senses in our dataset and the Wikipedia unbiased coverage of WordNet senses. So select-

ing the most frequent sense rather than any other sense for each target page represents a choice as arbitrary as picking a sense at random.

The final mapping contains 81,533 pairs of Wikipages and word senses they map to, covering 55.7% of the noun senses in WordNet.

Using our best performing mapping we are able to extend WordNet with 1,902,859 semantic edges: of these, 97.93% are deemed novel, i.e. no direct edge could previously be found between the synsets. In addition, we performed a stricter evaluation of the novelty of our relations by checking whether these can still be found indirectly by searching for a connecting path between the two synsets of interest. Here we found that 91.3%, 87.2% and 78.9% of the relations are novel to WordNet when performing a graph search of maximum depth of 2, 3 and 4, respectively.

4.2 Coarse-grained WSD

Experimental setting. We extrinsically evaluate the impact of WordNet++ on the Semeval-2007 coarse-grained all-words WSD task (Navigli et al., 2007). Performing experiments in a coarse-grained setting is a natural choice for several reasons: first, it has been argued that the fine granularity of WordNet is one of the main obstacles to accurate WSD (cf. the discussion in Navigli (2009b)); second, the meanings of Wikipedia pages are intuitively coarser than those in Word-Net[5]. For instance, mapping TRAVEL to the first or the second sense in WordNet is an arbitrary choice, as the Wikipage refers to both senses. Finally, given their different nature, WordNet and Wikipedia do not fully overlap. Accordingly, we expect the transfer of semantic relations from Wikipedia to WordNet to have sometimes the side effect to penalize some fine-grained senses of a word.

We experiment with two simple knowledge-based algorithms that are set to perform coarse-grained WSD on a sentence-by-sentence basis:

- **Simplified Extended Lesk (ExtLesk):** The first algorithm is a simplified version of the Lesk

---

[4]We leave out the evaluation of different contexts for a Wikipage for the sake of brevity. During prototyping we found that the best results were given by using the largest context available, as reported in Table 1.

[5]Note that our polysemy rates from Section 4.1 also include Wikipages whose lemma is contained in WordNet, but which have out-of-domain meanings, i.e. encyclopedic entries referring to specialized named entities such as e.g., DISCOVERY (SPACE SHUTTLE) or FIELD ARTILLERY (MAGAZINE). We computed the polysemy rate for a random sample of 20 polysemous words by manually removing these NEs and found that Wikipedia's polysemy rate is indeed lower than that of WordNet – i.e. average polysemy of 2.1 vs. 2.8.

algorithm (Lesk, 1986), that performs WSD based on the overlap between the context surrounding the target word to be disambiguated and the definitions of its candidate senses (Kilgarriff and Rosenzweig, 2000). Given a target word w, this method assigns to w the sense whose gloss has the highest overlap (i.e. most words in common) with the context of w, namely the set of content words co-occurring with it in a pre-defined window (a sentence in our case). Due to the limited context provided by the WordNet glosses, we follow Banerjee and Pedersen (2003) and expand the gloss of each sense s to include words from the glosses of those synsets in a semantic relation with s. These include all WordNet synsets which are directly connected to s, either by means of the semantic pointers found in WordNet or through the unlabeled links found in WordNet++.

- **Degree Centrality (Degree):** The second algorithm is a graph-based approach that relies on the notion of vertex degree (Navigli and Lapata, 2010). Starting from each sense s of the target word, it performs a depth-first search (DFS) of the WordNet(++) graph and collects all the paths connecting s to senses of other words in context. As a result, a sentence graph is produced. A maximum search depth is established to limit the size of this graph. The sense of the target word with the highest vertex degree is selected. We follow Navigli and Lapata (2010) and run Degree in a weakly supervised setting where the system attempts no sense assignment if the highest degree score is below a certain (empirically estimated) threshold. The optimal threshold and maximum search depth are estimated by maximizing Degree's F_1 on a development set of 1,000 randomly chosen noun instances from the SemCor corpus (Miller et al., 1993). Experiments on the development dataset using Degree on WordNet++ revealed a performance far lower than expected. Error analysis showed that many instances were incorrectly disambiguated, due to the noise from weak semantic links, e.g. the links from SODA (SOFT DRINK) to EUROPE or AUSTRALIA. Accordingly, in order to improve the disambiguation performance, we developed a filter to rule out weak semantic relations from WordNet++. Given a WordNet++ edge $(\mu(w), \mu(w'))$ where w and w' are both Wikipages and w links to w',

Resource	Algorithm	Nouns only		
		P	R	F_1
WordNet	ExtLesk	83.6	57.7	68.3
	Degree	86.3	65.5	74.5
Wikipedia	ExtLesk	82.3	64.1	72.0
	Degree	**96.2**	40.1	57.4
WordNet++	ExtLesk	82.7	69.2	75.4
	Degree	87.3	72.7	**79.4**
	MFS BL	77.4	**77.4**	77.4
	Random BL	63.5	63.5	63.5

Table 2: Performance on Semeval-2007 coarse-grained all-words WSD (nouns only subset).

we first collect all words from the category labels of w and w' into two bags of words. We remove stopwords and lemmatize the remaining words. We then compute the degree of overlap between the two sets of categories as the number of words in common between the two bags of words, normalized in the $[0, 1]$ interval. We finally retain the link for the DFS if such score is above an empirically determined threshold. The optimal value for this category overlap threshold was again estimated by maximizing Degree's F_1 on the development set. The final graph used by Degree consists of WordNet, together with 152,944 relations from our semantic relation enrichment method (cf. Section 3.3).

Results and discussion. We report our results in terms of precision, recall and F_1-measure on the Semeval-2007 coarse-grained all-words dataset (Navigli et al., 2007). We first evaluated ExtLesk and Degree using three different resources: (1) WordNet only; (2) Wikipedia only, i.e. only those relations harvested from the links found within Wikipedia pages; (3) their union, i.e. WordNet++. In Table 2 we report the results on nouns only. As common practice, we compare with random sense assignment and the most frequent sense (MFS) from SemCor as baselines. Enriching WordNet with encyclopedic relations from Wikipedia yields a consistent improvement against using WordNet (+7.1% and +4.9% F_1 for ExtLesk and Degree) or Wikipedia (+3.4% and +22.0%) alone. The best results are obtained by using Degree with WordNet++. The better performance of Wikipedia against WordNet when using ExtLesk (+3.7%) highlights the quality of the relations extracted. However, no such improvement is found with De-

Algorithm	Nouns only P/R/F_1	All words P/R/F_1
ExtLesk	81.0	79.1
Degree	**85.5**	**81.7**
SUSSX-FR	81.1	77.0
TreeMatch	N/A	73.6
NUS-PT	82.3	**82.5**
SSI	**84.1**	**83.2**
MFS BL	77.4	78.9
Random BL	63.5	62.7

Table 3: Performance on Semeval-2007 coarse-grained all-words WSD with MFS as a back-off strategy when no sense assignment is attempted.

gree, due to its lower recall. Interestingly, Degree on WordNet++ beats the MFS baseline, which is notably a difficult competitor for unsupervised and knowledge-lean systems.

We finally compare our two algorithms using WordNet++ with state-of-the-art WSD systems, namely the best unsupervised (Koeling and Mc-Carthy, 2007, SUSSX-FR) and supervised (Chan et al., 2007, NUS-PT) systems participating in the Semeval-2007 coarse-grained all-words task. We also compare with SSI (Navigli and Velardi, 2005) – a knowledge-based system that participated out of competition – and the unsupervised proposal from Chen et al. (2009, TreeMatch). Table 3 shows the results for nouns (1,108) and all words (2,269 words): we use the MFS as a back-off strategy when no sense assignment is attempted. Degree with WordNet++ achieves the best performance in the literature[6]. On the noun-only subset of the data, its performance is comparable with SSI and significantly better than the best supervised and unsupervised systems (+3.2% and +4.4% F_1 against NUS-PT and SUSSX-FR). On the entire dataset, it outperforms SUSSX-FR and TreeMatch (+4.7% and +8.1%) and its recall is not statistically different from that of SSI and NUS-PT. This result is particularly interesting, given that WordNet++ is extended only with relations between nominals, and, in contrast to SSI, it does not rely on a costly annotation effort to engineer the set of semantic relations. Last but not least, we achieve state-of-the-art performance with a much simpler algorithm that is based on the notion of vertex degree in a graph.

Algorithm	Sports P/R/F_1	Finance P/R/F_1
k-NN [†]	30.3	43.4
Static PR [†]	20.1	39.6
Personalized PR [†]	35.6	46.9
ExtLesk	40.1	45.6
Degree	**42.0**	**47.8**
MFS BL	19.6	37.1
Random BL	19.5	19.6

Table 4: Performance on the Sports and Finance sections of the dataset from Koeling et al. (2005): [†] indicates results from Agirre et al. (2009).

4.3 Domain WSD

The main strength of Wikipedia is to provide wide coverage for many specific domains. Accordingly, on the Semeval dataset our system achieves the best performance on a domain-specific text, namely d004, a document on computer science where we achieve 82.9% F_1 (+6.8% when compared with the best supervised system, namely NUS-PT). To test whether our performance on the Semeval dataset is an artifact of the data, i.e. d004 coming from Wikipedia itself, we evaluated our system on the Sports and Finance sections of the domain corpora from Koeling et al. (2005). In Table 4 we report our results on these datasets and compare them with Personalized PageRank, the state-of-the-art system from Agirre et al. (2009)[7], as well as Static PageRank and a k-NN supervised WSD system trained on SemCor.

The results we obtain on the two domains with our best configuration (Degree using WordNet++) outperform by a large margin k-NN, thus supporting the findings from Agirre et al. (2009) that knowledge-based systems exhibit a more robust performance than their supervised alternatives when evaluated across different domains. In addition, our system achieves better results than Static and Personalized PageRank, indicating that competitive disambiguation performance can still be achieved by a less sophisticated knowledge-based WSD algorithm when provided with a rich amount of high-quality knowledge. Finally, the results show that WordNet++ enables competitive performance also in a fine-grained domain setting.

---

[6]The differences between the results in bold in each column of the table are not statistically significant at $p < 0.05$.

[7]We compare only with those system configurations performing token-based WSD, i.e. disambiguating each instance of a target word separately, since our aim is not to perform type-based disambiguation.

5 Conclusions

In this paper, we have presented a large-scale method for the automatic enrichment of a computational lexicon with encyclopedic relational knowledge[8]. Our experiments show that the large amount of knowledge injected into WordNet is of high quality and, more importantly, it enables simple knowledge-based WSD systems to perform as well as the highest-performing supervised ones in a coarse-grained setting and to outperform them on domain-specific text. Thus, our results go one step beyond previous findings (Cuadros and Rigau, 2006; Agirre et al., 2009; Navigli and Lapata, 2010) and prove that knowledge-rich disambiguation is a competitive alternative to supervised systems, even when relying on a simple algorithm. We note, however, that the present contribution does not show *which* knowledge-rich algorithm performs best with WordNet++. In fact, more sophisticated approaches, such as Personalized PageRank (Agirre and Soroa, 2009), could be still applied to yield even higher performance. We leave such exploration to future work. Moreover, while the mapping has been used to enrich Word-Net with a large amount of semantic edges, the method can be reversed and applied to the encyclopedic resource itself, that is Wikipedia, to perform disambiguation with the corresponding sense inventory (cf. the task of wikification proposed by Mihalcea and Csomai (2007) and Milne and Witten (2008b)). In this paper, we focused on English Word Sense Disambiguation. However, since WordNet++ is part of a multilingual semantic network (Navigli and Ponzetto, 2010), we plan to explore the impact of this knowledge in a multilingual setting.

References

Eneko Agirre and Oier Lopez de Lacalle. 2004. Publicly available topic signatures for all WordNet nominal senses. In *Proc. of LREC '04*.

Eneko Agirre and David Martinez. 2001. Learning class-to-class selectional preferences. In *Proceedings of CoNLL-01*, pages 15–22.

Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proc. of EACL-09*, pages 33–41.

Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. 2009. Knowledge-based WSD on specific domains: performing better than generic supervised WSD. In *Proc. of IJCAI-09*, pages 1501–1506.

Satanjeev Banerjee and Ted Pedersen. 2003. Extended gloss overlap as a measure of semantic relatedness. In *Proc. of IJCAI-03*, pages 805–810.

Razvan Bunescu and Marius Paşca. 2006. Using encyclopedic knowledge for named entity disambiguation. In *Proc. of EACL-06*, pages 9–16.

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Yee Seng Chan, Hwee Tou Ng, and Zhi Zhong. 2007. NUS-ML: Exploiting parallel texts for Word Sense Disambiguation in the English all-words tasks. In *Proc. of SemEval-2007*, pages 253–256.

Ping Chen, Wei Ding, Chris Bowes, and David Brown. 2009. A fully unsupervised Word Sense Disambiguation method using dependency knowledge. In *Proc. of NAACL-HLT-09*, pages 28–36.

Tim Chklovski and Rada Mihalcea. 2002. Building a sense tagged corpus with Open Mind Word Expert. In *Proceedings of the ACL-02 Workshop on WSD: Recent Successes and Future Directions at ACL-02*.

Martin Chodorow, Roy Byrd, and George E. Heidorn. 1985. Extracting semantic hierarchies from a large on-line dictionary. In *Proc. of ACL-85*, pages 299–304.

Philipp Cimiano, Siegfried Handschuh, and Steffen Staab. 2004. Towards the self-annotating Web. In *Proc. of WWW-04*, pages 462–471.

Montse Cuadros and German Rigau. 2006. Quality assessment of large scale knowledge resources. In *Proc. of EMNLP-06*, pages 534–541.

Montse Cuadros and German Rigau. 2008. KnowNet: building a large net of knowledge from the Web. In *Proc. of COLING-08*, pages 161–168.

Philip Edmonds. 2000. Designing a task for SENSEVAL-2. Technical report, University of Brighton, U.K.

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

Evgeniy Gabrilovich and Shaul Markovitch. 2006. Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge. In *Proc. of AAAI-06*, pages 1301–1306.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proc. of IJCAI-07*, pages 1606–1611.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.

Sanda M. Harabagiu, George A. Miller, and Dan I. Moldovan. 1999. WordNet 2 – a morphologically and semantically enhanced resource. In *Proceedings of the SIGLEX99 Workshop on Standardizing Lexical Resources*, pages 1–8.

---

[8]The resulting resource, WordNet++, is freely available at `http://lcl.uniroma1.it/wordnetplusplus` for research purposes.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of COLING-92*, pages 539–545.

Adam Kilgarriff and Joseph Rosenzweig. 2000. Framework and results for English SENSEVAL. *Computers and the Humanities*, 34(1-2).

Rob Koeling and Diana McCarthy. 2007. Sussx: WSD using automatically acquired predominant senses. In *Proc. of SemEval-2007*, pages 314–317.

Rob Koeling, Diana McCarthy, and John Carroll. 2005. Domain-specific sense distributions and predominant sense acquisition. In *Proc. of HLT-EMNLP-05*, pages 419–426.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual Conference on Systems Documentation,* Toronto, Ontario, Canada, pages 24–26.

Diana McCarthy and John Carroll. 2003. Disambiguating nouns, verbs and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, 29(4):639–654.

Rada Mihalcea and Andras Csomai. 2007. Wikify! Linking documents to encyclopedic knowledge. In *Proc. of CIKM-07*, pages 233–242.

Rada Mihalcea. 2007. Using Wikipedia for automatic Word Sense Disambiguation. In *Proc. of NAACL-HLT-07*, pages 196–203.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, pages 303–308, Plainsboro, N.J.

David Milne and Ian H. Witten. 2008a. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, pages 25–30.

David Milne and Ian H. Witten. 2008b. Learning to link with Wikipedia. In *Proc. of CIKM-08*, pages 509–518.

Vivi Nastase and Michael Strube. 2008. Decoding Wikipedia category names for knowledge acquisition. In *Proc. of AAAI-08*, pages 1219–1224.

Vivi Nastase. 2008. Topic-driven multi-document summarization with encyclopedic knowledge and activation spreading. In *Proc. of EMNLP-08*, pages 763–772.

Roberto Navigli and Mirella Lapata. 2010. An experimental study on graph connectivity for unsupervised Word Sense Disambiguation. *IEEE Transactions on Pattern Anaylsis and Machine Intelligence*, 32(4):678–692.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proc. of ACL-10*.

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*, 27(7):1075–1088.

Roberto Navigli, Kenneth C. Litkowski, and Orin Hargraves. 2007. Semeval-2007 task 07: Coarse-grained English all-words task. In *Proc. of SemEval-2007*, pages 30–35.

Roberto Navigli. 2009a. Using cycles and quasi-cycles to disambiguate dictionary glosses. In *Proc. of EACL-09*, pages 594–602.

Roberto Navigli. 2009b. Word Sense Disambiguation: A survey. *ACM Computing Surveys*, 41(2):1–69.

Marco Pennacchiotti and Patrick Pantel. 2006. Ontologizing semantic relations. In *Proc. of COLING-ACL-06*, pages 793–800.

Simone Paolo Ponzetto and Roberto Navigli. 2009. Large-scale taxonomy mapping for restructuring and integrating Wikipedia. In *Proc. of IJCAI-09*, pages 2083–2088.

Simone Paolo Ponzetto and Michael Strube. 2007a. Deriving a large scale taxonomy from Wikipedia. In *Proc. of AAAI-07*, pages 1440–1445.

Simone Paolo Ponzetto and Michael Strube. 2007b. Knowledge derived from Wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, 30:181–212.

Nils Reiter, Matthias Hartung, and Anette Frank. 2008. A resource-poor approach for linking ontology classes to Wikipedia articles. In Johan Bos and Rodolfo Delmonte, editors, *Semantics in Text Processing*, volume 1 of *Research in Computational Semantics*, pages 381–387. College Publications, London, England.

German Rigau, Horacio Rodríguez, and Eneko Agirre. 1998. Building accurate semantic taxonomies from monolingual MRDs. In *Proc. of COLING-ACL-98*, pages 1103–1109.

Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets. In *Advances in Web Intelligence*, volume 3528 of *Lecture Notes in Computer Science*. Springer Verlag.

Christina Sauper and Regina Barzilay. 2009. Automatically generating Wikipedia articles: A structure-aware approach. In *Proc. of ACL-IJCNLP-09*, pages 208–216.

Eyal Shnarch, Libby Barak, and Ido Dagan. 2009. Extracting lexical reference rules from Wikipedia. In *Proc. of ACL-IJCNLP-09*, pages 450–458.

Rion Snow, Dan Jurafsky, and Andrew Ng. 2006. Semantic taxonomy induction from heterogeneous evidence. In *Proc. of COLING-ACL-06*, pages 801–808.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. Yago: A large ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 6(3):203–217.

Fei Wu and Daniel Weld. 2007. Automatically semantifying Wikipedia. In *Proc. of CIKM-07*, pages 41–50.

Fei Wu and Daniel Weld. 2008. Automatically refining the Wikipedia infobox ontology. In *Proc. of WWW-08*, pages 635–644.