BabelNet: Building a Very Large Multilingual Semantic Network

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

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

Abstract

In this paper we present BabelNet – a very large, wide-coverage multilingual semantic network. The resource is automatically constructed by means of a methodology that integrates lexicographic and encyclopedic knowledge from WordNet and Wikipedia. In addition Machine Translation is also applied to enrich the resource with lexical information for all languages. We conduct experiments on new and existing gold-standard datasets to show the high quality and coverage of the resource.

1 Introduction

In many research areas of Natural Language Processing (NLP) lexical knowledge is exploited to perform tasks effectively. These include, among others, text summarization (Nastase, 2008), Named Entity Recognition (Bunescu and Paşca, 2006), Question Answering (Harabagiu et al., 2000) and text categorization (Gabrilovich and Markovitch, 2006). Recent studies in the difficult task of Word Sense Disambiguation (Navigli, 2009b, WSD) have shown the impact of the amount and quality of lexical knowledge (Cuadros and Rigau, 2006): richer knowledge sources can be of great benefit to both knowledge-lean systems (Navigli and Lapata, 2010) and supervised classifiers (Ng and Lee, 1996; Yarowsky and Florian, 2002).

Various projects have been undertaken to make lexical knowledge available in a machine readable format. A pioneering endeavor was Word-Net (Fellbaum, 1998), a computational lexicon of English based on psycholinguistic theories. Subsequent projects have also tackled the significant problem of multilinguality. These include EuroWordNet (Vossen, 1998), MultiWordNet (Pianta et al., 2002), the Multilingual Central Repository (Atserias et al., 2004), and many others. However, manual construction methods inherently suffer from a number of drawbacks. First, maintaining and updating lexical knowledge resources is expensive and time-consuming. Second, such resources are typically lexicographic, and thus contain mainly concepts and only a few named entities. Third, resources for non-English languages often have a much poorer coverage since the construction effort must be repeated for every language of interest. As a result, an obvious bias exists towards conducting research in resource-rich languages, such as English.

A solution to these issues is to draw upon a large-scale collaborative resource, namely Wikipedia[1]. Wikipedia represents the perfect complement to WordNet, as it provides multilingual lexical knowledge of a mostly encyclopedic nature. While the contribution of any individual user might be imprecise or inaccurate, the continual intervention of expert contributors in all domains results in a resource of the highest quality (Giles, 2005). But while a great deal of work has been recently devoted to the automatic extraction of structured information from Wikipedia (Wu and Weld, 2007; Ponzetto and Strube, 2007; Suchanek et al., 2008; Medelyan et al., 2009, *inter alia*), the knowledge extracted is organized in a looser way than in a computational lexicon such as WordNet.

In this paper, we make a major step towards the vision of a wide-coverage multilingual knowledge resource. We present a novel methodology that produces a very large multilingual semantic network: BabelNet. This resource is created by linking Wikipedia to WordNet via an automatic mapping and by integrating lexical gaps in resource-

---

[1] `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.
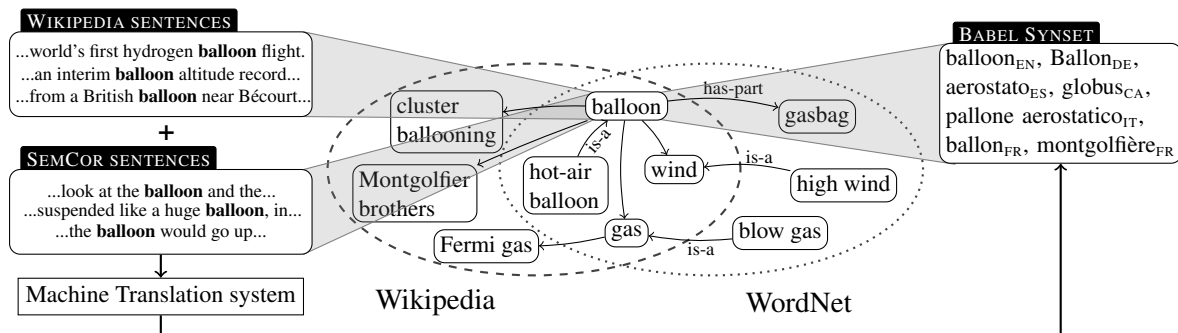
216

Figure 1: An illustrative overview of BabelNet.

poor languages with the aid of Machine Translation. The result is an "encyclopedic dictionary", that provides concepts and named entities lexicalized in many languages and connected with large amounts of semantic relations.

2 BabelNet

We encode knowledge as a labeled directed graph $G = (V, E)$ where V is the set of *vertices* – i.e. *concepts*[2] such as balloon – and $E \subseteq V \times R \times V$ is the set of *edges* connecting pairs of concepts. Each edge is labeled with a *semantic relation* from R, e.g. {*is-a*, *part-of*, ..., ϵ}, where ϵ denotes an unspecified semantic relation. Importantly, each vertex $v \in V$ contains a set of lexicalizations of the concept for different languages, e.g. { balloon_EN, Ballon_DE, aerostato_ES, ..., montgolfière_FR }.

Concepts and relations in BabelNet are harvested from the largest available semantic lexicon of English, WordNet, and a wide-coverage collaboratively edited encyclopedia, the English Wikipedia (Section 3.1). We collect (a) from WordNet, all available word senses (as *concepts*) and all the semantic pointers between synsets (as *relations*); (b) from Wikipedia, all encyclopedic entries (i.e. pages, as *concepts*) and semantically unspecified *relations* from hyperlinked text.

In order to provide a *unified resource*, we merge the intersection of these two knowledge sources (i.e. their concepts in common) by establishing a mapping between Wikipedia pages and WordNet senses (Section 3.2). This avoids duplicate concepts and allows their inventories of concepts to complement each other. Finally, to enable multilinguality, we collect the lexical realizations of the available concepts in different languages by

using (a) the human-generated translations provided in Wikipedia (the so-called *inter-language* links), as well as (b) a machine translation system to translate occurrences of the concepts within sense-tagged corpora, namely SemCor (Miller et al., 1993) – a corpus annotated with WordNet senses – and Wikipedia itself (Section 3.3). We call the resulting set of multilingual lexicalizations of a given concept a *babel synset*. An overview of BabelNet is given in Figure 1 (we label vertices with English lexicalizations): unlabeled edges are obtained from links in the Wikipedia pages (e.g. BALLOON (AIRCRAFT) links to WIND), whereas labeled ones from WordNet[3] (e.g. $balloon_n^1$ *has-part* $gasbag_n^1$). In this paper we restrict ourselves to concepts lexicalized as nouns. Nonetheless, our methodology can be applied to all parts of speech, but in that case Wikipedia cannot be exploited, since it mainly contains nominal entities.

3 Methodology

3.1 Knowledge Resources

WordNet. The most popular lexical knowledge resource in the field of NLP is certainly WordNet, a computational lexicon of the English language. A concept in WordNet is represented as a synonym set (called *synset*), i.e. the set of words that share the same meaning. For instance, the concept wind is expressed by the following synset:

$$\{ \text{wind}_n^1, \text{air current}_n^1, \text{current of air}_n^1 \},$$

where each word's subscripts and superscripts indicate their parts of speech (e.g. n stands for noun)

---

[2]Throughout the paper, unless otherwise stated, we use the general term *concept* to denote either a concept or a named entity.

[3]We use in the following WordNet version 3.0. We denote with w_p^i the i-th sense of a word w with part of speech p. We use word senses to unambiguously denote the corresponding synsets (e.g. $plane_n^1$ for { $airplane_n^1$, $aeroplane_n^1$, $plane_n^1$ }). Hereafter, we use *word sense* and *synset* interchangeably.

and sense number, respectively. For each synset, WordNet provides a textual definition, or gloss. For example, the gloss of the above synset is: "air moving from an area of high pressure to an area of low pressure".

Wikipedia. Our second resource, Wikipedia, is a Web-based collaborative encyclopedia. A Wikipedia page (henceforth, Wikipage) presents the knowledge about a specific concept (e.g. BALLOON (AIRCRAFT)) or named entity (e.g. MONTGOLFIER BROTHERS). The page typically contains hypertext linked to other relevant Wikipages. For instance, BALLOON (AIRCRAFT) is linked to WIND, GAS, and so on. The title of a Wikipage (e.g. BALLOON (AIRCRAFT)) is composed of the lemma of the concept defined (e.g. balloon) plus an optional label in parentheses which specifies its meaning if the lemma is ambiguous (e.g. AIRCRAFT VS. TOY). Wikipages also provide inter-language links to their counterparts in other languages (e.g. BALLOON (AIRCRAFT) links to the Spanish page AEROSTATO). Finally, some Wikipages are redirections to other pages, e.g. the Spanish BALÓN AEROSTÁTICO redirects to AEROSTATO.

3.2 Mapping Wikipedia to WordNet

The first phase of our methodology aims to establish links between Wikipages and WordNet senses. We aim to acquire a mapping μ such that, for each Wikipage w, we have:

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

where $Senses_{\text{WN}}(w)$ is the set of senses of the *lemma* of w in WordNet. For example, if our mapping methodology linked BALLOON (AIRCRAFT) to the corresponding WordNet sense balloon_n^1, we would have $\mu(\text{BALLOON (AIRCRAFT)}) = \text{balloon}_n^1$.

In order to establish a mapping between the two resources, we first identify the 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 Wikipage w, we use the following information as disambiguation context:

- **Sense labels**: e.g. given the page BALLOON (AIRCRAFT), the word aircraft is added to the disambiguation context.
- **Links**: the titles' lemmas of the pages linked from the target Wikipage (i.e., outgoing links). For instance, the links in the Wikipage BALLOON (AIRCRAFT) include wind, gas, etc.
- **Categories**: Wikipages are typically classified according to one or more categories. For example, the Wikipage BALLOON (AIRCRAFT) is categorized as BALLOONS, BALLOONING, etc. While many categories are very specific and do not appear in WordNet (e.g., SWEDISH WRITERS or SCIENTISTS WHO COMMITTED SUICIDE), we use their syntactic heads as disambiguation context (i.e. writer and scientist, respectively).

Given a Wikipage w, we define its disambiguation context $Ctx(w)$ as the set of words obtained from 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 collect the following information:

- **Synonymy**: all synonyms of s in S. For instance, given the sense airplane_n^1 and its corresponding synset { airplane_n^1, aeroplane_n^1, plane_n^1 }, the words contained therein are included in the context.
- **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 balloon_n^1, we include the words from its hypernym { $\text{lighter-than-air craft}_n^1$ } and all its hyponyms (e.g. { $\text{hot-air balloon}_n^1$ }).
- **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 balloon_n^1, it can be found that { balloon_n^1 } and { airship_n^1, dirigible_n^1 } are sisters. Thus airship and dirigible are included in the disambiguation context of s.
- **Gloss**: the set of lemmas of the content words occurring within the WordNet gloss of S.

We thus define the disambiguation context $Ctx(s)$ of sense s as the set of words obtained from all of the four sources above.

3.2.3 Mapping Algorithm

In order to link each Wikipedia page to a WordNet sense, we perform the following steps:

- Initially, 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 we map w to its only WordNet sense.
- For each remaining Wikipage w for which no mapping was previously found (i.e., $\mu(w) = \epsilon$), we assign the most likely sense to w based on the maximization of the conditional probabilities $p(s|w)$ over the senses $s \in Senses_{\text{WN}}(w)$ (no mapping is established if a tie occurs).

To find the mapping of a Wikipage w, we need to compute the conditional probability $p(s|w)$ of selecting the WordNet sense s given w. The sense s which maximizes this probability is determined as follows:

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

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, determining the most appropriate sense s consists of finding the sense s that maximizes the joint probability $p(s,w)$. 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. More details on the mapping algorithm can be found in Ponzetto and Navigli (2010).

3.3 Translating Babel Synsets

So far we have linked English Wikipages to WordNet senses. Given a Wikipage w, and provided it is mapped to a sense s (i.e., $\mu(w) = s$), we create a babel synset $S \cup W$, where S is the WordNet synset to which sense s belongs, and W includes:

(i) w; (ii) all its inter-language links (that is, translations of the Wikipage to other languages); (iii) the redirections to the inter-language links found in the Wikipedia of the target language. For instance, given that $\mu(\text{BALLOON}) = \text{balloon}_n^1$, the corresponding babel synset is { balloon$_{\text{EN}}$, Ballon$_{\text{DE}}$, aerostato$_{\text{ES}}$, balón aerostático$_{\text{ES}}$, ..., pallone aerostatico$_{\text{IT}}$ }. However, two issues arise: first, a concept might be covered only in one of the two resources (either WordNet or Wikipedia), meaning that no link can be established (e.g., FERMI GAS or gasbag$_n^1$ in Figure 1); second, even if covered in both resources, the Wikipage for the concept might not provide any translation for the language of interest (e.g., the Catalan for BALLOON is missing in Wikipedia).

In order to address the above issues and thus guarantee high coverage for all languages we developed a methodology for translating senses in the babel synset to missing languages. Given a WordNet word sense in our babel synset of interest (e.g. balloon$_n^1$) we collect its occurrences in SemCor (Miller et al., 1993), a corpus of more than 200,000 words annotated with WordNet senses. We do the same for Wikipages by retrieving sentences in Wikipedia with links to the Wikipage of interest. By repeating this step for each English lexicalization in a babel synset, we obtain a collection of sentences for the babel synset (see left part of Figure 1). Next, we apply state-of-the-art Machine Translation[4] and translate the set of sentences in all the languages of interest. Given a specific term in the initial babel synset, we collect the set of its translations. We then identify the most frequent translation in each language and add it to the babel synset. Note that translations are sense-specific, as the context in which a term occurs is provided to the translation system.

3.4 Example

We now illustrate the execution of our methodology by way of an example. Let us focus on the Wikipage BALLOON (AIRCRAFT). The word is polysemous both in Wikipedia and WordNet. In the first phase of our methodology we aim to find a mapping $\mu(\text{BALLOON (AIRCRAFT)})$ to an appropriate WordNet sense of the word. To

---

[4]We use the Google Translate API. An initial prototype used a statistical machine translation system based on Moses (Koehn et al., 2007) and trained on Europarl (Koehn, 2005). However, we found such system unable to cope with many technical names, such as in the domains of sciences, literature, history, etc.

this end we construct the disambiguation context for the Wikipage by including words from its label, links and categories (cf. Section 3.2.1). The context thus includes, among others, the following words: aircraft, wind, airship, lighter-than-air. We now construct the disambiguation context for the two WordNet senses of balloon (cf. Section 3.2.2), namely the aircraft (#1) and the toy (#2) senses. To do so, we include words from their synsets, hypernyms, hyponyms, sisters, and glosses. The context for $balloon_n^1$ includes: aircraft, craft, airship, lighter-than-air. The context for $balloon_n^2$ contains: toy, doll, hobby. The sense with the largest intersection is #1, so the following mapping is established: $\mu($BALLOON (AIRCRAFT)$) = balloon_n^1$. After the first phase, our babel synset includes the following English words from WordNet plus the Wikipedia inter-language links to other languages (we report German, Spanish and Italian): { $balloon_{EN}$, $Ballon_{DE}$, $aerostato_{ES}$, $balón\ aerostático_{ES}$, $pallone\ aerostatico_{IT}$ }.

In the second phase (see Section 3.3), we collect all the sentences in SemCor and Wikipedia in which the above English word sense occurs. We translate these sentences with the Google Translate API and select the most frequent translation in each language. As a result, we can enrich the initial babel synset with the following words: $mongolfière_{FR}$, $globus_{CA}$, $globo_{ES}$, $mongolfiera_{IT}$. Note that we had no translation for Catalan and French in the first phase, because the inter-language link was not available, and we also obtain new lexicalizations for the Spanish and Italian languages.

4 Experiment 1: Mapping Evaluation

Experimental setting. We first performed an evaluation of the quality of our mapping from Wikipedia to WordNet. To create a gold standard for evaluation we considered all lemmas whose senses are contained both in WordNet and Wikipedia: the intersection between the two resources contains 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 then selected a random sample of 1,000 Wikipages and asked an annotator with previous experience in lexicographic annota-

	P	R	F_1	A
Mapping algorithm	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.

tion to provide the correct WordNet sense for each page (an empty sense label was given, if no correct mapping was possible). The gold-standard dataset includes 505 non-empty mappings, i.e. Wikipages 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. Our annotators achieved a κ inter-annotator agreement (Carletta, 1996) of 0.9, indicating almost perfect agreement.

Results and discussion. Table 1 summarizes the performance of our mapping algorithm against the manually annotated dataset. Evaluation is performed in terms of standard measures of precision, recall, and F_1-measure. In addition we calculate accuracy, which also takes into account empty sense labels. As baselines we use the most frequent WordNet sense (MFS), and a random sense assignment.

The results show that our method achieves almost 80% F_1 and it improves over the baselines by a large margin. The final mapping contains 81,533 pairs of Wikipages and word senses they map to, covering 55.7% of the noun senses in WordNet. 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 statistical 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 selecting the first WordNet sense rather than any other sense for each target page represents a choice as arbitrary as picking a sense at random.

5 Experiment 2: Translation Evaluation

We perform a second set of experiments concerning the quality of the acquired concepts. This is assessed in terms of coverage against gold-standard resources (Section 5.1) and against a manually-validated dataset of translations (Section 5.2).

Language	Word senses	Synsets
German	15,762	9,877
Spanish	83,114	55,365
Catalan	64,171	40,466
Italian	57,255	32,156
French	44,265	31,742

Table 2: Size of the gold-standard wordnets.

5.1 Automatic Evaluation

Datasets. We compare BabelNet against gold-standard resources for 5 languages, namely: the subset of GermaNet (Lemnitzer and Kunze, 2002) included in EuroWordNet for German, Multi-WordNet (Pianta et al., 2002) for Italian, the Multilingual Central Repository for Spanish and Catalan (Atserias et al., 2004), and WOrdnet Libre du Français (Benoît and Fišer, 2008, WOLF) for French. In Table 2 we report the number of synsets and word senses available in the gold-standard resources for the 5 languages.

Measures. Let $\mathcal{B}$ be BabelNet, $\mathcal{F}$ our gold-standard non-English wordnet (e.g. GermaNet), and let $\mathcal{E}$ be the English WordNet. All the gold-standard non-English resources, as well as Babel-Net, are linked to the English WordNet: given a synset $S_\mathcal{F} \in \mathcal{F}$, we denote its corresponding babel synset as $S_\mathcal{B}$ and its synset in the English Word-Net as $S_\mathcal{E}$. We assess the coverage of BabelNet against our gold-standard wordnets both in terms of synsets and word senses. For synsets, we calculate coverage as follows:

$$\text{SynsetCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_\mathcal{F} \in \mathcal{F}} \delta(S_\mathcal{B}, S_\mathcal{F})}{|\{S_\mathcal{F} \in \mathcal{F}\}|},$$

where $\delta(S_\mathcal{B}, S_\mathcal{F}) = 1$ if the two synsets $S_\mathcal{B}$ and $S_\mathcal{F}$ have a synonym in common, 0 otherwise. That is, synset coverage is determined as the percentage of synsets of $\mathcal{F}$ that share a term with the corresponding babel synsets. For word senses we calculate a similar measure of coverage:

$$\text{WordCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_\mathcal{F} \in \mathcal{F}} \sum_{s_\mathcal{F} \in S_\mathcal{F}} \delta'(s_\mathcal{F}, S_\mathcal{B})}{|\{s_\mathcal{F} \in S_\mathcal{F} : S_\mathcal{F} \in \mathcal{F}\}|},$$

where $s_\mathcal{F}$ is a word sense in synset $S_\mathcal{F}$ and $\delta'(s_\mathcal{F}, S_\mathcal{B}) = 1$ if $s_\mathcal{F} \in S_\mathcal{B}$, 0 otherwise. That is we calculate the ratio of word senses in our gold-standard resource $\mathcal{F}$ that also occur in the corresponding synset $S_\mathcal{B}$ to the overall number of senses in $\mathcal{F}$.

However, our gold-standard resources cover only a portion of the English WordNet, whereas the overall coverage of BabelNet is much higher. We calculate extra coverage for synsets as follows:

$$\text{SynsetExtraCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_\mathcal{E} \in \mathcal{E} \setminus \mathcal{F}} \delta(S_\mathcal{B}, S_\mathcal{E})}{|\{S_\mathcal{F} \in \mathcal{F}\}|}.$$

Similarly, we calculate extra coverage for word senses in BabelNet corresponding to WordNet synsets not covered by the reference resource $\mathcal{F}$.

Results and discussion. We evaluate the coverage and extra coverage of word senses and synsets at different stages: (a) using only the inter-language links from Wikipedia (WIKI Links); (b) and (c) using only the automatic translations of the sentences from Wikipedia (WIKI Transl.) or Sem-Cor (WN Transl.); (d) using all available translations, i.e. BABELNET.

Coverage results are reported in Table 3. The percentage of word senses covered by BabelNet ranges from 52.9% (Italian) to 66.4 (Spanish) and 86.0% (French). Synset coverage ranges from 73.3% (Catalan) to 76.6% (Spanish) and 92.9% (French). As expected, synset coverage is higher, because a synset in the reference resource is considered to be covered if it shares at least one word with the corresponding synset in BabelNet.

Numbers for the extra coverage, which provides information about the percentage of word senses and synsets in BabelNet but not in the gold-standard resources, are given in Figure 2. The results show that we provide for all languages a high extra coverage for both word senses – between 340.1% (Catalan) and 2,298% (German) – and synsets – between 102.8% (Spanish) and 902.6% (German).

Table 3 and Figure 2 show that the best results are obtained when combining *all* available translations, i.e. both from Wikipedia and the machine translation system. The performance figures suffer from the errors of the mapping phase (see Section 4). Nonetheless, the results are generally high, with a peak for French, since WOLF has been created semi-automatically by combining several resources, *including* Wikipedia. The relatively low word sense coverage for Italian (55.4%) is, instead, due to the lack of many common words in the Italian gold-standard synsets. Examples include whip$_{EN}$ translated as staffile$_{IT}$ but not as the more common frusta$_{IT}$, playboy$_{EN}$ translated as vitaiolo$_{IT}$ but not gigolò$_{IT}$, etc.
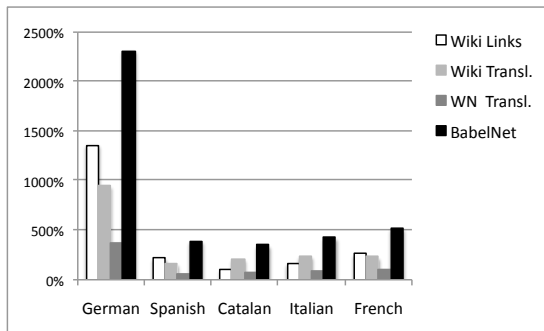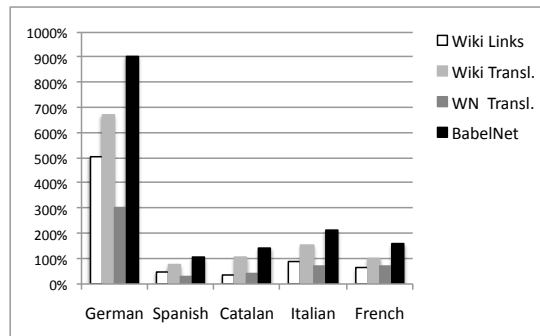
(a) word senses   (b) synsets

Figure 2: Extra coverage against gold-standard wordnets: word senses (a) and synsets (b).

	Resource	Method	SENSES	SYNSETS
German	WIKI {	Links	39.6	50.7
		Transl.	42.6	58.2
	WN	Transl.	21.0	28.6
	BABELNET	All	**57.6**	**73.4**
Spanish	WIKI {	Links	34.4	40.7
		Transl.	47.9	56.1
	WN	Transl.	25.2	30.0
	BABELNET	All	**66.4**	**76.6**
Catalan	WIKI {	Links	20.3	25.2
		Transl.	46.9	54.1
	WN	Transl.	25.0	29.6
	BABELNET	All	**64.0**	**73.3**
Italian	WIKI {	Links	28.1	40.0
		Transl.	39.9	58.0
	WN	Transl.	19.7	28.7
	BABELNET	All	**52.9**	**73.7**
French	WIKI {	Links	70.0	72.4
		Transl.	69.6	79.6
	WN	Transl.	16.3	19.4
	BABELNET	All	**86.0**	**92.9**

Table 3: Coverage against gold-standard wordnets (we report percentages).

5.2 Manual Evaluation

Experimental setup. The automatic evaluation quantifies how much of the gold-standard resources is covered by BabelNet. However, it does not say anything about the precision of the *additional* lexicalizations provided by BabelNet. Given that our resource has displayed a remarkably high extra coverage – ranging from 340% to 2,298% of the national wordnets (see Figure 2) – we performed a second evaluation to assess its precision. For each of our 5 languages, we

selected a random set of 600 babel synsets composed as follows: 200 synsets whose senses exist in WordNet only, 200 synsets in the intersection between WordNet and Wikipedia (i.e. those mapped with our method illustrated in Section 3.2), 200 synsets whose lexicalizations exist in Wikipedia only. Therefore, our dataset included $600 \times 5 = 3,000$ babel synsets. None of the synsets was covered by any of the five reference wordnets. The babel synsets were manually validated by expert annotators who decided which senses (i.e. lexicalizations) were appropriate given the corresponding WordNet gloss and/or Wikipage.

Results and discussion. We report the results in Table 4. For each language (rows) and for each of the three regions of BabelNet (columns), we report precision (i.e. the percentage of synonyms deemed correct) and, in parentheses, the overall number of synonyms evaluated. The results show that the different regions of BabelNet contain translations of different quality: while on average translations for WordNet-only synsets have a precision around 72%, when Wikipedia comes into play the performance increases considerably (around 80% in the intersection and 95% with Wikipedia-only translations). As can be seen from the figures in parentheses, the number of translations available in the presence of Wikipedia is higher. This quantitative difference is due to our method collecting many translations from the redirections in the Wikipedia of the target language (Section 3.3), as well as to the paucity of examples in SemCor for many synsets. In addition, some of the synsets in WordNet with no Wikipedia counterpart are very difficult to translate. Examples include terms like stammel, crape fern, baseball clinic, and many others for which we could

222

Language	WN	WN ∩ Wiki	Wiki
German	73.76 (282)	78.37 (777)	97.74 (709)
Spanish	69.45 (275)	78.53 (643)	92.46 (703)
Catalan	75.58 (258)	82.98 (517)	92.71 (398)
Italian	72.32 (271)	80.83 (574)	99.09 (552)
French	67.16 (268)	77.43 (709)	96.44 (758)

Table 4: Precision of BabelNet on synonyms in WordNet (WN), Wikipedia (Wiki) and their intersection (WN ∩ Wiki): percentage and total number of words (in parentheses) are reported.

not find translations in major editions of bilingual dictionaries. In contrast, good translations were produced using our machine translation method when enough sentences were available. Examples are: chaudrée de poisson$_{FR}$ for fish chowder$_{EN}$, grano de café$_{ES}$ for coffee bean$_{EN}$, etc.

6 Related Work

Previous attempts to manually build multilingual resources have led to the creation of a multitude of wordnets such as EuroWordNet (Vossen, 1998), MultiWordNet (Pianta et al., 2002), Balka-Net (Tufiş et al., 2004), Arabic WordNet (Black et al., 2006), the Multilingual Central Repository (Atserias et al., 2004), bilingual electronic dictionaries such as EDR (Yokoi, 1995), and fully-fledged frameworks for the development of multilingual lexicons (Lenci et al., 2000). As it is often the case with manually assembled resources, these lexical knowledge repositories are hindered by high development costs and an insufficient coverage. This barrier has led to proposals that acquire multilingual lexicons from either parallel text (Gale and Church, 1993; Fung, 1995, *inter alia*) or monolingual corpora (Sammer and Soderland, 2007; Haghighi et al., 2008). The disambiguation of bilingual dictionary glosses has also been proposed to create a bilingual semantic network from a machine readable dictionary (Navigli, 2009a). Recently, Etzioni et al. (2007) and Mausam et al. (2009) presented methods to produce massive multilingual translation dictionaries from Web resources such as online lexicons and Wiktionaries. However, while providing lexical resources on a very large scale for hundreds of thousands of language pairs, these do not encode semantic relations between concepts denoted by their lexical entries.

The research closest to ours is presented by de Melo and Weikum (2009), who developed a Universal WordNet (UWN) by automatically acquiring a semantic network for languages other than English. UWN is bootstrapped from WordNet and is built by collecting evidence extracted from existing wordnets, translation dictionaries, and parallel corpora. The result is a graph containing 800,000 words from over 200 languages in a hierarchically structured semantic network with over 1.5 million links from words to word senses. Our work goes one step further by (1) developing an even larger multilingual resource including *both* lexical semantic and encyclopedic knowledge, (2) enriching the structure of the 'core' semantic network (i.e. the semantic pointers from WordNet) with topical, semantically unspecified relations from the link structure of Wikipedia. This result is essentially achieved by complementing WordNet with Wikipedia, as well as by leveraging the multilingual structure of the latter. Previous attempts at linking the two resources have been proposed. These include associating Wikipedia pages with the most frequent WordNet sense (Suchanek et al., 2008), extracting domain information from Wikipedia and providing a manual mapping to WordNet concepts (Auer et al., 2007), a model based on vector spaces (Ruiz-Casado et al., 2005), a supervised approach using keyword extraction (Reiter et al., 2008), as well as automatically linking Wikipedia categories to WordNet based on structural information (Ponzetto and Navigli, 2009). In contrast to previous work, BabelNet is the first proposal that integrates the relational structure of WordNet with the semi-structured information from Wikipedia into a unified, wide-coverage, multilingual semantic network.

7 Conclusions

In this paper we have presented a novel methodology for the automatic construction of a large multilingual lexical knowledge resource. Key to our approach is the establishment of a mapping between a multilingual encyclopedic knowledge repository (Wikipedia) and a computational lexicon of English (WordNet). This integration process has several advantages. Firstly, the two resources contribute different kinds of lexical knowledge, one is concerned mostly with named entities, the other with concepts. Secondly, while Wikipedia is less structured than WordNet, it provides large

amounts of semantic relations and can be leveraged to enable multilinguality. Thus, even when they overlap, the two resources provide complementary information about the same named entities or concepts. Further, we contribute a large set of sense occurrences harvested from Wikipedia and SemCor, a corpus that we input to a state-of-the-art machine translation system to fill in the gap between resource-rich languages – such as English – and resource-poorer ones. Our hope is that the availability of such a language-rich resource[5] will enable many non-English and multilingual NLP applications to be developed.

Our experiments show that our fully-automated approach produces a large-scale lexical resource with high accuracy. The resource includes millions of semantic relations, mainly from Wikipedia (however, WordNet relations are labeled), and contains almost 3 million concepts (6.7 labels per concept on average). As pointed out in Section 5, such coverage is much wider than that of existing wordnets in non-English languages. While BabelNet currently includes 6 languages, links to freely-available wordnets[6] can immediately be established by utilizing the English WordNet as an interlanguage index. Indeed, BabelNet can be extended to virtually any language of interest. In fact, our translation method allows it to cope with any resource-poor language.

As future work, we plan to apply our method to other languages, including Eastern European, Arabic, and Asian languages. We also intend to link missing concepts in WordNet, by establishing their most likely hypernyms – e.g., à la Snow et al. (2006). We will perform a semi-automatic validation of BabelNet, e.g. by exploiting Amazon's Mechanical Turk (Callison-Burch, 2009) or designing a collaborative game (von Ahn, 2006) to validate low-ranking mappings and translations. Finally, we aim to apply BabelNet to a variety of applications which are known to benefit from a wide-coverage knowledge resource. We have already shown that the English-only subset of BabelNet allows simple knowledge-based algorithms to compete with supervised systems in standard coarse-grained and domain-specific WSD settings (Ponzetto and Navigli, 2010). We plan in the near future to apply BabelNet to the challenging task of cross-lingual WSD (Lefever and Hoste, 2009).

---

[5]BabelNet can be freely downloaded for research purposes at http://lcl.uniroma1.it/babelnet.
[6]http://www.globalwordnet.org.

References

Jordi Atserias, Luis Villarejo, German Rigau, Eneko Agirre, John Carroll, Bernardo Magnini, and Piek Vossen. 2004. The MEANING multilingual central repository. In *Proc. of GWC-04*, pages 80–210.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ive. 2007. Dbpedia: A nucleus for a web of open data. In *Proceedings of 6th International Semantic Web Conference joint with 2nd Asian Semantic Web Conference (ISWC+ASWC 2007)*, pages 722–735.

Sagot Benoît and Darja Fišer. 2008. Building a free French WordNet from multilingual resources. In *Proceedings of the Ontolex 2008 Workshop*.

William Black, Sabri Elkateb Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, and Adam Pease. 2006. Introducing the Arabic WordNet project. In *Proc. of GWC-06*, pages 295–299.

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

Chris Callison-Burch. 2009. Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proc. of EMNLP-09*, pages 286–295.

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

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

Gerard de Melo and Gerhard Weikum. 2009. Towards a universal wordnet by learning from combined evidence. In *Proc. of CIKM-09*, pages 513–522.

Oren Etzioni, Kobi Reiter, Stephen Soderland, and Marcus Sammer. 2007. Lexical translation with application to image search on the Web. In *Proceedings of Machine Translation Summit XI*.

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

Pascale Fung. 1995. A pattern matching method for finding noun and proper noun translations from noisy parallel corpora. In *Proc. of ACL-95*, pages 236–243.

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.

William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102.

Jim Giles. 2005. Internet encyclopedias go head to head. *Nature*, 438:900–901.

Aria Haghighi, Percy Liang, Taylor Berg-Kirkpatrick, and Dan Klein. 2008. Learning bilingual lexicons from monolingual corpora. In *Proc. of ACL-08*, pages 771–779.

Sanda M. Harabagiu, Dan Moldovan, Marius Paşca, Rada Mihalcea, Mihai Surdeanu, Razvan Bunescu, Roxana Girju, Vasile Rus, and Paul Morarescu. 2000. FALCON: Boosting knowledge for answer engines. In *Proc. of TREC-9*, pages 479–488.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. In *Comp. Vol. to Proc. of ACL-07*, pages 177–180.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X*.

Els Lefever and Veronique Hoste. 2009. Semeval-2010 task 3: Cross-lingual Word Sense Disambiguation. In *Proc. of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, pages 82–87, Boulder, Colorado.

Lothar Lemnitzer and Claudia Kunze. 2002. GermaNet – representation, visualization, application. In *Proc. of LREC '02*, pages 1485–1491.

Alessandro Lenci, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, and Antonio Zampolli. 2000. SIMPLE: A general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13(4):249–263.

Mausam, Stephen Soderland, Oren Etzioni, Daniel Weld, Michael Skinner, and Jeff Bilmes. 2009. Compiling a massive, multilingual dictionary via probabilistic inference. In *Proc. of ACL-IJCNLP-09*, pages 262–270.

Olena Medelyan, David Milne, Catherine Legg, and Ian H. Witten. 2009. Mining meaning from Wikipedia. *Int. J. Hum.-Comput. Stud.*, 67(9):716–754.

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.

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. 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.

Hwee Tou Ng and Hian Beng Lee. 1996. Integrating multiple knowledge sources to disambiguate word senses: An exemplar-based approach. In *Proc. of ACL-96*, pages 40–47.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proc. of GWC-02*, pages 21–25.

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 Roberto Navigli. 2010. Knowledge-rich Word Sense Disambiguation rivaling supervised system. In *Proc. of ACL-10*.

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

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.

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.

Marcus Sammer and Stephen Soderland. 2007. Building a sense-distinguished multilingual lexicon from monolingual corpora and bilingual lexicons. In *Proceedings of Machine Translation Summit XI*.

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.

Dan Tufiş, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. a general overview. *Romanian Journal on Science and Technology of Information*, 7(1-2):9–43.

Luis von Ahn. 2006. Games with a purpose. *IEEE Computer*, 6(39):92–94.

Piek Vossen, editor. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer, Dordrecht, The Netherlands.

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

David Yarowsky and Radu Florian. 2002. Evaluating sense disambiguation across diverse parameter spaces. *Natural Language Engineering*, 9(4):293–310.

Toshio Yokoi. 1995. The EDR electronic dictionary. *Communications of the ACM*, 38(11):42–44.

225