Aligning an Italian WordNet with a Lexicographic Dictionary: Coping with limited data

Tommaso Caselli
TrentoRISE
Via Sommarive, 18
Povo (TN) IT-38123
t.caselli@trentorise.eu

Carlo Strapparava
FBK-HLT
Via Sommarive, 18
Povo (TN) IT-38123
strappa@fbk.eu

Laure Vieu
LOA-CNR & IRIT-CNRS
118 route de Narbonne
Toulouse F-31062
vieu@irit.fr

Guido Vetere
IBM CAS
P.zza Manci, 17
Povo (TN) IT-38123
gvetere@it.ibm.com

Abstract

This work describes the evaluations of two approaches, Lexical Matching and Sense Similarity, for word sense alignment between MultiWordNet and a lexicographic dictionary, Senso Comune De Mauro, when having few sense descriptions (MultiWordNet) and no structure over senses (Senso Comune De Mauro). The results obtained from the merging of the two approaches are satisfying, with F1 values of 0.47 for verbs and 0.64 for nouns.

1 Introduction

This work is situated in the field of word sense alignment, a research area which has seen an increasing interest in recent years and which is a key requirement for achieving semantic interoperability between different lexical-semantic resources (Matuschek and Gurevych, 2013). Our goal is to automatically import high-quality glosses in Italian in MultiWordNet (Pianta et al., 2002) (MWN) by aligning its synsets to the entries of a lexicographic dictionary, namely the Senso Comune De Mauro (SCDM), thus providing Italian with a more complete and robust version of MWN. For SCDM, the linking of the entries with MWN plays a double role. On the one hand, it will introduce lexical-semantic relations, thus facilitating its use for NLP tasks in Italian, and, on the other hand, it will make SCDM a structurally and semantically interoperable resource for Italian, to which other lexical-semantic resources (both in Italian, such as PAROLE-SIMPLE-CLIPS (Ruimy et al., 2003), and in English, such as VerbNet (Kipper Schuler, 2005), among others), sense annotated corpora (e.g. the MultiSemCor corpus (Bentivogli and Pianta, 2005)), and Web-based encyclopedia (e.g. Wikipedia) can be connected.

At this stage of development we focused on the alignment of verbs and nouns. The remaining of this paper is organized as follows. Section 2 will state the task and describe the characteristics of the two lexica. In Section 3 some related works and the perculiarities of our work are discussed. The approaches we have adopted are described in Section 4. The evaluation is carried out in Section 5, including an error analysis. Finally, in Section 6 conclusions and future works are reported.

2 Problem Description and Resources

Following (Matuschek and Gurevych, 2013), word sense alignment (WSA) can be formally defined as a list of pairs of senses from two lexical-semantic resources. A pair of aligned senses denotes the same meaning. For instance, taken the two senses of the word "*day*" "`amount of hours of work done in one day`" and "`the recurring hours established by contract or usage for work`" (taken from translated SCDM and MWN, respectively), they must be aligned as they are clearly equivalent.

2.1 MultiWordNet

MWN is a computational multilingual lexicon perfectly aligned to Princeton WN 1.6. As in WN, concepts are organized in synonym sets (*synsets*) which are hierarchically connected by means of hypernym relations (*is_a*). Additional semantic relations such as meronymy, troponymy, nearest synonym and others are encoded as well. The Italian section of MWN is composed of 38,653 synsets, with 4,985 synsets for verbs and 28,517 synsets for nouns. Each synset is accompanied by a gloss describing its meaning and, when present, one or more examples of use. Only 3,177 glosses (8,21%) are in Italian and, in particular, 402 for verbs and 2,481 for nouns.

2.2 Senso Comune De Mauro

The SCDM lexicon is part of a larger research initiative, *Senso Comune*[1] (Oltramari et al. (2013)).

[1] `http://www.sensocomune.it`

Senso Comune aims at building an open knowledge base for the Italian language, designed as a crowd-sourced initiative that stands on the solid ground of an ontological formalization and well-established lexical resources. The lexicon entries have been obtained from the De Mauro GRADIT dictionary and consists in the 2,071 most frequent Italian words, for a total of 11,939 fundamental senses. As for verbs we have 3,827 senses, corresponding to 643 lemmas, with an average polysemy of 5.9 senses per lemma. As for nouns we have 4,586 senses, corresponding to 1,111 lemmas with an average polysemy of 4.12 senses per lemma. In SCDM, word senses are encoded following lexicographic principles and are associated with lexicographic examples of usage.

Senso Comune comprises three modules: i.) a top level module for basic ontological concepts; ii.) a lexical module for linguistic and lexicographic structures; and iii.) a frame module for modeling the predicative structure of verbs and nouns. The top level ontology is inspired by DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) (Masolo et al., 2002). All nominal entries have been manually classified according to the ontological concepts and an ontological classification of verb entries will start in the near future. With respect to MWN, word senses are not hierarchically structured and no semantic relation is encoded. Senses of polysemous entries have a flat representation, one following the other.

3 Related Works

Previous works in word sense alignment can be divided into two main groups: a.) approaches and frameworks which aim at linking lexica based on different models to WN synsets (Rigau and Eneko (1995); Navigli (2006); Roventini et al. (2007)) or language resources, such as Wikipedia (Ruiz-Casado et al. (2005); Mihalcea (2007); Niemann and Gurevych (2011)), and b.) approaches towards the merging of different language resources (Gurevych et al. (2012); Navigli and Ponzetto (2012)). Our work clearly fits into the first group. While different methods are employed (similarity-based approaches *vs.* graph-based approaches), common elements of these works are: i.) the extensive use of lexical knowledge based on the sense descriptions such as the WN glosses or an article first paragraph as in the case of Wikipedia; and ii.) the extension of the basic sense descriptions with additional information such as hypernyms for WN entries, domains labels or categories for dictionaries or Wikipedia entries so as to expand the set of available information, thus improving the quality of the alignments.

As for our task, the most similar work is (Navigli, 2006) where entries from a lexicographic dictionary, namely the Oxford English Dictionary (OED), are mapped to WN. The author adopts and compares two methods: a.) a pure lexical matching function based on the notion of lexical overlap (Lesk, 1986) of the lemmas in the sense descriptions; and b.) a semantic matching based on a knowledge-based WSD system, Structural Semantic Interconnections (SSI), built upon WN and enriched with collocation information representing semantic relatedness between sense pairs. In this latter approach, first each sense description in WN and in the OED is disambiguated by means of SSI with respect to the WN sense inventory, thus obtaining a semantic description as a bag of concepts. Then, two senses are matched if a relation edge is identified between the concepts in the description of each sense in the two lexica. Both approaches are evaluated with respect to a manually created gold standard. The author reports an overall F1 measure of 73.84% for lexical matching, and of 83.11% for semantic matching.

With respect to the SCDM, the OED has some advantages, namely i.) the distinction between core senses and subsenses for polysemous entries; ii.) the presence of hypernyms explicitly signalled; and iii.) domain labels associated with word senses. Such kind of information is not present in the SCDM where senses are presented as a flat list and no enrichment of the sense descriptions with additional information is available, except for the ontological tagging of nouns. Moreover, the low number of MWN glosses in Italian prevents a straightforward application of state-of-the-art methods for sense alignment. MWN sense descriptions must be built up from other sources. Thus, the main issue we are facing is related to data sparseness, that is how to tackle sense alignment when we have few descriptions in Italian (MWN side) and few meta-data and no structure over senses (SCDM side).

4 Methodology

The automatic alignment of senses has been conducted by applying two approaches for constructing the sense representations of the resources and evaluation.

4.1 Lexical Match

In the first approach, Lexical Match, for each word *w* and for each sense *s* in the given resources *R* $\in$ {MWN, SCDM} we constructed a sense descriptions d_R(s) as a bag of words in Italian. Provided the different characteristics of the two resources, two different types of bag of words have been built. As for the SCDM, the bag of words is represented by the lexical items in the textual definition of s_w, automatically lemmatized and part-of-speech analyzed with the TextPro tool suite (Pianta et al., 2008) with standard stopword removal. On the other hand, for each synset, *S*, and for each part of speech in analysis, the sense description of each MWN synset was built by optionally exploiting:

- the set of synset words in a synset excluding *w*;
- the set of direct hypernyms of *s* in the taxonomy hierarchy in MWN;
- the set of synset words in MWN standing in the relation of *nearest synonyms* with *s*;
- the set of synset words in MWN composing the manually disambiguated glosses of *s* from the "Princeton Annotated Gloss Corpus"[2]. To extract the corresponding Italian synset(s), we have ported MWN to WN 30;
- the set of synset words in MWN composing the gloss of *s* in Italian (when available);
- for verbs, the set of synset words in MWN standing in the relations of *entailment/is_entailed*, *causes/is_caused* with *s*;
- for nouns, the set of synset words in MWN standing in the relations of *part_of/has_part*, *has_member/is_member* with *s*.

The alignment of senses is based on the notion of lexical overlap. We used `Text::Similarity` v.0.09 module[3], and in particular the method `Text::Similarity::Overlaps`, to obtain the overlap value between two bags of words of s_w. Text similarity is based on counting the number of overlapping tokens between the two strings, normalized by the length of the strings.

One of the well known limitation of the Lexical Match approach is the so called "lexical gap" problem (Meyer and Gurevych, 2011), i.e. a reduced number of overlapping words. To overcome this limit, we have exploited a newly developed multilingual resource, BabelNet (Navigli and Ponzetto, 2012), which has been obtained by merging together WN synsets and Wikipedia pages with an accuracy of 83%. It contains 4,683,031 nominal glosses (2,985,243 of which are in English). In BabelNet English WN 3.0 synsets have been aligned to their corresponding Wikipedia pages and then extended to other languages, including Italian, by exploiting Wikipedia language links and WN mappings. As for our task, we have retained only those BabelNet entries which have a corresponding synset word in MWN. In this way, we have extended the bag of words representation of nominal entries for MWN synsets by adding the Italian Wikipedia glosses from BabelNet.

4.2 Sense Similarity

In the second approach, Sense Similarity, the basis for sense alignment is the Personalized Page Rank (PPR) algorithm (Eneko and Soroa, 2009) relying on a lexical-semantic knowledge base model as a graph *G* = (V, E) as available in the UKB tool suite[4]. As knowledge base we have used WN 3.0 extended with the "Princeton Annotated Gloss Corpus". Each vertex *v* of the graph is a synset, and the edges represent semantic relations between synsets (e.g. hyperonymy, hyponymy, etc.). The PPR algorithm ranks the vertices in a graph according to their importance within the set and assigns stronger initial probabilities to certain kinds of vertices in the graph. The result of the PPR algorithm is a vector whose elements denotes the probability for the corresponding vertex that a jumper ends on that vertex if randomly following the edges of the graph.

To obtain the PPR vector for a sense *s* of the

[2] See `http://wordnet.princeton.edu/glosstag.shtml`

[3] `http://www.d.umn.edu/~tpederse/text-similarity.html`

[4] See `http://ixa2.si.ehu.es/ukb/`

SCDM, we have translated the Italian textual definitions in English by means of a state-of-the-art Machine Translation system[5], automatically lemmatized and part-of-speech analyzed with the TextPro tool suite, remove standard stopwords and applied the UKB tool suite. The PPR vector is a thus semantic representation overall the entire WN synsets of the textual definition of s in SCDM.

As for the MWN synsets, we have exploited its conversion to WN 3.0. Instead of building the PPR vector by means of the lexical items, we have passed to the UKB tool suite the WN synset id, thus assuming that the MWN synset is already disambiguated.

Given two PPR vectors, namely ppr_{mwn} and ppr_{scdm} for the MWN synset w_{syn} and for the SCDM sense w_{scdm}, we calculated their cosine similarity. On the basis of the similarity score, the sense pair is considered as aligned or not.

5 Experiments and Evaluation

5.1 Gold Standards

To evaluate the reliability of the two approaches with respect to our data, we developed two different gold standards, one for verbs and one for nouns.

The verb gold standard is composed by 44 lemmas selected according to corpus frequency (highly frequent lemmas in the La Repubblica Corpus (Baroni et al., 2004)) and patterns in terms of semantic and syntactic features[6]. It is composed by 350 aligned sense pairs obtained by manually mapping the MWN synsets to their corresponding senses in the SCDM lexicon. These verbs corresponds to 279 synsets and 424 senses in the SCDM. Overall, 211 of the 279 MWN synsets have a corresponding sense in the SCDM (i.e. SCDM covers 84.22% of the MWN senses in the data set), while 235 out of 424 SCDM senses have a correspondence in MWN (i.e MWN covers 49.76% of the SCDM senses). Average degree of polysemy for MWN entries is 6.34, while for the SCDM is 9.63.

The noun gold standard is composed by 46 lemmas selected according to frequency and polysemy with respect to the fundamental senses in the SCDM (each lemma must have at least two fundamental senses in the SCDM). On the basis of the manual alignment, we have obtained 166 aligned sense pairs. The noun lemmas correspond to 229 synsets and 216 senses in the SCDM. Overall, 134 of the 229 MWN synsets have a corresponding sense in the SCDM (i.e. SCDM covers 53.71% of the MWN senses in the data set), while 123 out of 216 SCDM senses have a correspondence in MWN (i.e MWN covers 62.03% of the SCDM senses). Average degree of polysemy for MWN entries is 4.97, while for the SCDM is 4.69. The difference in terms of coverage with respect to the verbs is clearly due to two aspects, namely i.) the restrictions of the SCDM entries to the fundamental senses; ii.) the higher coverage in terms of nouns synsets of MWN with respect to the verbal ones.

Though small, the size of the gold standards is representative of the two lexica. In particular, the 279 verbs synsets yield 3,319 possible sense pairs, i.e. 11.8 SCDM senses per synset on average. As for nouns, the 229 nominal synsets yield 1,414 sense pairs, i.e. 6.13 SCDM senses on average.

5.2 Results

The evaluation has been performed by computing Precision (the ratio of the correct alignment with respect to all proposed alignments), Recall (the ratio of extracted correct alignment with respect to the alignments in the gold standard), F-measure (the harmonic mean of Precision and Recall calculated as $2PR/P + R$) and Accuracy (the precentage of the correctly identifed alignments and non alignments). As baseline, we have implemented a random match algorithm, *rand*, which for the same word *w* in SCDM and in MWN assigns a random SCDM sense to each synset with *w* as synset word, returning a one-to-one alignment. The selection of the correct alignments has been obtained by applying two types of thresholds with respect to all proposed alignments (the "no_threshold" row in the tables): i.) a simple cut-off at specified values (0.1; 0.2); ii.) the selection of the maximum score (either lesk measure or cosine; row "max_score" in the tables) between each synset *S* and the proposed aligned senses of the SCDM. As for the maximum score threshold, we have retained as good alignments also instances of a tie, thus allowing the possibility of having one MWN synset aligned to more than one SCDM sense.

[5] We use Google Translate API.

[6] A subset of these verbs have been taken from (Jezek and Quochi, 2010)

Lexical Match	P	R	F1	Acc.
Verb SYN - no_threshold	0.41	0.29	**0.34**	0.864
Verb SYN - ≥ 0.1	0.42	0.26	0.32	0.874
Verb SYN - ≥ 0.2	0.54	0.11	0.18	0.901
Verb SYN - max_score	0.59	0.19	**0.29**	**0.909**
Verb SREL - no_threshold	0.38	0.32	**0.35**	0.786
Verb SREL - ≥ 0.1	0.40	0.27	0.32	0.781
Verb SREL - ≥ 0.2	0.53	0.11	0.18	0.863
Verb SREL - max_score	0.60	0.20	**0.30**	**0.908**
Verb - `rand`	0.15	0.06	0.08	

Lexical Match	P	R	F1	Acc
Noun SYN - no_threshold	0.52	0.59	**0.55**	0.885
Noun SYN - ≥ 0.1	0.58	0.41	0.48	0.901
Noun SYN - ≥ 0.2	0.71	0.16	0.26	0.904
Noun SYN - max_score	0.69	0.42	**0.52**	**0.920**
Noun SREL - no_threshold	0.49	0.60	**0.54**	0.877
Noun SREL - ≥ 0.1	0.60	0.40	0.48	0.905
Noun SREL - ≥ 0.2	0.71	0.13	0.22	0.902
Noun SREL - max_score	0.69	0.42	**0.52**	**0.921**
Noun - `rand`	0.17	0.12	0.14	

Table 1: Results for automatic alignment based on Lexical Match for SYN and SREL sense representations.

5.2.1 Lexical Match Results

We have analyzed different combinations of the sense representation of a synset. We developed two basic representations: SYN, which is composed by the set of synset words excluding the target word *w* to be aligned, all of its direct hypernyms, the set of synset words in MWN standing in the relation of *nearest synonyms* and the synset words obtained from the "Princeton Annotated Gloss Corpus"; and SREL, which contains all the items of SYN plus the the synset words included in the selected set of semantic relations. The results are reported in Table 1.
As the figures show, all synset configurations outperform the baseline `rand` for both parts of speech in analysis. However, it is interesting to observe that the alignment of noun senses performs much better than that for verbs in both sense representations and with all filtering methods. On the basis of the alignment method (i.e. lexical overlap) such a difference in performance provides interesting data on the two resources in analysis. A manual exploration of the data in the configurations both for verbs and nouns has highlighted that, on the one hand, we suffer from data sparseness on the SCDM side as no extension of the sense description of the glosses is possible, and, on the other hand, that senses are described in ways that are semantically equivalent but with different lexical items.

As for verbs the Recall with no filtering (no_threshold) has extremely low levels, ranging from 0.32 for SREL to 0.29 for SYN. The SREL sense representation outperforms SYN when no filtering is applied only in terms of Recall (+0.03), thus signaling that the additional semantic relations play a very limited role in the description of verb senses without providing real additional information to match data in the SCDM glosses. Furthermore, the difference in performance of the SREL configuration is not statistically significant with respect to the SYN configuration ($p > 0.05$).

The situation looks different for nouns where, although low, the no threshold Recall values range between 0.60 (SREL) to 0.59 (for SYN). As for the two basic configurations, SYN and SREL, the results show that SYN is more accurate and that the impact of additional semantic relations, though it slightly improves the Recall, is not statistically signiticant ($p > 0.05$).

Both for verbs and nouns we decided to select the SYN basic configuration as the best sense representation because it has a simpler bag-of-words and better Precision. To improve the results, we have extended this basic representation with the lexical items in the corresponding glosses of BabelNet (+BABEL) (only for nouns) and the lexical items of the MWN Italian glosses (+IT) (for verbs and nouns)[7]. The results are illustrated in Table 2.

In both cases, the extension of the basic sense representations with additional data is positive, namely for Recall. Notice that for verbs the presence of Italian MWN glosses improves the alignment results (for the no-threshold filter, F1=0.37 vs. F=0.35 for SREL and F1=0.34 for SYN) as they introduce information which better represents the sense definition than the synset words in the bag of words representations and overcomes missing information in the WN 3.0 annotated glosses. For instance, consider the following example for the verb "*rendere*" [to make]. In example 1a) the two senses are aligned with a very low lexical overlap score as there is only one word in com-

[7] The Italian MWN glosses for the items in the Golds are present for 24% senses of verbs and 30% senses of nouns, respectively

Lexical Match	P	R	F1
Verb SYN+IT - no_threshold	0.36	0.38	**0.37**
Verb SYN+IT - ≥ 0.1	0.38	0.31	0.34
Verb SYN+IT - ≥ 0.2	0.51	0.13	0.20
Verb SYN+IT - max_score	**0.63**	0.23	**0.34**
Noun SYN+BABEL - no_threshold	0.47	0.66	**0.56**
Noun SYN+BABEL - ≥ 0.1	0.58	0.40	0.47
Noun SYN+BABEL - ≥ 0.2	0.69	0.12	0.21
Noun SYN+BABEL - max_score	**0.69**	0.44	**0.55**
Noun SYN+BABEL+IT - no_threshold	0.47	0.66	**0.55**
Noun SYN+BABEL+IT - ≥ 0.1	0.53	0.43	0.48
Noun SYN+BABEL+IT - ≥ 0.2	0.71	0.18	0.28
Noun SYN+BABEL+IT - max_score	**0.66**	0.45	**0.54**

Table 2: Results for Lexical Match alignment with extensions with BabelNet data and MWN Italian glosses.

mon ("*fare*"), while in 1b) the presence of the Italian glosses in the synset sense increases the lexical match score as it matches both words in the gloss in the SCDM. The lexical items of the sense descriptions are reported in Italian, matching words are in bold.

1a. **fare** essere mettere [synset_id v—00080274]
fare diventare [SCDM_id 243356]

1b. **fare** essere mettere **diventare** [synset_id v—00080274]
fare diventare [SCDM_id 243356]

The positive effect of the original Italian data for verbs points out a further issue for our task, namely that the derivation of sense representations of MWN synsets by means of synset words (including the sense annotated glosses of WN 3.0) is not as powerful as having at disposal original glosses.

Similarly, for nouns we register an improvement in Recall at a low or null cost for Precision for all filtering methods, with the exclusion of the no threshold filtering. Precision for SYN+BABEL+IT with maximum score filtering is lowered with respect to the extension with the BabelNet data only (P=0.66 for SYN+BABEL+IT *vs.* P=0.69 for SYN+BABEL)[8]. To better clarify these results, consider the following example for the noun "*palla*" [ball]. In the example 2a) the two senses are not aligned as there are no matching words, while in 2b) the extension by means of the BabelNet data provides a sufficient number of matching items for aligning the two senses. As for the previous example, the lexical items of the sense descriptions are reported in Italian, matching words are in bold.

2a. pallone oggetto cosa balocco partita battere bocciare circolare rotondo tondo [synset_id n—02240791]
sfera dimensione variabile materiale diverso cuoio gomma avorio pieno gonfiare aria usare numeroso gioco sport [SCDM_id 241637]

2b. pallone oggetto cosa balocco partita battere bocciare circolare rotondo tondo palla essere oggetto sferico **usare** vario **sport gioco** esempio calcio pallacanestro pallavolo biliardo bowling [synset_id n—02240791]
sfera dimensione variabile materiale diverso cuoio gomma avorio pieno gonfiare aria **usare** numeroso **gioco sport** [SCDM_id 241637]

Concerning the filtering of the proposed alignments, the maximum score filter provides the best results for Precision at a low cost in terms of Recall, with F1 scores for verbs ranging from 0.34 (SYN+IT) to 0.29 (SYN), and from 0.55 (SYN+BABEL) to 0.52 (SYN and SREL) for nouns. It is interesting to point out a further difference in performance between verbs and nouns. In particular, for verbs we can observe that the filtering based on maximum score has lower F1 values with respect to the no threshold baseline in all sense descriptions. As for nouns, on the contrary, both the two basic sense descriptions, SYN and SREL, and the SYN+BABEL configuration have comparable F1 values between the no threshold and the maximum score data. Nevertheless, the filtering based on the maximum score improves the quality of the proposed alignment by removing lots of false positives both for verbs and nouns (for verbs P=0.59 for SYN, P=0.60

[8] Excluding the BabelNet data and running the alignment only with the Italian glosses, SYN+IT, with maximum score filtering, gives F1=0.52 which is the same as SYN and SREL, and lower that SYN+BABEL.

for SREL, and P=0.63 for SYN+IT; for nouns, P=0.69 for SYN, SREL, and SYN+BABEL, P=0.66 for SYN+BABEL+IT) without impacting on the number of good instances retrieved (for verbs R=0.19 for SYN, R=0.20 for SREL, and R=0.23 for SYN+IT; for nouns R=0.42 for SYN and SREL, R=0.44 for SYN+BABEL; R=0.45 for SYN+BABEL+IT).

5.2.2 Similarity Measure Results

The results for the Similarity Measure obtained from the Personalized Page Rank algorithm on the basis of the vectors described in Section 4.2 are illustrated in Table 3.

Similarity Measure	P	R	F1
Verb - no_threshold	0.10	0.9	0.19
Verb - $\geq$ 0.1	0.47	0.25	**0.32**
Verb - $\geq$ 0.2	**0.66**	0.16	0.26
Verb - max_score	0.42	0.20	0.27
Verb - `rand`	0.15	0.06	0.08
Noun - no_threshold	0.12	0.94	0.21
Noun - $\geq$ 0.1	0.52	0.32	**0.40**
Noun - $\geq$ 0.2	**0.77**	0.21	0.33
Noun - max_score	0.42	0.38	0.40
Noun - `rand`	0.17	0.12	0.14

Table 3: Results for automatic alignment based on Similarity Score.

Similarly to the Lexical Match, the Personalized Page Rank approach outperforms the baseline `rand`. Overall, the differences in performance with the Lexical Match results are not immediate. In general, as the Recall values for no threshold filtering show, almost all aligned sense pairs of the gold are retrieved, outperforming the Lexical Match. Clearly, this difference is strictly related to the different nature of the sense descriptions, i.e. a *semantic* representation based on a lexical knowledge graph, which is able to catch semantically related items out of the scope for the Lexical Match approach.

By observing the figures for verbs, we notice that the simple cut-off thresholds provide better results with respect to the maximum score. The best F1 score (F1=0.32) is obtained when setting the cosine similarity to 0.1, though Precision is less than 0.50 (namely, 0.47). When compared with threshold value of 0.1 of the Lexical Match, the Personalized Page Rank method yields the best Precision (P=0.47 *vs.* P=0.42 for Verb SYN, P=0.38 for Verb SYN+IT, and P=0.40 for Verb SREL). Similar observations can be done when the threshold is set to 0.2. In this latter case, Personalized Page Rank yields the best Precision score for verbs with respect to all other filtering methods and the Lexical Match results obtained with maximum score (P=0.66 *vs.* P=0.59 for Verb SYN, P=0.63 for Verb SYN+IT, and P=0.60 for Verb SREL).

The analysis for nouns is more complex. Apparently, the Personalized Page Rank approach has lower F1 scores with respect to all Lexical Match sense configurations and filtering methods, including the no threshold score of the basic sense descriptions (respectively, F1=0.55 for SYN, F1=0.54 for SREL, F1=0.21 for Personalized Page Rank). However, when maximizing Precision for the Personalized Page Rank (threshold 0.2), the algorithm provides better performances (F1=0.33) with respect to Lexical Match on the same filtering method, minimizing the drop of Recall (R=0.21; +0.09 with respect to SYN+BABEL with same threshold; + 0.08 with respect to SREL; +0.05 with respect to SYN, respectively).

The better performance of the simple cut-off thresholds with respect to the maximum score is due to the fact that aligning senses by means of semantic similarity provides a larger set of alignments and facilitates the identification of multiple alignments, i.e. one-to-many.

5.2.3 Merging Lexical Match and Sense Similarity

As the two approaches are different in nature both with respect to the creation of the sense descriptions (simple bag of words *vs.* semantic representation) and to the methods with which the alignment pairs are extracted and computed, we have developed a further set of experiments by merging together the results obtained from the best sense descriptions and best filtering methods for Lexical Match and Semantic Similarity. As parameters for the identification of the best results we have taken into account the Precision and F1 values. Excluding the presence of Italian data from the sense descriptions of the Lexical Match approach due to their sparseness, we selected the SYN sense description filtered with maximum score for verbs (P=0.59, F1=0.29) and the SYN+BABEL sense description filtered with maximum score for nouns (P=0.69; F1=0.55). As for the Personalized Page Rank approach, we have selected both for verbs and nouns the cut-off threshold at 0.2. The results are reported in Table 4.

Merged	P	R	F1
Verb - SYN+ppr02	0.61	0.38	**0.47**
Noun - SYN+BABEL+ppr02	0.67	0.61	**0.64**

Table 4: Results for automatic alignment merging the best results from Lexical Match and Sense Similarity.

The combination of the best results yields the best performance for both parts of speech compared to the stand-alone approaches. In particular, for verbs we obtain an F1=0.47, with an improvement of 0.18 points with respect to SYN and of 21 points with respect to Personalized Page Rank with threshold 0.2. Similar improvements can be observed for nouns, where SYN+BABEL+ppr02 has an F1=0.64, with an improvement of 9 points with respect to SYN+BABEL and of 31 points with respect to Personalized Page Rank with threshold 0.2. In both cases the performance gains originate from the higher precision of the Personalized Page Rank approach which minimizes the data sparseness of the SCDM lexicon.

6 Conclusion and Future Work

This paper focuses on the automatic alignment of senses from two different resources when few data are available. In particular, the lack of Italian glosses in MWN and the absence of any kind of structured information in the SCDM dictionary posed a serious issue for the application of state-of-the-art techniques for sense alignment.

We experimented with two different approaches: Lexical Match and Sense Similarity obtained from Personalized Page Rank. In all cases, when filtering the data we are facing low scores for Recall which point out issues namely related to data sparseness in our lexica. By comparing the results of the two approaches, we can observe that: i.) the Personalized Page Rank yields the best Precision with respect to Lexical Match; ii.) Lexical Match, with a simple sense description configuration (i.e. the SYN configurations for verbs and nouns), is still a powerful approach for this kind of tasks; the exploitation of additional semantically related items (e.g. SREL for verbs) or additional sense descriptors (e.g. SYN+BABEL for nouns), though good in principle, has a limited contribution to solve the "lexical gap" problem in our case and points out differences in the way word senses are encoded in the two lexica; and iii.) Personalized Page Rank vectors and Lexical Match appears to qualify as complementary methods for achieving reliable sense alignments, namely when dealing with few data. Our approach provides satisfying results both for verb and noun sense alignment, with an overall F1=0.47 for verbs and an F1=0.64 for nouns. The better results for nouns are strictly related to the definitions of the senses which mainly relies on synonym words and hypernyms. On the other hand, verbs tend to have more abstract definitions and the contribution of additional semantic relations (i.e. the SREL configuration) is poor.

Future work will concentrate on two aspects by exploiting the sense alignment results. The aligned sense pairs will be used for sense clustering as a strategy to reduce the sense descriptions in MWN and in SCDM. Existing clustering of WN senses (e.g. Navigli (2006)) will be used as a starting point and for subsequent evaluation. Furthermore, we aim at importing the ontological classes of SCDM in MWN. This aspect will be useful for the identification of possible taxonomical errors in the MWN hierarchy and boostrap better sense alignments.

References

Marco Baroni, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston, and Marco Mazzoleni. 2004. Introducing the "`la Repubblica`" corpus: A large, annotated, `TEI(XML)`-compliant corpus of newspaper italian. In *Proceedings of the Fourth International conference on Language Resources and Evaluation (LREC-04)*.

Luisa Bentivogli and Emanuele Pianta. 2005. Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus. *Natural Language Engineering*, 11:247–261, 8.

Agirre Eneko and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009)*, Athens, Greece.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. Uby - a large-scale unified lexical-semantic resource based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012).*

Elisabetta Jezek and Valeria Quochi. 2010. Capturing coercions in texts: a first annotation exercise. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 1464–1471, Valletta, Malta. European Language Resources Association (ELRA).

Karin Kipper Schuler. 2005. *Verbnet: a broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, Philadelphia, PA, USA. AAI3179808.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine code from an ice cream cone. In *Proc. of 5th Conf. on Systems Documentation*. ACM Press.

Claudio Masolo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari, and Luc Schneider. 2002. Wonderweb deliverable D17: the wonderweb library of foundational ontologies. Technical report.

Michael Matuschek and Iryna Gurevych. 2013. Dijkstra-wsa: A graph-based approach to word sense alignment. *Transactions of the Association for Computational Linguistics (TACL)*, 2:to appear.

Michael Meyer and Iryna Gurevych. 2011. What psycholinguists know about chemistry: Aligning Wiktionary and WordNet for increased domain coverage. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*.

Rada Mihalcea. 2007. Using Wikipedia for automatic word sense disambiguation. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, Rochester, New York.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Roberto Navigli. 2006. Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 44^{th} Annual Meeting of the Association for Computational Linguistics joint with the 21^{st} International Conference on Computational Linguistics (COLING-ACL)*, Sydney, Australia.

Elisabeth Niemann and Iryna Gurevych. 2011. The peoples web meets linguistic knowledge: Automatic sense alignment of Wikipedia and WordNet. In *Proceedings of the 9th International Conference on Computational Semantics*, pages 205–214, Singapore, January.

Alessandro Oltramari, Guido Vetere, Isabella Chiari, Elisabetta Jezek, Fabio Massimo Zanzotto, Malvina Nissim, and Aldo Gangemi. 2013. Senso Comune: A collaborative knowledge resource for italian. In I. Gurevych and J. Kim, editors, *The Peoples Web Meets NLP, Theory and Applications of Natural Language Processing*, pages 45–67. Springer-Verlag, Berlin Heidelberg.

Emanuele Pianta, Luisa Bentivogli, and Cristian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *First International Conference on Global WordNet*, Mysore, India.

Emanuele Pianta, Cristian Girardi, and Roberto Zanoli. 2008. TextPro Tool Suite. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-08)*, volume CD-ROM, Marrakech, Morocco. European Language Resources Association (ELRA).

German Rigau and Agirre Eneko. 1995. Disambiguating bilingual nominal entries against WordNet. In *Proceedings of workshop The Computational Lexicon, 7th European Summer School in Logic, Language and Information*, Barcelona, Spain.

Adriana Roventini, Nilda Ruimy, Rita Marinelli, Marisa Ulivieri, and Michele Mammini. 2007. Mapping concrete entities from PAROLE-SIMPLE-CLIPS to ItalWordNet: Methodology and results. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague, Czech Republic, June.

Nilda Ruimy, Monica Monachini, Elisabetta Gola, Nicoletta Calzolari, Maria Cristina Del Fiorentino, Marisa Ulivieri, and Sergio Rossi. 2003. A computational semantic lexicon of italian: *SIMPLE. Linguistica Computazionale XVIII-XIX, Pisa*, pages 821–64.

Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets. In *Proceedings of the Third international conference on Advances in Web Intelligence*, AWIC'05, Berlin, Heidelberg. Springer-Verlag.