Aligning Verb Senses in Two Italian Lexical Semantic Resources

Tommaso Caselli
Trento RISE
Via Sommarive, 18
Povo I-38123
t.caselli@trentorise.eu

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

Strapparava Carlo
HLT-FBK
Via Sommarive, 18
Povo I-38123
strappa@fbk.eu

Guido Vetere
IBM - CAS Trento
P.zza Manci, 17
Povo I-38123
gvetere@it.ibm.com

Abstract

This work describes the evaluations of three different approaches, Lexical Match, Sense Similarity based on Personalized Page Rank, and Semantic Match based on Shallow Frame Structures, for word sense alignment of verbs between two Italian lexical-semantic resources, MultiWordNet and the Senso Comune Lexicon. The results obtained are quite satisfying with a final F1 score of 0.47 when merging together Lexical Match and Sense Similarity.

1 Introduction

Lexical-semantic resources play a key role in many Natural Language Processing tasks, such as Word Sense Disambiguation, Information Extraction, and Question-Answering, among others. The creation of lexical-sematic resources is costly in terms of manual efforts and time, and often important information is scattered in different lexica and difficult to use. Semantic interoperability between resources could represent the viable solution to allow reusability and develop more robust and powerful resources. Word sense alignment (WSA), a research area which has seen an increasing interest in recent years, qualifies as the preliminary requirement for achieving this goal (Matuschek and Gurevych, 2013).

The purpose of this work is to merge two Italian lexical-semantic resources, namely MultiWordNet (Pianta et al., 2002) (MWN) and Senso Comune Lexicon (SCL) (Oltramari et al., 2013), by automatically linking their entries. The final result will be two-folded. On the MWN side, this will provide Italian with a more complete and robust version of this lexicon. On the SCL side, the linking with MWN entries will introduce lexical-semantic relations, thus facilitating its use for NLP tasks in Italian, and it will make SCL a structurally and semantically interoperable resource for Italian, allowing its connection to other lexical-semantic resources, sense annotated corpora (e.g. the MultiSemCor corpus (Bentivogli and Pianta, 2005)), and Web-based encyclopedia (e.g. Wikipedia).

This work will focus on our experience on the alignment of verb senses. 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. Finally, in Section 6 conclusions and future work are reported.

2 Task and Resources

Following (Matuschek and Gurevych, 2013), WSA can be defined as the identification of pairs of senses from two lexical-semantic resources which denote the same meaning. For instance, the two senses of the verb "*love*", "`feel love or affection for someone`" and "`have a great affection or liking for`" (taken from translated SCL and MWN, respectively), must be aligned as they are clearly equivalent in meaning.

2.1 MultiWordNet

MWN is a 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. Each synset is accompanied by a gloss describing its meaning and, when present, one or more examples of use. Overall 3,177 glosses (8,21%) are in Italian and, in particular, 402 for verbs.

2.2 Senso Comune Lexicon

SCL is part of a larger research initiative (Oltramari et al. (2013)) which aims at building an open knowledge base for the Italian language. The lexicon entries have been obtained from the De Mauro GRADIT dictionary and consist 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. In SCL, word senses are encoded following lexicographic principles and are associated with lexicographic examples of usage.

SCL comprises three modules: i.) a 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). 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 has been encoded yet. Senses of polysemous entries have a flat representation, one following the other.

3 Related Works

Previous works in WSA can be divided into two main groups: a.) approaches and frameworks which aim at linking entries to WN from lexica based on different models (Rigau and Agirre (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 (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 the lexical knowledge of the sense descriptions; e.g. 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 to expand the set of available information, thus improving the quality of the alignments. The large of these works focuses on aligning noun senses. The only work which also tackles verb sense alignment is (Navigli, 2006) where entries from the Oxford English Dictionary (OED) are mapped to WN. The author explores 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. 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 (accuracy 66.08%), and of 83.11% for semantic matching (accuracy 77.94%). Alignment performances on single parts of speech are not reported.

With respect to the SCL, 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 SCL where senses are presented as a flat list and no enrichment of the sense descriptions with additional information is available. 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 in different ways. Summing up, 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 structuration over senses (SCL side).

4 Methodology

The automatic alignment of senses has been conducted by applying three approaches Lexical Match, Sense Similarity and Semantic Match.

4.1 Lexical Match

In the first approach, Lexical Match, for each word w and for each sense s in the given resources $R \in \{\text{MWN, SCL}\}$ we constructed a sense description

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 SCL, 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*, 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 (if available);
- the set of synset words in MWN standing in the relation of *nearest synonyms* with *s* (if available);
- the set of synset words in MWN composing the manually disambiguated glosses of *s* from the "Princeton Annotated Gloss Corpus"[1]. To extract the corresponding Italian synset(s), we have ported MWN to WN 3.0;
- the set of synset words in MWN composing the gloss of *s* in Italian (when available);
- the set of synset words in MWN standing in the relations of *entailment/is_entailed*, *causes/is_caused* with *s*;

The alignment of senses is based on the notion of lexical overlap. We used the `Text::Similarity` v.0.09 module[2] to obtain the overlap value between two bags of words. Text similarity is based on counting the number of overlapping tokens between the two strings, normalized by the length of the strings.

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[3]. 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 denote 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 SCL, we translated the Italian textual definitions in English by means of a state-of-the-art Machine Translation system[4], automatically lemmatized and part-of-speech analyzed with the TextPro tool suite, removed standard stopwords, and applied the UKB tool suite. The PPR vector is, thus, a semantic representation overall the entire WN synsets of the textual definition of *s* in SCL.

As for the MWN synsets, instead of building the PPR vector by means of the lexical items composing the sense description, 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 SCL sense w_{scdm}, we calculated their cosine similarity. On the basis of the similarity score, the sense pair is considered as aligned or not.

4.3 Semantic Match: Exploiting Shallow Frames Structures

On the basis of (Roland and Jurafsky, 2002) and current research activities in Senso Comune (Chiari et al., 2013), we assume as working hypothesis that different verb senses tend to correlate with different shallow frame patterns. Thus, we consider two verb senses to be aligned if the shallow frames structures (SFS) of their examples of use are the same. We assume as a SF structure the syntactic complements of the verb, with no distinction between arguments and adjuncts, and the semantic type of the complement filler(s). An example of an SFS is reported in example 1.

1. *Marco ha comprato un libro.*
   [Marco bought a book.]
   Verb: *comprare* [to buy]
   SFS: SUBJ[person] OBJ[artifact]

To obtain the SFSs, two different strategies have been used. For the SCL, we have extracted all

[1] http://wordnet.princeton.edu/glosstag.shtml

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

[3] http://ixa2.si.ehu.es/ukb/

[4] We use Google Translate API.

the lexicographic examples of use associated to each verb sense. For MWN, to recover a larger number of examples of use in Italian, we have exploited the data in the MultiSemCor corpus v1.0, a parallel corpus of English and Italian annotated with WN senses. For each sense annotated verb in the Italian section of MultiSemCor, we extracted all available corpus-based examples and obtain the SFS to be compared with the SCL instances. The acquisition of the SFSs has been obtained as follows:

- the SCL examples and the MultiSemCor data have been parsed with a state-of-the-art dependency parser (Attardi and Dell'Orletta, 2009);
- for each verb, we have automatically extracted all syntactic complements standing in a dependency relation of argument or complement, together with the lemma of the slot filler;
- nominal lemmas of syntactic complements have been automatically assigned with one of the 26 semantic types composing the WN supersenses (i.e. *noun.artifact*; *noun.object* etc. (Ciaramita and Johnson, 2003)) on the line of (Lenci et al., 2012). For each nominal filler, we selected the most frequent WN supersense. Sense frequency had been computed on the basis of MultiSemCor. In case a polysemous noun lemma was not present in the MultiSemCor data or its senses have the same frequency, all associated WN supersenses were assigned. As for verbal fillers, we assigned the generic semantic type of "*verb.eventuality*". Finally, in case a lemma filler of a syntactic complement is not attested in MWN such as a pronoun or a missing synset word, no values is assigned and the SFS is excluded from the possible matches. Optionally, when the noun filler was annotated with a synset in MultiSemCor, we have associated it to its corresponding WN supersense.

To clarify how this type of sense alignment works, consider the data in example 2. In 2a., we report the SFSs for the examples of use associated with the sense "`vivere abitualmente in un luogo`" [to live habitually in a place] of the verb "*abitare*' [to live] in the SCL. In 2b., we report the SFSs extracted from the MultiSemCor corpus for the MWN synset v#01809405, with gloss "`make one's home or live in`"[5].

2a. COMP-PREP$_{IN}$ [noun.location].
COMP-PREP$_{CON}$ [noun.group]
COMP-PREP$_{A}$ [noun.location]

2b. COMP-PREP$_{DA}$ [noun.person]
SUBJ[noun.person] COMP-PREP$_{DA}$[noun.group]
COMP-PREP$_{IN}$ [noun.location]

By comparing the SFSs, the COMP-PREP$_{IN}$ [noun.location] structure is the same in both senses, thus pointing to the alignment of the two entries.

5 Experiments and Evaluation

5.1 Gold Standard

To evaluate the reliability of the approaches with respect to our data, we developed a gold standard. The gold standard is composed by 44 lemmas selected according to frequency and patterns in terms of semantic and syntactic features[6]. It is composed by 350 sense pairs obtained by manually mapping the MWN synsets to their corresponding senses in the SC lexicon. These verbs correspond to 279 synsets and 424 senses in the SCL. Overall, 211 of the 279 MWN synsets have a corresponding sense in the SCL (i.e. SCL covers 84.22% of the MWN senses in the data set), while 235 out of 424 SCL senses have a correspondence in MWN (i.e MWN covers 49.76% of the SCL senses). Average degree of polysemy for MWN entries is 6.34, while for the SCL is 9.63.

5.2 Results

The evaluation is based on 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), and F-measure (the harmonic mean of Precision and Recall calculated as $2PR/P+R$). As baseline, we implemented a random match algorithm, *rand*, which for the same word w in SCL and in MWN assigns a random

[5] No Italian gloss available for this synset.

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

SCL sense to each synset with *w* as synset word, returning a one-to-one alignment. For the Lexical Match and Sense Similarity approaches, 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 overlap measure or cosine; row "max_score" in the tables) between each synset *S* and the proposed aligned senses of the SCL. For the maximum score threshold, we retained as good alignments also instances of a tie, allowing the possibility of having one MWN synset aligned to more than one SCL sense.

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 synset words included in the selected set of semantic relations. The results are reported in Table 1.

Lexical Match	P	R	F1
SYN - no_threshold	0.41	0.29	0.34
SYN - ≥ 0.1	0.42	0.26	0.32
SYN - ≥ 0.2	0.54	0.11	0.18
SYN - max_score	0.59	0.19	**0.29**
SREL - no_threshold	0.38	0.32	**0.35**
SREL - ≥ 0.1	0.40	0.27	0.32
SREL - ≥ 0.2	0.53	0.11	0.18
SREL - max_score	0.60	0.20	**0.30**
`rand`	0.15	0.06	0.08

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

Both sense configurations, SYN and SREL, outperform the baseline `rand`. However, the Recall with no filtering (no_threshold) has extremely low levels, ranging from 0.32 for SREL to 0.29 for SYN, pointing out that the two resources use different ways to encode the verb senses. Globally, the SREL sense representation does not perform better than SYN. When no filtering is applied the SREL configuration has an improvement in the Recall (+0.03) but not in Precision (-0.03), signaling that the semantic relations have a limited role in the description of verb senses and for identifying key information encoded in the SCL glosses. The difference in performance of the SREL configuration is not statistically significant with respect to the SYN configuration ($p > 0.05$). Provided this limited effect of the extended semantic relations, we have decided to select the SYN configuration as the best since it is simpler and with better values for Precision.

To improve the results, we have extended the SYN basic representations with the lexical items of the MWN Italian glosses (+IT)[7]. The results are illustrated in Table 2.

Lexical Match	P	R	F1
SYN+IT - no_threshold	0.36	0.38	0.37
SYN+IT - ≥ 0.1	0.38	0.31	0.34
SYN+IT - ≥ 0.2	0.51	0.13	0.20
SYN+IT - max_score	0.63	0.23	**0.34**
`rand`	0.15	0.06	0.08

Table 2: Results for Lexical Match alignment adding the Italian MWN glosses.

The extension of the basic sense representations with additional data is positive. In particular, it improves the alignment (for the no-threshold results, 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 and overcomes missing information in the WN 3.0 annotated glosses. The positive effect of the original Italian data 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 equivalent to having at disposal the original glosses.

Concerning the filtering methods, the maximum score filter provides the best results for Precision at a low cost in terms of Recall, with F1 scores ranging between 0.34 (SYN+IT) to 0.29 (SYN).

5.2.2 Sense Similarity Results

The results for the Sense Similarity obtained from the Personalized Page Rank algorithm are illustrated in Table 3.

Similarly to the Lexical Match, the Sense Similarity approach outperforms the baseline `rand`. Overall, the differences in performance with the

[7]The Italian MWN glosses for the items in the Golds are present for 24% senses of the verbs

Semantic Match	P	R	F1
Most Frequent Sense	0.21	0.05	0.08
Most Frequent + Correct Sense	0.33	0.05	**0.09**
Most Frequent + Correct + Vector Similarity	0.34	0.02	0.04
`rand`	0.15	0.06	0.08

Table 4: Results for Semantic Match experiments.

Similarity Measure	P	R	F1
PPR - no_threshold	0.10	0.9	0.19
PPR - $\geq$ 0.1	0.47	0.25	**0.32**
PPR - $\geq$ 0.2	**0.66**	0.16	0.26
PPR - max_score	0.42	0.20	0.27
`rand`	0.15	0.06	0.08

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

Lexical Match results are not immediate. In general, as the Recall value for no threshold filtering shows, almost all aligned sense pairs of the gold are retrieved, outperforming the Lexical Match approach. This difference is 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, we can 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, Sense Similarity 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, Sense Similarity yields the best Precision score 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 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 alignment pairs and facilitates the identification of multiple alignments, i.e. one-to-many.

5.2.3 Semantic Match Results

In Semantic Match we ran three different experiments, namely Most Frequent Sense, where the assignment of the semantic type of the SF slot fillers is based on the most frequent sense; Most Frequent + Correct Sense, where the assignment of the semantic type of the SF slot fillers is based on the most frequent sense and on the annotated sense for the MultiSemCor data, where available, and, finally, Most Frequent + Correct + Vector Similarity, where the assignment of the semantic type of the SF slot fillers is the same as in Most Frequent + Correct Sense plus an additional filtering for nominal SF fillers based on the vector pair WN similarity measure implemented in the `WordNet::Similarity` package[8].
The results obtained are disappointing. With the exception of Precision, all experiment configurations obtain Recall values lower than the baseline `rand`, suggesting that this approach, though linguistically and theoretically sound, suffers from serious flaws. Both Lexical Match and Sense Similarity outperforms this methods even when no filtering is applied.

For this approach, the low levels for Precision and Recall cannot be explained by means of "lexical gaps" or filtering methods. On the basis of manual analysis of the false negative and false positive data, we could claim that the main reasons for these results are due to:

- the reduced number of examples of in the SCL and their nature as "lexicographic" examples of use;
- the high variability in the syntactic realizations of the complements;
- missing annotated senses in the MultiSemCor corpus;
- parsing errors; and
- the difficulty in acquiring complete SFSs from the MultiSemCor data due to the pres-

[8]http://wn-similarity.sourceforge.net/

ence of SF slot fillers realized by pronouns whose assigment of the semantic type depends on their (anaphoric) resolutions.

In addition to this, the low levels of Precision are also due to the coarse-grained categories of the semantic types of the nominal slot fillers. For instance, the SCL examples of use of two different fundamental senses of the verb "*aprire*" [to open], namely "*aprire il rubinetto*" [to open the tap] and "*aprire la porta*" [to open the door] were all wrongly mapped to the same MWN synset, i.e. v#00920424 "*cause to open or to become open; "Mary opened the car door"*". To keep these senses separated, finer-grained semantic features for describing the semantic types of their nominal fillers, here both "noun.artifact", should be employed. The use of vector pairs WN similarity is an attempt into this direction which, however, resulted unsuccesful.

5.2.4 Merging the Approaches

As the three approaches are different in nature both with respect to the creation of the sense descriptions (simple bag of words *vs.* semantic representation *vs.* frame structures) and to the methods with which the alignment pairs are extracted, we have developed a further set of experiments by merging together the results obtained from the Lexical Match, Sense Similarity, and Semantic Match. As parameters for the identification of the best results we have taken into account the Precision and F1 values. We have excluded the presence of Italian data from the sense descriptions of the Lexical Match approach due to their sparseness. As for the Sense Similarity approach, we have selected the cut-off threshold at 0.2. For the Semantic Match we have selected the Most Frequent + Correct configuration. As for the merging we obtained four data sets: SYN+ppr02, which meges the Lexical Match and Sense Similarity methods, SYN+SM, which merges Lexical Match and Semantic Match, ppr02+SM, which merges Sense Similarity and Semantic Match, and SYN+ppr02+SM, which merges all three methods. The results are reported in Table 4.

The combination of the best result yields the best performance with respect to the stand-alone approaches. In particular, we obtain an F1=0.47 for SYN-ppr02, with an improvement of 18 points with respect to SYN, of 21 points with respect to Sense Similarity with threshold 0.2, and of 38 points with respect to Semantic Match. Furthermore, it is interesting to observe that the F1 score for SYN+SM (F1=0.33) and ppr02+SM (F1=0.31) are higher that those of SYN with maximum score filter (F1=0.29) and PPR - 0.2 (F1=0.26), suggesting that there is a kind of complementarity among the three alingment methods. However, the alignments from the Semantic Match method are noisy with respect to those obtained from Sense Similarity and Lexical Match. When merging the three methods together, SYN+ppr02+SM, we do not register any improvement but a lowering of the performances with the exception of Recall. This calls for a careful use of such data in this task, suggesting that simpler aligning methods are more robust.

Merged	P	R	F1
SYN+ppr02	**0.61**	0.38	**0.47**
SYN+SM	0.48	0.25	0.33
ppr02+SM	0.52	0.22	0.31
SYN+ppr02+SM	0.50	0.38	0.43

Table 4: Results for automatic alignment merging the best results from the three approaches.

6 Conclusion and Future Work

This paper reported on experiments on the automatic alignment of verb 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 SC lexicon posed a serious issue for the straightforward application of state-of-the-art techniques for sense alignment.

We explored three different methods for achieving sense alignment: Lexical Match, Sense Similarity, and Semantic Match. In all cases, we are facing low scores for Recall which point out issues related to data sparseness in our lexica. By comparing the results of the three approaches, we can observe that i.) the Sense Similarity yields the best Precision; ii.) Lexical Match, including minimal semantically related items (i.e. SYN) is a dumb but powerful approach for this kind of tasks; iii.) Semantic Match suffers from data sparseness and also from a certain mismatch between corpus data and lexicographic examples. This latter aspect impacts on the application of more complex approaches grounded on linguist theories to automatic methods for sense alignment. It also calls for an extension of the amount of manually annotated data and better methods of semantic typing of the SF slot fillers, as the poor results of Most Frequent + Correct + Vector Similarity show. Furthermore, lexicographic examples of use from SCL, and probably most of the other lexicographic dictionaries, are rather simple and not always prototypical with respect to the actual sense realization in real corpus data. Distributional approaches on SFS acquisition could be helpful to improve this method, provided that reliable ways for assigning SFSs to verb senses encoded in existing resources are developed.

Finally, Sense Similarity based on PPR and Lexical Match qualify as real complementary methods for achieving reliable sense alignments in a simple way and when dealing with few data. Our merged approach provides satisfying results with an overall F1=0.47. The alignment of verb senses is not a simple task as verbs tend to have more abstract definitions than nouns and rely on semantic relations such as entailment which are still poorly encoded in existing resources. Future work will concentrate on the aligned sense pairs obtained by SYN+ppr02 to experiment techniques to reduce the sense descriptions in MWN and in SCL to boostrap better sense alignments, and on the exploitation of crowdsourcing on pre-aligned data to collect additional information on SF structures.

References

G. Attardi and F. Dell'Orletta. 2009. Reverse revision and linear tree combination for dependency parsing. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, Boulder, Colorado.

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

I. Chiari, A. Gangemi, E. Jezek, A. Oltramari, G. Vetere, and L. Vieu. 2013. An open knowledge base for italian language in a collaborative perspective. In *Proceedings of DH-case13, Collaborative Annotations in Shared Environments: metadata, vocabularies and techniques in the Digital Humanities*.

M. Ciaramita and M. Johnson. 2003. Supersense tagging of unknown nouns in WordNet. In Michael Collins and Mark Steedman, editors, *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 168–175.

A. Eneko and A. 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.

E. Jezek and V. 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).

A. Lenci, G. Lapesa, and G. Bonansinga. 2012. Lexit: A computational resource on italian argument structure. In *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC'12)*, pages 3712–3718.

M. 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.

C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. 2002. Wonderweb deliverable D17: the wonderweb library of foundational ontologies. Technical report.

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

R. 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.

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

R. 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.

E. Niemann and I. 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.

A. Oltramari, G. Vetere, I. Chiari, E. Jezek, F.M. Zanzotto, M.Nissim, and A. 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.

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

E Pianta, C. Girardi, and R. 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).

G. Rigau and E. Agirre. 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.

D. Roland and D. Jurafsky. 2002. Verb sense and verb subcategorization probabilities. In S. Stevenson and P. Merlo, editors, *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, pages 325–346. John Benjamins, Amsterdam.

A. Roventini, N. Ruimy, R. Marinelli, U. Marisa, and M. Michele. 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.

M. Ruiz-Casado, E. Alfonseca, and P. 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.