

LIMSI: Translations as Source of Indirect Supervision for Multilingual All-Words Sense Disambiguation and Entity Linking

Marianna Apidianaki
LIMSI-CNRS
Rue John von Neumann
91405 Orsay Cedex, France
marianna@limsi.fr

Li Gong
IMMI-CNRS
Rue John von Neumann
91405 Orsay Cedex, France
gong@limsi.fr

Abstract

We present the LIMSI submission to the Multilingual Word Sense Disambiguation and Entity Linking task of SemEval-2015. The system exploits the parallelism of the multilingual test data and uses translations as source of indirect supervision for sense selection. The LIMSI system gets best results in English in all domains and shows that alignment information can successfully guide disambiguation. This simple but effective method can serve to generate high quality sense annotated data for WSD system training.

1 Introduction

This paper describes the LIMSI system at the Multilingual Word Sense Disambiguation (WSD) and Entity Linking (EL) task of SemEval-2015 (Moro and Navigli, 2015). The system performs sense selection by combining translation information obtained through alignment of the multilingual test set with sense ranking. It can thus be described as semi-supervised given the indirect supervision provided by the translations. The alignment correspondences serve as constraints for reducing the search space for each word to BabelNet synsets (hereafter, BabelSynsets) containing the translation and the retained synsets are sorted according to the BabelNet sense ranking. Our goal is to test the contribution of translations in multilingual WSD with no recourse to context information. The system needs no training and can be applied directly to parallel data.

The evaluation results show that the LIMSI system outperforms all systems in all domains in En-

glish and highlight the important role of translations in guiding disambiguation. This simple yet effective approach can serve to generate high quality sense annotations for WSD system training. In what follows, we provide a detailed description of the system, an analysis of the results and a discussion of the factors that determine the efficiency of the method.

2 Task Description

The SemEval-2015 Multilingual WSD and EL task (Moro and Navigli, 2015) aims to promote joint research in these two closely-related topics. WSD refers to the task of assigning meanings to occurrences of words in texts (Navigli, 2009) and its multilingual counterpart involves the identification of semantically adequate translations (Resnik and Yarowsky, 1997; Ide et al., 2002; Apidianaki, 2009). EL, on the other side, aims at linking entities in a text to the most suitable entry in a knowledge base. The systems participating in the Multilingual WSD and EL task can make a choice between different options (WSD, EL or both) and one or several WSD settings (all-words or specific part-of-speech disambiguation). Contrary to previous tasks (Navigli et al., 2013), the SemEval-2015 task addresses the disambiguation of words of all content parts of speech. No training data is provided and the test set consists of parallel texts in three languages (English, Italian and Spanish) pertaining to both open and closed domains (biomedical, math and computer, and a broader (social issues) domain). For evaluation, the data is manually annotated with senses from BabelNet (version 2.5.1), a wide-coverage multilin-

gual semantic network.¹ Senses in BabelNet are described by synsets which contain lexicographic and encyclopedic knowledge extracted from various sources² in many languages, and are linked between them with different types of relations (Navigli and Ponzetto, 2012). The LIMSI system disambiguates words of all parts of speech in the three languages. No multi-word units are extracted. However, although only WSD is addressed explicitly, the system is also assigned EL scores as it manages to annotate several Named Entities with the correct synset.

3 System Description

3.1 Alignment of the Evaluation Dataset

The test data contains four parallel documents in English, Spanish and Italian. Our system exploits the parallelism of the test set, a feature overlooked by previous systems (Navigli et al., 2013). In order to avoid some discrepancies observed at the level of sentence correspondences, we first align the texts pairwise using the Hunalign sentence aligner (Varga et al., 2005). Then we run GIZA++ (Och and Ney, 2003) in both directions at the lemma level and retain only intersecting alignments to rule out spurious correspondences. For each instance of an English content word in the test set we identify its Spanish translation in context and, alternatively, the English translations of Spanish and Italian words. We use the lemma and part-of-speech information provided by the task organizers.

3.2 Sense Selection

The established alignment correspondences serve as constraints to retrieve the BabelSynsets that are relevant for words in the test set, based on the assumption of a semantic correspondence between a word and its translation in context (Diab and Resnik, 2002). BabelSynsets group synonymous English words and their translations in different languages. Polysemous words are found in different synsets, as in WordNet (Miller et al., 1990), and are associated to different translations.

The procedure for selecting the most adequate BabelSynset for an occurrence of a word (w) in context is described in Figure 1. First, we find the synsets of

¹The resource is available at <http://babelnet.org/>

²WordNet, wiki resources and automatic translations.

Notation:

S_w : the set of BabelSynsets for w

t : a translation of w in context

S_w^t : the set of synsets in which t appears

The Sense Selection Algorithm:

```

 $S_w^t \leftarrow \emptyset$ 
 $S_w \leftarrow \text{getBabelSynsets}(w)$ 
for each BabelSynset  $s \in S_w$  do
  if  $t \in s$  then
    add  $s$  to  $S_w^t$ 
if  $|S_w^t| \geq 1$  then
  return  $\text{getBFS}(S_w^t)$ 
else
  return  $\text{getBFS}(S_w)$ 

```

Figure 1: The `getBabelSynsets` function retrieves the synsets available for w in BabelNet. The `getBFS` function ranks synsets according to importance. If the aligned translation is contained in different synsets of w , the most frequent one among this set of synsets is returned. If no synset is retained through alignment, the system falls back to the BFS baseline.

w (S_w) in BabelNet 2.0 and filter them to keep only synsets that contain both w and its aligned translation t in this context ($S_w^t \subseteq S_w$).³ If more than one synsets are retained, we rank them using the default sense comparator integrated within the BabelNet-API 2.5 (`BabelSynsetComparator`) and keep the highest ranked synset. Otherwise, if t is found in only one synset, this constitutes the sense tag for the word. The system falls back to the BabelNet First Sense (BFS)⁴ for unaligned instances or in cases where t is not found in any synset. As the alignment constraint does not apply in this case, the BFS corresponds to the highest ranked among *all* synsets of w . Note that the sense selected by our method for a word might correspond to its BFS or not. As selection is done among the subset of senses that satisfy the alignment constraint, if this is the case for the BFS it remains among the candidate synsets and can

³In these experiments, we only use translations in one language. We would expect the use of translations in different languages to increase the accuracy of the filtering but as a downside, it could reduce the recall as synsets should contain all translations.

⁴The most frequent sense (MFS) for a word in BabelNet.

All domains			Biomedical			Math & computer			Social issues		
System	All	WSD	System	All	WSD	System	All	WSD	System	All	WSD
LIMSI	65.8	64.7	LIMSI	71.3	68.9	LIMSI	54.1	53.9	LIMSI	67.2	67.7
SUDOKU-2	61.6	59.9	SUDOKU-3	71.2	68.8	SUDOKU-2	53.2	53.1	vua-background	60.8	61.1
SUDOKU-3	60.7	58.9	SUDOKU-2	68.9	66.4	SUDOKU-3	49.4	49.1	SUDOKU-1	56.4	56.2
vua-background	58.4	60.3	vua-background	63.6	66.4	EBL-Hope	41.7	39.8	SUDOKU-2	55.6	54.5
SUDOKU-1	55.8	57.5	SUDOKU-1	62.4	65.0	TeamUFAL	29.8	28.4	WSD-games-1-2	53.5	53.8
BFS	67.5	66.3	BFS	72.1	69.9	BFS	55.3	55.2	BFS	70.8	70.5

Table 1: Best performing systems at the SemEval-2015 Multilingual WSD and Entity Linking task for English.

System	All domains		Biomedical		Math & computer		Social issues	
	ES	IT	ES	IT	ES	IT	ES	IT
LIMSI	45.0	48.4	51.0	53.1	34.8	44.6	43.1	42.9
SUDOKU 1/2	57.1	59.9	62.7	65.1	49.7	52.1	57.0	61.0
BFS	37.5	40.2	43.7	44.3	28.7	36.7	34.0	35.7

Table 2: LIMSI, best system and BFS scores in Spanish and Italian.

be selected, otherwise it is discarded. For instance, the noun *side* has 21 BabelSynsets but its Spanish translation in this context:

The tablets are pale-orange and have a score line on both sides so that they can be halved.

cara, is found in only two synsets: 00032604n and 00071434n. These are semantically close and describe fine-grained nuances of the “outer surface of an object” meaning of *side*, also expressed by *cara*.⁵ Sense ranking correctly suggests 00032604n (“a surface forming part of the outside of an object”) as the most adequate sense annotation for this instance of the word. In this case our method improves over the BFS baseline which proposes 00071431n (“a place within a region identified relative to a center or reference location”), a synset that our system rules out from the beginning as it does not contain the translation *cara*.

4 Evaluation Results

Table 1 gives an overview of the results obtained for English.⁶ The systems are evaluated using standard WSD evaluation metrics. Precision measures the percentage of the sense assignments provided by

⁵BabelSynsets often correspond to WordNet synsets describing fine-grained nuances of meaning.

⁶A full presentation of the results is available in the task description paper (Moro and Navigli, 2015).

the system that are identical to the gold standard; recall measures the percentage of instances that are correctly labeled by the system. Results in the table are reported in F1 score. The five best performing systems in both tasks (WSD & EL) and WSD only are compared to the BFS baseline.

The LIMSI alignment-based system yields the top performance in English among the 17 submitted systems, in all domains. This result is very interesting given that our method is very simple: it needs no training and is very easy to compute as it only relies on alignment and sense ranking. Note that the BFS baseline for English is a very strong one that none of the systems manages to beat. As the test set is very small (~ 138 parallel sentences), we expect the method to perform even better on larger corpora where the automatic alignment will have higher accuracy and coverage.

Our system performs poorly in Spanish and Italian in comparison to English, and is ranked in the fourth position. The scores obtained in these languages are given in Table 2 and are compared to the best performing system and the baseline. A close analysis of the results reveals that the weaker system performance is due to the way the BabelNet API carries out sense ranking in these languages. In English, WordNet senses are ranked first sorted

System		EN	ES	IT
LIMSI		596	596	592
LIMSI = BFS	both ✓	396	231	236
	both ×	150	218	198
LIMSI ≠ BFS	LIMSI ✓	37	136	142
	BFS ✓	13	11	16
BFS		563	500	499
	✓	363	158	182
	×	200	342	317

Table 3: The top part of the table gives the # of correct/wrong annotations made by the LIMSI system. The lower part shows the # of correct/wrong predictions when the system falls back to the BFS.

by sense number⁷ and are followed by Wikipedia senses in lexicographic order (Navigli, 2013). For languages other than English where frequency information is not available, senses are sorted in lexicographic order,⁸ a criterion that often fails to reflect their relevance (i.e. rare senses might be placed higher than more frequent ones). This certainly affects our system which relies on sense ranking a) when multiple senses are retained after filtering by alignment, and b) when the BFS is needed.⁹

The low values of the Spanish and Italian BFS baseline reported by the task organizers confirm this finding. As the first sense retained by the BabelNet API in these languages often is not the most frequent sense, the baseline is outperformed by almost all participating systems. The higher scores obtained by our system compared to the baseline show that the alignment-based filtering remains beneficial in spite of the problematic sense ranking, as the aligned translation might occur in only one BabelSynset. Table 3 provides a detailed analysis of the results. The top part of the table shows the accuracy of the alignment-based predictions, which might coincide with the BFS or not. Our system improves over the BFS in 37 cases in English, 136 in Spanish and 142 in Italian. On the contrary, the BFS does better only

⁷Sense numbers in WordNet reflect the frequency of the senses in the SemCor corpus (Miller et al., 1993).

⁸An additional criterion applies to Wikipedia senses according to which pages that contain a parenthetical explanation, as in disambiguation pages, are ranked lower than ones that do not.

⁹For example, in cases of unaligned words or where the aligned translation is not found in some synset.

13, 11 and 16 times in the three languages. The system falls back to the BFS in case of unaligned words or when the translations are not found in some BabelNet synset. As shown in the lower part of Table 3, the BFS predictions are often wrong, especially in Spanish and Italian (342 and 317 wrong predictions, respectively). This analysis shows the limited impact of the BFS on the performance of the LIMSI system which manages to improve over the baseline in numerous cases.

The system fails to provide the correct sense in cases of parallel ambiguities where a word and its translation carry the same senses. For example, this instance of *window*:

Here's a screenshot of kalgebra main window.

is aligned to *ventana* in the Spanish text, which translates both the “opening” and the “computer” sense of the word. Although the Spanish translation helps to rule out 11 of the 15 BabelSynsets of *window*, ranking the remaining four synsets puts forward the more frequent “opening” sense (00081285n) which is incorrect for this instance. Using translations in multiple languages could improve accuracy in these cases.

5 Conclusion

We have described the LIMSI system submitted to the SemEval-2015 Multilingual All-Words Sense Disambiguation and Entity Linking task. The system is based on automatic translation alignment and sense ranking, it needs no training and is directly applied to the evaluation data. By exploiting the indirect supervision provided through alignment, this simple approach gives top performance in English. The high quality semantic annotations provided by our system can serve as training data for supervised WSD algorithms.

Based on these encouraging results, we see a number of research directions for future work. As the method in its current form is bound to be used on parallel data, we would like to experiment with alignments provided by Machine Translation systems and disambiguate monolingual texts. Moreover, we intend to explore alternative sense ranking solutions to improve the performance of the method in languages other than English.

References

- Marianna Apidianaki. 2009. Data-driven Semantic Analysis for Multilingual WSD and Lexical Selection in Translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*, pages 77–85, Athens, Greece.
- Mona Diab and Philip Resnik. 2002. An Unsupervised Method for Word Sense Tagging using Parallel Corpora. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 255–262, Philadelphia, USA.
- Nancy Ide, Tomaz Erjavec, and Dan Tufiş. 2002. Sense Discrimination with Parallel Corpora. In *Proceedings of the ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pages 54–60, Philadelphia, USA.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3:235–244.
- George A. Miller, Claudia Leacock, Randee Teng, and Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of a Workshop on Human Language Technology*, pages 303–308, Plainsboro, New Jersey, USA.
- Andrea Moro and Roberto Navigli. 2015. SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, USA.
- 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, David Jurgens, and Daniele Vannella. 2013. SemEval-2013 Task 12: Multilingual Word Sense Disambiguation. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 222–231, Atlanta, Georgia, USA.
- Roberto Navigli. 2009. Word Sense Disambiguation: a Survey. *ACM Computing Surveys*, 41(2):1–69.
- Roberto Navigli. 2013. A Quick Tour of BabelNet 1.1. In *Computational Linguistics and Intelligent Text Processing - 14th International Conference, CICLing 2013, Part I*, pages 25–37, Samos, Greece.
- Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29:19–51.
- Philip Resnik and David Yarowsky. 1997. A Perspective on Word Sense Disambiguation Methods and Their Evaluation. In *Proceedings of the SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What and How?*, Washington, DC, USA.
- Dániel Varga, Péter Halácsy, András Kornai, Viktor Nagy, László Németh, and Viktor Trón. 2005. Parallel corpora for medium density languages. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2005)*, pages 590–596, Borovets, Bulgaria.