

# EBL-Hope: Multilingual Word Sense Disambiguation Using A Hybrid Knowledge-Based Technique

**Eniafe Festus Ayetiran**

CIRSFID, University of Bologna  
Via Galliera, 3 - 40121  
Bologna, Italy  
eniafe.ayetiran2@unibo.it

**Guido Boella**

Department of Computer Science  
University of Turin  
Turin, Italy  
boella@di.unito.it

## Abstract

We present a hybrid knowledge-based approach to multilingual word sense disambiguation using BabelNet. Our approach is based on a hybrid technique derived from the modified version of the Lesk algorithm and the Jiang & Conrath similarity measure. We present our system's runs for the word sense disambiguation subtask of the Multilingual Word Sense Disambiguation and Entity Linking task of SemEval 2015. Our system ranked 9th among the participating systems for English.

## 1 Introduction

The computational identification of the meaning of words in context is called Word Sense Disambiguation (WSD), also known as Lexical Disambiguation. There have been a significant amount of research on WSD over the years with numerous different approaches being explored. Multilingual word sense disambiguation aims to disambiguate the target word in different languages. This, however, involves a different scenario compared to monolingual WSD in the sense that a single word in a language might have varying number of senses in other languages with significant differences in the semantics of some of the available senses.

Approaches to word sense disambiguation may be: (1) knowledge-based which depends on some knowledge dictionary or lexicon (2) supervised machine learning techniques which train systems from labelled training sets and (3) unsupervised which

is based on unlabelled corpora, and do not exploit any manually sense-tagged corpus to provide a sense choice for a word in context.

We present a hybrid knowledge-based approach based on the Modified Lesk algorithm and the Jiang & Conrath similarity measure using BabelNet (Navigli and Ponzetto, 2012). The system presented here is an adaptation of our earlier work on monolingual word sense disambiguation in English (Ayetiran et al., 2014).

## 2 Methodology

Figure 1 illustrates the general architecture of our hybrid disambiguation system.

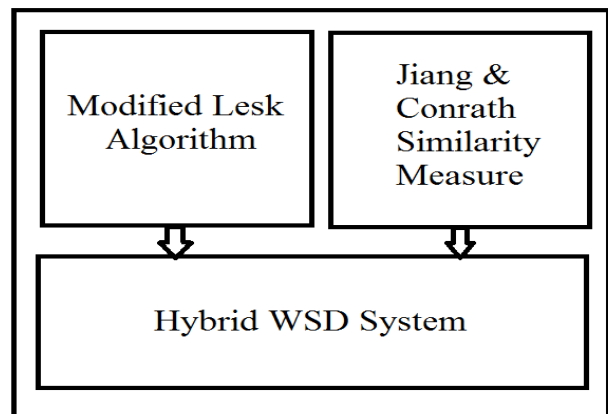


Figure 1: The Hybrid Word Sense Disambiguation System - A system that combines two distinct disambiguation submodules.

## 2.1 The Lesk Algorithm

Micheal Lesk (1986) invented this approach named gloss overlap or the Lesk algorithm. It is one of the first algorithms developed for the semantic disambiguation of all words in unrestricted texts. The only resource required by the algorithm is a set of dictionary entries, one for each possible word sense, and knowledge about the immediate context where the sense disambiguation is performed. The idea behind the Lesk algorithm represents the seed for today's corpus-based algorithms. Almost every supervised WSD system relies one way or the other on some form of contextual overlap, with the overlap being typically measured between the context of an ambiguous word and contexts specific to various meanings of that word, as learned from previously annotated data.

The main idea behind the original definition of the algorithm is to disambiguate words by finding the overlap among their sense definitions. Namely, given two words,  $W_1$  and  $W_2$ , each with  $NW_1$  and  $NW_2$  senses defined in a dictionary, for each possible sense pair  $W_{1_i}$  and  $W_{2_j}$ ,  $i = 1, \dots, NW_1$ ,  $j = 1, \dots, NW_2$ , we first determine the overlap of the corresponding definitions by counting the number of words they have in common. Next, the sense pair with the maximum overlap is selected, and therefore the sense is assigned to each word in the text as the appropriate sense. Several variations of the algorithm have been proposed after the initial work of Lesk. Ours follow the work of Banerjee and Pedersen (2002) who adapted the algorithm using WordNet (Miller, 1990) and the semantic relations in it.

## 2.2 Jiang & Conrath Similarity Measure

Jiang & Conrath similarity (Jiang & Conrath, 1997) is a similarity metric derived from corpus statistics and the WordNet lexical taxonomy. The method makes use of information content (IC) scores derived from corpus statistics (Reisnik 1995) to weight edges in the taxonomy. Edge weights are set to the difference in IC of the concepts represented by the two connected nodes.

For this algorithm, Reisnik (1995)'s IC measure is augmented with the notion of path length between

concepts. This approach includes the information content of the concepts themselves along with the information content of their lowest common subsumer. A lowest common subsumer is a concept in a lexical taxonomy which has the shortest distance from the two concepts compared. They argue that the strength of a child link is proportional to the conditional probability of encountering an instance of the child sense  $s_i$  given an instance of its parent sense. The resulting formula can be expressed in Equation (1) below:

$$Dist(w_1, w_2) = IC(s_1) + IC(s_2) - 2 \times IC(Lsuper(s_1, s_2)) \quad (1)$$

Where  $s_1$  and  $s_2$  are the first and second senses respectively and LSuper (lowest common subsumer) is the lowest super-ordinate of  $s_1$  and  $s_2$ . IC is the information content given by equation (2):

$$IC(c) = \log^{-1}P(s) \quad (2)$$

$P(s)$  is the probability of encountering an instance of sense  $s$ .

## 3 The Hybrid WSD System

For monosemous words, the sense is returned as disambiguated based on the part of speech. For polysemous words, we followed the Adapted Lesk approach of Banerjee and Pederson (2002) but instead of a limited window size used by Banerjee and Pederson, we used all context words as the window size.

Most prior work has not made use of the antonymy relation for WSD. But according to Ji (2010), if two context words are antonyms and belong to the same semantic cluster, they tend to represent the alternative attributes for the target word. Furthermore, if two words are antonymous, the gloss and examples of the opposing senses often contain many words that are mutually useful for disambiguating both the original sense and its opposite. Therefore, we added the glosses of antonyms in addition to hypernyms, hyponyms, meronyms etc. used by Banerjee and Pedersen (2002). Also, for verbs we have added the glosses of entailment and causes relations of each word sense to their vectors. For adjectives and adverbs, we added the morphologically related nouns to the vectors of each word sense in computing the similarity score.

The similarity score for the Modified Lesk algorithm is computed using the Cosine similarity. The vectors are composed using the glosses of the word senses, that of their hypernyms, hyponyms, and antonyms. We then compute the cosine of the angle between the two vectors. This metric is a measurement of orientation and not magnitude. The magnitude of the score for each word is normalized by the magnitude of the scores for all words within the vector. The resulting normalized scores reflect the degree the sense is characterized by each of the component words.

Cosine similarity can be trivially computed as the dot product of vectors normalized by their Euclidean length:

$$\vec{a} = (a_1, a_2, a_3, \dots, a_n) \quad \text{and} \quad \vec{b} = (b_1, b_2, b_3, \dots, b_n)$$

Here  $a_n$  and  $b_n$  are the components of vectors containing length normalized TF-IDF scores for either the words in a context window or the words within the glosses associated with a sense being scored. The dot product is then computed as follows:

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

The dot product is a simple multiplication of each component from the both vectors added together. The geometric definition of the dot product given by equation (3):

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\theta \quad (3)$$

Using the the cummutative property, we have equation (4):

$$\vec{a} \cdot \vec{b} = \|\vec{b}\| \|\vec{a}\| \cos\theta \quad (4)$$

where  $\|\vec{a}\| \cos\theta$  is the projection of  $\vec{a}$  into  $\vec{b}$  in which solving the dot product equation for  $\cos\theta$  gives the cosine similarity in equation (5):

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (5)$$

where  $a \cdot b$  is the dot product and  $\|a\|$  and  $\|b\|$  are the vector lengths of  $a$  and  $b$ , respectively.

We disambiguated each target word in a sentence using the Jiang & Conrath similarity measure using all the context words as the window size. We did this by computing Jiang & Conrath similarity score for each candidate senses of the target word and select the sense that has the highest sum total similarity score to all the words in the context window.

For each context word  $w$  and candidate word senses  $c_{eval}$ , we compute individual similarity scores using equation (6):

$$sim(w, c_{eval}) = \max_{c \in sen(w)} [sim(c, c_{eval})] \quad (6)$$

where  $sim(w, c_{eval})$  function computes the maximum similarity score obtained by computing Jiang & Conrath similarity for all the candidate senses in a context word. The aggregate summation of the individual similarity scores is given in equation (7):

$$\text{argmax}_{c_{eval} \in sen(w)} = \sum_{w \in context(W)} \max_{c \in sen(w)} [sim(c, c_{eval})] \quad (7)$$

An agreement between the results produced by each of the two algorithms means the word under consideration has been likely correctly disambiguated and the sense on which they agreed is returned as the correct sense. Whenever one module fails to produce any sense that can be applied to a word but the other succeeds, we just return the sense computed by the successful module. Module failures occur when all of the available senses receive a score of 0 according to the module's underlying similarity algorithm (e.g., due to lack of overlapping words for Modified Lesk).

Finally, in a situation where the two modules select different senses, we heuristically resolved the disagreement. Our heuristic first computes the derivationally related forms of all of the words in the context window and adds each of them the vector representation of the word being assessed. Then for the senses produced by the Modified Lesk and Jiang & Conrath algorithms, we obtain the similarity score between the vector representations of the two competing senses and the new expanded context vector. The algorithm returns the sense selected

by the module whose winning vector is most similar to the augmented context vector.

The intuition behind this notion of validation is that the glosses of a word sense, and that of their semantically related ones in the WordNet lexical taxonomy should share words in common as much as possible with words in context with the target word. Adding the derivationally related forms of the words in the context window increases the chances of overlap when there are mismatches caused by changes in word morphology. When both modules fail to identify a sense, the Most Frequent Sense (MFS) in the Semcor corpus is used as the appropriate sense.

## 4 Experimental Setting

The SemEval 2015 Multilingual Word Sense Disambiguation and Entity Linking task provides datasets in English, Spanish and Italian. BabelNet (Navigli and Ponzetto, 2012) which provides automatic translation of each word sense in other languages have been employed. To enrich the glosses used by the Modified Lesk algorithm, the glosses provided by BabelNet from Wikipedia in the 3 subtask languages have been used to extend the initial glosses available in WordNet (Miller, 1990).

Furthermore, BabelNet contains some word senses which are not available in WordNet. These senses and their glosses were used directly without any reference to WordNet translation since it does not have any. For English, we disambiguate all the open target words while for Spanish and Italian, we disambiguate all noun target words. Due to some challenges we faced close to our task’s evaluation deadline, we were unable to obtain BabelNet 2.5 which is the official resource for the task. Instead, we used BabelNet 1.1.1 from the SemEval 2013 Multilingual Word Sense Disambiguation Task, which we initially used to develop our system but unfortunately contains only noun words for Spanish and Italian and does not include some English words found in the test set.

## 5 Results and Discussion

Table 1 compares the performance of our system with other participating systems on the English subtask. Table 2 shows the result of our system for the

System	Precision	Recall	F1
LIMSI	68.7	63.1	65.8
SUDOKU-Run2	62.9	60.4	61.6
SUDOKU-Run3	61.9	59.4	60.6
vua-background	67.5	51.4	58.4
SUDOKU-Run1	60.1	52.1	55.8
WSD-games-Run2	58.8	50.0	54.0
WSD-games-Run1	57.4	48.8	52.8
WSD-games Run3	53.5	45.4	49.1
<i>EBL-Hope</i>	48.4	44.4	46.3
TeamUFAL	40.4	36.5	38.3

Table 1: Performance of All Participating Systems for English Subtask. Our *EBL-Hope* System ranked 9th out of the submitted systems.

Spanish and Italian subtask where we submitted a run for only nouns and named entities.

Subtask	Precision	Recall	F1
Spanish	52.5	44.6	48.2
Italian	43.1	35.3	38.8

Table 2: *EBL-Hope*’s hybrid system performance on the Spanish and Italian subtasks.

Our system performs noticeably better in Spanish than Italian. Further analysis shows that the weakest area of our system for the English subtask are the verbs, which achieve 35.8 F1 score. We achieve high scores on named-entities with an F1 scores of 80.2 in English, 48.5 in Italian and the highest F1 score across all participating systems on Spanish with 70.8.

Table 3 and Table 4 give the performance obtained when using the Modified Lesk and Jiang & Conrath modules independently. Our hybrid system outperforms the individual component modules on both English and Spanish. On Italian, the Hybrid system performs comparably to Jiang & Conrath, which is the best individual module.

Subtask	Precision	Recall	F1
English	43.6	41.3	42.4
Spanish	48.1	41.2	44.3
Italian	46.3	33.5	38.9

Table 4: Performance of the Jiang & Conrath module in isolation on the 3 subtasks.

Subtask	Precision	Recall	F1
English	44.2	40.6	42.3
Spanish	47.6	40.1	43.5
Italian	40.3	31.7	35.4

Table 3: Performance of the Modified Lesk module in isolation on the 3 subtasks.

## 6 Conclusion

In this work, we have combined two algorithms for word sense disambiguation, Modified Lesk and an approach based on Jiang & Conrath similarity. The resulting hybrid system improves performance by heuristically resolving disagreements in the word sense assigned by the individual algorithms. We observe the results of the combined algorithm do consistently outperform each of the individual algorithms used in isolation. However, our poor performance on the official evaluation could likely have been improved by making use of the more recent 2.5 version of BabelNet as recommended by the task organizers.

## Acknowledgement

This work has been supported by European Commission scholarship under the Erasmus+ doctoral scholarship programmes. We would like to thank the anonymous reviewers for their helpful suggestions and comments. Special thanks to Daniel Cer for his great and useful editorial input on the final manuscript.

## References

- Satanjeev Banerjee and Ted Pedersen. 2002. An adapted Lesk Algorithm for Word Sense Disambiguation using WordNet. In *Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing (CICLING)*, Mexico City, Mexico, 17 - 23 February, 2002, pp. 136 - 145.
- George Miller. 1990. An Online Lexical Database. *International Journal of Lexicography*, 3(4): 235 - 244.
- Jay J. Jiang and David W. Conrath. 1997. Semantic similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proceedings of the 10th International Conference on Research in Computational Linguistics*, Taipei, Taiwan, 2 - 4 August 1998, pp. 19 - 33.
- Michael E. Lesk. 1986. Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to

Tell a Pine Cone from an Ice Cream Cone. In *Proceedings of the 5th ACM-SIGDOC Conference*, Toronto, Canada, 8 - 11 June 1986, pp. 24 - 26.

- Eniafe F. Ayetiran, Guido Boella, Luigi Di Caro, Livio Robaldo. 2014. Enhancing Word Sense Disambiguation Using A Hybrid Knowledge-Based Technique. In *Proceedings of 11th international workshop on natural language processing and cognitive science*, Venice, Italy 27 - 29, October, pp. 15 - 26.
- Heng Ji. 2010. One Sense per Context Cluster: Improving Word Sense Disambiguation Using Web-Scale Phrase Clustering. In *Proceedings of the 4th Universal Communication Symposium (IUCS)*, Beijing, China, 18 -19 October 2010, pp. 181 -184.
- Roberto Navigli and Simone P. Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. In *Artificial Intelligence*, 193(2012) 217-250.
- Philip Reisman. 1995. One Sense per Context Cluster: Using Information Content to Evaluate Semantic Similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada, 20 - 25 August 1995, pp. 448-453.