

Thesis title

Osman Başkaya

September 2014



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Related Works</b>	<b>7</b>
<b>3</b>	<b>System</b>	<b>9</b>
3.1	Pipeline . . . . .	9
3.2	Different Representations . . . . .	9
3.3	Different approaches . . . . .	9
<b>4</b>	<b>Evaluation Framework for WSI Systems</b>	<b>11</b>
4.1	Agirre’s method . . . . .	11
4.2	Complex Classifiers for mapping and some discussion . . . . .	11
<b>5</b>	<b>Experiments on various tasks</b>	<b>13</b>
5.1	WSI tasks . . . . .	13
5.1.1	SemEval 2010 WSI Task . . . . .	13
5.1.2	SemEval 2013 WSI Task . . . . .	13
5.1.3	Pseudoword Dataset . . . . .	14
5.2	All-Word task . . . . .	14
5.3	Semantic Similarity task . . . . .	14
<b>6</b>	<b>Sensitivity Analysis</b>	<b>15</b>
<b>7</b>	<b>Conclusion</b>	<b>17</b>



# Chapter 1

## Introduction



## Chapter 2

### Related Works





# Chapter 3

## System

### 3.1 Pipeline

### 3.2 Different Representations

### 3.3 Different approaches



## Chapter 4

# Evaluation Framework for WSI Systems

### 4.1 Agirre's method

### 4.2 Complex Classifiers for mapping and some discussion



# Chapter 5

## Experiments on various tasks

### 5.1 WSI tasks

#### 5.1.1 SemEval 2010 WSI Task

#### 5.1.2 SemEval 2013 WSI Task

##### Dataset

The test data for the graded word sense induction task in SemEval-2013 includes 50 terms containing 20 verbs, 20 nouns and 10 adjectives. There are a total of 4664 test instances provided. All evaluation was performed on test instances only. In addition, the organizers provided sense labeled trial data which can be used for tuning. This trial data is a redistribution of the Graded Sense and Usage data set provided by Katrin Erk, Diana McCarthy, and Nicholas Gaylord [Erk and McCarthy \[2009\]](#). It consists of 8 terms; 3 verbs, 3 nouns, and 2 adjectives all with moderate polysemy (4-7 senses). Each term in trial data has 50 contexts, in total 400 instances provided. Lastly, participants can use ukWaC [Ferraresi et al. \[2008\]](#), a 2- billion word web-gathered corpus, for sense induction. Furthermore, unlike in previous WSI tasks, organizers allow participants to use additional contexts not found in the ukWaC under the condition that they submit systems for both using only the ukWaC and with their augmented corpora. The gold-standard of test data was prepared using WordNet 3.1 by 10 annotators. Since WSI systems report their annotations in a different sense inventory than WordNet 3.1, a mapping procedure should be used first. The organizers use the sense mapping procedure explained in [Jurgens \[2012\]](#). This procedure has adopted the supervised evaluation setting of past SemEval WSI Tasks, but the main difference is that the former takes into account applicability weights for each sense which is a necessary for graded word sense. Although the data contains graded senses for some instances, I used only the

instances that labeled one sense. This exclusion decreased the number of instances used in the experiments to 4122.

The following section contains baseline scores for the test data on single sense instances.

### Baselines

**Most Frequent Baseline:** Task organizers provided the FScore of most frequent baseline for single sense data, which is 0.578.

**Random Baselines:** [Random baselines can be seen here](#)

**kNN-baseline for substitute vectors:** As explained in 5.1.1, I followed the same procedure to calculate this baseline. [Results can be seen in details.](#)

**kNN-baseline for embeddings:** [Scores for embeddings can be seen here.](#)

### 5.1.3 Pseudoword Dataset

## 5.2 All-Word task

## 5.3 Semantic Similarity task

## Chapter 6

# Sensitivity Analysis





# Chapter 7

## Conclusion



# References

Katrin Erk and Diana McCarthy. Graded word sense assignment. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1*, pages 440–449. Association for Computational Linguistics, 2009.

Adriano Ferraresi, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. Introducing and evaluating ukwac, a very large web-derived corpus of english. In *In Proceedings of the 4th Web as Corpus Workshop (WAC-4)*, 2008.

David Jurgens. An Evaluation of Graded Sense Disambiguation using Word Sense Induction. In *Proceedings of \*SEM, the First Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, 2012.