

Abstract

Fourth Year Project

Johnny Scanlon- 17328352

The aim of this project was to investigate the effects of target word position in unsupervised word sense disambiguation. The EM algorithm is used to perform this word sense disambiguation on these N-grams.

To examine the effects of position the project starts by fixing target word position in these 5-grams. Using a diachronic model, the per year probabilities of detected senses can be graphed. By graphing these sense probabilities, the emergence of neologisms and relative sense probabilities can be seen. By comparing these to known values, the effect of target word position can be examined and compared. The project then examines more complex models, examining how position is influencing these results.

Pseudo-neologisms are used to make fairer comparison between different models and perform more detailed exploration of edge cases.

There was an observed difference between n-grams of different target word positions. Methods for using these differences to improve results in more complex models are then explored with positive results.

Ultimately it is shown that target word position is relevant to the success or failure of a word sense disambiguation, and knowledge of it can be used to improve the model and get better results.