

## Abstract

The goal of this thesis is to extract synonym pairs from a large collection of text, using minimal supervision. The initial motivation is to use the extracted synonyms to improve machine translation evaluation. The approach is based on the word embeddings popularized by Mikolov et al. (2013a). We analyze how distributional word vectors can be used to extract synonyms for English and German, and what are the frequent error categories. We propose the measure *relative cosine similarity*, to increase the precision of the extraction. Furthermore, we show that by combining differently trained word embeddings, or using a part-of-speech tagger, the performance of the extraction can be improved. The final system is evaluated manually, and in the task of machine translation evaluation for both languages. We show our system can extract synonyms from part-of-speech tagged text that can be used to improve machine translation evaluation.

**Keywords:** synonym extraction, minimal supervision, machine translation evaluation, word embeddings