ABSTRACT

**Evaluating Bilingual Embeddings in Bilingual Dictionary Alignment**

**Yiğit Sever**

**Master of Science, Department of Computer Engineering**

**Supervisor: Dr. Gönenç Ercan**

**September 2019, 97 Pages**

Dictionaries catalog and describe the semantic information of a lexicon. WordNet provides an edge by presenting distinct concepts with the hierarchy information among them. Research in computer science has been using this hand crafted tool in natural language applications such as text summarization and machine translation. Original WordNet has been compiled for English yet counterparts for other languages are not as readily available nor as comprehensive. In order for research on languages other than English to benefit from the power of a WordNet, machine assisted creation and evaluation methods are essential.

Word embeddings can provide a mapping between words and points in a real valued vector space. Using these vectors, representing documents as well as forming geometric relationships between them is a well studied area of research. In this thesis we start by hypothesizing that a dictionary definition captures the semantic basis of the described word. We used word embeddings as building blocks to map dictionary definitions into a multidimensional space. These spaces can be aligned to accommodate two languages, allowing the transfer of information from one language to another. We investigate the success of retrieving and matching discrete senses across languages by employing supervised and unsupervised methods. Our experiments show that dictionary alignment can be evaluated successfully by using both unsupervised and supervised methods but corpora sizes should be taken into consideration. We further argue that some methods are not viable considering their poor performance.

**Keywords: dictionary alignment, word embeddings, semantic encoder, short text similarity**