

## **Abstract**

The lack of hand curated data is a major impediment to developing statistical semantic processors for many of the world languages. A major issue of semantic processors in Natural Language Processing (NLP) is that they require manually annotated data to perform accurately. Our work aims to address this issue by leveraging existing annotations and semantic processors from multiple source languages by projecting their annotations via statistical word alignments traditionally used in Machine Translation. Taking the Named Entity Recognition (NER) task as a use case of semantic processing, this work presents a method to automatically induce Named Entity taggers using parallel data, without any manual intervention. Our method leverages existing semantic processors and annotations to overcome the lack of annotation data for a given language. The intuition is to transfer or project semantic annotations, from multiple sources to a target language, by statistical word alignment methods applied to parallel texts (Och and Ney, 2000; Liang et al., 2006). The projected annotations can then be used to automatically generate semantic processors for the target language. In this way we would be able to provide NLP processors without training data for the target language. The experiments are focused on 4 languages: German, English, Spanish and Italian, and our empirical evaluation results show that our method obtains competitive results when compared with models trained on gold-standard out-of-domain data. This shows that our projection algorithm is effective to transport NER annotations across languages via parallel data thus providing a fully automatic method to obtain NER taggers for as many as the number of languages aligned via parallel corpora.