

Abstract

Semantic Role Labelling (SRL), the task of automatically identifying and labelling a predicate-argument structure at the sentence level, has been shown to be important for a broad spectrum of natural language processing (NLP) applications, such as information extraction, summarization, plagiarism detection, question answering, and machine translation. Due to the high costs of manual annotation for SRL, Akbik *et al.* (2016b) proposed a method to generate Proposition Banks (PBs) for novel languages by means of annotation projection in parallel corpora, followed by a manual correction step in order to filter and merge the created semantic frames. In this project, we propose a method to perform the correction process semi-automatically by using a multilingual distributional semantic model and a learning algorithm for classification. Although the project is aimed at creating a Corrected Spanish PB, the method is language-independent and will be used to correct PBs in other languages as well. The method was evaluated on the manually Curated French, German and Chinese PB, and obtained promising results in the Projected Spanish PB, which are expected to help speed up the manual correction process overall.