Abstract

Iustina Ilisei, Diana Inkpen, Gloria Corpas Pastor and Ruslan Mitkov. Identification of Translational Sublanguage: A Machine Learning Approach

This paper presents a machine learning approach to the study of translational sublanguage. The goal is to train a computer system to distinguish between translated and non-translated text, in order to determine the characteristic features that influence the classifiers. Several algorithms reach up to 97.62% success rate on a technical dataset. Moreover, the SVM classifier consistently reports a statistically significant improved accuracy when the learning system benefits from the addition of simplification features to the basic translational classifier system. Therefore, these findings may be considered an argument for the existence of the Simplification Universal.