L1 Influence on Content Word errors in Learner English Corpora: Insights from Distributed Representation of Words.

Kanishka Misra
Purdue University, West Lafayette, Indiana, United States

Hemanth Devarapalli
Purdue University, West Lafayette, Indiana, United States

Julia Rayz
Purdue University, West Lafayette, Indiana, United States

Abstract

The first language of a person has been shown to influence the processing of words when they learn a new language. This has been previously researched in behavioral studies, as well as by using lexical distributions or co-occurrence counts between word combinations to detect errors. In this paper we follow the findings of two recent studies and test their hypotheses within the framework of two different word embedding models, based on their representation of the erroneous usage of concent words. Using an error-annotated corpus of essays written by learnings bellowing to 16 different first languages, we compare incorrect words and their correct replacements as vectors in English and the learners first language. The results are consistent with previous findings that the first language has an influence on errors in the second language. The relationships between typologically similar languages differed between the models of embedding, suggesting an avenue for future explorations.