

Abs-143

Hui-Chuan Lu and Cheng-Yu Chang (National Cheng Kung University, Taiwan)

A corpus-based error detection and improvement suggestion tool for learners of Spanish

In this paper, we will present a newly-developed tool for error detection and possible candidate suggestion based on two heterogeneous corpora in order to make up the lack of similar resources in learning Spanish and benefit foreign language learning in the future.

Our error detector is mainly based on an error-annotated learners' corpus (Corpus of Taiwanese Learners of Spanish, CATE (Lu, 2010)) and it can detect basic errors on several linguistic levels. The main idea is, given an error segment in an original text written by Taiwanese learners, there must exist a mapping in the revised one by native speakers of Spanish. Based on this mapping relation, we not only construct an error detector but also label the disorder permutation as an error for any incoming new text.

On the other hand, the improvement suggestions result from two complementary directions for learners. In the implementation, for each beginning of the error, we cluster the related phrases from revised texts in CATE. These clusters will be one of our two types of suggestion candidates. Besides, we also provide suggestion candidates from the corpus of online resources (e.g. wikipedia, Spanish news and academic articles). Once the segment of an incoming new article has been reported as an error, the top-n candidates from dual directions will appear simultaneously for the correction.

Finally, we will show experimental results of efficiency by undertaking an evaluation task from the point of view of users including learners and educators.

Reference:

Lu, Hui-Chuan. 2010. "An Annotated Taiwanese Learners' Corpus of Spanish, CATE." *Corpus Linguistics and Linguistic Theory*. 6(2), 297-300.