

Building bilingual semantic representations based on a corpus-based statistical learning algorithm

Xiaowei Zhao
Emmanuel College

Abstract: In the current study, we applied a corpus-based statistical learning algorithm to derive semantic representations of words under bilingual situations (English and Chinese). The algorithm relies on the analyses of contextual information extracted from a text corpus, specifically, analyses of word co-occurrences in a large-scale electronic database of text. Particularly, we examined how the semantic structure of L2 words can be built based on and influenced by the semantic representations of L1 words in a sequential L2 learning situation. We got the semantic representations under various conditions and the results were processed and illustrated on self-organizing maps, an unsupervised neural network model that projects the statistical structure of the context onto a 2-D space. We further discussed a couple of factors that affected the validity of the representations.