

Computational explanation of “fiction text effectivity” for vocabulary improvement: Corpus analyses using latent semantic analysis

Keisuke Inohara

The University of Electro-Communications

Akira Utsumi

The University of Electro-Communications

Abstract: Previous studies have suggested that fiction book reading has a stronger positive effect on vocabulary development than nonfiction. In this study, we examined this phenomenon in terms of word appearance information in fiction (story texts), nonfiction (explanation texts), and web text using latent semantic analysis (LSA). In a human experiment with Japanese undergraduates, we replicated fiction (story) text effectivity. Participants who often read story texts achieved the highest vocabulary test scores. Then, in a corpus experiment, we constructed a story text corpus, explanation text corpus, and web text corpus of identical size. Based on these corpora, we calculated the LSA similarities between words, and simulated answering the same vocabulary test as used in the human experiment. The corpus experiment demonstrated the nonfiction (explanation) text effectivity, that is, the explanation corpus was the highest. The cause of discrepancy in the results and the educational implications of this study were also discussed.