

The relationship between similarities computed by LSA and several types of association

Keisuke Inohara

Kyoto University

Takashi Kusumi

Kyoto University

Abstract: Latent Semantic Analysis (LSA) is a computational theory of meaning. Meanings of words are extracted from a large corpus of texts by a statistical method and represented as vectors in a semantic space. It is known that similarities of word-pairs computed by LSA can explain various language processing of human(e.g., Landauer and Dumais, 1997). To know features of LSA, we created four semantic spaces from Japanese corpses: news paper, novels, books (except for novels), and both of novels and books. The similarities of word pairs were compared with scores in different types of association tasks. Participants were asked to associate several concepts(e.g., subordinate categories, synonyms, or action concepts) with stimulus words. As a result, a correlation coefficient between the similarities and association scores of the action concept task is the highest. This finding is consistent with Hare et al.(2009), LSA and similar models reflect people's knowledge about daily events.