

Latent semantic analysis accounts for semantic relatedness of Chinese reversible words

Yuhtsuen Tzeng

National Chung Cheng University

Minglei Chen

National Central University

Walter Kintsch

University of Colorado

Abstract: Latent Semantic Analysis has been a successful theory of language and memory representation for many languages. It is yet to be tested under more distant language systems such as Chinese. A Chinese LSA system was previously built and its validity is sought in this study. We are particularly interested in reversible two-character Chinese words because they feature a new word when character order is switched, e.g., Shihgu means incident but its reversible counterpart Gushih becomes story. We extracted 1980 possible pairs from a large corpus and 750 of them survived as legal reversible word pairs as rated by 32 native speakers. Chinese LSA cosine values were computed for all pairs and their semantic relatedness was categorized as same, related, or unrelated. ANOVA results indicated that average cosine values differ for different semantic relatedness words and the patterns correspond well to human judgment. This supports the validity of Chinese LSA system.