Using Hierarchically Structured Lexicon as Key Clues Solving Data Sparseness Problems in Word Sense Disambiguation: a Case for Korean and its Applications to English and Chinese

Aesun Yoon
Korean Language Processing Laboratory
Dept. of French
Pusan National University
Busan, 609-735, Rep of Korean
asyoon@pusan.ac.kr

Minho Kim, Hyuk-Chul Kwon
Korean Language Processing Laboratory
Dept. of French
Pusan National University
Busan, 609-735, Rep of Korean
{karma, hckwon}pusan.ac.kr

Abstract

Word sense disambiguation (WSD) determines the accuracy of almost all tasks in natural language processing. Korean Processing Laboratory of Pusan National University has been working on efficient automatic WSD methods, especially for Korean language. This paper presents our unsupervised model using hierarchically-structured lexicon, i.e. Korean WordNet (KorLex). KorLex can provide us with key clues for solving data sparseness problems, which are inherent in the unsupervised WSD. The proposed model shows 91.14% average accuracy, which is 26.95% higher than the best performance obtained by a supervised method (Lesk's dictionary-based WSD). Our model obtains also a higher accuracy for English and Chinese, using Princeton WordNet and HowNet.