A computational approach to resolving the polysemy of postpositions in Korean

Seongmin Mun
MoDyCo (UMR 7114), CNRS, Paris Nanterre University
seongmin.mun@parisnanterre.fr

The current on-going project is for the resolution of polysemy involving the Korean postpositions. Korean is a Subject-Object-Verb language, which marks case with dedicated postpositions [9]. In this research project, we investigate the polysemy of postpositions in Korean under the framework of Construction Grammar [5]. A postposition is defined as a function word indicating grammatical information to which it is associated [9]. As a form-function pairing, a postposition can be polysemous in that one form delivers multiple functions [4]. An adverbial postposition -(u)lo, for instance, is either directional or instrumental, the two major functions of this particle [2] (1, 2).

(1) -(u)lo as directional (`(I) went to the road.`)

도로-(으)로 갔다.
tolo-(u)lo ka-ass-ta.
road-DIR go-PST-SE

(2) -(u)lo as instrumental (`(I) went by bicycle´)

자전거-(으)로 갔다.
cacenke-(u)lo ka-ass-ta.
bicycle-INS use-PST-SE

We pose key questions as to what is polysemy of postpositions in Korean and how can computer identify the polysemy of the word? The meaning of a word in a sentence can be approximated by it’s relation to the co-occurring words (dubbed the Distributional Hypothesis, [8]). It is thus assumed that we can identify the polysemy of a word based on information obtained from surrounding words and their network. This account has been implemented by way of NLP methods [6]. In this project, we use several NLP methods (such as SVD [3], PPMI SVD [12, 1], and SGNS [10, 11]) for the analysis of the Sejong corpus [7] which is made by large-scale corpus project in Korea in other to reveal the nature of polysemy involving postpositions in Korean. Currently, we are simultaneously progressing to develop the visualization and to make Gold standard from Sejong corpus. Gold standard includes target postpositions (-ey, -eyse, and -(u)lo) in the sentence and is designed to represent the functional semantic role of postposition (such as agent, experience, mental agent, companion, theme, location, direction, goal, final state, source, instrument, effector, criterion, purpose, content, etc.). With the visualization, we explore the information obtained from surrounding words and their network of a selected postposition. We will share the final system that can explore the distribution of polysemous postpositions and uses the distribution to automatically recognize the functional semantic role of postpositions in the upcoming conference.
Références


