

Zero Anaphors in Chinese Discourse Processing

Chin-Chuan Cheng
City University of Hong Kong
ctcheng@cityu.edu.hk

Abstract

In Chinese, zero anaphors occur frequently. This talk discusses several issues in identification of zero anaphors. Syntactic tagging of words in a running text for detection of anaphora requires specification of argument structure for verbal elements. Verbs requiring a subject or an object can be used to locate the position of zero anaphors. Typical argument structures for verbs, however, often do not take into account the expansion of the structure in discourse such as placing the instrument in object position. Once the place of a zero anaphor is located, the discourse processor will attempt to recover the entity. There may be more than one entity in the previous context that can fit in the argument structure of the verb. It has been discussed in the literature that the recovery of zero anaphors requires extra-linguistic knowledge. However, I will discuss two recovery principles that utilize linguistic information as given in the text. The opening statement of a topic continuity normally provides the element for ellipsis in the statements that follow. So the first principle is to return to the opening statement for recovery. Perceptive verbs, on the other hand, often trigger the search of anaphors in the most recent element. This recency principle can be broken by certain adverbial expressions that mark the turn of the event. These principles apply within a topic continuity. The size of a topic continuity is empirically determined to be about 50 words with some variation.