Zero Anaphora Resolution in Chinese Discourse

Yuzhen Cui, Qinan Hu, Haihua Pan and Jianhua Hu

Department of Chinese, Translation and Linguistics
City University of Hong Kong, Hong Kong
```
{50007840@student., qinan.hu@student., cthpan@,
          ctjhu@}cityu.edu.hk
```

Abstract. This paper explores various factors involved in the resolution of zero anaphora in Chinese discourse. Our study differs from previous ones in distinguishing three types of utterances and using clauses as the unit of resolution. The hierarchical structures of utterances enable us to process inter- and intra-utterance anaphora uniformly. Experimental results show that (1) clauses function significantly better than sentences as the unit of resolution, providing an improvement of precision from 36.0% to 63.4%; (2) the inclusion of cataphors and the use of NP forms as a criterion in Cf ranking do not lead to significant improvement of precision; and (3) when assigning antecedents to more than one zero pronoun in the same utterance, the criterion based on grammatical functions gives rise to better performance than that with linear orders.

1 Introduction

Several studies were conducted on zero anaphora for languages like Chinese [1], Japanese [2], Italian [3] and Turkish [4]. The Chinese study resolves zero pronouns in a part-of-speech tagged and shallow-parsed corpus, focusing on pronouns in topic, subject, or object positions in main clauses. All these studies employ Centering Theory (CT) [5, 6] as their framework.

Several problems are found in previous studies. First, it is not clear what counts as an utterance in Chinese discourse. Previous studies either provide no specification or simply use commas and periods as the indicators of utterance ending. Second, the resolution of zero pronouns in subordinate clauses has not been well studied. Third, when two zero pronouns or more occur in the same utterance, it is unclear when they share the same antecedent and when they do not. Finally, cataphora is often not discussed in previous studies.

2 Zero Anaphora in Chinese Discourse

In this study, the term *utterance* refers to an instance of a sentence which is delimited by periods, exclamations, or question marks, and three types of utterances are distinguished, i.e. simple, compound, and complex utterances. A simple utterance consists