An Empirical Study of Zero Anaphora Resolution in Chinese Based on Centering Model

Ching-Long Yeh and Yi-Jun Chen
Department of Computer Science and Engineering
Tatung University
40 Chungshan N. Rd. 3rd. Section
Taipei 104
R.O.C.
chingyeh@cse.ttu.edu.tw d8806005@cseserv.cse.ttu.edu.tw

Abstract

In this paper, we describe the creation of Chinese zero anaphora resolution rules by performing experiments. The rules were constructed based on the centering model. In the experiments, we selected several texts as testing examples. We compared the referents of zero anaphors in the testing texts identified by hand with the ones resolved by using an algorithm employing a resolution rule. Three rules were used to carry out the experiment. The results show that the rule considering grammatical role criteria and domain knowledge obtained the best result: 85% of zero anaphors in the test texts were correctly resolved. We investigate problems of miss-resolution of zero anaphors in the test text and propose solution to deal with them.

1. Introduction

In Chinese text, anaphors are frequently eliminated, termed zero anaphor (ZA) hereafter, due to their prominence in discourse [LT81]. For example in (1), the topic of the utterance (1a) is 'Electronics stocks,' which is eliminated in the second utterance and the topic of utterance (1c), 'Securities stocks,' is eliminated in the utterance (1d).

(1) a. i

    Electronics stocks were affected by high-tech stocks fallen heavily in America.

    b. i

    (Electronics stocks) continued falling down today.

    c. j

    Securities stocks also had respondence.

    d. j

    (Securities stocks) fell by close one after another on the market.

A simple rule, Rule 1, can be formulated by observing the phenomenon of topic chain in Chinese text. This rule can be used to correctly resolve the referent of the ZA in

(1a), for example.

Rule 1: If a ZA occurs in the topic position of utterance *i*, then its antecedent is the topic of utterance *i-1*.

In general, zero anaphors in Chinese can occur in any grammatical slot with an antecedent that may occur in any grammatical slot, regardless of their distance [LT79]. Thus Rule 1 is obviously insufficient to account for the resolution of ZAs.

Within the theories of discourse, Centering is a computational model, which has been developed as a methodology for the explanation of the local coherence and its relationship to attentional state at the local level and focuses on pronominal and nominal anaphora [GJW83, GJW95]. It is formalized as a system of constraints and rules, which can, as part of a computational discourse model, act to control inference [JW81]. In the centering model, each utterance in a discourse segment has two structures associated with it, called Forward-Looking and Backward-Looking centers, which correspond approximately to Sidner's potential foci and discourse focus [Sid79]. Forward-Looking Centers, C_f, is a set of discourse entities in an utterance, and Backward-Looking Center, C_b, is a special member of this set, which is the discourse entity that the utterance most centrally concerns. Our analysis is based on this computational model to resolve the intersentential ZAs.

In this paper, we aim at formulating rules for the resolution of zero anaphors in Chinese. We start with a rule, Rule 2, formulated by employing the centering model.

Rule 2: For each utterance U_i in a discourse segment $U_1, \ldots, U_m$: If $C_b(U_i)$ is realized by a zero anaphor in U_{i+1} then the $C_b(U_{i+1})$ must be realized by $C_b(U_i)$.

We performed an experiment by using an algorithm employing this rule to see how the ZAs in news text are resolved. The initial result showed that about half of the ZAs could not be correctly resolved. Consequently we considered adding other constraints, such as grammatical role criteria and semantic knowledge, to enhance the rule and get better results. We repeated the experiment and the result showed that about 85% can be correctly resolved by using the new rule. The remaining 15% errors of the ZAs resolution occur because of the lack of sufficient semantic knowledge and the character of locality of centering model. We further investigate these situations and propose an approach to solve the problem.

In the next section we describe this nature of zero anaphora in Chinese. In Section 3, we describe the centering model, and we illustrate the result of the

empirical study we made by observing the industry news we collected in Section 4. The discussion and implementation are in Section 5 and 6, respectively, and finally conclusions and future works are made.

2. Zero Anaphora in Chinese

In Chinese, anaphors can be classified as zero, pronominal and nominal forms, as exemplified in (2) by ^{i}, and , respectively [Chen87][1]. Zero anaphors are generally noun phrases that are understood from the context and do not need to be specified.

(2) a. ^{i}

Zhangsan frightened and ran outside.

   b. ^{i} ^{j}

(He) bumped into a person.

   c. ^{i} ^{j}

He saw clearly that person's appearance.

   d. ^{i} ^{j}

(He) recognized who that man is.

According to [LT81], zero anaphors can be classified as intrasentential or intersentential. Intrasentential zero anaphora occur mainly in topic-prominent constructions, namely, sentences having a topic but not a subject such as the in (3). In this sentence, the noun phrase, (house), is the topic while the subject is not present. In sentences of this sort, subjects, in general, refer to general classes or unspecified noun phrases. In English, *you*, *they* (or more formally *one* is used in this function. This kind of zero anaphor occurs specifically in topic-prominent constructions; they have nothing to do with entities in previous sentences in discourse.

(3)

   The house, (someone) has finished building it.

   In the intersentential case, antecedent and anaphors are located in different sentences. Depending upon the distance between the sentences containing antecedent

---

[1] We use a f_a^b to denote a zero anaphor, where the subscript a is the index of the zero anaphor itself and the superscript b is the index of the referent. A single without any script represents an intrasentential zero anaphor. Also note that a superscript attached to an NP is used to represent the index of the referent.

and anaphor, it can further be divided into two types: immediate and long distance. The former is where the sentence containing the antecedent is immediately followed by the one containing the anaphor, such as f_1^j in (4b) and f_1^k in (4d). For the long distance type, the sentence containing the antecedent and anaphors, on the other hand, are not in immediately succeeding order, such as f_1^i in (4e).

(4) a. ^{i} ^{j}
A crab has four pairs of feet.

    b. f_1^j
(They) are commonly called "tuier."

    c. ^{k}
Since every "tuier"'s joint can only bend downwards,

    d. f_1^k
(it) can't bend backward or forwards.

    e. f_1^i
(When) (it) crawls,

    f. f_2^i
(it) must use the tips of feet on one side to grasp the ground.

    g. f_3^i
(It) then uses the feet on the other side to move upwards.

    h. f_4^i
(It) pushes (its) body towards one side.

   Since Chinese has no inflection, conjugation, or case markers, the pronominal system is relatively simple, as shown in Table 1 [LT81]. A third-person pronoun can be used to replace an intersentential zero anaphor, except for first- and second-person pronouns, without changing the meaning of the sentence. Though the resulting meaning of each sentence is unchanged, the whole discourse becomes less coherent.

Table 1: Pronominal system in Chinese

Number	Person	Pronoun
singular	first	
singular	second	,
singular	third	, ,
plural	first	
plural	second	,
plural	third	, ,

3. Centering Model

Centering has its computational foundations established by Grosz and Sidner [Gro77, Sid79] and were further developed by Grosze, Joshi and Weinstein [GJW83, GJW95]. Within the framework of the centering model, each utterance U in a discourse segment has two structures associated with it, called forward-looking centers, $C_f(U)$, and backward-looking center, $C_b(U)$. The forward-looking centers of U_n, $C_f(U_n)$, depend only on the expressions that constitute that utterance. They are not constrained by features of any previous utterance in the discourse segment (DS), and the elements of $C_f(U_n)$ are partially ordered to reflect relative prominence in U_n. The more highly ranked an element of $C_f(U_n)$, the more likely it is to be $C_b(U_{n+1})$. The highest ranked element of $C_f(U_n)$ that is realized[2] in U_{n+1} is the $C_b(U_{n+1})$.

The set of forward-looking centers, C_f, is ranked according to discourse salience. The highest ranked member of the set of forward-looking centers is referred to as the preferred center, C_p.[3] The preferred center of the utterance U_n represents a prediction about the C_b of the following utterance U_{n+1} and is the most preferred antecedent of an anaphoric or elliptical expression in U_{n+1}. Hence, the most important single construct of the centering model is the ordering of the list of forward-looking centers [WIC94, SH96].

3.1 Constraints and rules

In addition to the structures for centers, C_b, and C_f, the theory of centering specifies a set of constraints and rules [WIC94, GJW95].

Constraints

For each utterance U_i in a discourse segment $U_1, \ldots, U_m$:

1. U_i has exactly one C_b.

2. Every element of $C_f(U_i)$ must be realized in U_i.

3. Ranking of elements in $C_f(U_i)$ guides determination of $C_b(U_{i+1})$.

4. The choice of $C_b(U_i)$ is from $C_f(U_{i-1})$, and can not be from $C_f(U_{i-2})$ or other prior sets of C_f.

---

[2] An utterance U, realizes c if c is an element of the situation described by U, or c is the semantics interpretation of come subpart of U.
[3] The notion of preferred center corresponds to Sider's notion of expected focus [Sid83]

Backward-looking centers, C_bs, are often omitted or pronominalized and discourses that continue centering the same entity are more coherent than those that shift from one center to another. This means that some transitions are preferred over others. These observations are encapsulated in two rules [WIC90, WIC94, GJW95]:

Rules

For each utterance U_i in a discourse segment U_1, …, U_m:

I. I. If any element of $C_f(U_i)$ is realized by a pronoun in U_{i+1} then the $C_b(U_{i+1})$ must be realized by a pronoun also.

II. Sequences of continuation are preferred over sequence of retaining; and sequences of retaining are to be preferred over sequences of shifting.

Rule I represents one function of pronominal reference: the use of a pronoun to realize the C_b signals the hearer that the speaker is continuing to talk about the same thing. Psychological research and cross-linguistic research have validated that the C_b is preferentially realized by a pronoun in English and by equivalent forms (i.e. zero anaphora) in other languages [GJW95].

Rule II reflect the intuition that continuation of the center and the use of retentions when possible to produce smooth transitions to a new center provide a basis for local coherence. The transition states are further described in the next section.

3.2 Transition states

The typology of transitions from U_{i-1} to U_i is based on two factors: whether the $C_b(U_i)$ is the same as $C_b(U_{i-1})$, and whether this discourse entity, $C_b(U_i)$, is the same as the $C_p(U_i)$:

1. $C_b(U_i) = C_b(U_{i-1})$, or $C_b(U_{i-1})$ is undefined.
2. $C_b(U_i) = C_p(U_i)$

If both (1) and (2) hold then a pair continuations across U_n and across U_{n+1}. If (1) holds but (2) does not then the utterances are in a retaining transition, which corresponds to a situation where the speaker is intending to shift onto a new entity in the next utterance. If (1) does not hold then the utterances are in one of the shifting transition states depending on whether or not (2) holds. The definition of transition states is summarized in Table 2 [WIC94].

Table 2: Transition states

	$C_b(U_i) = C_b(U_{i-1})$ or $C_b(U_{i-1})$ is undefined	$C_b(U_i)$ $C_b(U_{i-1})$
$C_b(U_i) = C_p(U_i)$	CONTINUE	SMOOTH-SHIFT
$C_b(U_i)$ $C_p(U_i)$	RETAIN	ROUGH-SHIFT

For illustration purpose, consider the example (1) in Section 1; in the Table 3, the centering structures contain C_b, C_f and C_p where the set of C_f are partially ordered to reflect relative prominence in each utterance. The first two transition states of (1a) and (1b) are CONTINUE corresponding to the two factors, "$C_b(U_i) = C_b(U_{i-1})$, or $C_b(U_{i-1})$ is undefined" and "$C_b(U_i) = C_p(U_i)$, or $C_b(U_i)$ is undefined." In (1c), the transition state is RETAIN because of "$C_b(U_{1c})$ $C_p(U_{1c})$.". SMOOTH-SHIFT is the last transition state of example (1) while "$C_b(U_{1d}) = C_p(U_{1d})$" and "$C_b(U_{1d})$ $C_b(U_{1c})$" hold.

Table 3: Centering structures and transition states for example (1)

(1a)	C_b: undefined C_f: [,] C_p:	CONTINUE
(1b)	C_b: C_f: [ZA ( )] C_p ZA ( )	CONTINUE
(1c)	C_b: C_f: [ ] C_p:	RETAIN
(1d)	C_b: C_f: [ZA ( ),] C_p: ZA ( )	SMOOTH-SHIFT

4. Experiment and Result

This paper is concerned with resolving the problem of zero anaphora in Chinese using the centering model. In this section, we first describe the methodology of zero anaphora resolution we adopted based on centering. Second, we explain how to apply our rules and represent the results of applying the different rules to the test texts.

4.1 Experiment for zero anaphora resolution

The task of zero and nominal anaphora resolution is performed after the semantic interpretation phase that converts the syntactic structure of a sentence into a semantic representation form such as the logic form [JA94]. After semantic interpretation, an anaphor becomes a parameter in a logic form. For example, the logic form of the (5b) is ( ). The task of anaphora resolution is to find out the referent of the omission in the logic forms.

(5) a. i

    Zhangsan bought an apple.

  b. i

    (It) is very fresh.

Recall that the centering model, an utterance, U_i, is associated with a set of forward-looking centers, C_f, with each element an entity in U_i. The highest ranked element in the set, C_p, becomes the prediction of backward-looking centers, C_b, of the following utterance, which is zeroed if it does not violate syntactic constraints, such as the object of a prepositional phrase [LT81]. Therefore to apply the centering model for zero anaphora resolution, the essential task is to rank the elements in the set. The task of ranking elements is determined according to certain rules, for example Rule 2 described previously in Section 1. In this paper, our goal is to develop effective rules to obtain better result.

We performed an experiment to examine the effectiveness of using a rule for the resolution of zero anaphors. In the experiment, we selected a number of industry news as the test texts. Table 4 summarizes the total news, paragraphs, utterances, zero anaphors and words in the test texts.

Table 4: Summary of test texts

	Paragraphs	Utterances	Words	Zero Anaphors
1	4	36	199	25
2	3	26	229	9
3	4	31	213	13
4	4	29	213	15
5	3	27	208	11
6	4	35	282	15
7	3	28	234	14
8	3	27	289	12
Total	**28**	**239**	**1867**	**115**

In the experiment, we first of all identify by hand the referent of each zero

anaphor occurring in the texts. Then we compute the referents of zero anaphors identified by using an algorithm employing a resolution rule. The computed result is then compared with the one by hand to see the correction rate of the resolution rule. The correction rate of a resolution rule is defined as below.

> **Correction rate:** Assume that m ZAs occur in n utterances. The correction rate of a resolution rule is the number of referents of ZAs resolved by an algorithm employing the resolution rule that are identical to the ones identified by hand.

The experiment is performed repeatedly by replacing new rules and it is stopped until promising result is obtained. The initial result of using Rule 2 shows that only 55% of the ZAs are correctly resolved, which is obviously not effective enough. The errors occurs in the initial result may be that Rule 2 does contain enough semantic knowledge. In the following, we propose other rules to replace Rule 2 and compare the results.

4.2 Results of using other rules

Grosz *et al*., in their paper [GJW95], assume that grammatical roles are the major determinant for ranking the forward-looking centers, with the order *"Subject > Object(s) > Others"*. In Chinese, the concept of subject seems to be less significant while the topic in a sentence appears to be crucial in explaining the structure of ordinary sentences in the language [LT81]. By adopting the concept of grammatical roles and topic-prominence in Chinese, we order the grammatical roles in Chinese with topic having the highest priority as shown in Figure 1. The subject and objects occurring in an embedded clause, that is, *Secondary Subject* and *Secondary Objects*, are give lower priority.

> *Topic > Main Subject > Direct Object >*
> *Secondary Subject > Secondary Objects*

Figure 1: Grammatical role criteria

By adding the grammatical role criteria to Rule 2, we obtain a new rule, Rule 3:

> **Rule 3:** For each utterance U_i in a discourse segment $U_1, \ldots , U_m$: If $C_b(U_i)$ is realized by a ZA in U_{i+1} and no other noun phrase having higher priority of grammatical role criteria than the ZA then the $C_b(U_{i+1})$ must be realized by $C_b(U_i)$.

Rule 3 is used to verify if the order of the elements in grammatical role criteria we assumed is helpful to raise the correction rate of zero anaphora resolution. We further developed another rule, Rule 4, by considering the domain knowledge corresponding to the test texts.

> **Rule 4:** For each utterance U_i in a discourse segment $U_1, \ldots, U_m$: If $C_b(U_i)$ is realized by both specific nouns in the lexicon and a ZA having the highest priority of grammatical role criteria in U_{i+1} then the $C_b(U_{i+1})$ must be realized by $C_b(U_i)$.

In Rule 4, in addition to grammatical role criteria, we further add the lexical semantic knowledge to the nouns specified in the lexicon. The experiment results of using these rules are investigated as follows.

4.3 Experiment results using three rules

The experiment is performed three times by using Rule2, 3 and 4, respectively. The first experiment employs the simplest rule, Rule 2, as described in Section 1. Since Rule 2 does not have constraint to order elements in C_f, here we take the surface order of entities from left to right in the utterance. After performing the experiment, the correction rate is 55%, which is obviously not satisfied. In the second experiment, we employed an enhanced rule, Rule 3, and the correction rate is 62%. The result is better but it is still not significant. In the third experiment, we used a further enhanced rule, Rule 4, the correction rate becomes 85%, which is more promising. The results are summarized in Table 5.

Table 5: Summary of experiment results using three rules

	Rule 2	**Rule 3**	**Rule 4**
ZAs correctly resolved	63	71	98
Correction Rate	55%	62%	85%

5. Discussions

We have performed experiments on ZA resolution by using three rules with different complexities. The result is promising to some extent; however, there are still 15% of ZAs in the test texts can not be correctly resolved. In the following, we investigate the problems and propose methods to deal with them. One problem is because of insufficient semantic knowledge, namely domain ontology. In the lexical database,

one word may have several word senses and there is a set of synonyms for each sense [MBF+90]. Besides, one word may have hypernyms, hyponyms, coordinate sisters, and other relationship to another word, e.g.,　　　　'Tatung,' is a hyponym of 'Electronics stocks,' and　　　　　　　'listed securities,' is a hypernym of 'Electronics stocks.' If the domain ontology contains sufficient lexical and semantic knowledge, it would be helpful to analyze a discourse by understanding the context.

Another problem is with the locality of $C_b(U_i)$ as mentioned in Constraint 4 in Section 3.1. The centering model only accounts for local coherence, that is, the computation of C_b and C_f is confined within successive utterances. Thus the rules we proposed in Section 4 can only deal with immediate zero anaphors. For zero anaphors having their antecedents outside this scope, the rules would be ineffective. Worse yet, the miss-resolution of a long distance zero anaphora would fail to resolve the following zero anaphors, or *error chaining* [SH96]. To solve this problem, we extend the referent set of $C_b(U_i)$ to be the collection of entities occurring in utterances previously in the discourse, that is, $U_1 \ldots U_{i-1}$. The referent of a long distance zero anaphor is then determined by examining the elements in the extended referent set. The algorithm for resolving long distance zero anaphors is described as below.

> **Description:** A long distance zero anaphor z is found in the current utterance U_i and then it enters the following procedure. Assume that the extended referent set is E. A temporary set, *temp_set*, is used to record the elements in E that satisfy the semantic constraints of z.
>
> **Procedure:**
> For each element e in E do
> 　　If e satisfies the semantic constraints of z, then add e to *temp_set*.
> end for;
> If there is one element in *temp_set* then return the element as the result;
> else return the element in *temp_set* having longest distance from z as the result.

The semantic constraints we used in the above procedure come from the selectional restrictions of the main verb in utterance U_i [JA94]. This kind of restrictions can be used to select the referents of zero anaphors in the topic position. On the one hand, in the sentence which the topic and subject are identical, the zero anaphor in the topic position is restricted by the semantics of the main verb. On the other hand, for sentences with both topic and subject, the topic is frequently moved from the object position of the sentence. Thus zero anaphors of this sort are restricted by the main verb as well. We ignore the selectional restrictions of other syntactic

constructs such as coverb and adjective phrases because the objects or heads of these kinds of phrases can not be zeroed according to syntactic constraints in Chinese [LT81]. Consider, for example, the long distance zero anaphor f_2^i in (6d). Before entering the above procedure, assume that the extended referent set, { ^{i}, ^{j}, ^{k}, ^{l} }, was obtained, where the first two elements satisfy the selectional restrictions of the main verb of (6d), . Here the first one is selected because it is in a more prominent position.

(6) a. $\quad^i\qquad\qquad^j\qquad\qquad\qquad^k$

People on the market worry that Central Bank will intervene the exchange rate again.

b. f_1^i

(They) are afraid to enter the exchange market.

c. $\quad^l\qquad^k$

The NTD's exchange rate stops to slowly fall down.

d. $f_2^i\qquad\qquad^j\qquad\qquad\qquad\qquad^j$

(They) expect that Central Bank of China will not let NTD be depreciated.

6. Implementation

The goal of this paper is to resolve zero anaphors occurring in discourses based on the centering model. A discourse is a sequence of utterances exhibiting coherence [GJW95]. The resolution of zero anaphors in a discourse is therefore divided into two parts. First, we process each utterance in turn and identify zero anaphors occurring in the utterance. Then we apply a zero anaphor resolution algorithm to resolve the referents of the zero anaphors.

The first part consists of tasks of word segmentation, parsing and semantic interpretation. An input utterance is fragmented into word sequence, and after parsing and semantic interpretation, the semantic form is obtained. Therefore, in this part, the input is a sequence of utterances and the output is the corresponding sequence of semantic forms. Zero anaphors with the information of either immediate or long distance are represented as arguments in the semantic forms. Basically, a zero anaphor is considered an immediate one. But if there are linguistic cues accompanied with the utterance, such as the utterance is the beginning of a new full sentence, and it has initial adverbial connectives, *etc.*, then the zero anaphor is considered a long distance

case. In the second part, the resolution procedure examines each zero anaphor in turn. If an immediate zero anaphor is found, then apply the resolution rules described in Section 4. Otherwise, if it is a long distance zero anaphor, then apply the procedure as described in Section 5. The system architecture is shown in Figure 2.
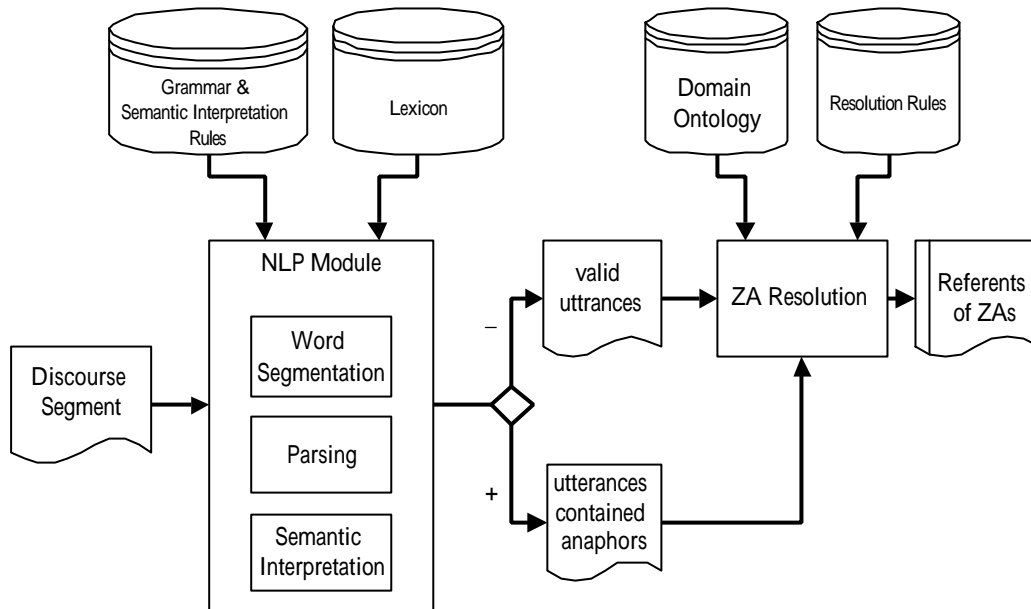

Figure 2: System architecture

In the system the NLP Module carries out in order the work of word segmentation, parsing and semantic interpretation by consulting the lexicon and the syntactic and semantic rules. This module corresponds to the first part described previously in this section. According to the segmentation standard proposed by Academia Sinica [HCC96, HC+97], we built a small lexicon for the test texts, and employ a simple algorithm of word segmentation. The algorithm is according to a strategy that prioritizes the longest word first. The syntactic grammar rules we construct are from the utterances of the test texts and refer to Sinica Corpus and Auto tag program [SC01, CKIP99] and then the parser corresponding to the grammar rules is build as a sentence-level parser in DCG [GM89]. Each utterance within an input discourse segment is converted into a syntactical structure by the parser and the output structure is interpreted to produce the semantic form, which includes the entities in the utterance and is also used to judge whether the utterance contains zero anaphors or not.

ZA resolution by consulting the domain ontology and resolution rules is the

second part of our system. If an input utterance contains a zero anaphor, then apply the resolution rules described in Section 4 to obtain the referent of the zero anaphor. Currently, the ZA Resolution only deals with the immediate zero anaphors. We will extend the algorithm to include the resolution of long distance zero anaphors described in Section 5.

7. Conclusions

In this paper, we performed the experiments on zero anaphora resolution in Chinese based on centering model. In the experiments, 85% of zero anaphors in the test texts were correctly resolved. The remaining zero anaphors were miss-resolved because of lack of sufficient domain knowledge and occurrence of long distance zero anaphors. Since the centering model only focuses on local coherence in discourse, we therefore propose to extend the referent set of a zero anaphor to include all entities occurring previously in the discourse. Though the experiment results are promising to some extent, we found that there are problems that are worth further study. First we need to build domain ontology to get better resolution. Second, the phenomenon of error chaining is inherent in zero anaphors resolution. Thus an effective method is needed to account for this problem. The method we proposed in Section 5 is a step towards solving this problem. Third, the test texts used in this paper were selected from industry news. We will further extend our experiment to include texts from other domains.

References

[Chen87] P. Chen. 1987. *Hanyu lingxin huizhi de huayu fenxi* (a discourse approach to zero anaphora in chinese) (in chinese). Zhongguo Yuwen (Chinese Linguistics), pages 363-378.
[CKIP99] CKIP. 1999.   (Auto tag), Academic Sinica.
[GJW83] B. J. Grosz. A. K. Joshi and S. Weinstein, 1983. Providing a unified account of definite noun phrases in discourse. *Proc. of 21st Annual Meeting of the ACL*
[GS86] B. J. Grosz and C. L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, No 3 Vol 12, pp. 175-204.
[GJW95] B. J. Grosz, A. K. Joshi and S. Weinstein.1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 21(2), pp. 203-225.
[GM89] G. Gazdar and C. Mellish. 1989. *Natural Language Processing in PROLOG – An Introduction to Computational Linguistics*, Addison-

Wesley.

[Gro77]   B. J. Grosz. 1977. The representation and use of focus in dialogue understanding *Technical Report 151*, SRI International.

[HCC96]   Chu-Ren Huang, Keh-Jiann Chen and Li-li Chang. 1996. Segmentation Standard for Chinese Natural Language Processing. *Proceedings of the 1996 International Conference on Computational Linguistics (COLING 96')*, pp.1045-1048. Copenhagen, Denmark.

[HH76]   Halliday, M. A. K. and Hasan, R. 1976. *Cohesion in English.* (English Language Series, 9). London, Longman.

[JA94]   James Allen. 1994. *Natural Language Understanding 2nd ed.*, The Benjamin/Cummings Publishing Company, Inc.

[JW81]   Aravind K. Joshi and Scott Weinstein. 1981. Control of inference: Role of some aspects of discourse structure – centering. In *Proc. International Joint Conference on Artificial Intelligence*.

[Kat97]   Boris Katz. 1997. From Sentence Processing to Information Access on the World Wide Web. *1997 AAAI Spring Symposium*.

[LT79]   Charles N. Li and Sandra A. Thompson. 1979. Third-person pronouns and zero-anaphora in Chinese discourse. In T. Givon, editor, *Syntax and Semantics: Discourse and Syntax,* volume 12, pages 311-335. Academic Press.

[LT81]   Charles N. Li and Sandra A. Thompson. 1981. *Chinese Chinese – A Functional Reference Grammar*, University of California Press.

[MBF+90] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography,* vol. 3(4), pp. 235--244.

[SC01]   (Sinica Corpus). 2001. Academic Sinica. *http://www.sinica.edu.tw/*

[SH96]   Strube, M. and U. Hahn. 1996. *Functional Centering. Proc. Of ACL ' 96*, Santa Cruz, Ca., pp.270-277.

[Sid79]   C. L. Sider. 1979. *Toward a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Ph.D. thesis, MIT.

[Sid83]   C. L. Sider. 1983. Focusing in the comprehension of definite anaphora. *Computational Models of Discourse*, MIT Press.

[WIC90]   Walker, M. A., M. Iida and S. Cote. 1990. Centering in Japanese discourse. *Proc. Of COLING-90*, Appendix, 6pp.

[WIC94]   Walker, M. A., M. Iida and S. Cote, 1994. Japan Discourse and the Process of Centering. *Computational Linguistics,* 20(2): 193-233.

[Yeh95]   Ching-Long Yeh. 1995. *Generation of Anaphors in Chinese*, Ph.D. dissertation, University of Edinburgh