Topic Identification in Chinese Based on Centering Model

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

Abstract

In this paper we are concerned with identifying the topics of sentences in Chinese texts. The key elements of the centering model of local discourse coherence are employed to identify the topic which is the most salient element in a Chinese sentence. Due to the phenomenon of zero anaphora occurring in Chinese texts frequently, in addition to the centering model, we further employ the constraint rules to identify the antecedents of zero anaphors. Unlike most traditional approaches to parsing sentences based on the integration of complex linguistic information and domain knowledge, we work on the output of a part-of-speech tagger and use shallow parsing instead of complex parsing to identify the topics from sentences.

1 Introduction

One of the most striking characteristics in a topic-prominent language like Chinese is the important element, "topic," in a sentence which can represent what the sentence is about (Li and Thompson, 1981). That is, if we can identify topics from Chinese sentences, we can obtain the most information embedded in the text. In this paper, we tend to identify the topic of each utterance within a discourse based on the centering model. However, in many natural languages, elements that can be easily deduced by the reader are frequently omitted from expressions in texts. The elimination of anaphoric expressions is termed zero anaphor (ZA) which often occurs in topic position in a Chinese sentence, due to their prominence in discourse. Accordingly, to accomplish the task of topic identification, we have to solve the problem of zero anaphora resolution.

There are several methods of anaphora resolution. One method is to integrate different knowledge sources or factors (e.g. gender and number agreement, c-command constraints, semantic information) that discount unlikely candidates until a minimal set of plausible candidates is obtained (Grosz et al., 1995; Lappin and Leass, 1994; Okumura and Tamura, 1996; Walker et al., 1998; Yeh and Chen, 2001). Anaphoric relations between anaphors and their antecedents are identified based on the integration of linguistic and domain knowledge. However, it is very labor-intensive and time-consuming to construct a domain knowledge base. Another method employs statistical models or AI techniques, such as machine learning, to compute the most likely candidate (Aone and Bennett, 1995; Connoly et al., 1994; Ge et al., 1998; Seki et al., 2002). This method can sort out the above problems. However, it heavily relies upon the availability of sufficiently large text corpora that are tagged, in particular, with referential information (Stuckardt, 2002).

Our method is an inexpensive, fast and reliable procedure for anaphora resolution, which relies on cheaper and more reliable NLP tools such as part-of-speech (POS) tagger and shallow parsers (Baldwin, 1997; Ferrández et al., 1998; Kennedy and Boguraev, 1996; Mitkov, 1998; Yeh and Chen, 2003). The resolution process works from the output of a POS tagger enriched with annotations of grammatical function of lexical items in the input text stream. The shallow parsing technique is used to detect zero anaphors and identifies the noun phrases preceding the anaphors as antecedents.

In the following sections we first describe the centering model which including the key elements of the centering model of local discourse coherence. In Section 3 we describe the details of shallow parsing we employed. In Section 4 we explain our ZA resolution method based on the centering model and the constraint rules. The method of topic identification in Chinese sentences is illustrated in Section 5. In the last section the conclusions are made.

2   Centering Model

In the centering theory (Grosz and Sidner, 1986; Grosz *et al*, 1995; Walker *et al.*, 1994; Strube and Hahn, 1996), the 'attentional state' was identified as a basic component of discourse structure that consisted of two levels of focusing: global and local. For Grosz and Sidner, the centering theory provided a model for monitoring local focus and yielded the centering model which was designed to account for the difference in the perceived coherence of discourses. In 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. Grosz *et al.*, in their paper (Grosz *et al*, 1995), assume that grammatical roles are the major determinant for ranking the forward-looking centers, with the order "*Subject > Object(s) > Others*". The superlative element of $C_f(U_n)$ may become the C_b of the following utterance, $C_b(U_{n+1})$.

In addition to the structures for centers, C_b, and C_f, the centering model specifies a set of constraints and rules (Grosz *et al*, 1995; Walker *et al*. 1994).

Constraints

For each utterance U_i in a discourse segment consisting of utterances $U_1, \ldots, U_m$:
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.

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:

Rules

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

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 (Grosz *et al*, 1995).

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.

For example in (1), the subjects of the utterance (1b) and (1d) are eliminated, and their antecedents are identified as the subjects of the preceding utterances (1a) and (1c) respectively[1] according to the centering model.

(1) a. 電子股 ^{i} 受 美國高科技股 重挫 影響，
   Electronics stock^i receive USA high-tech stock heavy-fall affect
   Electronics stocks^i were affected by high-tech stocks fallen heavily in America.

   b. ϕ^i 持續 下跌。
   (Electronics stocks)^i continue fall
   (Electronics stocks)^i continued falling down.

   c. 證券股 ^{j} 也 有 相對 回應，
   Securities stocks^j also have relative respondence
   Securities stocks^j also had respondence.

   d. ϕ^j 陸續 下殺 至 跌停。
   (Securities stocks)^j continue fall by close.
   (Securities stocks)^j fell by close one after another.

3   Shallow Parsing

Shallow (or partial) parsing which is an inexpensive, fast and reliable method does not deliver full syntactic analysis but is limited to parsing smaller constituents such as noun phrases or verb phrases (Abney, 1996; Li and Roth, 2001; Mitkov, 1999). For example, the sentence (2) can be divided as follows:

(2) 花蓮 成爲 熱門 的 旅遊 地點。
   Hualien became the popular tourist attraction.
   → [NP 花蓮] [VP 成爲] [NP 熱門 的 旅遊 地點]

---

[1] We use a ϕ_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 ϕ 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.

[NP Hualien] [VP became] [NP the popular tourist attraction]

Given a Chinese sentence, our method of shallow parsing is divided into the following steps: First the sentence is divided into a sequence of POS-tagged words by employing a segmentation program, AUTOTAG, which is a POS tagger developed by CKIP, Academia Sinica (CKIP, 1999). Second the sequence of words is parsed into smaller constituents such as noun phrases and verb phrases with phrase-level parsing. Each phrase is represented as a word list. Then the sequence of word lists is transformed into *triples*, [*S,P,O*]. For example in (3), (3b) is the output of sentence (3a) produced by AUTOTAG, and (3c) is the *triple* representation.

(3) a. [花蓮(Nc) 成爲(VG) 熱門(VH) 的(DE) 旅遊(VA) 地點(Na)]

b. [NP,[花蓮]], [VP,[成爲]], [NP,[熱門,的,旅遊,地點]]

c. [[花蓮], [成爲], [熱門,的,旅遊,地點]]

The definition of *triple* representation is illustrated in Definition 1.The *triple* here is a simple representation which consists of three elements: *S*, *P* and *O* which correspond to the *Subject* (noun phrase), *Predicate* (verb phrase) and *Object* (noun phrase) respectively in a clause.

Definition 1:

A Triple T is characterized by a 3-tuple:

T = [*S*, *P*, *O*] where

- *S* is a list of nouns whose grammatical role is the subject of a clause.
- *P* is a list of verbs or a preposition whose grammatical role is the predicate of a clause.
- *O* is a list of nouns whose grammatical role is the object of a clause.

In the step of *triple* transformation, the sequence of word lists as shown in (3b) is transformed into triples by employing the Triple Rules. The Triple Rules is built by referring to the Chinese syntax. There are four kinds of Triples in the Triple Rules, which corresponds to five basic clauses: subject + transitive verb + object, subject + intransitive verb, subject + preposition + object, and a noun phrase only. The rules listed below are employed in order:

Triple Rules:

Triple1(S,P,O) → np(S), vtp(P), np(O).

Triple2(S,P,*none*) → np(S), vip(P).

Triple3(S,P,O) → np(S), prep(P), np(O).

Triple4(S,*none,none*) → np(S).

The vtp(P) denotes the predicate is a transitive verb phrase, which contains a transitive verb in the rightmost position in the phrase; likewise the vip(P) denotes the predicate is an intransitive verb phrase, which contains an intransitive verb in the rightmost position in the phrase. In the rule Triple3, the prep(P) denotes the predicate is a preposition. The Triple4 is employed if only a sentence contains only one noun phrase and no other constituent. If all the rules in the Triple Rules failed, the ZA Triple Rules are employed to detect zero anaphor (ZA) candidates.

ZA Triple Rules:

Triple1^{z1}(*zero*,P,O)→ vtp(P), np(O).

Triple1^{z2}(S,P,*zero*)→ np(S), vtp(P).

Triple1^{z3}(*zero*,P,*zero*)→ vtp(P).

Triple2^{z1} (*zero*,P,*none*)→ vip(P).

Triple3^{z1}(*zero*,P,O) → prep(P), np(O).

Triple4^{z1}(*zero*,P,O) → co-conj(P), np(O).

The zero anaphora in Chinese generally occurs in the topic, subject or object position. The rules Triple1^{z1}, Triple2^{z1}, and Triple3^{z1} detect the zero anaphora occurring in the topic or subject position. The rule Triple1^{z2} detects the zero anaphora in the object position and Triple1^{z3} detect the zero anaphora occurring in both subject and object positions. In the Triple4, the co-conj(P) denotes a coordinating conjunction appearing in the initial position of a clause. For example in (4), there are two *triples* generated. In the second *triple*, *zero* denotes a zero anaphor according to Triple1^{z1}.

(4) 張三 參加 比賽 贏得 冠軍。
   Zhangsan entered a competition and won the champion.

→ [[[張三], [參加], [比賽]], [[zero], [贏得], [冠軍]]]
   [[[Zhangsan], [enter], [competition]], [[zero], [win], [champion]]]

4 Zero Anaphora Resolution

4.1 ZA Resolution Method

The process of analyzing Chinese zero anaphora is different from general pronoun resolution in English because zero anaphors are not expressed in discourse. Therefore, the ZA resolution method we develop is divided into three phases. First each sentence of an input document is translated into *triples* as described in Section 3. Second is ZA identification that verifies each ZA candidates annotated in *triples* by employing ZA

identification constraints. Third is antecedent identification that identifies the antecedent of each detected ZA using rules based on the centering model.

In the ZA detection phase, we use the ZA Triple Rules described in the Section 3 to detect omitted cases as ZA candidates denoted by *zero* in *triples*. Table 1 shows some examples corresponding to the ZA Triple Rules.

ZA Triple Rule	Example
Triple1^{z1} (*zero*,P,O)	∅ 撞到 一 個 人 (1b) zhuangdao yi ge ren (he) bump-to a person (He) bumped into a person.
Triple1^{z2} (S,P,*zero*)	張三 喜歡 ∅ 嗎 Zhangsan xihuan ma Zhangsan like (somebody or something) Q^2 Does Zhangsan like (somebody or something)?
Triple1^{z3} (*zero*,P,*zero*)	∅ 喜歡 ∅ xihuan (he) like (somebody or something) (He) likes (somebody or something).
Triple2^{z1} (*zero*,P,*none*)	∅ 去 購物 了 qu gouwu le (he) go shopping ASPECT (He) has gone shopping.
Triple3^{z1} (*zero*,P,O)	∅ 在 那邊 zai nabian (he) in there (He) is there.
Triple4^{z1} (*zero*,P,O)	∅ 跟 小朋友 玩 gen xiaopengyou wan (he) with child play (He) is playing with little children.

Table 1: Examples of zero anaphora

After ZA candidates are detected by employing the ZA Triple Rules, the ZA identification constraints are utilized to filter out non-anaphoric cases. In the ZA identification constraints, the constraint 1 is employed to exclude the exophora[3] or cataphora[4] which is different from anaphora in

---

2 We use a Q to denote a question (ma); a ASPECT to denote aspect marker.

3 Exophora is reference of an expression directly to an extralinguistic referent and the referent does not require another expression for its interpretation.

4 Cataphora arises when a reference is made to an entity mentioned subsequently.

texts. The constraint 2 includes some cases might be incorrectly detected as zero anaphors, such as passive sentences or inverted sentences (Hu, 1995).

ZA identification constraints
For each ZA candidate *c* in a discourse:
1. *c* can not be in the first utterance in a discourse segment
2. ZA does not occur in the following case:
NP + *bei* + NP + VP + *c*
NP (topic) + NP (subject) + VP + *c*

In the antecedent identification phase, we employ the concept, 'backward-looking center' of centering model to identify the antecedent of each ZA. First we use noun phrase rules to obtain noun phrases in each utterance, and then the antecedent is identified as the most prominent noun phrase of the preceding utterance (Yeh and Chen, 2001):

Antecedent identification rule:

For each zero anaphor *z* in a discourse segment consisting of utterances $U_1, \dots, U_m$:
If *z* occurs in U_i, and no zero anaphor occurs in U_{i-1}
  then choose the noun phrase with the corresponding grammatical role in U_{i-1} as the antecedent
Else if only one zero anaphor occurs in U_{i-1}
  then choose the antecedent of the zero anaphor in U_{i-1} as the antecedent of *z*
Else if more than one zero anaphor occurs in U_{i-1}
  then choose the antecedent of the zero anaphor in U_{i-1} as the antecedent of *z* according to grammatical role criteria: *Topic > Subject > Object > Others*
End if

Due to topic-prominence in Chinese (Li and Thompson, 1981), topic is the most salient grammatical role. In general, if the topic is omitted, the subject will be in the initial position of an utterance. If the topic and subject are omitted concurrently, the ZA occurs. Since the antecedent identification rule is corresponding to the concept of centering model.

4.2 ZA Resolution Experiment

In the experiment of ZA resolution, we use a test corpus which is a collection of 150 news articles contained 998 paragraphs, 4631 utterances, and 40884 Chinese words. By employing the ZA Triple Rules and ZA identification constraints mentioned previously, zero anaphors occur in topic or subject, and object positions can be detected. Because the ZA Triple Rules cover each possible topic or subject, and object omission cases, the result shows that the zero anaphors are over detected and

the precision rate (PR) is 84% calculated using equation 1.

$$\text{PR of ZA detection} = \frac{\text{No. of ZA correctly detected}}{\text{No. of ZA candidates}} \quad ....(1)$$

The main errors of ZA detection occur in the experiment when parsing inverted sentences and non-anaphoric cases (e.g. exophora or cataphora) (Hu, 1995; Mitkov, 2002). Cataphora is similar to anaphora, the difference being the direction of the reference. In this paper, we do not deal with the case that the referent of a zero anaphor is in the following utterances, but we can detect about 60% cataphora in the test corpus by employing ZA identification constraint 1.

In the phase of antecedent identification, we take the output of employing the ZA Triple Rules and ZA identification constraints, and further to identify the antecedents of zero anaphors by using antecedent identification rule based on the centering model. For example, in the discourse segment (5), the zero anaphors are detected in the utterances (5b) and (5c). According to the antecedent identification rule, the noun phrase, 基隆醫院 'Kee-lung General Hospital,' whose grammatical role is corresponding to the zero anaphor ϕ_1^i in (5b) is identified as the antecedent. Subsequently, the antecedent of the zero anaphor ϕ_2^i in (5c) is identified as the antecedent of ϕ_1^i in (5b), 基隆醫院.

(5) a. 基隆醫院^i 為 擴大 服務 範圍，
　　 Jilong yiyuan wei kuoda fuwu fanwei
　　 Kee-lung hospital for expand service coverage
　　 Kee-lung^i General Hospital aims to increase service coverage.

　 b. ϕ_1^i 積極 提升 醫療 服務 品質 及 標準化，
　　 jiji tisheng yiliao fuwu pinzhi ji biaozhunhua
　　 (Kee-lung General Hospital)^i active improve medical-treatment service quality and standardization
　　 (Kee-lung General Hospital)^i actively improves the service quality of medical treatment and standardization.

　 c. ϕ_2^i 獲 衛生署 認可 為 辦理 外勞體檢醫院。
　　 huo weishengshu renke wei banli wailao tijian yiyuan
　　 (Kee-lung General Hospital)^i obtain Department-of-Health certify to-be handle foreign-laborer physical-examination hospital
　　 (Kee-lung General Hospital)^i is certified by Department of Health as a hospital which

can handle physical examinations of foreign laborers.

The recall rates (RR) and precision rates (PR) of ZA resolution is 70% and 60.3% respectively calculated using equation 2 and equation 3. Errors occur in the phase when a zero anaphor refers to an entity other than the corresponding grammatical role or the antecedent of the zero anaphor in the preceding utterance.

$$\text{RR of ZA resolution} = \frac{\text{No. of ant. correctly identified}}{\text{No. of ZA candidates}} \quad ..(2)$$

$$\text{PR of ZA resolution} = \frac{\text{No. of ant. correctly identified}}{\text{No. of ZA occurred in text}} \quad ...(3)$$

5　Topic Identification

Topic identification is similar to theme identification in (Rambow, 1993). The theme clearly corresponds to the C_b: the theme, under a general definition, is what the current utterance is about; what utterances are about provides a link to previous discourse, since otherwise the text would be incoherent. The role of the C_b is precisely to provide such a link.

In our approach, in addition to the centering model, we further employ the antecedent identification rule to identify the topic. When a ZA occurs in the utterance U_i, the antecedent of the ZA is identified as the topic of U_i. Otherwise, if the transition relation, center shifting, occurs, topic will not be identified as any of the element in the preceding utterance but the element in the current utterance according to grammatical role criteria. The topic identification rule is described below:

Topic identification rule:

For identifying each topic t in a discourse segment consisting of utterances $U_1, \dots , U_m$:
If at least one ZA occurs in U_i
　　 then refer to grammatical role criteria to choose the antecedent of the ZA as the t
Else if no ZA occurs in U_i
　　 then refer to grammatical role criteria to choose one element of U_i as the t
End if

We now take the discourse segment (1) as an example to identify each topic of the utterances (1a) to (1d) by employing the topic identification rule. As illustrated in Table2, the topic of (1a) is 電子股 'Electronics stocks,' and the topic of (1b) is identified as the antecedent of ϕ^i, 電子股 'Electronics stocks.' Similarly, the topic of (1d) is 證券股 'Securities stocks,' which is the same as the topic of (1c).

Utterance	Topic
(1a) 電子股^i 受 美國高科技股 重挫 影響， Electronics stocks^i were affected by high-tech stocks fallen heavily in America	電子股 Electronics stocks
(1b) ϕ^i 持續 下跌。 (Electronics stocks)^i continued falling down.	電子股 Electronics stocks
(1c) 證券股^j 也 有 相對回應， Securities stocks^j also had respondence	證券股 Securities stocks
(1d) ϕ^j 陸續 下殺 至 跌停。 (Securities stocks)^j fell by close one after another	證券股 Securities stocks

Table 2: Examples of zero anaphora

6 Conclusion

In this paper, we propose a method of topic identification in Chinese based on the centering model. Based on observations on real texts, we found that to identify the topics in Chinese context is much related to the issue of zero anaphora resolution. We use a zero anaphora resolution method to resolve the problem of ellipsis in Chinese text. The zero anaphora resolution method works on the output of a part-of-speech tagger and employs a shallow parsing instead of a complex parsing to resolve zero anaphors in Chinese text. Due to time limit, we have not applied the result of topic identification to applications for evaluation. We will further continue improving the accuracy of zero anaphora resolution and develop the applications based on topic identification, such as information extraction/retrieval and text categorization.

7 Acknowledgements

References

Abney, Steven. 1996. Tagging and Partial Parsing. In: Ken Church, Steve Young, and Gerrit Bloothooft (eds.), *Corpus-Based Methods in Language and Speech*. An ELSNET volume. Kluwer Academic Publishers, Dordrecht.

Aone, Chinatsu and Bennett, Scott William. 1995. Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd Annual Meeting of the ACL*, Santa Cruz, CA, pages 122–129.

Baldwin B. 1997. CogNIAC: high precision coreference with limited knowledge and linguistic resources. ACL/EACL workshop on Operational factors in practical, robust anaphor resolution.

CKIP. 1999. 中文自動斷詞系統 Version 1.0 (Autotag), *http://godel.iis.sinica.edu.tw/CKIP/*, Academia Sinica.

Connoly, Dennis, John D. Burger & David S. Day. 1994. A Machine learning approach to anaphoric reference. *Proceedings of the International Conference on New Methods in Language Processing*, 255-261, Manchester, United Kingdom.

Eleni Miltsakaki. 1999. Locating Topics in Text Processing. In *Proceedings of CLIN '99*.

Ferrández, A., Palomar, Manuel and Moreno, Lidia. 1998. Anaphor Resolution in Unrestricted Texts with Partial Parsing. *Proceedings of the 18th International Conference on Computational Linguistics (COLING'98)/ACL'98 Conference*, pages 385-391. Montreal, Canada.

Ge, Niyu, Hale, John and Charniak, Eugene. 1998. A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora*, pages 161 –170.

Grosz, B. J. and Sidner, C. L.. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics,* No 3 Vol 12, pp. 175-204.

Grosz, B. J., Joshi, A. K. and Weinstein, S. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics,* 21(2), pp. 203-225.

Hu, Wenze. 1995. *Functional Perspectives and Chinese Word Order*. Ph. D. dissertation, The Ohio State University.

Kennedy, Christopher and Branimir Boguraev. 1996. Anaphora for everyone: pronominal anaphora resolution without a parser. *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*, 113-118. Copenhagen, Denmark.

Lappin, S. and Leass, H. 1994. An algorithm for pronominal anaphor resolution. *Computational Linguistics*, 20(4).

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

Li, X.; Roth D. 2001. Exploring Evidence for Shallow Parsing. *Proceedings of Workshop on Computational Natural Language Learning*, Toulouse, France.

Okumura, Manabu and Kouji Tamura. 1996. Zero pronoun resolution in Japanese discourse based on centering theory. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, 871-876.

Mitkov, Ruslan. 1998. Robust pronoun resolution with limited knowledge. *Proceedings of the 18th International Conference on Computational Linguistics (COLING'98)/ACL'98 Conference*. Montreal, Canada.

Mitkov, Ruslan. 1999. Anaphora resolution: the state of the art. Working paper (Based on the COLING'98/ACL'98 tutorial on anaphora resolution). University of Wolverhampton, Wolverhampton.

Mitkov, Ruslan. 2002. *Anaphora Resolution*, Longman.

Rambow, O. (1993). Pragmatic aspects of scrambling and topicalization in German: A Centering Approach. In *IRCS Workshop on Centering in Discourse*. Univ. of Pennsylvania, 1993.

Seki, Kazuhiro, Fujii, Atsushi, and Ishikawa, Tetsuya. 2002. A Probabilistic Method for Analyzing Japanese Anaphora Integrating Zero Pronoun Detection and Resolution. *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, pp.911-917.

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

Stuckardt, Roland. 2002. Machine-Learning-Based vs. Manually Designed Approaches to Anaphor Resolution: the Best of Two Worlds. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC2002)*, University of Lisbon, Portugal, pages 211-216.

Givón, T. 1983. Topic continuity in discourse: An introduction. *Topic continuity in discourse*. Amsterdam/Philadelphia [Pennsylvania]: John Benjamins.

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

Walker, M. A. 1998. Centering, anaphora resolution, and discourse structure. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, Centering in Discourse. Oxford University Press.

Yeh, Ching-Long and Chen, Yi-Chun. 2001. An empirical study of zero anaphora resolution in Chinese based on centering theory. In *Proceedings of ROCLING* XIV, Tainan, Taiwan.

Yeh, Ching-Long and Chen, Yi-Chun. 2003. Zero anaphora resolution in Chinese with partial parsing based on centering theory. In *Proceedings of IEEE NLP-KE03*, Beijing, China.