

Automatic Pronominal Anaphora Resolution in English Texts

Tyne Liang and Dian-Song Wu
Department of Computer and Information Science
National Chiao Tung University
Hsinchu, Taiwan
Email: tliang@cis.nctu.edu.tw; gis90507@cis.nctu.edu.tw

Abstract

Anaphora is a common phenomenon in discourses as well as an important research issue in the applications of natural language processing. In this paper, the anaphora resolution is achieved by employing WordNet ontology and heuristic rules. The proposed system identifies both intra-sentential and inter-sentential antecedents of anaphors. Information about animacy is obtained by analyzing the hierarchical relation of nouns and verbs in the surrounding context. The identification of animacy entities and pleonastic-it usage in English discourses are employed to promote the resolution accuracy.

1. Introduction

1.1 Problem description

Anaphora resolution is vital for areas such as machine translation, summarization, question-answering system and so on. In machine translating, anaphora must be resolved for languages that mark the gender of pronouns. One main drawback with most current machine translation systems is that the translation usually does not go beyond sentence level, and so does not deal with discourse understanding successfully. Inter-sentential anaphora resolution would thus be a great assistance to the development of machine translation systems. On the other hand, many of automatic text summarization systems apply a scoring mechanism to identify the most salient sentences. However, the task result is not always guaranteed to be coherent with each other. It could lead to errors if the selected sentence contains anaphoric expressions. To improve the accuracy of extracting important sentences, it is essential to solve the problem of anaphoric references in advance.

Pronominal anaphora is the most common phenomenon which the pronouns are substituted with previous mentioned entities. This type of anaphora can be further divided into four subclasses, namely,

Nominative: {he, she, it, they}

Reflexive: {himself, herself, itself, themselves}

Possessive: {his, her, its, their}

Objective: {him, her, it, them}

However, the usage of “it” can also be a non-anaphoric expression which does not refer to any items mentioned before and is called expletive or pleonastic-it [Lappin and Leass, 94]. Although pleonastic pronouns are not considered anaphoric since they do not have an antecedent to refer to, yet recognizing such occurrences is essential during anaphora resolution. In [Mitkov, 01], the non-anaphoric pronouns are in average of 14.2% from a corpus of 28,272 words.

Definite noun phrase anaphora occurs in the situation that the antecedent is referred by a general concept entity. The general concept entity can be a semantically close phrase such as synonyms or superordinates of the antecedent [Mitkov, 99]. The word *one* has a number of different uses apart from counting. One of the important functions is as an anaphoric form. For example:

Mike has a white shirt and Jane has a red **one**.

Intra-sentential anaphora means that the anaphor and the corresponding antecedent occur in the same sentence. Inter-sentential anaphora is where the antecedent occurs in a sentence prior to the sentence with the anaphor. In [Lappin and Leass, 94], there are 15.9% of Inter-sentential cases and 84.1% Intra-sentential cases in their testing result. In the report of [Mitkov, 01], there are 33.4% of Inter-sentential cases and 66.6% Intra-sentential cases.

1.2 Related works

Traditionally, anaphora resolution systems rely on syntactic, semantic or pragmatic clues to identify the antecedent of an anaphor. Hobbs’ algorithm [Hobbs, 76] is the first syntax-oriented method presented in this research domain. From the result of syntactic tree, they check the number and gender agreement between antecedent candidates and a specified pronoun. In RAP (Resolution of Anaphora Procedure) proposed by Lappin and Leass [94], the algorithm applies to the syntactic representations generated by McCord's Slot Grammar parser, and relies on salience measures derived from syntactic structure. It does not make use of semantic information or real world knowledge in choosing among the candidates. A modified version of RAP system is proposed by [Kennedy and Boguraev, 96]. It depends only on part-of-speech tagging with a shallow syntactic parse indicating grammatical role

of NPs and containment in an adjunct or noun phrase.

In [Cardie et al., 99], they treated coreference as a clustering task. Then a distance metric function was used to decide whether these two noun phrases are similar or not. In [Denber, 98], an algorithm called Anaphora Matcher (AM) is implemented to handle inter-sentential anaphora over a two-sentence context. It uses information about the sentence as well as real world semantic knowledge obtained from outer sources. The lexical database system WordNet is utilized to acquire the semantic clues about the words in the input sentences. He declared that most anaphora does not refer back more than one sentence in any case. Thus a two-sentence “window size” is sufficient for anaphora resolution in the domain of image queries.

A statistical approach was introduced by [Dagan and Itai, 90], in which the corpus information was used to disambiguate pronouns. It is an alternative solution to the syntactical dependent constraints knowledge. Their experiment makes an attempt to resolve references of the pronoun “it” in sentences randomly selected from the corpus. The model uses a statistical feature of the co-occurrence patterns obtained from the corpus to find out the antecedent. The antecedent candidate with the highest frequency in the co-occurrence patterns are selected to match the anaphor.

A knowledge-poor approach is proposed by [Mitkov, 98], it can also be applied to different languages (English, Polish, and Arabic). The main components of this method are so-called “antecedent indicators” which are used for assigning scores (2, 1, 0, -1) against each candidate noun phrases. They play a decisive role in tracking down the antecedent from a set of possible candidates. CogNIAC (COGnition eNIAC) [Baldwin, 97] is a system developed at the University of Pennsylvania to resolve pronouns with limited knowledge and linguistic resources. It presents a high precision pronoun resolution system that is capable of greater than 90% precision with 60% recall for some pronouns. [Mitkov, 02] presented a new, advanced and completely revamped version of Mitkov’s knowledge-poor approach to pronoun resolution. In contrast to most anaphora resolution approaches, the system MARS, operates in fully automatic mode. The three new indicators that were included in MARS are Boost Pronoun, Syntactic Parallelism and Frequent Candidates.

In [Mitkov, 01], they proposed an evaluation environment for comparing anaphora resolution algorithms which is illustrated by presenting the results of the comparative evaluation on the basis of several evaluation measures. Their testing corpus contains 28,272 words, with 19,305 noun phrases and 422 pronouns, out of which 362 are anaphoric expressions. The overall success rate calculated for the 422 pronouns found in the texts was 56.9% for Mitkov’s method, 49.72% for Cogniac and 61.6% for Kennedy and Boguraev’s method.

2. System Architecture

2.1 Proposed System Overview

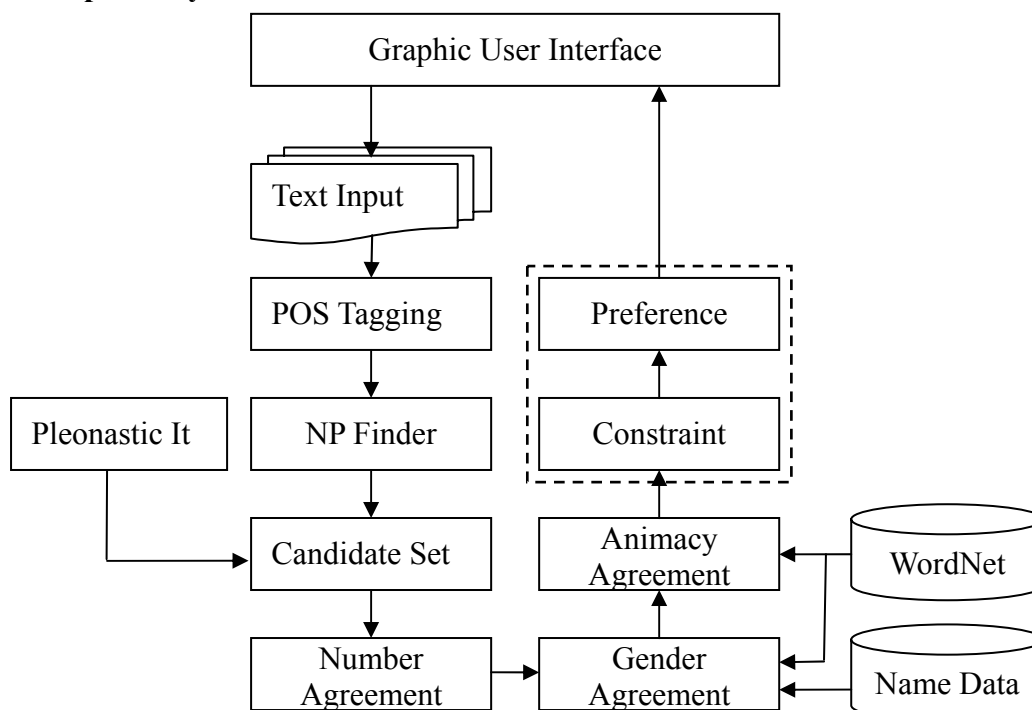


Figure 1: Architecture overview.

The procedure to identify antecedents is described as follows:

1. Each text is parsed into sentences and tagged by POS tagger. An internal representation data structure with essential information (such as sentence offset, word offset, word POS, base form, etc.) is stored.
2. Base noun phrases in each sentence will be identified by NP finder module and stored in a global data structure. Then the number agreement is implemented on the head noun. Testing capitalized nouns in the name gazetteer to find out the person names. The gender feature is attached to the name if it can be found uniquely in male or female class. In this phase, WordNet is also used to find out possible gender clues to increase resolution performance. The gender attribute is ignored to avoid the ambiguity while the noun can be masculine or feminine.
3. Anaphors are checked sequentially from the beginning of the first sentence. They are stored in the list with information of sentence offset and word offset in order. Then pleonastic-it is checked so that no further attempt for resolution is made.
4. The remaining noun phrases preceding the anaphor within predefined

window size are collected as antecedent candidates. Then the candidate set is furtherly filtered by the gender and animacy agreement.

5. The remaining candidates are evaluated by heuristic rules afterward. These rules can be classified into preference rules and constraint rules. A scoring equation (equation 1) is made to evaluate how likely a candidate will be selected as the antecedent.

$$score(can, ana) = (\sum_i rule_pre_i - \sum_j rule_con_j) \times \prod_k agreement_k \quad (1)$$

where

can: each candidate noun phrase for the specified anaphor

ana: anaphor to be resolved

rule_pre_i: the *i*th preference rule

rule_con_i: the *i*th constraint rule

agreement_k: denotes number agreement, gender agreement and animacy agreement

2.2 Main Components

2.2.1 POS Tagging

The TOSCA-ICLE tagger [Aarts et al., 97] was used for the lemmatization and tagging of English learner corpora. The TOSCA-ICLE tagset consists of 16 major wordclasses. These major wordclasses may further be specified by features for subclasses as well as for a variety of syntactic, semantic and morphological characteristics.

2.2.2 NP Finder

According to part-of-speech result, the basic noun phrase patterns are found as follows:

base NP → modifier + head noun

modifier → <article| number| present participle| past participle |adjective| noun>

In this paper, the proposed base noun phrase finder is implemented on the basis of a finite state machine (figure 2). Each state indicates a particular part-of-speech of a word. The arcs between states mean a word input from the sentence sequentially. If a word sequence can be recognized from the initial state and ends in a final state, it is accepted as a base noun phrase with no recursion, otherwise rejected. An example of base noun phrase output is illustrated in figure 3.

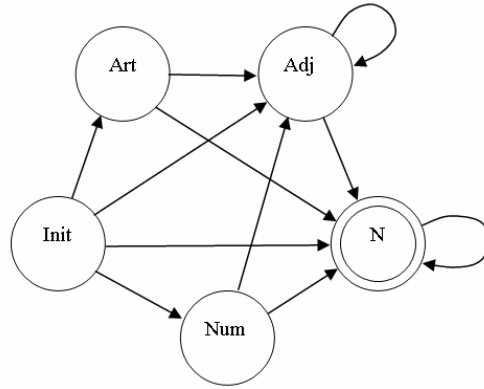


Figure 2: Finite state machine for a noun phrase.

The ministers agreed on a plan to boost screening at international departure points,
 bar travelers with SARS symptoms, and require health declaration forms for visitors
 from affected countries.

Figure 3: An Example output of base noun phrase.

2.2.3 Pleonastic-it Module

The pleonastic-it module is used to filter out those semantic empty usage conditions which is essential for pronominal anaphora resolution. A pronoun it is said to be pleonastic when it is used in a discourse where the pronoun has no antecedent.

The usage of “pleonastic-it” can be classified into state reference and passive reference [Denber, 98]. State references are usually used for assertions about the weather or the time, and it is furtherly divided into meteorological references and temporal references.

Passive references consist of modal adjectives and cognitive verbs. The modal adjectives (**Modaladj**) like advisable, convenient, desirable, difficult, easy, economical, certain, etc. are specified. The set of modal adjectives is extended with their comparative and superlative forms. Cognitive verbs (**Cogv**), on the other hand, are like anticipate, assume, believe, expect, know, recommend, think, etc.

Most of "pleonastic-it" can be described as the following patterns:

1. It is **Modaladj** that **S**
2. It is **Modaladj** (for **NP**) to **VP**
3. It is **Cogv-ed** that **S**
4. It seems/appears/means/follows (that) **S**
5. **NP** makes/finds it **Modaladj** (for **NP**) to **VP**
6. It is time to **VP**
7. It is thanks to **NP** that **S**

2.2.4 Number Agreement

Number is the quantity that distinguishes between singular (one entity) and plural (numerous entities). It makes the process of deciding candidates easier since they must be consistent in number. With the output of tagger, all the noun phrases and pronouns are annotated with number (single or plural). For a specified pronoun, we can discard those noun phrases whose numbers differ from the pronoun.

2.2.5 Gender Agreement

Gender recognition process can deal with words that have gender features. To distinguish the gender information of a person, we collect an English first name list from (<http://www.behindthename.com/>) covering 5,661 male first name entries and 5,087 female ones. Besides, we employ some useful clues from WordNet result by using keyword search around the query result. These keywords can be divided into two classes :

Class_Female= {feminine, female, woman, women}

Class_Male= {masculine, male, man, men}

2.2.6 Animacy Agreement

Animacy denotes the living entities which can be referred by some gender-marked pronouns (he, she, him, her, his, hers, himself, herself) in texts. Conventionally, animate entities include people and animals. Since we can hardly obtain the property of animacy with respect to a noun phrase by its surface morphology, we make use of WordNet [Miller, 93] for the recognition of animate entities. In which a noun can only have a hypernym but many hyponyms (an opposite relation to hypernym). In the light of twenty-five unique beginners, we can observe that two of them can be taken as the representation of animacy. These two unique beginners are {animal, fauna} and {person, human being}. Since all the hyponyms inherit the properties from their hypernyms, the animacy of a noun can be achieved by making use of this hierarchical relation. However, a noun may have several senses with the change of different contexts. The output result with respect to a noun must be employed to resolve this problem. First of all, a threshold value t_noun is defined (equation 2) as the ratio of the number of senses in animacy files to the number of total senses. This threshold value can be obtained by training on a corpus and the value is selected when the accuracy rate reaches the maximum.

$$t_noun = \frac{\text{the_number_of_senses_in_animacy_files}}{\text{the_total_senses_of_the_noun}} \quad (2)$$

$$t_verb = \frac{\text{the_number_of_senses_in_animacy_files}}{\text{the_total_senses_of_the_verb}} \quad (3)$$

$$\text{accuracy} = \frac{\text{the_number_of_animacy_entities_identified_correctly}}{\text{the_total_number_of_animacy_entities}} \quad (4)$$

Besides the utilization of noun hypernym relation, unique beginners of verbs are taken into consideration as well. These lexicographer files with respect to verb synsets are {cognition}, {communication}, {emotion}, and {social} (table 1). The sense of a verb, for example “read”, varies from context to context as well. We can also define a threshold value t_verb as the ratio of the number of senses in animacy files (table 1) to the number of total senses.

Table 1: Example of animate verb.

Unique beginners	Example of verb
{cognition}	Think, analyze, judge ...
{communication}	Tell, ask, teach ...
{emotion}	Feel, love, fear ...
{social}	Participate, make, establish ...

The training data from the Brown corpus consists of 10,134 words, 2,155 noun phrases, and 517 animacy entities. It shows that 24% of the noun phrases in the corpus refer to animacy entities whereas 76% of them refer to inanimacy ones. Threshold values can be obtained by training on the corpus and select the value when the accuracy rate (equation 4) reaches the maximum. Therefore t_noun and t_verb are achieved to be 0.8 and 0.9 respectively according to the observation in figure 4.

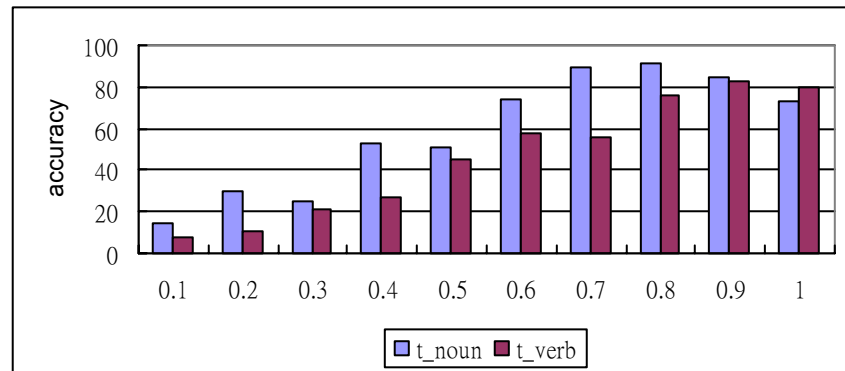
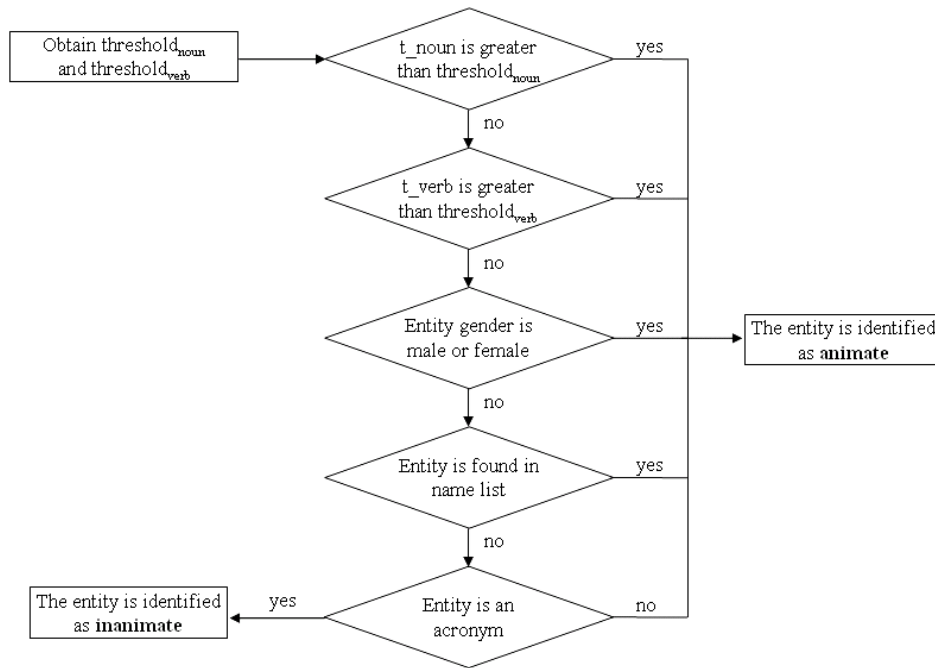


Figure 4: Thresholds of Animacy Entities.

The process of determining whether a noun phrase belong to animacy or not is described below :



2.2.7 Heuristic Rules

I. Syntactic parallelism rule

The syntactic parallelism could be an important clue while other constraints or preferences could not be employed to identify an unique unambiguous antecedent. It denotes the preference that correct antecedent has the same part-of-speech and grammatical function as the anaphor. The grammatical function of nouns can be subject, object or subject complement. The subject is the person, thing, concept or idea that is the topic of the sentence. The object is directly or indirectly affected by the nature of the verb. Words which follow verbs are not always direct or indirect objects. After a particular kind of verb, nouns remain in the subjective case. We call these subjective completions or subject complements.

For example:

The security guard took off **the uniform** after getting off duty.

He put **it** in the bottom of the closet.

The “**He**” (the subject) in the second sentence refers to “**The security guard**” which is also the subject of the first sentence. In the same way, the “**it**” refers to “**the uniform**” which is the object of the first sentence as well. Empirical evidence also shows that anaphors usually match their antecedents in their syntactic functions.

II. Semantic parallelism rule

This preference works with identifying collocation patterns in which anaphora took place. In this way, system can automatically identify semantic roles and employ them to select the most appropriate candidate. Collocation relations specify the relation between words that tend to co-occur in the same lexical contexts. It

emphasizes that noun phrases which have the same semantic role as the anaphor are favored.

III. Definiteness rule

Definiteness is a category concerned with the grammaticalization of identifiability and nonidentifiability of referents. A definite noun phrase is a noun phrase that starts with the word "the", for example, "the young lady" is a definite noun phrase. Definite noun phrases which can be identified uniquely are more likely to be the antecedent of anaphors than indefinite ones.

IV. Mention Frequency rule

Iterated items in the context are regarded as the likely candidates for the antecedent of an anaphor. Generally, the high frequent mentioned items denote the focus of the topic as well as the most likely candidate.

V. Sentence recency rule

Recency information is employed by most of the implementations for anaphora resolution. In [Lappin, 94] the recency factor is the one with highest weight among a set of factors that influence the choice of antecedent. The recency factor states that if there are two (or more) candidate antecedents for an anaphor and all of these candidates satisfy the consistency restrictions for the anaphor (i.e. they are qualified candidates) then the most recent one (the one closest to the anaphor) is chosen. In [Mitkov et al., 01], the average distance (in sentences) between the anaphor and the antecedent is 1.3, and the average distance in noun phrases is 4.3 NPs.

VI. Non-prepositional noun phrase rule

A noun phrase not contained in another noun phrase is favored as the possible candidate. This condition can be explained from the perspective of functional ranking: subject > direct object > indirect object. A noun phrase embedded in a prepositional noun phrase is usually an indirect object.

VII. Conjunction constraint rule

Conjunctions are usually used to link words, phrases and clauses. If the candidate is connected with the anaphor by a conjunction, they can hardly have anaphora relation.

For example:

Mr. Brown teaches in a high school. Both **Jane** and **he** enjoy watching the movies in the weekend.

2.3 The Brown Corpus

The training and testing text are selected randomly from the Brown corpus. The Corpus is divided into 500 samples of about 2000 words each. The samples represent a wide range of styles and varieties of prose. The main categories are listed in figure 5.

A. Press: Reportage	J. Learned
B. Press: Editorial	K. General Fiction
C. Press: Reviews	L. Mystery and Detective Fiction
D. Religion	M. Science Fiction
E. Skills and Hobbies	N. Adventure and Western Fiction
F. Popular Lore	P. Romance and Love Story
G. Biography, Memoirs, etc.	R. Humor
H. Miscellaneous	

Figure 5: Categories of the Brown corpus.

2.4 System functions

The main system window is shown in figure 6. The text editor is used to input raw text without any annotations and shows the analyzed result. The POS tagger component takes the input text and outputs tokens, lemmas, most likely tags and the number of alternative tags. NP chunker makes use of finite state machine (FSM) to recognize strings belong to a specified regular set.

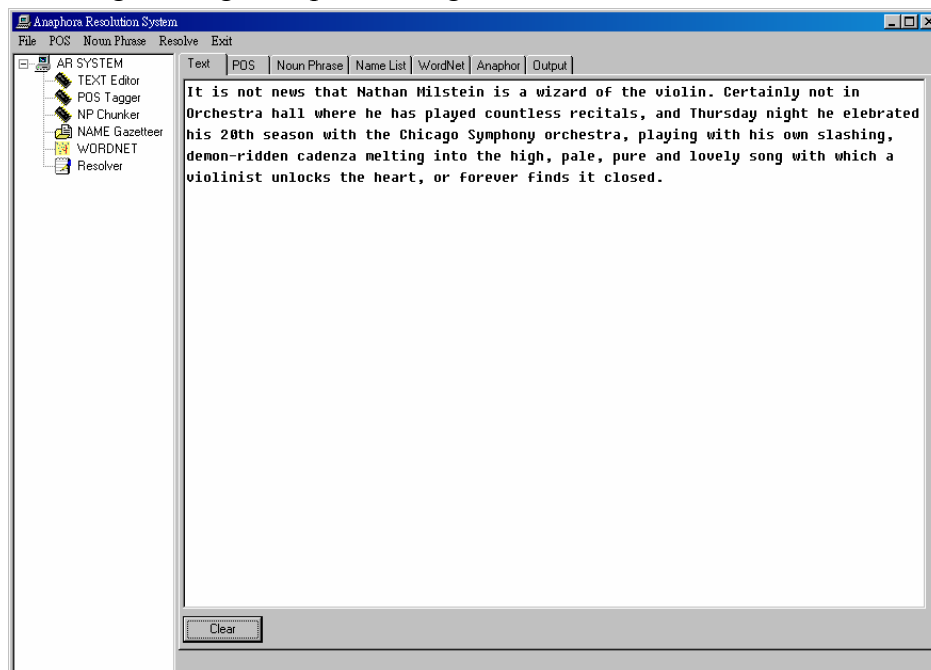


Figure 6: The main system window.

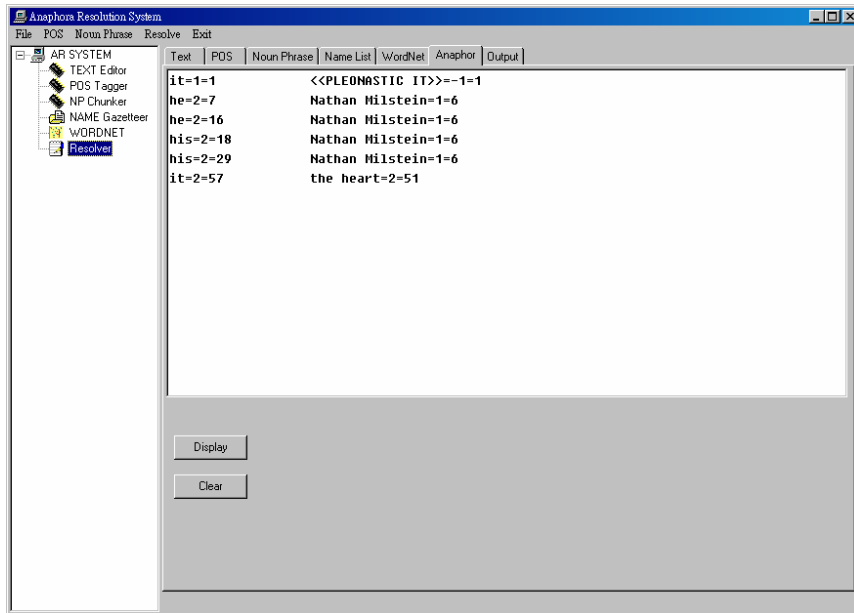


Figure 7: Anaphora pairs.

After performing the selection procedure, the most appropriate antecedent is chosen to match each anaphor in the text. Figure 7 illustrates the result of anaphora pairs in each line in which sentence number and word number are attached to the end of the entities. For example, the “it” in the first word of the first sentence denotes a pleonastic-it and the other “it” in the 57th word of the second sentence refers to “the heart”. Figure 8 shows the original text input with antecedent annotation followed each anaphor in the text. All the annotations are highlighted to make it easy to carry out the subsequent testing purposes.

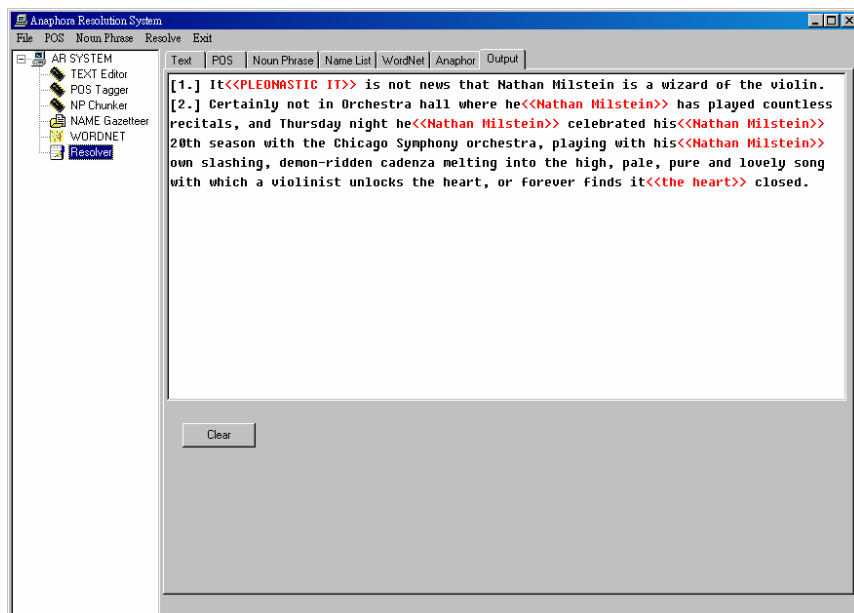


Figure 8: Anaphor with antecedent annotation.

3. Experimental Results and Analysis

The proposed system is developed in the following environment (table 2).

Table 2: System environment.

Operating System	Microsoft Windows 2000 Advanced Server
Main Processor	AMD Athlon K7 866MHZ
Main Memory	256 MB SDRAM
Graphic Card	NVIDIA Geforce2 Mx 32M
Programming language	Borland C++ Builder 5.0

The evaluation task is based on random texts selected from the Brown corpus of different genres. There are 14,124 words, 2,970 noun phrases and 530 anaphors in the testing data. Two baseline models are set up to compare the effectiveness with our proposed anaphora resolution (AR) system. The first baseline model (called baseline subject) performs the number and gender agreement between candidates and anaphors, and then chooses the most recent subject as the antecedent from the candidate set. The second baseline model (called baseline recent) performs a similar procedure but it selects the most recent noun phrase as the antecedent which matches the number and gender agreement with the anaphor. The measurement can be calculated as follows:

$$\text{Success Rate} = \frac{\text{number of correctly resolved anaphors}}{\text{number of all anaphors}} \quad (5)$$

In the result of our experiment baseline subject (table 3), there are 41% of antecedents can be identified by finding the most recent subject, however, only 17% of antecedents can be resolved by means of selecting the most recent noun phrase with the same gender and number agreement to anaphors.

Table 3: Success rate of baseline models.

Genre	Baseline subject	Baseline recent
Reportage	52%	26%
Editorial	48%	15%
Reviews	32%	13%
Religion	44%	22%
Skills	41%	13%
Lore	31%	11%
Average	41%	17%

Figure 9 presents the distribution of sentence distance between antecedents and anaphors. The value 0 denotes intra-sentential anaphora and other values mean inter-sentential anaphora. Figure 10 shows the average word distance distribution with respect to each genre. The identification of pleonastic-it can be achieved to 89% accuracy (table 4).

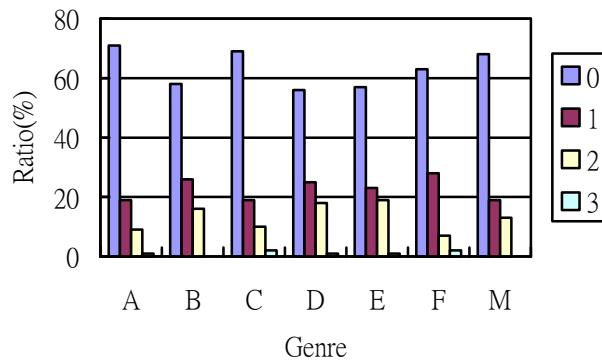


Figure 9: Referential sentence distance distribution.

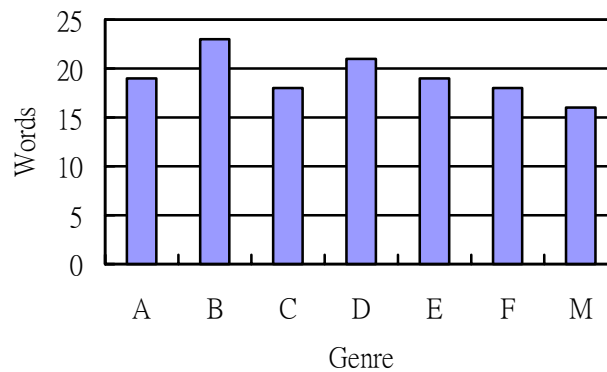


Figure 10: Referential word distance distribution.

Table 4: Pleonastic-it identification.

	Number of Anaphora	Anaphoric expression	Number of Pleonastic-it	Ratio of Pleonastic-it	Accuracy of identification
Total	530	483	47	9%	89%

The evaluation result of our system which applies animacy agreement and heuristic rules for resolution is listed in table 5. It also contains the results for each individual genre of testing data and the overall success rate reaches 77%.

Table 5: Success rate of AR system.

Genre	Words	Lines	NPs	Anims	Anaphors	Success Rate
Reportage	1972	90	488	110	52	80%
Editorial	1967	95	458	54	54	80%
Reviews	2104	113	480	121	92	79%
Religion	2002	80	395	75	68	76%
Skills	2027	89	391	67	89	78%
Lore	2018	75	434	51	69	69%
Fiction	2034	120	324	53	106	79%
Total	14124	662	2970	531	530	77%

4. Conclusion and Future Work

In this paper, the WordNet ontology and heuristic rules are adopted to the anaphora resolution. The recognition of animacy entities and gender features in the discourses is helpful to the promotion of resolution accuracy. The proposed system is able to deal with intra-sentential and inter-sentential anaphora in English text and includes an appropriate treatment of pleonastic pronouns. From experiment results, our proposed method is comparable with prior works using fully parsing of the text. In contrast to most anaphora resolution approaches, our system benefits from the recognition of animacy occurrence and operates in fully automatic mode to achieve optimal performance. With the growing interest in natural language processing and its various applications, anaphora resolution is worth considering for further message understanding and the consistency of discourses.

Our future work will be directed into following studies:

1. Extending the set of anaphor being processed:
This analysis aims at identifying instances (such as definite anaphor) that could be useful in anaphora resolution.
2. Resolving nominal coreference:
The language resource WordNet can be utilized to identify the coreference relation on the basis of synonymy/hypernym/hyponym relation.

References

- Aarts Jan, Henk Barkema and Nelleke Oostdijk (1997), "The TOSCA-ICLE Tagset: Tagging Manual", TOSCA Research Group for Corpus Linguistics.
- Baldwin, Breck (1997), "CogNIAC: high precision coreference with limited knowledge and linguistic resources", Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution, pp. 38-45.
- Bontcheva, Kalina, Marin Dimitrov, Diana Maynard and Valentin Tablan (2002), "Shallow Methods for Named Entity Coreference Resolution", Proceedings of TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES (TALN), pp. 24-32.
- Cardie, Claire and Kiri Wagstaff (1999), "Noun Phrase Coreference as Clustering", Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.
- Chen, Kuang-hua and Hsin-Hsi Chen (1994), "Extracting Noun Phrases from Large-Scale Texts: A Hybrid Approach and Its Automatic Evaluation", Proceedings of the 32nd ACL Annual Meeting, 1994, pp. 234-241.
- Dagan, Ido and Alon Itai (1990), "Automatic processing of large corpora for the resolution of anaphora references", Proceedings of the 13th International Conference on Computational Linguistics (COLING'90), Vol. III, 1-3, Helsinki, Finland.
- Denber, Michel (1998), "Automatic resolution of anaphora in English", Technical report, Eastman Kodak Co.
- Evans, Richard and Constantin Orasan (2000), "Improving anaphora resolution by identifying animate entities in texts", In Proceedings of DAARC-2000.
- Ge, Niyu, John Hale and Eugene Charniak (1998), "A Statistical Approach to Anaphora Resolution", Proceedings of the Sixth Workshop on Very Large Corpora (COLING-ACL98), pp.161-170.
- Kennedy, Christopher and Branimir Boguraev (1996), "Anaphora for everyone: Pronominal anaphora resolution without a parser", Proceedings of the 16th International Conference on Computational Linguistics, pp.113-118.
- Lappin, Shalom and Herbert Leass (1994), "An Algorithm for Pronominal Anaphora Resolution", Computational Linguistics, Volume 20, Part 4, pp. 535-561.
- Miller, George (1993), "Nouns in WordNet: A Lexical Inheritance System", Journal of Lexicography, pp. 245-264.
- 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. pp. 869-875.

- Mitkov, Ruslan (1999), “Anaphora Resolution: The State of the Art”, Working paper (Based on the COLING'98/ACL'98 tutorial on anaphora resolution)
- Mitkov, Ruslan and Catalina Barbu (2001), “Evaluation tool for rule-based anaphora resolution methods”, Proceedings of ACL'01, Toulouse, 2001.
- Mitkov, Ruslan, Richard Evans and Constantin Orasan (2002), “A new, fully automatic version of Mitkov's knowledge-poor pronoun resolution method”, In Proceedings of CICLing- 2000, Mexico City, Mexico.
- Wang, Ning, Chunfa Yuan, K.F. Wang and Wenjie Li (2002), “Anaphora Resolution in Chinese Financial News for Information Extraction”, Proceedings of 4th World Congress on Intelligent Control and Automation, June 2002, Shanghai, pp.2422-2426.