

## **Automatic Pronominal Anaphora Resolution in English Texts**

**Tyne Liang\* and Dian-Song Wu\***

### **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, 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 relations of nouns and verbs in the surrounding context. The identification of animacy entities and pleonastic-it usage in English discourses are employed to promote resolution accuracy.

Traditionally, anaphora resolution systems have relied on syntactic, semantic or pragmatic clues to identify the antecedent of an anaphor. Our proposed method makes use of WordNet ontology to identify animate entities as well as essential gender information. In the animacy agreement module, the property is identified by the hypernym relation between entities and their unique beginners defined in WordNet. In addition, the verb of the entity is also an important clue used to reduce the uncertainty. An experiment was conducted using a balanced corpus to resolve the pronominal anaphora phenomenon. The methods proposed in [Lappin and Leass, 94] and [Mitkov, 01] focus on the corpora with only inanimate pronouns such as “it” or “its”. Thus the results of intra-sentential and inter-sentential anaphora distribution are different. In an experiment using Brown corpus, we found that the distribution proportion of intra-sentential anaphora is about 60%. Seven heuristic rules are applied in our system; five of them are preference rules, and two are constraint rules. They are derived from syntactic, semantic, pragmatic conventions and from the analysis of training data. A relative measurement indicates that about 30% of the errors can be eliminated by applying heuristic module.

---

\* Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan  
Email: tliang@cis.nctu.edu.tw; gis90507@cis.nctu.edu.tw

## 1. Introduction

### 1.1 Problem description

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

Pronominal anaphora, where pronouns are substituted by previously mentioned entities, is a common phenomenon. 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, “it” can also be a non-anaphoric expression which does not refer to any previously mentioned item, in which case it is called an expletive or the pleonastic-it [Lappin and Leass, 94]. Although pleonastic pronouns are not considered anaphoric since they do not have antecedents to refer to, recognizing such occurrences is, nevertheless, essential during anaphora resolution. In [Mitkov, 01], non-anaphoric pronouns were found to constitute 14.2% of a corpus of 28,272 words.

Definite noun phrase anaphora occurs where the antecedent is referred by a general concept entity. The general concept entity can be a semantically close phrase, such as a synonym or super-ordinates of the antecedent [Mitkov, 99]. The word *one* has a number of different usages apart from counting. One of its 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 means the antecedent occurs in a sentence prior to the sentence with the anaphor. In [Lappin and Leass, 94], there were 15.9% inter-sentential cases and 84.1% intra-sentential cases in the testing results. In [Mitkov, 01], there were 33.4% inter-sentential cases and 66.6% intra-sentential cases.

Traditionally, anaphora resolution systems have relied on syntactic, semantic or pragmatic clues to identify the antecedent of an anaphor. Hobbs' algorithm [Hobbs, 76] was the first syntax-oriented method presented in this research domain. From the result of a syntactic tree, they checked the number and gender agreement between antecedent candidates and a specified pronoun. In RAP (Resolution of Anaphora Procedure) proposed by Lappin and Leass [94], an algorithm is applied to the syntactic representations generated by McCord's Slot Grammar parser, and salience measures are derived from the 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 was proposed by [Kennedy and Boguraev, 96]. It employed only part-of-speech tagging with a shallow syntactic parse indicating the grammatical roles of NPs and their containment in adjuncts or noun phrases.

Cardie *et al.* [99] treated coreferencing as a clustering task. Then a distance metric function was used to decide whether two noun phrases were similar or not. In [Denber, 98], an algorithm called Anaphora Matcher (AM) was implemented to handle inter-sentential anaphora in a two-sentence context. This method uses information about the sentence as well as real world semantic knowledge obtained from other sources. The lexical database system WordNet is utilized to acquire semantic clues about the words in the input sentences. It is noted that anaphora do not refer back more than one sentence in most cases. Thus, a two-sentence "window size" is sufficient for anaphora resolution in the domain of image queries.

A statistical approach to disambiguate pronoun "it" in sentences was introduced in [Dagan and Itai, 90]. The disambiguation is based on the co-occurring patterns obtained from a corpus to find the antecedent. The antecedent candidate with the highest frequency in the co-occurring patterns is selected as a match for the anaphor.

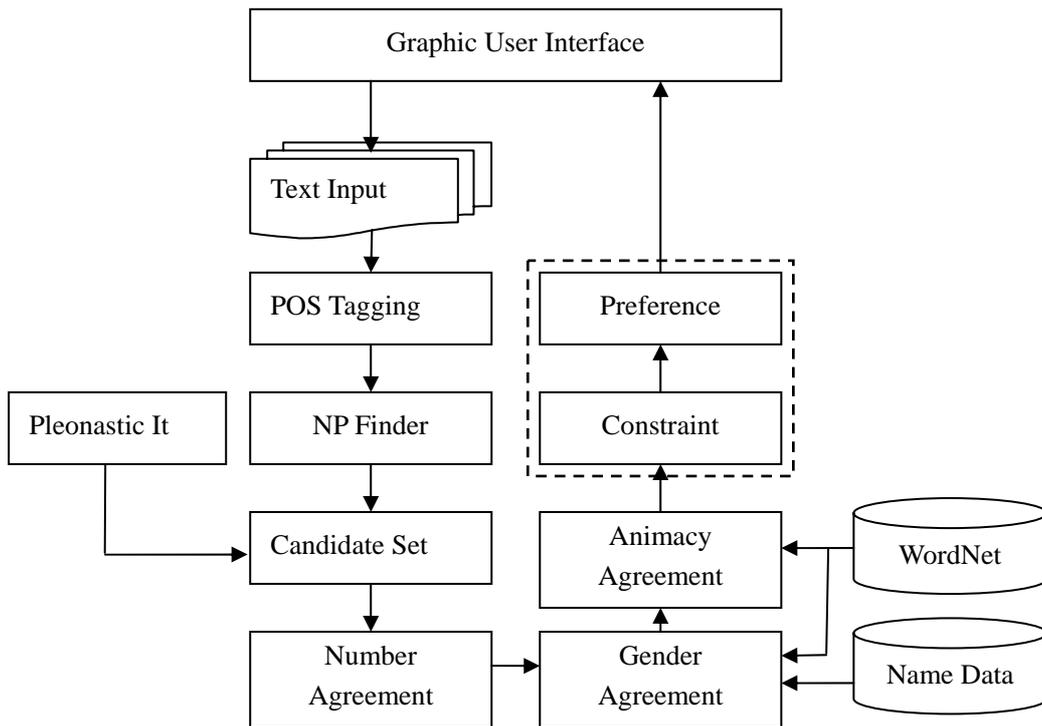
A knowledge-poor approach was proposed in [Mitkov, 98]; it can be applied to different languages (English, Polish, and Arabic). The main components of this method are the so-called "antecedent indicators" which are used to assign a score (2, 1, 0, -1) for each candidate noun phrase. The scores play a 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 using limited knowledge and linguistic resources. It is a high precision pronoun resolution system that is capable of achieving more than 90% precision with 60% recall for some pronouns. Mitkov [02] presented

a new, advanced and completely revamped version of his own knowledge-poor approach to pronoun resolution. In contrast to most anaphora resolution approaches, the system called MARS operates in the fully automatic mode. Three new indicators included in MARS are Boost Pronoun, Syntactic Parallelism and Frequent Candidates.

In [Mitkov, 01], the authors proposed an evaluation environment for comparing anaphora resolution algorithms. Performances are illustrated by presenting the results of a comparative evaluation conducted on the basis of several evaluation measures. Their testing corpus contained 28,272 words, with 19,305 noun phrases and 422 pronouns, of which 362 were 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 used 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 are identified by NP finder module and stored in a global data structure. Then the number agreement is applied to the head noun. Capitalized nouns in the name gazetteer are tested to find personal names. A name will be tagged with the gender feature if it can be found uniquely in male or female class defined in gender agreement module. In this phase, WordNet is also used to find possible gender clues for improving resolution performance. The gender attribute is ignored to avoid ambiguity when the person name can be masculine or feminine.
3. Anaphors are checked sequentially from the beginning of the first sentence. They are stored in a list with sentence offset and word offset information. Then pleonastic-it is checked so that no further attempts at resolution are made.
4. The remaining noun phrases preceding the anaphor within a predefined window size are collected as antecedent candidates. Then the candidate set is further filtered by means of gender and animacy agreement.
5. The remaining candidates are then evaluated by means of heuristic rules. These rules can be classified as preference rules or constraint rules. A scoring equation (equation 1) is used to evaluate how likely it is that a candidate will be selected as the antecedent. The scoring equation calculates the accumulated score of each possible candidate. The parameter  $agreement_k$  denotes number agreement, gender agreement and animacy agreement output. If one of these three outputs indicates disagreement, the score will be set to zero. The parameter value enclosed in parentheses is the accumulated number of rules that fit our predefined heuristic rules:

$$score (can, ana) = \left( \sum_i rule\_pre_i - \sum_j rule\_con_j \right) \times \prod_k agreement_k, \quad (1)$$

where

$can$ : each candidate noun phrase for the specified anaphor;

$ana$ : anaphor to be resolved;

$rule\_pre_i$ : the  $i$ th preference rule;

$rule\_con_j$ : the  $j$ 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] has been used to lemmatize and tag English learner corpora. The TOSCA-ICLE tag set consists of 16 major word classes. These major word

classes may be further specified by means of features of subclasses as well as a variety of syntactic, semantic and morphological characteristics.

### 2.2.2 NP Finder

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

base NP  $\rightarrow$  modifier + head noun

modifier  $\rightarrow$  <article| number| present participle| past participle |adjective| noun>

At the beginning, our system identifies base noun phrases that contain no other smaller noun phrases within them. For example, *the chief executive officer of a financial company* is divided into *the chief executive officer* and *a financial company* for the convenience of judging whether the noun phrase is a prepositional noun phrase or not. This could be of help in selecting a correct candidate for a specific anaphor. Once the final candidate is selected, the entire modifier is combined together again.

The proposed base noun phrase finder is implemented based on a finite state machine (Figure 2). Each state indicates a particular part-of-speech of a word. The arcs between states indicate a word input from the first word of the sentence. 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, it is rejected. An example of base noun phrase output is illustrated in Figure 3.

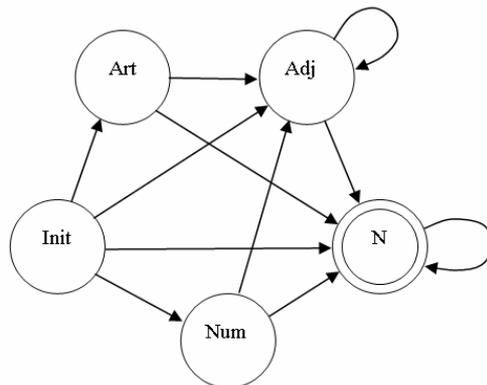


Figure 2. Finite state machine for a noun phrase.

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 a base noun phrase.

### 2.2.3 Pleonastic-it Module

The pleonastic-it module is used to filter out those semantic empty usage conditions which are essential for pronominal anaphora resolution. A word “it” is said to be pleonastic when it is used in a discourse where the word does not refer to any antecedent.

References of “pleonastic-it” can be classified as state references or passive references [Denber, 98]. State references are usually used for assertions about the weather or the time, and this category is further divided into meteorological references and temporal references.

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

Most instances of "pleonastic-it" can be described by 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

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

### 2.2.5 Gender Agreement

The gender recognition process can deal with words that have gender features. To distinguish the gender information of a person, we use an English first name list collected from (<http://www.behindthename.com/>) covering 5,661 male first name entries and 5,087 female ones. In addition, we employ some useful clues from WordNet results by conducting 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 to by some gender-marked pronouns (he, she, him, her, his, hers, himself, herself) in texts. Conventionally, animate entities include people and animals. Since it is hard to obtain the property of animacy with respect to a noun phrase by its surface morphology, we use WordNet [Miller, 93] to recognize animate entities in which a noun can only have one hypernym but can have many hyponyms. With twenty-five unique beginners, we observe that two of them can be taken as representations of animacy. These two unique beginners are {animal, fauna} and {person, human being}. Since all the hyponyms inherit properties from their hypernyms, the animacy of a noun can be determined by making use of this hierarchical relation. However, a noun may have several senses, depending on the context. 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 a corpus, and the value is selected when the accuracy rate reaches its maximum:

$$t\_noun = \frac{\text{the\_number\_of\_senses\_in\_animacy\_files}}{\text{the\_total\_senses\_of\_the\_noun}}, \quad (2)$$

$$t\_verb = \frac{\text{the\_number\_of\_senses\_in\_animacy\_files}}{\text{the\_total\_senses\_of\_the\_verb}}, \quad (3)$$

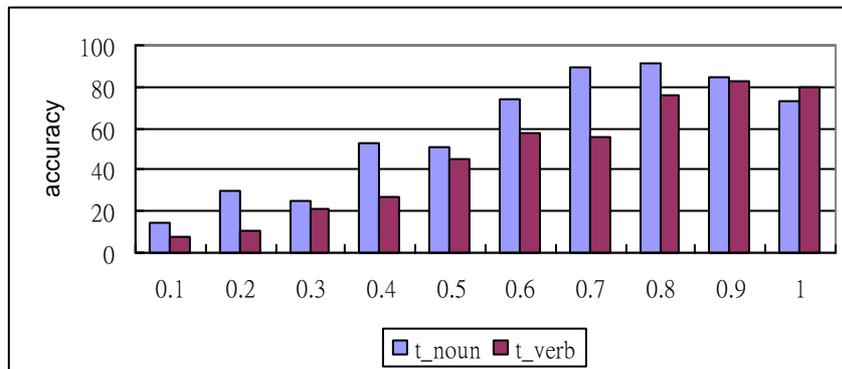
$$\text{accuracy} = \frac{\text{the\_number\_of\_animacy\_entities\_identified\_correctly}}{\text{the\_total\_number\_of\_animacy\_entities}}. \quad (4)$$

Besides the noun hypernym relation, unique beginners of verbs are also taken into consideration. These lexicographical 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 an animate verb.**

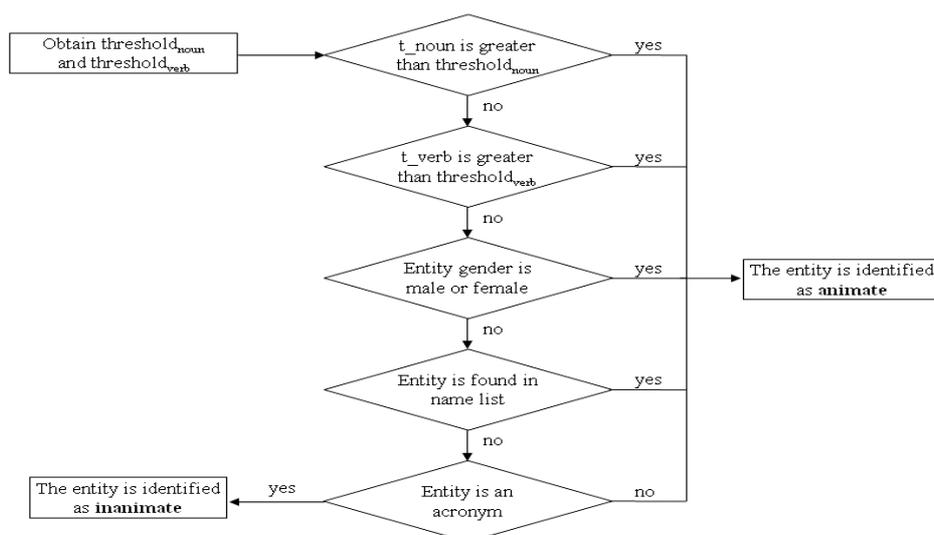
Unique beginners	Example of verb
{cognition}	Think, analyze, judge ...
{communication}	Tell, ask, teach ...
{emotion}	Feel, love, fear ...
{social}	Participate, make, establish ...

The training data that we obtained from the Brown corpus consisted of 10,134 words, 2,155 noun phrases, and 517 animacy entities. We found that 24% of the noun phrases in the corpus referred to animate entities, whereas 76% of them referred to inanimate ones. We utilized the ratio of senses from the WordNet output to decide whether the entity was an animate entity or not. Therefore, the ratio of senses in the noun and its verb is obtained in the training phase to achieve the highest possible accuracy. Afterwards, the testing phase makes use of these two threshold values to decide on the animate feature. Threshold values can be obtained by training on the corpus and selecting the value when the accuracy rate (equation 4) reaches its maximum. Therefore,  $t_{\text{noun}}$  and  $t_{\text{verb}}$  were found to be 0.8 and 0.9, respectively, according to the distribution in Figure 4.



**Figure 4. Thresholds of Animacy Entities.**

The process of determining whether a noun phrase is animate or inanimate is described below :



### 2.2.7 Heuristic Rules

#### I. Syntactic parallelism rule

The syntactic parallelism of an anaphor and an antecedent could be an important clue when other constraints or preferences can not be employed to identify a unique unambiguous antecedent. The rule reflects the preference that the correct antecedent has the same part-of-speech and grammatical function as the anaphor. Nouns can function grammatically as subjects, objects or subject complements. 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, such as verb “be”, 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.

“**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, “**it**” refers to “**the uniform,**” which is the object of the first sentence. Empirical evidence also shows that anaphors usually match their antecedents in terms of their syntactic functions.

#### II. Semantic parallelism rule

This preference works by identifying collocation patterns in which anaphora appear. In this way, the system can automatically identify semantic roles and employ them to select the most appropriate candidate. Collocation relations specify the relations between words that tend to co-occur in the same lexical contexts. The rule emphasizes that those noun phrases with the same semantic roles as the anaphor are preferred answer candidates.

#### III. Definiteness rule

Definiteness is a category concerned with the grammaticalization of the identifiability and non-identifiability 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 antecedents of anaphors than indefinite noun phrases.

#### IV. Mention Frequency rule

Recurring items in a context are regarded as likely candidates for the antecedent of an anaphor. Generally, high frequency items indicate the topic as well as the most likely candidate.

### V. Sentence recency rule

Recency information is employed in most of the implementations of anaphora resolution. In [Lappin, 94], the recency factor is the one with the 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 that 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 (within a sentence) between the anaphor and the antecedent was found to be 1.3, and the average distance for noun phrases was found to be 4.3 NPs.

### VI. Non-prepositional noun phrase rule

A noun phrase not contained in another noun phrase is considered a 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 a candidate is connected with an anaphor by a conjunction, the anaphora relation is hard to be constructed between these two entities.

For example:

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

## 2.3 The Brown Corpus

The training and testing texts were 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 to show the analysis result. The POS tagger component takes the input text and outputs tokens, lemmas, most likely tags and the number of alternative tags. The NP chunker makes use of a finite state machine (FSM) to recognize strings which belong to a specified regular set.

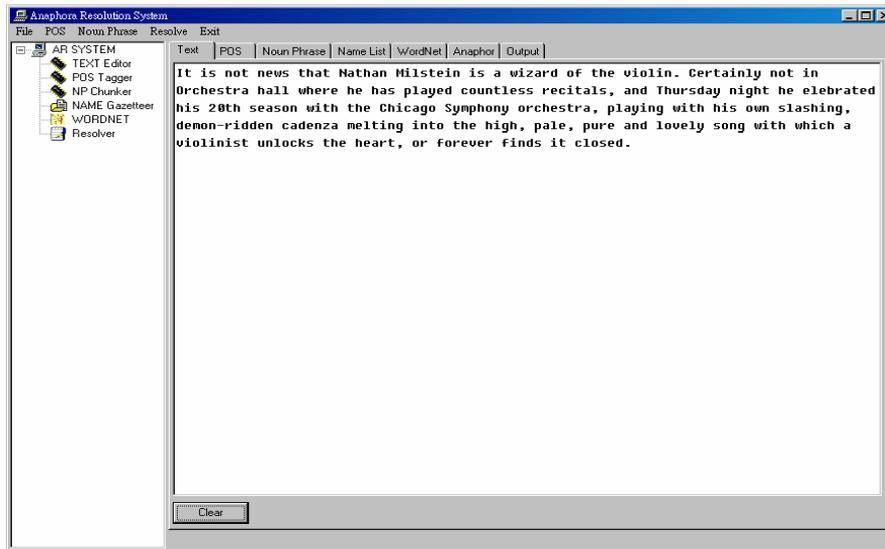


Figure 6. The main system window.

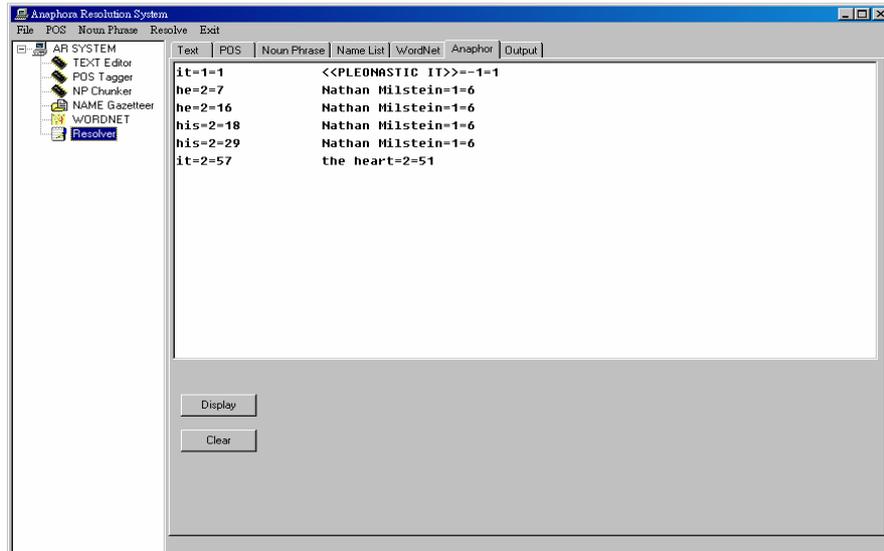


Figure 7. Anaphora pairs.

After the selection procedure is performed, 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 at the end of each entity. For example, “it” as the first word of the first sentence denotes a pleonastic-it, and the other “it,” the 57th word of the second sentence refers to “the heart.” Figure 8 shows the original text input with antecedent annotation following each anaphor in the text. All the annotations are highlighted to facilitate subsequent testing.

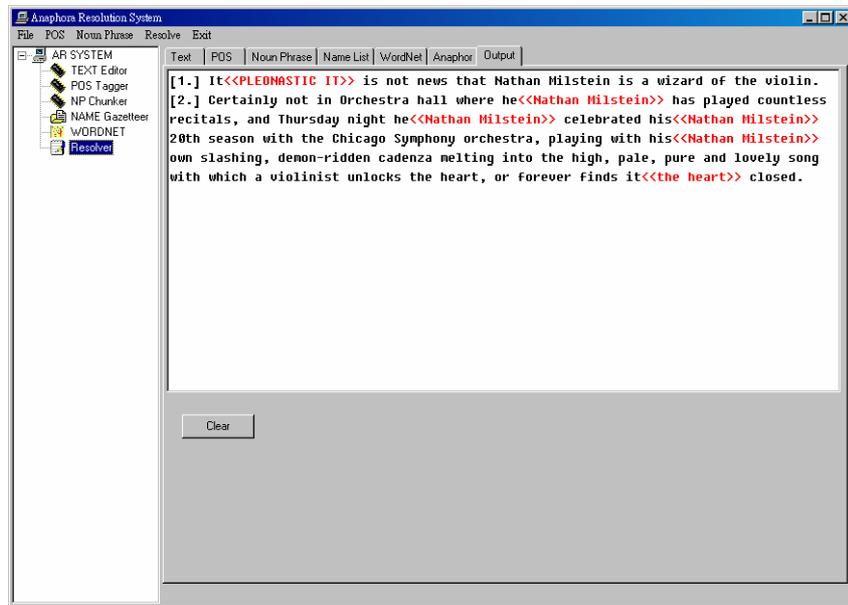


Figure 8. Anaphor with antecedent annotation.

### 3. Experimental Results and Analysis

The evaluation experiment employed random texts of different genres selected from the Brown corpus. There were 14,124 words, 2,970 noun phrases and 530 anaphors in the testing data. Two baseline models were established to compare the progress of performance with our proposed anaphora resolution (AR) system. The first baseline model (called the baseline subject) determined the number and gender agreement between candidates and anaphors, and then chose the most recent subject as the antecedent from the candidate set. The second baseline model (called baseline recent) performed a similar procedure, but it selected the most recent noun phrase as the antecedent which matched the anaphor in terms of number and gender agreement. The success rate was calculated as follows:

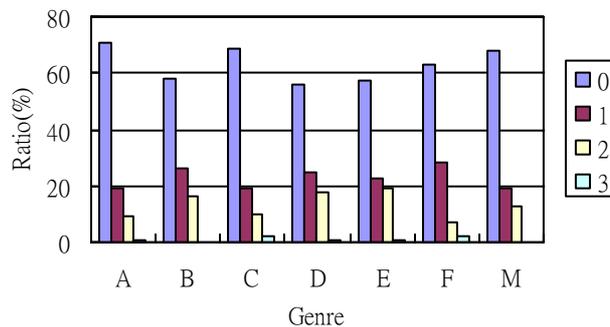
$$\text{Success Rate} = \frac{\text{number of correctly resolved anaphors}}{\text{number of all anaphors}} \quad (5)$$

The results obtained (Table 3) showed that there are 41% of the antecedents could be identified by finding the most recent subject; however, only 17% of the antecedents could be resolved by selecting the most recent noun phrase with the same gender and number agreement as the anaphor.

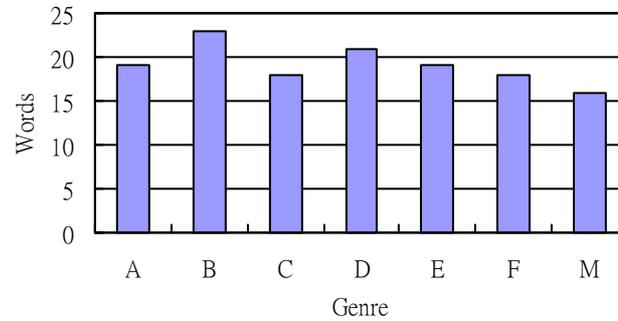
**Table 3. Success rate of baseline models.**

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 the sentence distance between antecedents and anaphors. The value 0 denotes intra-sentential anaphora and other values indicate inter-sentential anaphora. In the experiment, a balanced corpus was used to resolve the pronominal anaphora phenomenon. The methods proposed in [Lappin and Leass, 94] and [Mitkov, 01] employ corpora with only inanimate pronouns, such as “it” or “its.” Thus, the results for intra-sentential and inter-sentential anaphora distribution obtained using those methods are different. In our experiment on the Brown corpus, the distribution proportion of intra-sentential anaphora was about 60%. Figure 10 shows the average word distance distribution for each genre. The pleonastic-it could be identified with 89% accuracy (Table 4).



**Figure 9. Referential sentence distance distribution.**

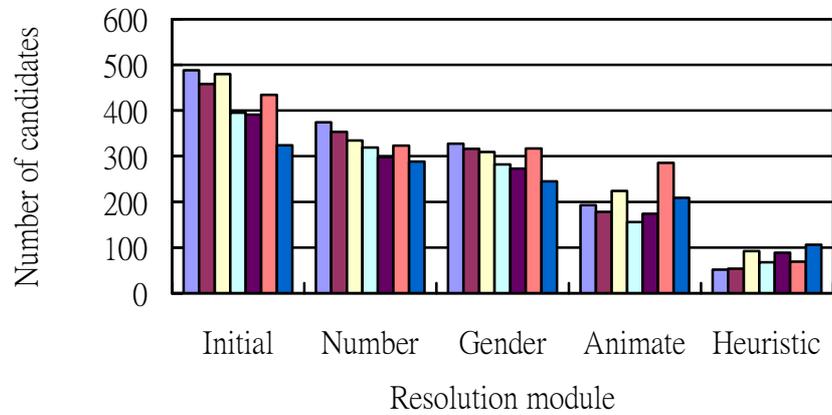


**Figure 10.** *Referential word distance distribution.*

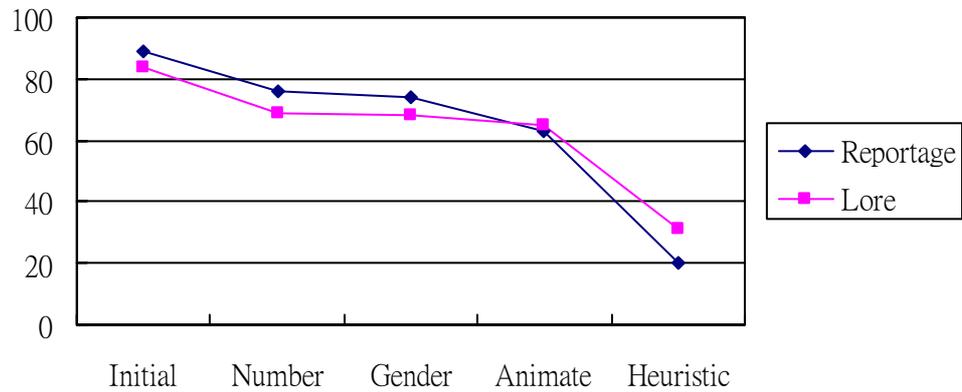
**Table 4.** *Pleonastic-it identification.*

	Number of anaphora	Number of Anaphoric expressions	Number of Pleonastic-its	Ratio of Pleonastic-it to pronoun	Accuracy of identification
Total	530	483	47	9%	89%

The next experiment provided empirical evidence showing that the enforcement of agreement constraints increases the system's chances of selecting a correct antecedent from an initial candidate set. To assess the effectiveness of each module, the total number of candidates in each genre was determined after applying the following four phases which include number agreement, gender agreement, animacy agreement, and heuristic rules (Figure 11). As shown in Figure 12, the error rates for two genres of testing data indicated the improvement in choosing correct antecedents following each resolution phase. Apparently, the animate module achieved more significant error rate reduction in the reportage domain than the other one.



**Figure 11.** Candidate distribution after applying resolution modules.



**Figure 12.** Error rate after applying resolution modules.

The final evaluation results obtained using our system, which applied animacy agreement and heuristic rules to resolution, are listed in Table 6. It also shows the results for each individual genre of testing data and the overall success rate, which reached 77%.

**Table 6. Success rate of the 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%

Our proposed method makes use of the WordNet ontology to identify animate entities as well as essential gender information. In the animacy agreement module, each property is identified by the hypernym relation between entities and their unique beginners defined in WordNet. In addition, the verb of the entity is also an important clue for reducing the uncertainty. An overall comparison is shown below:

	Our method	[Kennedy and Boguraev, 96]	[Baldwin, 97]	[Mitkov, 98]
Hypernym relation in Noun	Y	N	N	N
Verb relation	Y	N	N	N
Name database	Y	N	N	N
Gender feature from WordNet	Y	N	N	Y
Full parsing	N	N	N	N
Heuristic	Y	Y	Y	Y
Accuracy	77%	62%	50%	57%

In the preprocessing phase, the accuracy of the POS tagger was about 95%. If a noun is misclassified as another part-of-speech, for example, if the noun “patient” is tagged as an adjective, then there is no chance for it to be considered as a legal antecedent candidate of an anaphor. The other problems encountered in the system are multiple antecedents and unknown word phenomena. In the case of multiple antecedents, the correct answer is composed of more than one entity, such as “Boys and girls are singing with pleasure.” In this case, additional

heuristic are needed to decide whether the entities should be combined into one entity or not. In the case of an unknown word, the tagger may fail to identify the part of speech of the word so that in WordNet, no unique beginner can be assigned. This can lead to a matching failure if the entity turns out to be the correct anaphoric reference.

#### **4. Conclusion and Future Work**

In this paper, the WordNet ontology and heuristic rules have been adopted to perform anaphora resolution. The recognition of animacy entities and gender features in discourses is helpful for improving resolution accuracy. The proposed system is able to deal with intra-sentential and inter-sentential anaphora in English texts and deals appropriately with pleonastic pronouns. From the experiment results, our proposed method is comparable in performance with prior works that fully parse the text. In contrast to most anaphora resolution approaches, our system benefits from the recognition of animacy agreement and operates in a fully automatic mode to achieve optimal performance. With the growing interest in natural language processing and its various applications, anaphora resolution is essential for further message understanding and the coherence of discourses during text processing.

Our future works will be as follows:

1. Extending the set of anaphors to be processed:  
This analysis aims at identifying instances (such as definite anaphors) that could be useful in anaphora resolution.
2. Resolving nominal coreferences:  
The language resource WordNet can be utilized to identify coreference entities by their synonymy/hypernym/hyponym relations.

#### **Acknowledgement**

This research is partially supported by National Science Council, R.O.C., under NSC contract 91-2213-E-009-082 and by MediaTek Research Center, National Chiao Tung University, Taiwan.

#### **References**

- Jan, Aarts, Henk Barkema and Nelleke Oostdijk, "The TOSCA-ICLE Tagset: Tagging Manual," *TOSCA Research Group for Corpus Linguistics*, 1997.
- Baldwin, Breck, "CogNIAC: high precision coreference with limited knowledge and linguistic resources," *In Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 1997, pp. 38-45.

- Bontcheva, Kalina, Marin Dimitrov, Diana Maynard and Valentin Tablan, "Shallow Methods for Named Entity Coreference Resolution," *In Proceedings of TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES (TALN)*, 2002, pp. 24-32.
- Cardie, Claire and Kiri Wagstaff, "Noun Phrase Coreference as Clustering," *In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1999.
- Chen, Kuang-hua and Hsin-Hsi Chen, "Extracting Noun Phrases from Large-Scale Texts: A Hybrid Approach and Its Automatic Evaluation," *In Proceedings of the 32nd ACL Annual Meeting*, 1994, pp. 234-241.
- Dagan, Ido and Alon Itai, "Automatic processing of large corpora for the resolution of anaphora references," *In Proceedings of the 13th International Conference on Computational Linguistics (COLING'90), Vol. III, 1-3*, 1990.
- Denber, Michel, "Automatic resolution of anaphora in English," *Technical report, Eastman Kodak Co.*, 1998.
- Evans, Richard and Constantin Orasan, "Improving anaphora resolution by identifying animate entities in texts," *In Proceedings of DAARC*, 2000.
- Ge, Niyu, John Hale and Eugene Charniak, "A Statistical Approach to Anaphora Resolution," *In Proceedings of the Sixth Workshop on Very Large Corpora (COLING-ACL98)*, 1998, pp.161-170.
- Kennedy, Christopher and Branimir Boguraev, "Anaphora for everyone: Pronominal anaphora resolution without a parser," *In Proceedings of the 16<sup>th</sup> International Conference on Computational Linguistics*, 1996, pp.113-118.
- Lappin, Shalom and Herbert Leass, "An Algorithm for Pronominal Anaphora Resolution," *Computational Linguistics*, Volume 20, Part 4, 1994, pp. 535-561.
- Miller, George, "Nouns in WordNet: A Lexical Inheritance System," *Journal of Lexicography*, 1993, pp. 245-264.
- Mitkov, Ruslan, "Robust pronoun resolution with limited knowledge," *In Proceedings of the 18th International Conference on Computational Linguistics (COLING'98)/ACL'98 Conference Montreal, Canada*. 1998, pp. 869-875.
- Mitkov, Ruslan, "Anaphora Resolution: The State of the Art," Working paper (Based on the COLING'98/ACL'98 tutorial on anaphora resolution), 1999.
- Mitkov, Ruslan and Catalina Barbu, "Evaluation tool for rule-based anaphora resolution methods," *In Proceedings of ACL'01, Toulouse*, 2001.
- Mitkov, Ruslan, Richard Evans and Constantin Orasan, "A new fully automatic version of Mitkov's knowledge-poor pronoun resolution method," *In Proceedings of CICLing-2000, Mexico City, Mexico*.
- Wang, Ning Yuan, Chunfa, Wang, K.F. and Li, Wenjie "Anaphora Resolution in Chinese Financial News for Information Extraction," *In Proceedings of 4th World Congress on Intelligent Control and Automation*, June 2002, pp.2422-2426.

