Knowledge-based Sense Pruning using the HowNet: an Alternative to Word Sense Disambiguation

By

Chi-Yung Wang

A Thesis Submitted to
The Hong Kong University of Science & Technology
In Partial Fulfillment of the Requirements for
The Degree of Master of Philosophy
In Computer Science

December 2001, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science & Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science & Technology to reproduce the thesis by photocopy or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_____

Chi-Yung Wang

Knowledge-based Sense Pruning using the HowNet:
an Alternative to Word Sense Disambiguation

By

Chi-Yung Wang

This is to certify that I have examined the above MPhil thesis
and have found that it is complete and satisfactory in all respects,
And that any and all revisions required by
the thesis examination committee have been made.

_____

Dr. Brian Mak, Supervisor

_____

Prof. Derick Wood, Acting Head of Department

Computer Science Department
January, 2002

Acknowledgements

I would like to take this chance to express my sincere gratitude to my supervisors, Dr Gan Kok Wee and Dr. Brian Mak, for their patience and guidance for this thesis.

Thanks go to Dr. Gan, who accepted my MPhil application and supported me as Research Assistant in first year. At the beginning of my postgraduate study, Dr. Gan gave me an intensive set of lessons about linguistic knowledge, which I hadn't touched but is necessary for this thesis. Subsequently, he has been very helpful for my study and research, even after he left university. Also, Dr. Gan is a nice person, whose concern for me extends beyond academics.

Thanks also go to Dr. Mak, who accepted me in my second year study. Dr. Mak is nice and concise. He gave me much support in thesis writing.

I am very grateful to my ex-colleague, Mr. Ping-Wai Wong, for his instant help in the question of linguistic knowledge. He is very helpful.

It is great experience to study in HKUST. I enjoy the beautiful view and love the nice people there.

Table of Contents

Knowledge-based Sense Pruning using the HowNet: an Alternative to Word Sense Disambiguation

By Chi-Yung Wang

Computer Science Department

The Hong Kong University of Science & Technology

Abstract

In this thesis, we try to solve the problem of word sense disambiguation (WSD) in natural language processing by sense pruning using a knowledge-based approach. Traditional WSD methods provide only one meaning for each word in a passage. However, we believe that textual information alone may not be sufficient to determine the exact meaning of each word which has to be resolved when higher-level knowledge becomes available. Thus, we propose that the objective of WSD is to reduce the number of plausible meanings of a word as much as possible through "sense pruning". After sense pruning, we will associate a word with a list of plausible meanings. We would like to keep the truly correct sense of each word on its own meaning list and yet keep the number of possible meanings of a whole sentence as small as possible.

We applied sense pruning to Chinese WSD, making use of the HowNet. HowNet is a knowledge base that describes all entities in its database by a set of unambiguous sememes. It provides information about the relationship between concepts or their attributes, in which concepts are represented by the sememes. One of our contributions is integrating various knowledge from HowNet for sense pruning, such as, relations between the sememes, information structures in Chinese, relations of attributes and attribute values, and characteristics of functional words. Based on HowNet, four additional databases were developed for sense pruning in this thesis.

We evaluated our sense pruning algorithm on the Corpus of Sinica from Taiwan. Two criteria were used for the evaluation: recall rate and reduction of the number of possible meanings of a sentence. Effects of the size of the analytical window and the analytical unit, and the speed of the algorithm were fully studied. In summary, sense pruning achieves a recall rate of 91% while reducing the number of possible meanings of a sentence by 48% when a whole sentence is taken as an analytical unit.

Chapter 1 Introduction

We give a brief discussion about the motivation of the thesis in this chapter. The objective and of this thesis is stated and the outline of this thesis is described at the end of this chapter.

1.1 Motivation

Machine understanding is a tremendously difficult problem in Natural Language Processing (NLP). In general NLP research, there are several stages from raw materials (simple text without any information tagged) to fully interpreted context, such as segmentation, syntactic parsing, sense disambiguation and semantic interpretation. Though these steps usually are researched independently, we believe that the connection between steps can't be ignored. This linkage of stages is called the re-constructive approach to text understanding (Dong, 1999). Gan and Wong (2000) had classified the stages as (1) Sentence Breaking, (2) Concept Group Extraction, (3) Sense pruning, (4) Message Structure Identification and (5) Event Relation and Role-shifting. In past researches, researchers always provide an 'only one' solution in each step. However, if there is any mistake or incorrect output from previous steps, the subsequent steps will not get the good results. That is why we suggest to do sense pruning in this thesis instead of Word Sense Disambiguation (WSD) as usual. In other words, it is not assumed that the result is an 'only one' solution. After sense pruning, the result becomes the input of the semantic interpretation processes, such as Message Structure Identification and Event Relation and Role-shifting.

Traditionally, the methodology of sense disambiguation could be classified into rule-based and statistical approaches. In a rule-based approach, one sense has to be determined by one or more rules. Since the processing is confined within a distance scope, the techniques are rather poor. In a statistical approach, sense disambiguation is based on the probabilities of the appearance of that sense. This approach is a conventional method and the decision is rather coarse. An innovation will also be made for sense disambiguation on the condition that a new knowledge resource like HowNet is used. The new approach will take a Complete Sentence (defined in section 4.1) as the testing field and calculate the scores of the senses by comparing the information from the senses to be disambiguated and the other senses in the field. The detail of application of HowNet knowledge will discussed in Chapter 3 and 4. Compare to traditional methods, the new Knowledge-based approach is finer. And, another advantage of the approach is that its algorithm is language-independent and system-independent. A sense pruning tool for Machine Translation can also be used for some other applications.

1.2 Objective and Scope of this Thesis

We apply the new knowledge base HowNet in the research of sense pruning. The corpus of Sinica, Taiwan, is used for testing. The objective of the thesis is to implement the new approach of sense pruning in the research of text understanding. We will also try to evaluate and enrich the knowledge base HowNet.

As a middle stage of the re-constructive approach to text understanding, the goal of this thesis is to achieve a high recall rate. That means the correct answer will be retained in the output of the system. If the correct answer of sense is pruned away, then the output of the next stage, semantic interpretation, cannot be predicted with good results. The value of this stage research is determined by the complexity reduced. The more the complexity is reduced, the more the work load of next stage researchers is reduced. To balance two criteria, the recall rate will be in the first priority.

1.3 Outline of this Thesis

In this thesis, we present a brief survey of related works about Word Sense Disambiguation. Also, a short comparison of WordNet and HowNet is introduced in Chapter 2, Related Works. In Chapter 3, Dictionary and supplementary documents of HowNet and Information Structure are introduced before the detail of this thesis. The core of this thesis is the system of Sense Pruning. The knowledge sources for sense pruning are discussed in Chapter 4. Chapter 5 includes the preparation, methodology of the Sense Pruning system and the discussion of the results. Finally, the contributions, conclusions and some future works are discussed in Chapter 6.

Chapter 2 Related Works

In this chapter, we would like to introduce some similar researches. In the first section (2.1), the rule-based approach of Word Sense Disambiguation (WSD) is introduced. Then, we will introduce some research using HowNet in section 2.2. Finally, we compare HowNet to WordNet, which is a common lexical database in Natural Language Processing (NLP).

2.1 Rule-based Approach

We will classify the WSD into the rule-based and statistical approaches. The purpose of WSD is to identify the correct sense of a word token in context (Ng and Zelle, 1997). It is assumed that each word token in the input sentence is tagged with at least one sense or definition. And, the output is that sentence with each word taken tagged with one sense or definition only. The statistical approach uses the probability of appearance to disambiguate the sense. Since it is quite different to Sense Pruning, we will not discuss it there.

In general, the WSD system automated learning techniques is based on corpora of natural language examples in an attempt to automatically induce suitable language-processing models. It learns the disambiguation knowledge from a large sense-tagged corpus. After training, the WSD system can assign a correct sense or definition to each word token of a new sentence.

Before applying the system, some training examples are encoded in some rules by linguistic knowledge. Different knowledge is represented by different forms of rules, such as:

- *Surrounding words*, which are the unordered set of words surrounding the word token, are developed by common sense. For example, if 'bank' is surrounding 'interest', then the sense of 'interest' will tend to 'money paid for the use of money'.
- *Local collocations* are developed by some word phrases. This is a short sequence of words near the word token and the word order is taken into account. For example, in the phrase 'in the interest of', the sense of interest will tend to 'advantage, advancement, or favor'.
- *Syntactic relations* such as subject-verb, verb-object and adjective-noun are important sources of WSD.
- *Parts of speech* of the neighboring words
- *Morphological forms* of word are also useful in WSD.

After considering the basic form of rules, the next step is the learning algorithm. The common algorithms are Bayesian probabilistic algorithms, neural networks, decision lists and exemplar-based algorithms.

Mooney (1996) evaluated some widely used machine-learning algorithms for disambiguating the word 'line'. He reported that the naïve-Bayes algorithm gives the highest accuracy. Surrounding words were used in this research.

Ng (1997) improved the exemplar-based algorithm for implementation in the DSO National Laboratories corpus. He reported with a higher accuracy rate compared to the naïve-Bayes algorithm. In his study, only the local collocation of the feature vector was used. So, both algorithms are good for WSD. The performance depends on the combination of features and algorithms.

The similarity of this thesis and other rule-based approaches is that sense is disambiguated by some rules of linguistic information. The difference is that the rules of other rule-based approaches rely on the corpus, but the rules of this thesis are developed from the HowNet, which is independent of the corpus.

2.2 WSD Using HowNet

We have mentioned that HowNet is a new system. There is not much research using it. Yang, Zhang and Zhang (2000) use HowNet as an information source to do WSD research. They use the statistical approach. The disambiguation is based on a database, called a mutual information database. This database provides the information about the degree of a certain relation between a pair of sememes, which is the basic unit of HowNet's dictionary. The mutual information database is developed by the frequency of co-occurrence of sememes in the corpus. The implementation is on a corpus of 10,000 characters from Peoples' Daily with the mutual information database of 709,496 items. Before disambiguating, segmentation and sense tagging are done. The accuracy of the system is around 75%.

Yang, Zhang and Zhang applied one of HowNet's information characters, sememes which using a traditional algorithm to do the WSD. It found the advantage of using HowNet is that it can be easily applied to other kinds of corpa. Laborious hand tagging is also avoided. This research is a one of the pioneer research in Nature Language Processing based on HowNet. Actually, HowNet is a new and rich knowledge base. There is still much information useful for WSD or other research areas in NLP.

2.3 WordNet versus HowNet

WordNet is a popular database in Natural Language Processing. Actually, the semantic relations of nouns are quite similar in WordNet and HowNet (Wong and Fung, 2002). However, they are definitely different in meaning representation. In this part, we will give you a brief description to the similarities and differences between WordNet and HowNet.

WordNet (Miller, 1990; Miller and Felbaum, 1991; Fellbaum, 1998) is an on-line lexical database in which English nouns, verbs, adjectives and adverb are organized in terms of semantic relations such as synonymy, antonymy, hyponymy and meronymy. Such a lexical system was lacking in Chinese until the release of HowNet in 1999. But, HowNet (Dong, 1988) is not just a Chinese version of WordNet. It has its own structure in describing inter-concept relations and inter-attribute relations of concepts. Its design is to provide computer-readable knowledge that is crucial to text understanding and machine translation (Dong, 1999).

WordNet and HowNet share similar ideas in the definition of nouns. As mentioned in Miller (1993), the definition of a common noun typically consists of (i) its immediate superordinate term and (ii) some distinguishing features. These two components are used in the definition of nouns in WordNet and of concepts in HowNet. Superordinate terms (hypernyms) are organized in a hierarchical structure, in which the subordinates (hyponyms) inherit the distinguishing features of the superordinates (Miller, 1993). Hypernym gives a general classification of a concept and the distinguishing features provide more specific information to distinguish one concept from the other.

HowNet differs from WordNet in meaning representation. As mentioned in Miller *et al* (1993), meaning representation is either constructive or differential. HowNet uses the former whereas WordNet uses the latter. WordNet, using the differential approach, relies on the device that enables one to differentiate one concept from the other. It uses synsets to group similar concepts together and differentiate them. HowNet follows a different strategy. A close set of sememes (a base unit of meaning that cannot be further decomposed) is used to construct concept definitions. This is the difference between the differential approach and the constructive approach. As Chinese characters are monosyllabic and convey meaning, they are suitable sememe candidates to define concepts represented by Chinese words, of which most are polysyllabic. Using a bottom-up approach, a number of sememes were extracted after a meticulous examination of 6,000 Chinese characters. Similar sememes are combined and tested by using them to tag polysyllabic words. Eventually, a set of over 1400 sememes is found and organized hierarchically. Now, let use ‘teacher’ as an example.

Example 2.1: Meaning representation by WordNet and HowNet

1. Meaning representation in WordNet – synset:

{teacher, instructor} – (a person whose occupation is teaching)

2. Meaning representation in HowNet – combination of sememes and pointers:

Concept: 教師 (teacher)

Definition: DEF=human|人, *teach|教, education|教育

We can see that both WordNet and HowNet are organized by semantic relations. Semantic relations are relations between concepts and between their attributes. Concepts are represented by synsets in WordNet, but represented by a combination of sememes and pointers in HowNet. WordNet uses the synset {teacher, instructor} to represent the concept 'teacher'. HowNet decomposes this concept into sememes 'human|人', 'teach|教', and 'education|教育', and uses the pointer '*' to express the semantic relation between the concept 'teacher' and the event 'teach|教'. The sememe appearing in the first position of 'DEF' ('human|人') is the categorical attribute, which names the hypernym of the concept 'teacher'. Those sememes appearing in other positions ('teach|教', and 'education|教育') are additional attributes, which give more specific information to the concept: The sememe without pointer 'education|教育' is the specific attribute value of the concept 'teacher'. The one with the pointer '*' represents an event role relation, which states that the function of teacher is the agent of 'teach'.

The main difference of WordNet and HowNet lies in the theory of meaning representation. WordNet define the words by a differential approach. HowNet, using the constructive approach, use sememes (the basic unit of meaning) to build up the meaning of a concept. We will illustrate more about HowNet in following chapters.

Chapter 3 HowNet System

Before the introduction of the Sense Pruning system, we have to introduce HowNet first. HowNet version 1 was released in 1999. The most updated HowNet (version 2000) was released in October 2000. All the contents and some related research of HowNet are posted at (http://www.keenage.com), which is a bilingual (English – Chinese) web site. The HowNet system consists of HowNet and Information Structure. They will both be introduced in this chapter.

3.1 HowNet

HowNet is an on-line common-sense knowledge base (Dong, 1999). HowNet unveils inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. (The definition of concept and attributes will be illustrated later.) Dong (1988) believes that NLP ultimately requires the support of a powerful knowledge base. Dong defines knowledge as a system encompassing the varied relations amongst concepts or the attributes of concepts. That is if one acquires more concepts or captures more relations, one is more knowledgeable.

The design of HowNet is based on its ontological view of the objective world. All physical and non-physical matters undergo a continual process of motion and change in a specific space and time. The motion and change are usually reflected by a change in state that in turn, is manifested by a change in value of some attributes. The way we understand Attribute is that any one object necessarily carries a set of attributes. Similarities and differences between the objects are determined by the attributes they each carry. There can be no object without attributes. For instance,

human beings are attached with natural attributes such as race, colour, gender, age, ability to think, ability to use language as well as social attributes such as nationality, class origin, job, wealth etc.

HowNet is a fully computational knowledge-based providing computer-readable knowledge that is crucial to text understanding and machine translation (Dong, 1999). The knowledge structure of HowNet is a graph rather than a tree. It is devoted to demonstrating the general and specific properties of concepts. For instance, “human being” is the general property of concepts “doctor” and “patient”. (All the general properties will be documented in a file, called ‘Main Feature of Concepts’.) For concept “cure”, the agent of “cure” is the specific property of “doctor” and as the patient of “cure” is the specific property of “patient”. Figure 3.1 demonstrates a simple example for presentation of relations between concepts. This graph also illustrates HowNet makes the system computer-operable. The explicated relations of HowNet include hypernymy-hyponymy, synonymy, antonymy, metonymy, part-whole, attribute-host, material-product, converse, dynamic role and concept co-occurrence etc.

HowNet is not only a lexicon, but also a knowledge system support by a series of databases. Apart from the Knowledge Dictionary of concepts, HowNet has many supplementary databases such as the (1) Main Features of Concepts, (2) Secondary Features of Concepts, (3) Synonymous, Antonymous and Converse Relations (SACR) and (4) Event Relatedness and Role-shifting (ERRS). These are fundamental components of the system and not merely coding specifications and to be used in conjunction with the Knowledge Dictionary.

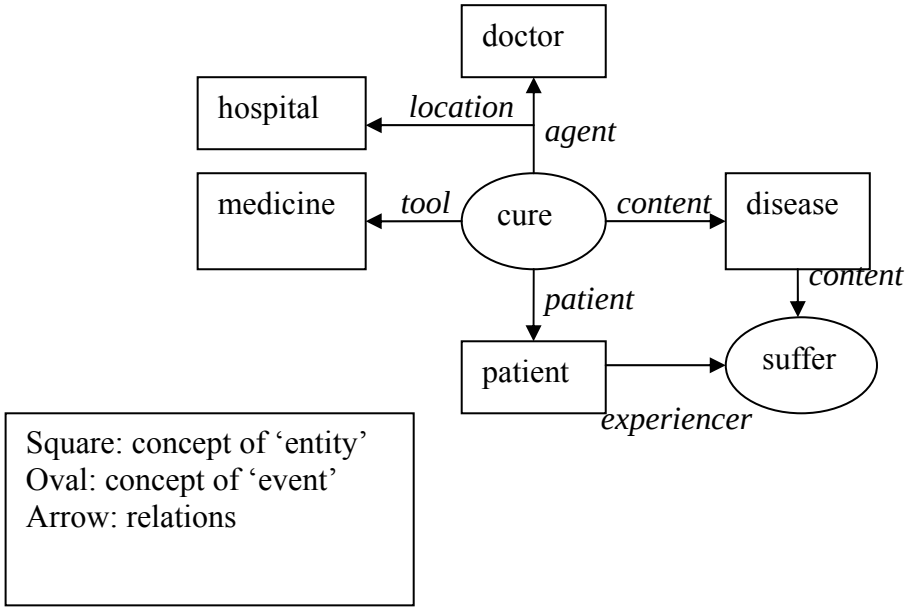


Figure 3.1 Graph for demonstrating the relations between concepts presented

HowNet constructs a graph structure of its knowledge base from relations. This is the fundamental distinction between HowNet and other tree-structure lexical databases (Dong, 1999). The basic unit of meaning in HowNet is called sememe that cannot be further decomposed. Though the defining of sememes is difficult, they are easily used and understood. The coverage of the set of sememes was tested against polysyllabic concepts to identify additional sememes. Eventually, a total of 1,503 sememes were found and organized hierarchically. The top-most level of classification in HowNet thus includes: entity|實體, event|事件, attribute|屬性 and attribute value|屬性值. It is important to point out that the classification is derived in a bottom-up manner. First, a set of sememes is extracted from about 6,000 Chinese characters. This is feasible because each Chinese character is monosyllabic and they are meaning-bearing. Similar sememes are grouped. This is a closed set from which all concepts are defined. Concept is an entry of the HowNet Knowledge Dictionary.

The bottom-up approach takes advantage of the fact that all concepts, either current or new, can be expressed using a combination of one or more existing Chinese characters. There has yet to be a new concept that has required the creation of a new Chinese character. Therefore, by deriving the set of sememes in a bottom-up fashion, it is believed that the set of sememes is stable and robust enough to describe all kinds of concepts, whether current or new. The fact is that HowNet has verified over 110,000 concepts. As yet, there has been no need to create new sememes to define new concepts, thus indicating the robustness of the sememe set of HowNet (Gan and Wong, 2000).

HowNet includes the Chinese-English Bilingual Knowledge Dictionary and HowNet Management System. We describe them in following sections.

3.1.1 Knowledge Dictionary

The Knowledge Dictionary is created by referring to the most common dictionaries. Dong (1999) believes knowledge is owned by all. The knowledge engineers shall first design the framework and suggest the prototype of the Knowledge Dictionary. Then, a common-sense Knowledge Dictionary constituting a knowledge system is constructed. This base describes general concepts and map out the relations among them. Based on this foundation, work can be extended to develop a vast and profound base and some specialized knowledge bases, which are created by on professionals in the respective fields. For instance, some professional words like the specialized words used in Medicine or Engineering, relevant professionals take the task of concept defining based on this prototype. The latest version (HowNet 2000) covers over 110,000 concepts in the Dictionary.

The format of each concept in the Knowledge Dictionary is as follows:

NO.= number of the entry

W_X = word / phrase form

G_X = word / phrase syntactic class

E_X = example of usage

DEF = concept definition

The first term is the concept number in the Knowledge Dictionary. The second term 'W_X' stores the name of that concept in word or phrase form. Then 'G_X' stores the syntactic class of that concept. For instance, a noun is represented by 'N', a verb is represented by 'V', etc. Sometimes, there is (are) some example(s) for that concept. It's (They are) stored in the 'E_X' term. It is not an essential term. The last term 'DEF' is the definition of that concept. This term can't be left out. 'DEF' consists of sememe(s), which are separated by a comma if there is more than one sememe. Each sememe is presented in both English and Chinese, which are separated by a bar ('|'). Sometimes, a pointer (it will be illustrated in section 4.2.1) is attached in front of the sememe. Example 3.1 is an example of concept "知網".

Example 3.1: Format of concept

Format of concept "知網" (HowNet) in Knowledge Dictionary

NO.=056352
W_C=知網
G_C=N
E_C=
DEF=software|軟件,#knowledge|知識

Explanation: Concept “知網” (HowNet) is the number 56352 entry in the Knowledge Dictionary. This concept is used as “知網” (HowNet). (Different concept entries may have the same appearance.) This concept is used as a noun (‘N’) in syntactic class. There is no example provided. The definition of this concept is that ‘software|軟件’ and it has the co-relationship (the kind of relations is indicated by the pointer ‘#’) with ‘knowledge|知識’.

3.1.2 Documents of HowNet Management System

The HowNet management system consists of a series of documents. They are (1) structure of the main features, (2) list of attributes, (3) list of attribute values, (4) list of quantity, (5) list of quantity values, (6) list of secondary features, (7) list of event role and attributes, (8) list of syntax, (9) list of antonyms, and (10) list of converse. With the conjunction of these documents, the knowledge dictionary of HowNet will be used in various areas automatically. That is why HowNet is computer-operable. Actually, all the documents are designed in a simple way, in list form, except the document of main features, which is well organized in a tree structure. The main feature of concepts will be stressed in this section.

The main feature of the concept is represented by the sememe, which is put in the first position in the definition (‘DEF’) of concept. The two main classes, ‘event|事件’ class and ‘entity|實體’ class, are organized in tree structure. The development of the hierarchy is based on HowNet’s ontology view.

The main feature of ‘event|事件’ class usually is attached with a set of necessary roles, which is expressed within curly brackets {}. These stipulated roles are described in the list of event roles and attributes. Those roles listed in the bracket are the necessary roles of the feature concerned. In other words, if any of the listed roles are missing, feature cannot constitute the named event. In some features, there is also a square bracket [] containing the relevant features. Under this hierarchy, four types of relationship could be found between main features. They are (1) hypernymy - hyponymy relation, (2) static - dynamic relation, (3) relatedness of events and (4) role-shifting. In this thesis, we start the research by use of relation (1).

In Figure 3.2, we can determine the hypernymy-hyponymy relation (上下位關係) between main features in the structure. We look in detail into the main features of ‘event|事件’ class by example 3.2.

Figure 3.2: The main features of ‘event|事件’ class

(In the above figure, ‘事件’ (‘event|事件’) is the head of the tree. The folder shape in yellow represents nodes and leaves of the tree. And if there is a small square in front of the folder shape, it is a node. Otherwise, it is a leaf. Further, if there is a ‘-’ inside the small square, it means that all ‘sons’ of that node exist. Otherwise, if there is a ‘+’ inside the small square, it means that all ‘sons’ of that node are hidden. The opened folder shape indicates that node is highlighted.)

Example 3.2: Illustration of the main feature for ‘event|事件’ class

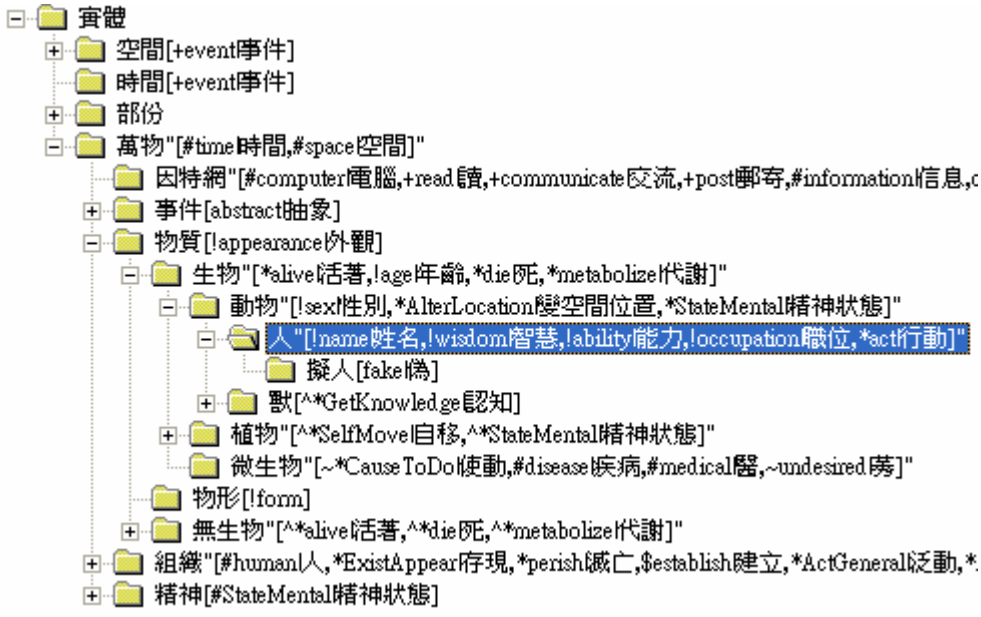
Concept: “取巧”
Definition: DEF=do|做,manner=sly|狡
Main feature: do|做 {agent,content,manner}

Explanation: We find that the ‘father’ of ‘do|做’ in the hierarchy of ‘event|事件’ class is ‘ActGeneral|泛動’. ‘do|做’ inherits all the properties of ‘ActGeneral|泛動’. So ‘sons’ of ‘do|做’ (which are hidden in the figure 3.2) inherit all its properties. Now, let us look at the curly bracket. That means when the ‘do|做’ takes place, it must involve the questions of who ‘do’ (the agent), what ‘do’ (the content), how ‘do’ (the manner).

The hierarchy in the ‘entity|實體’ class does not run as deep as in the ‘event|事件’ class and the descriptions are targeted at demonstrating both the general characteristics as well as the particular features. The general characteristics of each concept are listed in square brackets [] while the particular features are coded in the respective DEF.

Similar to ‘event|事件’, we also can determine the hypernymy-hyponymy relation (上下位關係) between main features in the structure. Figure 3.3 shows the structure of ‘entity|實體’ class. For example, the pair of main features ‘animal|動物’ and ‘people|人’. We look in detail into the main features of ‘entity|實體’ class by example 3.3.

Figure 3.3: The main features of ‘entity|實體’ class



(For the notation of Figure 3.3, please refer to Figure 3.2.)

Example 3.3: Illustration of the main feature for ‘entity|實體’ class

Concept: “警衛”
Definition: DEF= human|人,*protect|保護
Main feature: human|人 [!name|姓名,!wisdom|智慧,!ability|能力,!occupation|职位,*act|行动]

Explanation: We find that the ‘father’ of ‘human|人’ in the hierarchy of ‘entity|實體’ class is ‘AnimalHuman|動物’. ‘human|人’ inherits all the properties of ‘AnimalHuman|動物’. So as the son of ‘human|人’, it inherits all its properties. Now, let us look at the square bracket. The pointer ‘!’ indicates that ‘name’, ‘wisdom’, ‘ability’ and ‘occupation’ are properties of ‘human|人’. And, ‘human|人’ is the agent of concepts with main feature ‘act|行动’ or its descendants in the hierarchy of ‘event|事件’ class.

All the management documents make the HowNet system different from other lexicon databases. They support the Knowledge Dictionary to be extendable and more information in advance to definition (‘DEF’ of concept coding) only. HowNet aims to provide an ultimate solution to NLP. In this thesis, we just start to extract part of it to apply in WSD.

3.2 Information Structure

Information Structure is the extended research of HowNet in Chinese language. HowNet is a language independent system. But, Information Structure depends on particular language. That means, the same information will be presented in different structures in different languages. Based on the sememe of HowNet, the Information Structure Database describes the relationship of dynamic roles or properties between different words or phrases in Chinese. The Database allows us to know the way of description of a concept in Chinese. It is the principle of Chinese language structure.

So, what is “Information Structure”? It is a structure which consists of two or more words or combined words or phrases excluding the conjunction words. It is rational at syntactic and semantic levels. Also, it delivers particular information. We know that the basic unit of the syntactic level is part of speech, such as a noun or a verb. For Information Structure, the basic unit is a sememe or its properties, which are defined in HowNet.

Example 3.4: Illustration of information delivered in the message

“毒品走私集團”(Narcotic drugs smuggling group)

At the syntactic level of description, the analysis of Penn Treebank (Xue, 1999: 72-77) reveals the structure of this phrase that it is a noun phrase with the head of "集團" only modified by a relative clause "毒品走私" which involves operator movement. And, at the semantic level of description, we would indicate that "集團" (group) is the *agent* of the event "走私" (smuggle) and "毒品" (Narcotic drugs) is the *patient* of "走私" (smuggle).

The Information Structure of this example consists of two parts, the HowNet definitions and the dependency relations. The descriptions are as follows:

Word vs. Definitions:	毒品: medicine|葯物, ?addictive|嗜好物
	走私: transport|運送, manner= secret|秘, crime|罪
	集團: community|團體

Dependency relations:	毒品[patient]←走私←[agent]集團

(The description of Definitions is introduced in the previous chapter.)

In this example, the descriptions specify that a '*community*' is an *agent* involved in a '*transport*' event transporting the *patient* '*medicine*'. Furthermore, the '*transport*' event is a '*crime*' and the manner is '*secret*'. The '*medicine*' is a material of '*addictive*' products. The arrow between two word tokens is a dependency connection with the concept pointed to by the arrow denoting the dependent and the word token at the other end as the governor. The name of the dependency relation is enclosed in a square bracket and it could appear at either the dependent or the governor side.

Up to now, over 270 types of Information Structure patterns have been defined in the Information Structure Database. The pattern of Information Structure is specified in the following format:

(sememe) [DRel] → [DRel] (sememe),

where sememes are well defined by HowNet and *DRel* means the name of a dependency relation.

For the dependency relation to apply, the governor and the dependent must satisfy the requirement of the sememes. The following table shows a subset of the Information Structures. For instance, the type

"SEM_S = (時間|time) [時間|time] ← (事件|event)"

applies to the formation of the following units at various levels of linguistic structure:

Word level:	"午←睡" (afternoon nap)
Phrase level:	"暑期←補習" (summer study)
Sentence level:	"長期←商品短缺" (long-time shortage of commodities)
	"1999年12月9日星期四←發生洩漏" (leaking occurs on Thursday, December 9, 1999)

The above table also shows that Information Structures are derived in a bottom-up fashion from analyzing the mechanism used in the formation of words. That mechanism is also applicable for phrase and sentence construction in Chinese.

Chapter 4 SenPrune

In this thesis, we developed the program, called SenPrune. It will be presented in this chapter. First, there is an overview description of the program in section 4.1. Then we will illustrate how the knowledge of HowNet is extracted for disambiguating the senses or definitions. Within section 4.2, there are four sub-sections describing the different kinds of knowledge applied.

SenPrune is a Word Senses Pruning System. Before introducing the details, some terms have to be defined first:

Word token	The basic unit of a corpus. It can be a single or combined Chinese character. Also, it is matched to a concept in HowNet.
Complete Sentence	A sequence of word tokens limited between punctuation of Full Stop, Question Mark and Exclamation Mark.
Incomplete Sentence	A sequence of word tokens limited by punctuation of Full-stop, Question Mark, Exclamation Mark, Comma, Colon, Semi-colon and Slight-pause Mark.
Definition	The sense of word token by the HowNet Dictionary.

To further explain the above definitions, we show a part of the input text of SenPrune program.

Example 4.1 Sample of a sentence in input passage

…
。　{標點},
林文政　人,專,
小　屬性值,程度,欠,|人,幼,|屬性值,久暫,暫,|…
隊長　人,官,
昨天　時間,昔,日,
下午　時間,午,|屬性值,時間,午,
二　數量值,多少,基,眾,|數量值,次序,第,
時　時間,|時間,主,|時間,季,|時間,時,特,|…
，　{標點},
在　{condition},|{scope},|{進展},|…
簽出簿　讀物,@記錄,
簽　寫,|題寫,
上　{range},|{scope},|{動趨,上},|{發端},|{
查贓　調查,content=人工物,警,#罪,
。　{標點},
…

Explanation: There are 2 columns separated by a computer distance of 'tab' in the above sample. The content of the original corpus is on the left and the HowNet definition(s) (which are separated by '|' if there are more than one definition) of the related word token are on the right. The example of different definitions is shown in the following table:

Terms	Example(s)
Complete Sentence	"林文政小隊長…上查贓"
Incomplete Sentence	"林文政小隊長…二時"
	"在簽出簿…上查贓"
Work Token	"林文政"
	"二"
Definition(s)	"人,專," (definition of "林文政")
	"數量值,多少,基,眾,|數量值,次序,第," (definitions of "二")

4.1 Overall Design

SenPrune consists of two programs and 5 databases (Figure 4.1). In the following block diagram, the cylinder shape represents knowledge databases and the oval shape represents the C++ Programs. The rectangle represents the corpus.

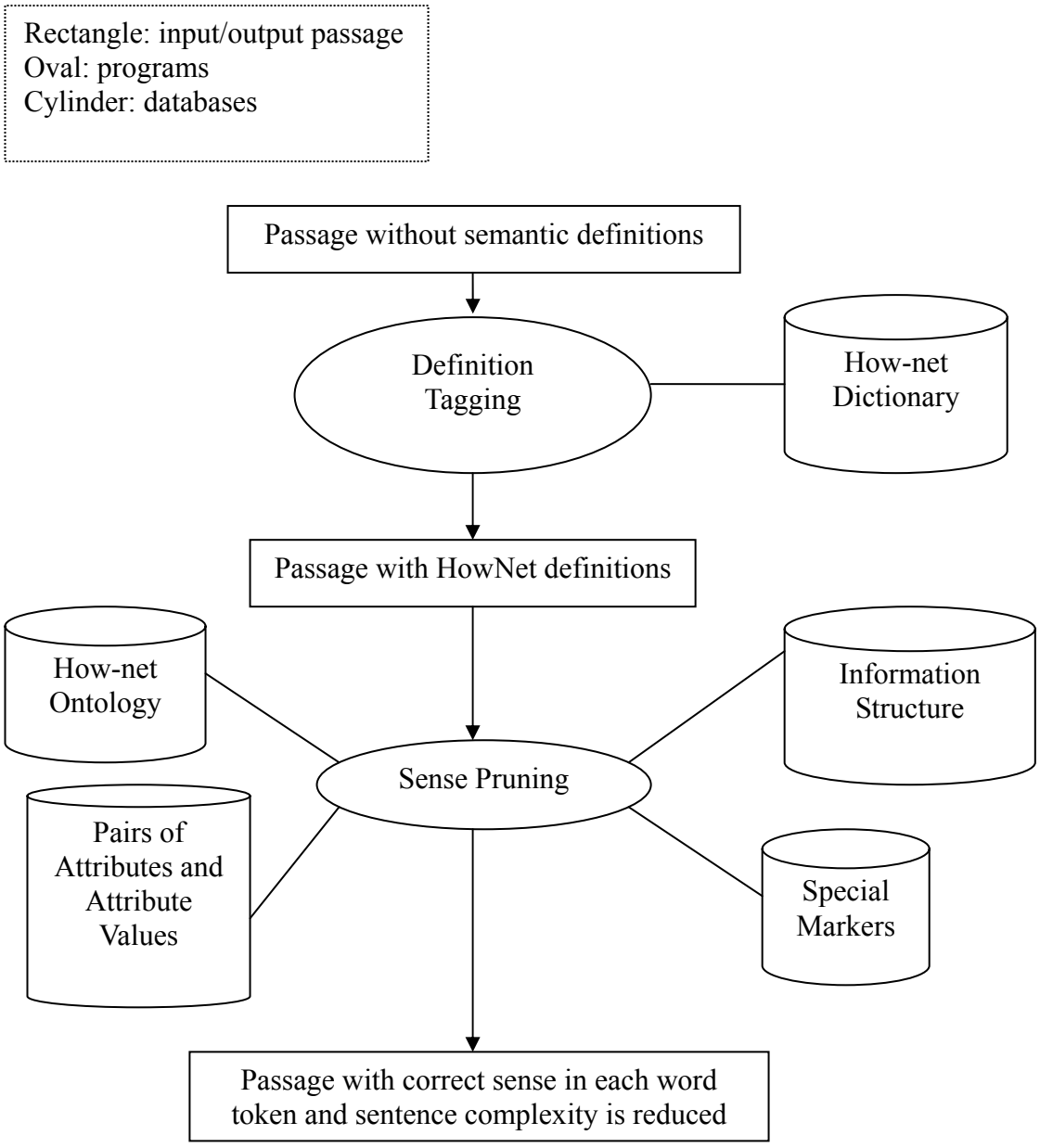


Figure 4.1: Block diagram of SenPrune.

From top to bottom of the block diagram, the input corpus is assumed to be segmented as the dictionary entries of HowNet. In the part “Definition Tagging”, the definition will be tagged to the word tokens in the corpus by an updating dictionary of HowNet. After that, the corpus is ready for disambiguation. In the program “SenPrune”, there are four databases that provide support for analysis the relationship between definitions of different word tokens. The details of databases will be discussed in the next chapter. The output of SenPrune is a corpus, where most definitions are pruned away form each word token except the correct definition.

4.2 Knowledge of SenPrune

After the introduction of fundamental knowledge of HowNet, we would like to deeply discuss what knowledge we have extracted from HowNet system for disambiguation.

In the block design, there are four databases to support SenPrune. They are created by different knowledge within HowNet. We will introduce them one by one.

4.2.1 Sememes Co-occurrence

The format of the concept in the Knowledge Dictionary is rich in information at the semantic level. It is a list of feature(s), which are separated by a Comma ‘,’. For the concepts of ‘event’ and ‘entity’, the main feature is put in the first position. Secondary features or dynamic roles or other main features (with special marks, which are defined by the HowNet management system) are put in other positions if there are any.

Example 4.2: Illustration of Sememes co-occurrence

Concept	Definition
病人	human|人,*SufferFrom|罹患,$cure|醫治,#medical|醫,undesired|莠
患	SufferFrom|罹患,medical|醫

Let us detail reveals the content of definition by these two examples:

First, 'human|人' is the main feature of 'patient|病人' that means patient is a human. Other features are '*SufferFrom|罹患', '$cure|醫治', '#medical|醫' and 'undesired|莠'. '*' mark indicates that 'patient|病人' is the agent of 'SufferFrom|罹患'. '$' mark indicates that 'patient|病人' is the patient of 'cure|醫治'. '#' mark indicates that 'patient|病人' is related to 'medical|醫'. Lastly, 'undesired|莠' is a term in the 'list of secondary features'. It indicates that 'patient|病人' is not a good thing.

For the other concept 'suffer|患', 'SufferFrom|罹患' is the main feature. The secondary feature is 'medical|醫', which is also a special term in 'list of secondary features'. It is attached with the general characteristics, [#cure|醫治,#disease|疾病].

Based on the information from the definition, we could find that the common sense of the concepts, e.g., a patient is a human who suffers from disease. We can conclude that 'patient|病人' and 'suffer|患' have a close relation at the semantic level. And, we observe that there is some co-occurrence in two closely related concepts.

Now, let us show how the information from the definition will benefit sense disambiguation. From the knowledge Dictionary, the total number of 'suffer|患' definitions is 5. They are:

Concept	Definition
患	1. SufferFrom|罹患,medical|醫
	2. emotion|情感,undesired|莠,#sad|憂愁
	3. phenomena|現象,undesired|莠,#unfortunate|不幸
	4. phenomena|現象,undesired|莠,hardship|艱,#unfortunate|不幸
	5. sad|憂愁

(The index is added to Definitions for easy reference. There is no index in the Knowledge Dictionary.)

Based on checking the co-occurrence of features between the concepts 'suffer|患' and 'patient|病人', both features in Definition 1 of concept 'suffer|患' occur in the definition of concept 'patient|病人'. Definition 2 to 4 of 'suffer|患' have one feature occurrence in the definition of the other concept. So, we can conclude that Definition 1 has a closer relationship with the definition of 'patient|病人' compared to Definition 2 – 5. Moreover, in a *complete sentence*, we could find more co-occurrence of features between definitions in different *word tokens* / concepts. Such as, in a Chinese complete sentence, "張三患了癌症，找名醫治療，醫生對這類病人已司空見慣。". The definition of word tokens with the co-occurrence features in Definition 1 of word token '患' are:

Word Token/ Concept	Definition
癌症	disease|疾病,
名醫	human|人,medical|醫,*cure|醫治,glorious|榮,
治療	cure|醫治,
	aValue|屬性值,property|特性,cure|醫治,
醫生	human|人,medical|醫,*cure|醫治,
病人	human| 人 ,*SufferFrom| 罹 患 ,$cure| 醫 治 ,#medical|醫,undesired|莠,

Obviously, it is easy to get the co-occurrence feature from the definition(s) of above word tokens and Definition 1 of ‘suffer|患’. (Feature ‘疾病’ of definition of ‘癌症’ is the occurrence in general characteristics of the secondary feature ‘醫’.) Up to now, we have a clear picture of how to get co-occurrence of features from HowNet’s definition construction. There is another characteristic of HowNet hierarchy that we shall use.

Let’s see another example:

Concept	Definition
好友	human|人,friend|友,
淚	part|部件,%AnimalHuman|動物,liquid|液,

In these two concepts, we can’t find any feature with co-occurrence in both definitions of concepts. However, we still find that these concepts have a relationship by the hierarchy of main features. Recall the information of ‘hypernymy-hyponymy relation’ relationship (上下位關係), ‘human|人’ is under the ‘AnimalHuman|動物’ that means ‘human’ is one kind of ‘AnimalHuman’. For the concept ‘tear|淚’, it is a part of ‘AnimalHuman|動物’. So, it is also a part of ‘human’. In another example, ‘animate|生物’ is above ‘AnimalHuman|動物’ in the hierarchy. Then, ‘AnimalHuman|動物’ is a kind of ‘animate|生物’. But in this case, the concept ‘tear|淚’ is not a part of ‘animate|生物’.

To conclude, whenever checking the co-occurrence of features in definitions, we have to apply the hierarchy structure of main features to determine whether items which exist upper position of the main feature of one definition occur in another definition or not.

4.2.2 Information Structure

We had introduced the HowNet Information Structure in section 3.2. In this section, we will describe how to create a database from extracting the original structure. HowNet Information Structure has defined more than 279 types of patterns. It is classified into 45 relationships, such as '合成', '修飾', '並列' etc. There is at least one type of pattern under each relationship.

There is a sample of part of the structure:

1. 合成

    1.1.1 SYN_S=N <- N

    SEM_S=(萬物) [合成] <- (萬物)

        例子：道-路，房-屋，艦-艇，方-法，方-向，知-識，壕-溝，皮-革，奴-隸，話-語，牙-齒，疾-病，三言-兩語，千方-百計，千軍-萬馬，罪魁-禍首，

    …

    2.1.1 SYN_S=A <- A

    SEM_S=(屬性) [合成] <- (屬性)

        例子：大-小，長-短，粗-細，深-淺，輕-重，厚-薄，薄-厚，高-低，高-矮，冷-暖，濃-淡，胖-瘦，難-易，安-危，貧-富，利-弊，益-害，
        輕重-緩急，天高-地厚，

    …

2. 修飾

…

Based on the above structure, we would like to extract the information of 'SEM_S=(sememe) [relationship] <- (sememe)' for creating a database. It is list of patterns in which a pattern is a pair of sememes.

The format of the database:

<features1><tab><features2>

The following is a sample of part of the Information Structure Pattern database file created:

…
人,官, <tab>人,專,
人,官, <tab>文字,姓,人,專,
人,專, <tab>人,官,
土石, <tab>物質
天變<tab>天體,
天體, <tab>天變
文字,姓,人,專, <tab>人,官,
文字,姓,人,專, <tab>文字,姓,人,專,
名量, <tab>名量,
名量, <tab>數量值,
地方, <tab>他移
地方, <tab>自移
地方, <tab>萬物
地方,專, <tab>人,官,
地方,專, <tab>地方,專
地方,專, <tab>物質
地方,專, <tab>組織
…

The pattern consists of a pair of features or a pair of feature lists. It indicates that the left feature(s) has a relationship to the right feature(s) or vice versa. In general, the features will be followed by a Comma ‘,’ except some cases. For instances, the pattern ‘地方,專, <tab> 組織’. ‘組織’ doesn’t have a Comma in the end. That means all items, which under the position of ‘組織’ in the hierarchy of main features, are valid for this pattern. Actually, in this stage, we would not care what relationship is it between the feature(s) in the pattern.

Example 4.3: Illustration of Information Structure Knowledge for Sense Pruning (1)

Pattern: ‘地方,專, <tab> 人,官,’

Concepts: ‘American|美國’ and ‘President|總統’

Concept	Definition
美國	place|地方,country|國家,ProperName|專,(North America|北美)
總統	human|人,official|官,country|國家

Explanation: By the pattern, we could say that ‘American|美國’ and ‘President|總統’ have a relationship. First, the features, ‘地方’ and ‘專’ (left side of pattern), occur in the definition of the concept ‘American|美國’. And the features, ‘人’ and ‘官’ (right side of pattern), occur in the definition of ‘President|總統’. So, both concepts satisfy the condition of pattern.

Example 4.4: Illustration of Information Structure Knowledge for Sense Pruning (2)

Pattern: ‘地方,專, <tab> 組織’

Concepts: ‘The People Republic of China|中華人民共和國’ and ‘the State Council| 國務院’

Concept	Definition
中華人民共和國	place|地方,country|國家,ProperName|專,(China|中國)
國務院	institution|機構,politics|政,ProperName|專,

Explanation: In this pattern, ‘organization|組織’ does not end with a Comma. That means the right side of the pattern is valid for the main feature of definition in concept is ‘organization|組織’ or those under the position of it in the hierarchy of the Main Features Structure, which is one of the HowNet management systems.

The definition of ‘The People’s Republic of China|中華人民共和國’ with features ‘place|地方’ and ‘ProperName|專’, which match the left side of pattern. The main feature ‘institution|機構’ of ‘the State Council|國務院’ definition is under the position of ‘organization|組織’ in the hierarchy of the Main Features Structure one of the HowNet management systems. So, two concepts satisfy the both sides of the pattern.

Now, we know how to implement the Information Structure to indicate the relationship between two concepts. One point is to emphasize that the Information Structure depends on language. The database we create is only for Chinese. And, the patterns of the database are designed in only one direction. For instance, if ‘the State Council|國務院’ occur in position before ‘The People’s Republic of China|中華人民共和國’ in a sentence, it does not satisfy the pattern ‘地方,專, <tab> 組織’.

4.2.3 Pair of Attributes and Attribute Values

‘Attribute’ is an important concept in the HowNet philosophy. Mr. Dong (1999) said that ‘there will be no object without attributes’. To understand the ‘Attribute’, we must know it is necessary for any object. And, object has a set of ‘Attributes’, which can be used to determine the similarities and differences between objects. Let’s use human being as an example. The set of natural ‘Attributes’ are race, color, gender, age etc. Also, human being has the set of social ‘Attribute’ such as nationality, profession, wealth etc. In coding specifications of HowNet, ‘Attributes’ are necessarily defined in terms of the concept or the possible classes of the concept.

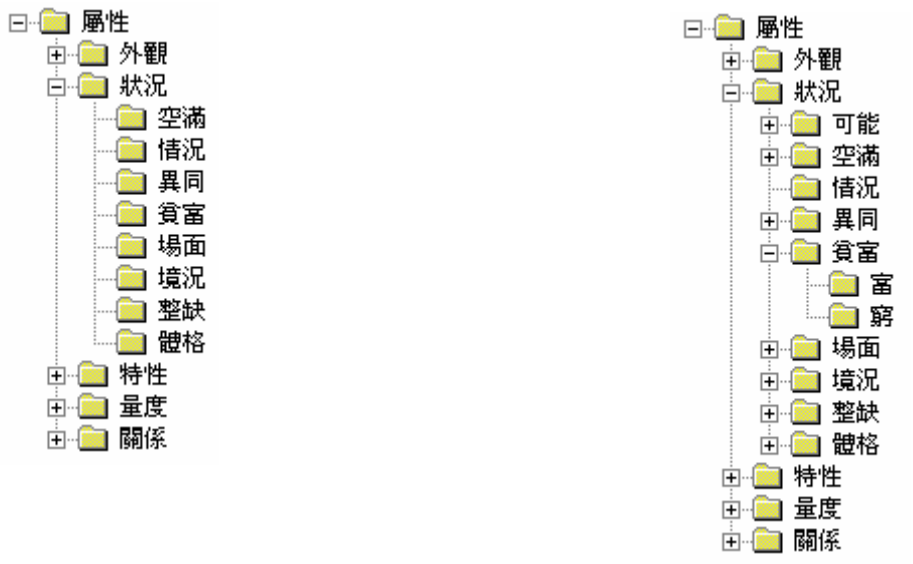


Figure 4.2: The structure of ‘Attribute’ and ‘Attribute Value’

(The left one is the list of ‘Attributes’ and right one is the list of ‘Attribute values’.)

In HowNet, the objective world is classified as a tree structure with the head ‘entity|實體’ (referring to hierarchy of ‘entity|實體’ class). Some ‘Attributes’ are attached as the general characteristics to the sememes, the items in hierarchy. For example, for sememe ‘human|人’, the Attributes (‘name|姓名’, ‘wisdom|智慧’, ‘ability|能力’, ‘occupation|職位’) are attached. Some general ‘Attributes’, such as ‘age|年齡’ and ‘sex|性別’, are attached to sememes ‘animate|生物’ and ‘AnimalHuman|動物’, which are in upper position of ‘entity|實體’ class hierarchy to sememe ‘human|人’. ‘human|人’ is supposed to inherit all the general characteristics from the father nodes in the hierarchy.

Example 4.5: Illustration of ‘Pair of Attributes and Attribute Values’ for Sense Pruning

Let’s see this Chinese phrase ‘顯見林小隊長死意甚堅’. There are 8 word tokens. (‘顯見-林-小-隊長-死-意-甚-堅’) We would like to focus on the analysis of ‘意’ and ‘堅’.

Concept	Definition
意	thought|念頭
堅	aValue|屬性值,will|意志,strong|強,desired|良

Both concepts have various definitions. The above definitions are the correct one. “堅” is an Attribute value of Attribute ‘will|意志’. We also find that Attribute ‘will|意志’ has relationship with sememe ‘mental|精神’ based on human experience. Tracing main features in hierarchy of ‘entity|實體’, ‘mental|精神’ is the grand parent of ‘thought|念頭’. So, we conclude that word tokens ‘意’ and ‘堅’ have a relationship.

Now, we would like to create a database that stores the pattern of the object and its attribute, which have a relationship. The database is developed based on this information and also enriches it if there is any item missing. For instance, we found ‘will|意志’ and ‘mental|精神’ have a relationship, but it is missed in the HowNet definition.

Now, we would like to describe the database in short. First, the format as:

<Attribute> <tab> <sememe>

The following is a sample of part of the database file:

…
久暫<tab>念頭
內容<tab>讀物
內容<tab>書刊
內容<tab>信息
內容<tab>符號
內容<tab>圖象
尺寸<tab>物質
尺寸<tab>房間
尺寸<tab>樂器
方向<tab>建築物
…

In the above format, the left side is the ‘Attributes’ and the right side is the related sememes in ‘entity|實體’ class. Actually, each ‘Attribute’ may have one or more related sememes, such as ‘內容’ has 5 related sememes (‘讀物’, ‘書刊’, ‘信息’, ‘符號’ and ‘圖象’). For simplicity, we developed each record in a one-to-one structure.

4.2.4 Special Markers for Functional Words

In HowNet, the functional definition of concepts is indicated by a curved bracket surrounding the feature(s). For instance, for the concept ‘的’, one of its definition is ‘{DeChinese|構助}’ (some functional definition of concepts does not have the Chinese interpretation). Unfortunately, there is not enough information related to the sememes of functional words. And, these functional words always have many definitions. And some relationships between the definitions are obvious in human experience but not found in HowNet.

Example 4.6: Illustration of functional word “在” for sense pruning(1)

For concept “在”, definition ‘{location}’ has a relation with those definitions with sememe ‘place|地方’.

In sentence of “在杭州南路口…”

在	杭州
{condition},	place|地方,city|市,ProperName|專,(China|中國)
{scope},	
{time},	
{location},	
{進展},	
exist|存在,	
depend|依靠,	
alive|活著,	
situated|處于,	
ResultFrom|緣于,	

(The first row of the above table is the concept and their definition(s) are list in remaining rows.)

In this case, concept “在” has 5 functional definitions. It is obvious that definition ‘{location}’ has a relationship with the definition of concept “杭州” by human experience.

Example 4.7: Illustration of functional word “在” for sense pruning (2)

For concept “在”, definition ‘{time}’ has a relation with those definitions with sememe ‘time|時間’.

In this case, the functional definition ‘{time}’ of concept “在” has a relationship with the definition of concept “十月”.

Sentence: “在十月慶典期間…”

在	十月
{condition},	time|時間,month|月
{scope},	
{time},	
{location},	
{進展},	
exist|存在,	
depend|依靠,	
alive|活著,	
situated|處于,	
ResultFrom|緣于,	

(The first row of the above table is the concept and their definition(s) are list in remaining rows.)

Both examples (4.6 & 4.7) are obvious by human experience. However, this information can’t be determined in the previous sections. So, we create a database which stores the above information. First of all, a collection of word data, which provides the frequency of Chinese word tokens, is used for selecting the concepts. We just focus on those word tokens that have a high frequency. And the principle of creating database is to keep it concise and simple. Now, there are 38 records in the database.

The key of the database is the Word Token, the Marker (such as ‘在’ in previous example.) Each marker is followed by its functional definition and the rest part is the related sememe/feature/concept.

The format of each record is:

<Marker(concept)><tab><definition><tab><feature/definition/concept>

The following is a sample of part of the database file:

…
在<tab>{location},<tab>地方,
在<tab>{location},<tab>位置,
在<tab>{time},<tab>時間,
在<tab>{time},<tab>數量值,次序,第,
…

Now, we have described the four databases of our Sense Pruning system. Actually, all the information of databases is extracted from the HowNet. We are exploring the power of HowNet to the research of computational linguistic.

Chapter 5 Evaluation

In this chapter, we will describe the implementation of our Sense Pruning system and evaluate the results. First, the corpus of implementation is introduced in section 5.1. The methodology and algorithm will be described in section 5.2. The criteria of result are defined in section 5.3. In section 5.4, there is the discussion of the result in different effects.

5.1 Corpus

In this research, we use a corpus of newspaper texts covering the crime domain. The corpus consists of 30,000 words that are extracted from the Sinica corpus, version 3.0 (CKIP, 1995). There are totally 103 passages. That corpus is a base of hypertext tagged with syntactical information. Since HowNet and Sinica are two different systems, they are different in segmentation. Before the experiment, those corpuses are re-segmented to match the concepts of HowNet. Afterward, they are tagged by the HowNet definitions. After that, all word tokens are annotated with the updated definitions from HowNet Knowledge Dictionary. Now, these passages are ready to be input into the program “SenPrune”.

There is the statistic of the 103 passages.

Item	Mean
Number of word tokens per corpus	234.1845
Number of word tokens per Complete Sentences	44.7514
Number of word tokens per Incomplete Sentences	7.6599

5.2 Methodology

Up to now, we have analyzed the information of HowNet for disambiguation. We apply this information to the system “SenPrune”.

First of all, we have to simplify the input corpus to a format as:

<word token><tab><HowNet definitions>

(‘word token’ includes all words and punctuation marks.)

The input passage will be divided into several analytical lengths, which may be a Complete Sentence or Incomplete Sentence or a certain length of window size.

The algorithm of the system SenPrune works as follows:

1. Initialize score of definition(s) of each word token in input passage to 0;
2. for each word within the analytical length,
    a. for each word token (W),
        i. for each definition (A)
            * for each definition (B), which belongs to later word tokens of (W).
                - scoring by ‘sememes co-occurrence’ (A, B);
                - scoring by ‘information structure pattern’ (A, B);
                - scoring by ‘Pair of Attributes and Attributes Values’ (A, B);
                - scoring by ‘Special Marker’ (A, B);
3. for each word token of input passage
    a. prune out the definition that its score is 0;
    b. write the word token tagged with updated definitions into output passage.

The above algorithm consists of three parts. The first part (1) is for initialize all word tokens from the input passage with a score zero. Second is scoring based on the four databases and the last part is for pruning and output. In the scoring, there are four parts from different databases and they are evenly rated. Each definition of a word token compares to definitions of other word tokens within the analytical length. After scoring, the system will prune out those definitions that have not scored. Finally, a new passage, in which the word tokens tagged with less definition, is created.

5.3 Criteria

The objective of this thesis is to keep the correct definition and reduce the sentence complexity. So, we consider two statistics to evaluate the result: One is the 'Recall' and another is 'Combination Reduced'.

Recall means the correct definition of each word token in the output passage.

Combination Reduced means reduction of sentence complexity. Sentence complexity is calculated by the product of the definition number of each word token in the sentence. We get the rate of combination reduced by the sentence complexity of output passage divides the sentence complexity of input passage.

5.4 Results

We did the experiment on 103 passages from Sinica of Taiwan. After the "SenPrune" program, the unrelated definitions of word tokens have been pruned out. The result is calculated by comparison to another set of this corpus which had been disambiguated by linguistic experts. Since there are 103 passages in total, we run SenPrune in each one respectively. In each passage, we get the value of 'Recall' and 'Combination reduced' rates. Eventually, we get the result by calculating the geometric mean of 103 sets of values.

5.4.1 Experiment 1: Complete Sentence

The standard experiment uses Complete Sentence as the analytical length. Each definition of each word token is compared to all definitions of other word token within the same sentence. After the implementation, we get the result:

Recall Rate	Combination Reduced Rate
97.13%	47.63%

The high Recall Rate shows that the information provided in the HowNet system is valued and useful for NLP. For our purpose, we want to reduce the workload of researchers in the next stage but the correct answer is not pruned away. Ideally, the high rate is target at 100%. However, in fact, we cannot achieve it since some information of HowNet has not been developed well. We have found that some definitions of function words are constructed by some sememes that are not well defined. For example, '與' has three definitions, which are '{partner},', '{target},' and '{和},'. All those sense units in brackets are belonging to the database of

'dynamic roles'. In disambiguating system of "SenPrune", we can't apply this information since they are not well defined right now.

5.4.2 Experiment 2: Effect of Window Size

We would like to show the effect when more information is supported in grading. We keep the Complete Sentence as a unit and then adjust the window size of grading. Each definition of the word token will compare to the definitions of other word tokens, which are within the distance of window size and also the same Completed Sentence. Four sets of implementation are done. The results are:

Window Size	Recall Rate	Combination Reduced Rate
1	89.84%	89.93%
3	91.20%	80.27%
5	92.84%	71.10%
7	93.53%	68.08%
9	94.14%	65.30%

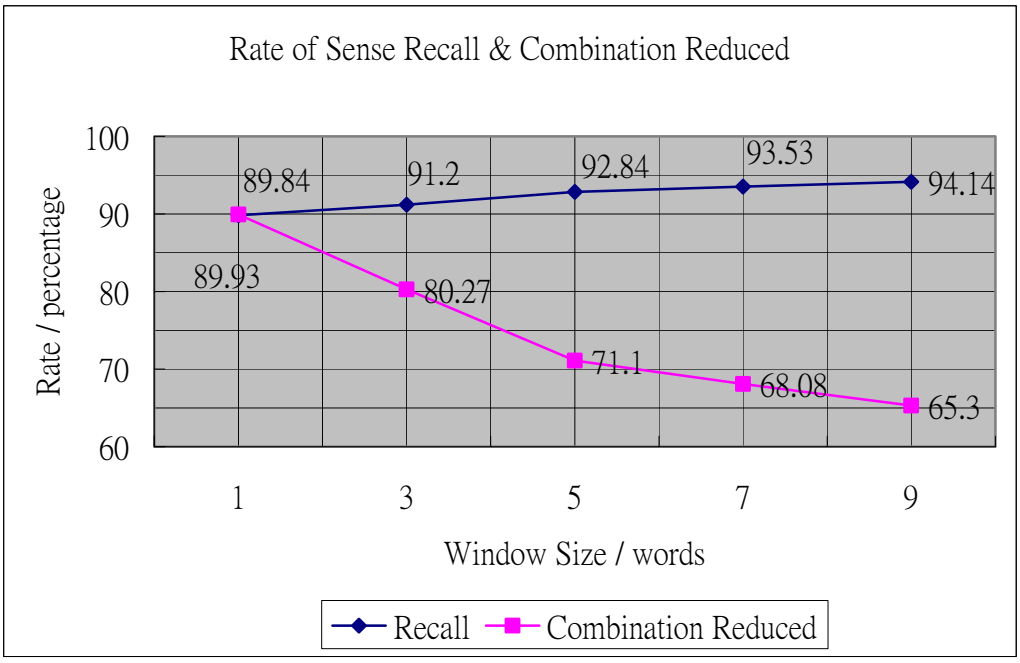


Figure 5.1: Rate of Recall and Combination Reduced in 'Effect of window size'

Figure 5.1 shows that the Recall Rate is proportional to the Window Size. As the Window Size increases, Recall Rate will increase. There is no doubt that the more information that is provided; the greater the accuracy is achieved. Otherwise, it shows that the related information between the definitions does not need to be in the neighbor position. It could be any position in the Complete Sentence. This is one of the philosophies of HowNet that the related information could be found in any position within a complete sentence.

Referring to the lower line in the figure 5.1, there is a trend that the rate of Combination Reduced is decreased as the Window Size is increased. Similar to the Recall Rate, the longer the window size, the more information that is provided to support the definitions in the word token. As a result, more definitions get scores and should not be pruned from the output file.

5.4.3 Experiment 3: Effect by Analytical Unit

The purpose of this implementation is to investigate the coverage of related information for definition comparison in a sentence. Each definition of one word token shall be compared to all definitions of other word tokens within the analytical length. In previous implementations, the analysis unit is limited by the Complete Sentence even with the windows of varying size. Now, we show the effect when the analysis unit for both Complete and Incomplete sentences. The result is:

Analysis Unit	Recall Rate	Combination Reduced Rate
Complete Sentence	97.13%	47.63%
Incomplete Sentence	92.48%	67.78%

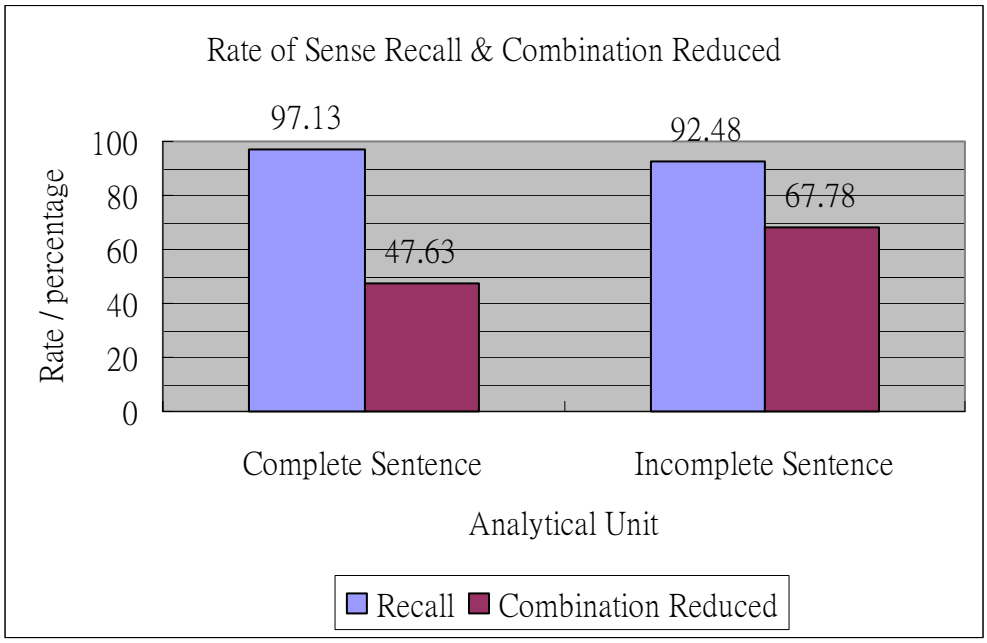


Figure 5.2: Rate of Recall and Combination Reduced in Effect by analysis unit

In figure 5.2, we found that the Recall Rate of an Incomplete Sentence is roughly same as that of Window Size 5 in Experiment 2, which is 92.84%. But, the combination reduced rate is smaller. In this experiment, it is 67.78% compared to 71.1% in experiment of Window Size 5 in the previous section. This shows that when the recall rate is the same, the longer the analytical length, the better the combination reduced result (higher rate). Because the analytical length of Experiment 2 is Complete Sentence.

And, the average of word tokens of the analysis length by Incomplete Sentence is 7.66. Again, refer to Figure 5.1. If the Window Size is 7.66, the Recall Rate is 94% by linear approximation. Compared to the experiment of Incomplete Sentence, it is 92.48% only. It also shows that when the window size is the same, the longer the analytical length, the better the Recall result (higher rate).

From the result of this section, we found that longer analytical length returns better result of Recall Rate (Increase). The results show that the information, which is useful for disambiguating senses, can be found in different positions within the same sentence. This result can be as the footnote of Dong's idea (1999) that the bigger the context the better the understanding.

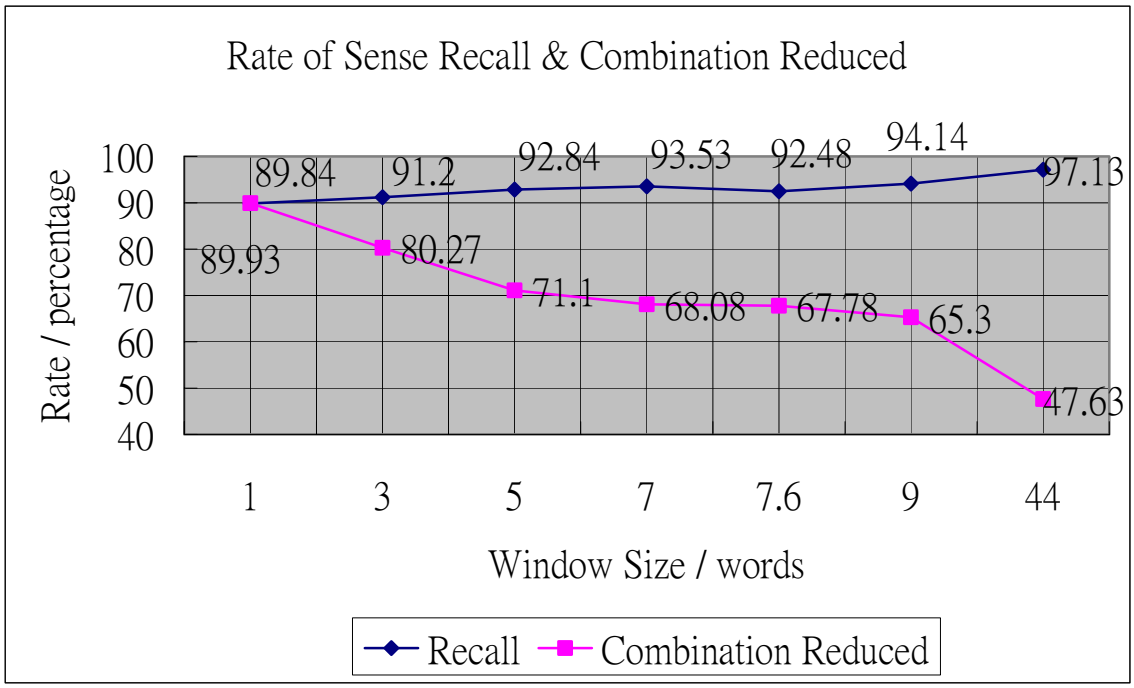


Figure 5.3: Rate of Recall and Combination Reduced in Effect by window size & analysis unit

In figure 5.3, we merge the result of figure 5.1 and figure 5.2. It provides a better view of the conclusion in this section.

5.4.4 Experiment 4: Effect of Databases

In this thesis, there are totally four databases developed for sense pruning. In this experiment, we would like to show the effect as the number of databases

involved. For the analytical length, we use the Incomplete Sentence. So, the result of all databases involved is same as experiment 2.

There are 4 databases: (1) main features of 'event|事件' & 'entity|實體', (2) information structure, (3) attributes and attribute values & (4) markers for functional words. The results are:

Database(s)	Recall	Combination Reduced
(1) only	89.82	94.04
(1) & (2)	92.08	71.38
(1), (2) & (3)	92.13	71.33
ALL	92.48	67.79

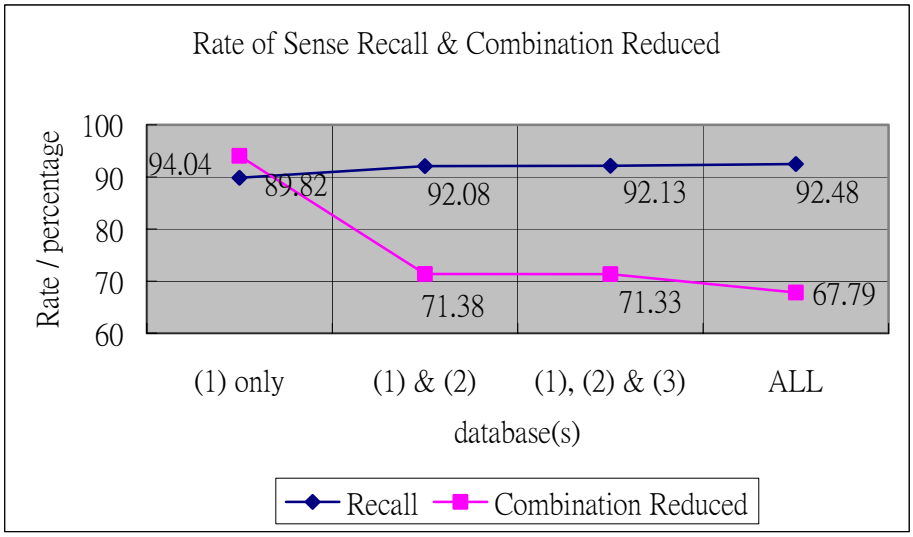


Figure 5.4: Rate of Recall and Combination Reduced in Effect by different databases

In Figure 5.4, we found that if only database (1), the recall rate can achieve around 90 per cent. This result is higher than that of baseline in next experiment too. This database is useful for the checking the sememes co-occurrence. It shows that the concept definition of HowNet is well defined. As the databases increased, the recall rate is increased slightly. It also shows that other databases are useful for the sense pruning.

5.4.5 Experiment 5: Baseline

Since it is difficult to find some other similar research, we try to implement the system in a simple way. We implemented a simply Word Sense Disambiguation by statistic method. A database, which stores the appearance frequency of concepts' definitions, is created. Since there is not a similar corpus or database available, we divide the corpus into Training and Testing parts. For instance, we use one fifth of 103 passages as the Testing corpus and the rest four fifth as the Training corpus. The database is created by counting the definitions from the Training corpus. WSD is implemented by the Testing corpus. The methodology is that WSD of each word token in the Testing corpus is based on selecting the highest frequency definition from the statistic of the database. If the word token doesn't exist in the database, the first definition of the word token in the Training corpus will be selected. To maintain the fairness, the sequences of definitions in the word tokens are not done by any rules. We assume it is in randomness.

We divide the corpus into 5 groups, such as 20, 20, 20, 21, 22 passages per group. In each group, those passages are the Testing corpus and the remaining passages are the Training corpus. The mean of 5 groups' results are:

Group	Recall Rate
1	83.92%
2	85.65%
3	86.49%
4	83.57%
5	85.67%
mean	85.07%

Since there is only one definition in the result by the modified system, the rate of sentence meaning combination reduced is approximately 100 per cent. For the rate

of recall, the result is 86.80%. We find that it is lower than all the previous result. The previous lowest result is 89.80%, from the implementation of 'full stop' analysis length with window size 1. In this baseline implementation, since the Testing and Training corpus are from the same domains, a relatively good result can be expected. For our research, the system is independent of corpus. So, it shows that this sense pruning system using HowNet provides a valued result for the researchers in further steps.

5.5 Speed Issue

In the SenPrune system, we do the scoring by all the probabilities of the definitions' combinations. It is a very large work of computation. We try to investigate the running time of the above implementation. The implementation is done by PC with configuration:

CPU: Intel P2 300Mhz

RAM: 128M SDRAM

OS: Microsoft NT4.

The result is:

Window Size	Running Time (second)
1	200.33
3	551.05
5	888.96
7	1214.56
9	1561.83

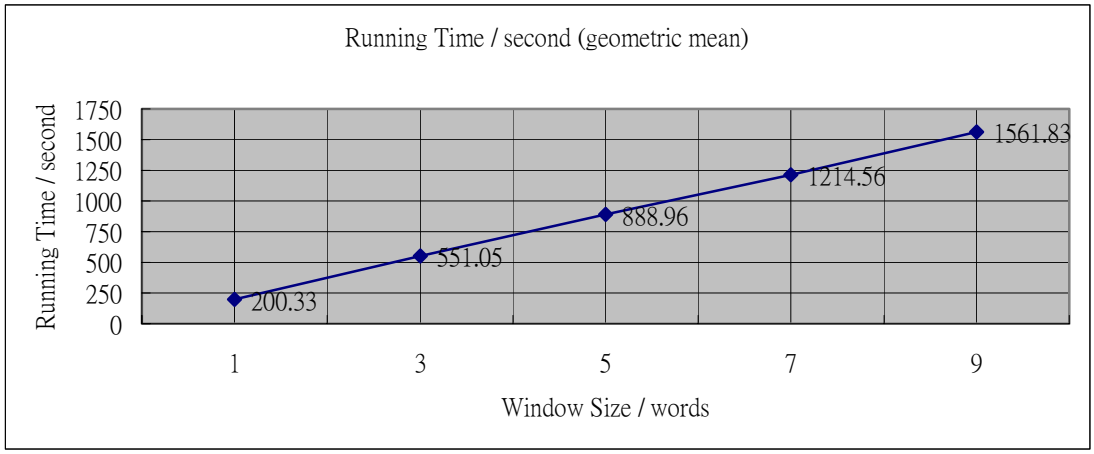


Figure 5.5: Speed Issue

In this part, we would like to show the relation of performance and cost of time. Now, we can find that if the analysis length is longer, more running time is required for same rate of recall to be improved. Actually, the running cost is one of obstacle in this project.

Chapter 6 Conclusion

6.1 Contributions

From the results of our program "SenPrune", we found that the knowledge-based sense pruning can achieve a satisfactory result. Though the rate of recall is still not 100%, we believe that it is a correct way to perform sense pruning using HowNet. We are the first to carry out the approach of sense pruning. We hope this result will bring more attention to sense pruning or even Re-constructive Text Understanding. Second, we are the first to totally apply the HowNet philosophy and knowledge instead of using the Knowledge Dictionary only. HowNet is a powerful knowledge base. We believe there is still more useful knowledge for NLP yet to be explored. We consider ourselves to be pioneer of Re-constructive Text Understanding Approach using HowNet.

6.2 Limitations of this Approach

Based on the algorithm of program SenPrune, we find that HowNet is a powerful system for NLP. But, still some information is missed making it incomplete. In definition of concepts, the hierarchy of the main features provides strong analysis of the world. We could use it for finding the relations between the concepts and also between the features and concepts. The information structure is the extended research of HowNet. It is a language-depended information database. We reorganized it as a database that presents the characteristics of Chinese for disambiguating the word

tokens. This database could be more powerful since the information structure is not released officially. There will be greater coverage and more information provided in coming released version. The rest two databases that we use for disambiguating are ‘Pair of Attributes and Attribute Values’ and ‘Special Markers for Functional Words’. We enhance the research by creating two databases. Database ‘Pair of Attributes and Attribute Values’ focuses on the attributes of features. It enriches the entries of HowNet and makes it more efficient for programming. Database ‘Special Marker’ is an idea to get some special words in Chinese. We call that concept is ‘Marker’. It works in most functional words, which is not clear enough in HowNet system. Actually, these two databases are created roughly. We could do it better by doing more research in attributes of HowNet and the functional words.

6.3 Future Work

Besides the functional words, there is still information of HowNet that we cannot apply directly now, such as the dynamic roles of the main features in ‘event|事件’ class hierarchy. Also, the algorithm of system SenPrune could be developed for more implementation, such as the rating of different scoring functions. We hadn’t done it now. It was because the running time is huge. And also, a more important reason is that we have to get more information before judge the rating of different scoring function. For doing that, more linguistic knowledge is involved. In future, we believe Sense Pruning shall be integrated to other parts of Reconstructive Text Understanding.

References

CKIP (1995) *中央研究院平衡語料庫的內容與說明，技術報告 95-02* [Technical Report no. 95-02, the content and illustration of Sinica corpus of Academia Sinica]. Institute of Information Science, Academia Sinica.

Dong, Zhendong (1988) *Knowledge Description: What, How and Who?* In: Proceedings of International Symposium on Electronic Dictionary, Tokyo, Japan.

Dong, Zhendong (1999) *HowNet*. http://www.keenage.com.

Dong, Zhendong (1999) *Bigger Context and Better Understanding – Expectation on Future MT Technology*. In: Proceedings of International Conference on Machine Translation & Computer Language Information Processing, 26-28 June 1999, Beijing, China, pp. 17-25.

Dong Zhendong (2000) *中文信息結構模式* [The pattern of Chinese information structure], ms.

Erhong Yang, Guoqing Zhang and Yongkui Zhang (2000) *The Research of Word Sense Disambiguation Method Based on Co-occurrence Frequency of HowNet*. In: Proceedings of the 2nd Chinese Language Processing Workshop, Association for Computational Linguistics 2000 Conference, October 2000, Hong Kong, pp.60-65.

Gan, Kok-Wee and Ping-Wai Wong (2000) *Annotating information structures in Chinese text using HowNet*, In: Proceedings of the 2nd Chinese Language Processing Workshop, Association for Computational Linguistics 2000 Conference, October 2000, Hong Kong, pp.85-92.

Gan, Kok-Wee and Wai-Mun Tham (1999) *基於知網的常識知識標註* [General Knowledge Annotation Based on HowNet], Computational Linguistics and Chinese Language Processing, vol. 4, 2, pp. 39-86.

Miller, George A. (1990) *WordNet: An Online Lexical Database*. In: Special Issue of International Journal of Lexicography, Vol 3, No. 4.

Miller, George A. (1993) *Nouns in WordNet: a lexical inheritance system*. Five Papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton University.

Miller, George A. and Christiane Fellbaum (1991) *Semantic Networks of English*. In: Beth Levin and Steven Pinker (eds.) Lexical and Conceptual Semantics, Cognition Special Issue. Amsterdam, The Netherlands: Elsevier Science Publishers, B.V., pp. 197-229.

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller (1993) *Introduction to WordNet: An On-line Lexical Database*. Five Papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton University.

Mooney, R. J. (1996) *Comparative Experiments on Disambiguating Word Senses: An Illustration of the Role of Bias in Machine Learning.* In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp 82-91. Somerset, N.J.: Association for Computational Linguistics.

Ng, H. T. (1997) *Exemplar-Based Word-Sense Disambiguation: Some Recent Improvements.* In Proceedings of the Second Conference on Empirical Method s in Natural Language Processing, pp 208-213. Somerset, N.J.: Association for Computational Linguistics.

Ng, H. T. and Zelle, J. M. (1997) *Corpus-Based Approaches to Semantic Interpretation in Natural Language Processing.* AI Magazine, Winter 1997, pp 45-64. Menlo Park, Calif.: American Association for Artificial Intelligence.

Wong, Ping-Wai and P. Fung (2002) *Nouns in WordNet and HowNet: an Analysis and Comparison of Semantic Relations.* To appear in the 1st International Conference on Global WordNet, 21st-25th, January 2002, Mysore, India.

Xue, Nianwen, Fei Xia, Shizhe Huang, Anthony Kroch (1999) *The Bracketing Guidelines for the Penn Chinese Treebank (Draff II)*, http://www.upenn.edu/ctb/.

Appendix A: Original Corpus Format

<doc>
<2906><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax><syntax class="Nh"><semantic class="{他},男,"><relation head="5" name="agent">他</relation></semantic></syntax><syntax class="Nd"><semantic class="{遞進},"><relation head="6" name="dummy">同時</relation></semantic></syntax><syntax class="D"><semantic class="{也},"><relation head="5" name="{also}">也</relation></semantic></syntax><syntax class="VE"><semantic class="表示,"><relation head="3" name="besides">指出</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2906>
<2907><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Na"><semantic class="人,家,眾,"><relation head="10" name="dummy">父母</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標點},"><relation head="2" name="dummy">、</relation></semantic></syntax><syntax class="Na"><semantic class="事務,教育,"><relation head="5" name="restrictive">教育</relation></semantic></syntax><syntax class="Na"><semantic class="事務,管理,"><relation head="6" name="restrictive">行政</relation></semantic></syntax><syntax class="Na"><semantic class="機構,統稱,"><relation head="3" name="coordinate">機關</relation></semantic></syntax><syntax class="Caa"><semantic class="{和},"><relation head="6" name="dummy">與</relation></semantic></syntax><syntax class="Na"><semantic class="實施,patient=律法,警,"><relation head="9" name=",agent">司法</relation></semantic></syntax><syntax class="Na"><semantic class="機構,統稱,"><relation head="7" name="coordinate">機關</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2907>
<2908><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="P"><semantic class="{scope},"><relation head="11" name="dummy">在</relation></semantic></syntax><syntax class="VE"><semantic class="阻止,"><relation head="6" name="restrictive">防制</relation></semantic></syntax><syntax class="Na"><semantic class="人,幼,眾,"><relation head="5|3" name="agent|patient">青少年</relation></semantic></syntax><syntax class="VA"><semantic class="做,content=罪,"><relation head="3" name="ResultEvent">犯罪</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation head="9" name="dummy">的</relation></semantic></syntax><syntax class="Nep"><semantic class="屬性值,類型,特,"><relation head="9" name="restrictive">這</relation></semantic></syntax><syntax class="Nf"><semantic class="名量,&萬物,"><relation head="7" name="quantity,restrictive">件</relation></semantic></syntax><syntax class="Na"><semantic class="事情,"><relation head="10" name="scope">事情</relation></semantic></syntax><syntax class="Ng"><semantic class="{scope},"><relation head="2" name="scope">上</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2908>
<2909><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="D"><semantic class="{語氣},"><relation head="4" name="{modality}">應</relation></semantic></syntax><syntax class="D"><semantic class="屬性值,舉止,共同,"><relation head="4" name="manner">分別</relation></semantic></syntax><syntax class=""><semantic class="承擔,"><relation head="14" name="dummy">負</relation></semantic></syntax><syntax class=""><semantic class="{發端},"><relation head="4" name="{Vprocess}">起</relation></semantic></syntax><syntax class="Neu"><semantic class="數量值,多少,基,"><relation head="7" name="coordinate">一</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標

點},"><relation head="8" name="dummy">、</relation></semantic></syntax><syntax class="Neu"><semantic class="數量值,多少,基,眾,"><relation head="9" name="coordinate">二</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標點},"><relation head="10" name="dummy">、</relation></semantic></syntax><syntax class="Neu"><semantic class="數量值,多少,基,眾,"><relation head="11" name="restrictive">三</relation></semantic></syntax><syntax class="Na"><semantic class="位置,"><relation head="12" name="restrictive">線</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation head="13" name="dummy">的</relation></semantic></syntax><syntax class="Na"><semantic class="責任,"><relation head="4" name="content">責任</relation></semantic></syntax><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax></2909>
<2912><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax><syntax class="P"><semantic class="{scope},"><relation head="5" name="dummy">在</relation></semantic></syntax><syntax class="Na"><semantic class="人,家,眾,"><relation head="4" name="restrictive">父母</relation></semantic></syntax><syntax class="Na"><semantic class="部件,%實體,側,"><relation head="2" name="scope">方面</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2912>
<2913><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Nh"><semantic class="{他},眾,"><relation head="6" name="agent">他們</relation></semantic></syntax><syntax class="P"><semantic class="{concerning},"><relation head="6" name="dummy">對於</relation></semantic></syntax><syntax class="Na"><semantic class="人,家,眾,幼,"><relation head="3" name="concerning">子女</relation></semantic></syntax><syntax class="D"><semantic class="{語氣},"><relation head="6" name="{modality}">應該</relation></semantic></syntax><syntax class="VE"><semantic class="實現,"><relation head="12" name="dummy">做到</relation></semantic></syntax><syntax class="D"><semantic class="{否},"><relation head="8" name="neg">不</relation></semantic></syntax><syntax class="VC"><semantic class="愛惜,"><relation head="6" name="patient">溺愛</relation></semantic></syntax><syntax class="Caa"><semantic class="{和},"><relation head="8" name="dummy">及</relation></semantic></syntax><syntax class="D"><semantic class="{否},"><relation head="11" name="neg">不</relation></semantic></syntax><syntax class="VJ"><semantic class="放縱,"><relation head="9" name="coordinate">放縱</relation></semantic></syntax><syntax class="SEMICOLONCATEGORY"><semantic class="{標點},"><relation head="0" name="主">；</relation></semantic></syntax></2913>
<2914><syntax class="SEMICOLONCATEGORY"><semantic class="{標點},"><relation head="0" name="主">；</relation></semantic></syntax><syntax class="P"><semantic class="{scope},"><relation head="5" name="dummy">在</relation></semantic></syntax><syntax class="Na"><semantic class="事務,教育,"><relation head="4" name="restrictive">教育</relation></semantic></syntax><syntax class="Na"><semantic class="部件,%實體,側,"><relation head="2" name="scope">方面</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2914>
<2915><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax feature="nom" class="VJ"><semantic class="屬性值,特性,有關,"><relation head="3" name="modifier">有關</relation></semantic></syntax><syntax class="Na"><semantic class="部件,%組織,統稱,"><relation head="4" name="experiencer">單位</relation></semantic></syntax><syntax class="VK"><semantic class="需求,"><relation head="8" name="dummy">需要</relation></semantic></syntax><syntax class="VC"><semantic class="優化,"><relation head="4" name="content">加強</relation></semantic></syntax><syntax class="Na"><semantic class="屬性,境況,&人,"><relation head="7" name="content">生活</relation></semantic></syntax><syntax feature="nom" class="VC"><semantic class="教,"><relation head="5" name="patient">輔導

</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2915>
<2916><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="VC"><semantic class="實施,"><relation head="5" name="dummy">實行</relation></semantic></syntax><syntax class="Na"><semantic class="屬性,境況,常態,&實體,"><relation head="4" name="content">常態</relation></semantic></syntax><syntax class=""><semantic class="分類,"><relation head="2" name="content">分班</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2916>
<2917><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Cbb"><semantic class="{和},"><relation head="6" name="dummy">並且</relation></semantic></syntax><syntax class="D"><semantic class="屬性值,程度,很,"><relation head="4" name="degree">多多</relation></semantic></syntax><syntax class=""><semantic class="注意,愛惜,"><relation head="2" name="coordinate">關愛</relation></semantic></syntax><syntax class="Na"><semantic class="人,*學,教育,"><relation head="4" name="target">學生</relation></semantic></syntax><syntax class="SEMICOLONCATEGORY"><semantic class="{標點},"><relation head="0" name="主">；</relation></semantic></syntax></2917>
<2918><syntax class="SEMICOLONCATEGORY"><semantic class="{標點},"><relation head="0" name="主">；</relation></semantic></syntax><syntax class="P"><semantic class="{scope},"><relation head="5" name="dummy">在</relation></semantic></syntax><syntax class="Na"><semantic class="屬性值,性能,*實施,#律法,警,"><relation head="4" name="restrictive">司法</relation></semantic></syntax><syntax class="Na"><semantic class="部件,%實體,側,"><relation head="2" name="scope">方面</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2918>
<2919><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax feature="nom" class="VJ"><semantic class="屬性值,特性,有關,"><relation head="3" name="modifier">有關</relation></semantic></syntax><syntax class="Na"><semantic class="機構,統稱,"><relation head="8" name="unresolved">機關</relation></semantic></syntax><syntax class="P"><semantic class="{scope},"><relation head="8" name="dummy">在</relation></semantic></syntax><syntax class="VC"><semantic class="改正,醫,"><relation head="7" name="scope">矯治</relation></semantic></syntax><syntax class="VA"><semantic class="做,content=罪,"><relation head="5" name="patient">犯罪</relation></semantic></syntax><syntax class="Ng"><semantic class="{scope},"><relation head="4" name="scope">上</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2919>
<2920><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="D"><semantic class="{但},"><relation head="12" name="dummy">則</relation></semantic></syntax><syntax class="D"><semantic class="{語氣},"><relation head="4" name="{modality}">應</relation></semantic></syntax><syntax class="VE"><semantic class="思考,"><relation head="2" name="coordinate">考慮</relation></semantic></syntax><syntax class="Na"><semantic class="人,幼,眾,"><relation head="6" name="agent">青少年</relation></semantic></syntax><syntax feature="nom" class="VA"><semantic class="做,content=罪,"><relation head="7" name="relation">犯罪</relation></semantic></syntax><syntax class="Abs"><semantic class="Abs"><relation head="4" name="content">{{</relation></semantic></syntax><syntax class="Na"><semantic class="屬性,舉止,&人,"><relation head="7" name="content">行爲</relation></semantic></syntax><syntax class="Caa"><semantic class="{和},"><relation head="8" name="dummy">及</relation></semantic></syntax><syntax class="Na"><semantic class="精神,"><relation head="9" name="coordinate">心理</relation></semantic></syntax><syntax class="Abs"><semantic class="Abs"><relation head="7"

name="cooccur">}}</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2920>
<2921><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class="P"><semantic
class="{source},"><relation head="4" name="dummy">從</relation></semantic></syntax><syntax
class="Na"><semantic class="屬性值,主次,主,"><relation head="2" name="source">根本
</relation></semantic></syntax><syntax class="VC"><semantic class="改正,"><relation head="5"
name="dummy">矯正</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2921>
<2922><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class=""><semantic
class="{purpose},"><relation head="6" name="dummy">以</relation></semantic></syntax><syntax
class=""><semantic class="阻止,"><relation head="2" name="purpose">防
</relation></semantic></syntax><syntax class="D"><semantic class="屬性值,頻率,再,"><relation
head="5" name="frequency">再</relation></semantic></syntax><syntax class="VC"><semantic
class="違背,罪,"><relation head="3" name="ResultEvent">犯
</relation></semantic></syntax><syntax class="PERIODCATEGORY"><semantic class="{標
點},"><relation head="0" name="主">。</relation></semantic></syntax></2922>
<2923><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0"
name="主">。</relation></semantic></syntax><syntax class="P"><semantic
class="{source},"><relation head="8" name="dummy">從</relation></semantic></syntax><syntax
class="Ncd"><semantic class="屬性值,類型,特,"><relation head="4" name="restrictive">以上
</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation
head="7" name="dummy">的</relation></semantic></syntax><syntax class="VE"><semantic
class="調查,"><relation head="6" name=",succeeding">調查
</relation></semantic></syntax><syntax class="Na"><semantic class="分析,"><relation head="7"
name="content">分析</relation></semantic></syntax><syntax class="Na"><semantic class="語
文,"><relation head="2" name="source">報告</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2923>
<2924><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class="D"><semantic
class="{否},"><relation head="3" name="neg">不</relation></semantic></syntax><syntax
class="VH"><semantic class="屬性值,難易,難,莠,"><relation head="4" name="manner">難
</relation></semantic></syntax><syntax class="VE"><semantic class="感知,"><relation head="5"
name="dummy">看出</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2924>
<2925><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class="P"><semantic
class="{scope},"><relation head="10" name="dummy">在</relation></semantic></syntax><syntax
class="Neqa"><semantic class="屬性值,幅度,全,"><relation head="8" name="range">整
</relation></semantic></syntax><syntax class="Nf"><semantic class="屬性值,類型,特,"><relation
head="3" name="quantity">個</relation></semantic></syntax><syntax class="Na"><semantic
class="人,幼,眾,"><relation head="6" name="agent">青少年
</relation></semantic></syntax><syntax class="VA"><semantic class="做,content=罪,"><relation
head="7" name="restrictive">犯罪</relation></semantic></syntax><syntax class="DE"><semantic
class="{構助},"><relation head="8" name="dummy">的</relation></semantic></syntax><syntax
class="Na"><semantic class="現象,不幸,莠,"><relation head="9" name="scope">問題
</relation></semantic></syntax><syntax class="Ng"><semantic class="{scope},"><relation
head="2" name="scope">上</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，

</relation></semantic></syntax></2925>
<2926><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Na"><semantic class="團體,家,"><relation head="4" name="relevant">家庭</relation></semantic></syntax><syntax class="D"><semantic class="{信度},"><relation head="4" name="{modality}">的確</relation></semantic></syntax><syntax class="SHI"><semantic class="是,"><relation head="9" name="dummy">是</relation></semantic></syntax><syntax class="Dfa"><semantic class="屬性值,程度,最,"><relation head="6" name="degree">最</relation></semantic></syntax><syntax class="VH"><semantic class="屬性值,主次,主,"><relation head="7" name="restrictive">重要</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation head="8" name="dummy">的</relation></semantic></syntax><syntax class="Na"><semantic class="部件,%實體,"><relation head="4" name="descriptive">因素</relation></semantic></syntax><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax></2926>
<2927><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax><syntax class="VK"><semantic class="懊悔,"><relation head="3" name="restrictive">遺憾</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation head="4" name="dummy">的</relation></semantic></syntax><syntax class="SHI"><semantic class="是,"><relation head="5" name="dummy">是</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2927>
<2928><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Nes"><semantic class="屬性值,類型,特,"><relation head="4" name="restrictive">該</relation></semantic></syntax><syntax class="Na"><semantic class="新聞,"><relation head="4" name="restrictive">新聞</relation></semantic></syntax><syntax class="Na"><semantic class="新聞,"><relation head="5" name="restrictive">報導</relation></semantic></syntax><syntax class="Ng"><semantic class="位置,間,"><relation head="9" name="dummy">中</relation></semantic></syntax><syntax class="D"><semantic class="{強調},"><relation head="7" name="{modality}">並</relation></semantic></syntax><syntax class=""><semantic class="{否},昔,"><relation head="8" name="neg">未</relation></semantic></syntax><syntax class="VH"><semantic class="屬性值,內容,詳,良,"><relation head="9" name="manner">詳細</relation></semantic></syntax><syntax class="VE"><semantic class="說明,"><relation head="26" name="dummy">說明</relation></semantic></syntax><syntax class="Nd"><semantic class="時間,今,日,"><relation head="11" name="time">今日</relation></semantic></syntax><syntax class="Na"><semantic class="團體,家,"><relation head="12" name="scope">家庭</relation></semantic></syntax><syntax class="Na"><semantic class="屬性,性能,&實體,"><relation head="13" name="experiencer">功能</relation></semantic></syntax><syntax class="VHC"><semantic class="衰敗,"><relation head="14" name="coordinate">退化</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標點},"><relation head="17" name="dummy">、</relation></semantic></syntax><syntax class=""><semantic class="人,家,眾,"><relation head="16" name="host">父母</relation></semantic></syntax><syntax class="Na"><semantic class="屬性,勢力,&人,&組織,&信息,"><relation head="17" name="experiencer">權威</relation></semantic></syntax><syntax class=""><semantic class="衰敗,"><relation head="18" name="coordinate">式微</relation></semantic></syntax><syntax class="Caa"><semantic class="{和},"><relation head="19" name="dummy">及</relation></semantic></syntax><syntax class="D"><semantic class="無能,"><relation head="23" name="restrictive">不能</relation></semantic></syntax><syntax class=""><semantic class="承擔,"><relation head="19" name="content">負</relation></semantic></syntax><syntax class=""><semantic class="{發端},"><relation head="20" name="{Vprocess}">起</relation></semantic></syntax><syntax class="VC"><semantic class="教,"><relation head="20" name="content">教化</relation></semantic></syntax><syntax class="DE"><semantic class="{構助},"><relation head="24" name="dummy">的</relation></semantic></syntax><syntax

class="Na"><semantic class="原因,"><relation head="25" name="restrictive">原因
</relation></semantic></syntax><syntax class="VH"><semantic class="位置,疑問,"><relation
head="9" name="content">何在</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2928>
<2930><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class="Cbb"><semantic
class="{但},"><relation head="5" name="dummy">但</relation></semantic></syntax><syntax
class="D"><semantic class="{語氣},"><relation head="4" name="{modality}">可以
</relation></semantic></syntax><syntax class="VK"><semantic class="屬性值,能力,能,猜
測,"><relation head="2" name="coordinate">想像</relation></semantic></syntax><syntax
class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，
</relation></semantic></syntax></2930>
<2931><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0"
name="主">，</relation></semantic></syntax><syntax class="VK"><semantic class="屬性值,程度,
欠,"><relation head="3" name="degree">不外乎</relation></semantic></syntax><syntax
class="SHI"><semantic class="是,"><relation head="30" name="dummy">是
</relation></semantic></syntax><syntax class="VJ"><semantic class="緣于,"><relation head="6"
name="coordinate">來自</relation></semantic></syntax><syntax class="VH"><semantic class="屬
性,貧富,窮,&人,&組織,"><relation head="4" name="cause">貧窮
</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標
點},"><relation head="10" name="dummy">、</relation></semantic></syntax><syntax
class="Na"><semantic class="人,家,眾,"><relation head="8" name="restrictive">父母
</relation></semantic></syntax><syntax class="Na"><semantic class="人,莠,愚,"><relation
head="9" name="coordinate">文盲</relation></semantic></syntax><syntax class="Caa"><semantic
class="{或},"><relation head="10" name="dummy">或</relation></semantic></syntax><syntax
class="VJ"><semantic class="經受,"><relation head="14" name="coordinate">受
</relation></semantic></syntax><syntax class=""><semantic class="屬性值,程度,極,"><relation
head="12" name="degree">極</relation></semantic></syntax><syntax class=""><semantic class="
數量值,多少,少,"><relation head="13" name="quantity">少
</relation></semantic></syntax><syntax class="Na"><semantic class="教,"><relation head="10"
name="沒有輸入!">教育</relation></semantic></syntax><syntax
class="CAESURACATEGORY"><semantic class="{標點},"><relation head="15"
name="coordinate">、</relation></semantic></syntax><syntax class="Caa"><semantic
class="{或},"><relation head="20" name="dummy">或者</relation></semantic></syntax><syntax
class="D"><semantic class="屬性值,舉止,勻,良,"><relation head="18" name="range">均
</relation></semantic></syntax><syntax class="P"><semantic class="{進展},"><relation head="18"
name="{Vprocess}">在</relation></semantic></syntax><syntax class="VA"><semantic class="從
事,content=事務,#責任,"><relation head="19" name="coordinate">上班
</relation></semantic></syntax><syntax class="Caa"><semantic class="{或},"><relation head="20"
name="dummy">或</relation></semantic></syntax><syntax class="VA"><semantic class="從
事,content=事務,商,"><relation head="21" name="coordinate">做生意
</relation></semantic></syntax><syntax class="CAESURACATEGORY"><semantic class="{標
點},"><relation head="22" name="coordinate">、</relation></semantic></syntax><syntax
class="Caa"><semantic class="{或},"><relation head="27" name="dummy">或者
</relation></semantic></syntax><syntax class="D"><semantic class="{時態},昔,"><relation
head="24" name="{tense}">已經</relation></semantic></syntax><syntax class="VA"><semantic
class="分離,"><relation head="25" name="coordinate">離婚
</relation></semantic></syntax><syntax class="Caa"><semantic class="{或},"><relation head="27"
name="dummy">或</relation></semantic></syntax><syntax class="D"><semantic class="屬性值,
頻率,經常,"><relation head="27" name="frequency">經常</relation></semantic></syntax><syntax
class="VA"><semantic class="爭吵,"><relation head="29" name="restrictive">爭吵
</relation></semantic></syntax><syntax class="Cab"><semantic class="{和},"><relation head="0"
name="主">等等</relation></semantic></syntax><syntax class="Na"><semantic class="原

因,"><relation head="3" name="descriptive">原因</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2931>
<2935><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="Cbb"><semantic class="{但},"><relation head="7" name="dummy">而</relation></semantic></syntax><syntax class="Na"><semantic class="屬性值,主次,主,"><relation head="4" name="manner">根本</relation></semantic></syntax><syntax class="VC"><semantic class="處理,"><relation head="5" name="restrictive">解決</relation></semantic></syntax><syntax class="Ng"><semantic class="{構助},"><relation head="6" name="dummy">之</relation></semantic></syntax><syntax class="Na"><semantic class="方法,"><relation head="2" name="coordinate">道</relation></semantic></syntax><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax></2935>
<2936><syntax class="COMMACATEGORY"><semantic class="{標點},"><relation head="0" name="主">，</relation></semantic></syntax><syntax class="VG"><semantic class="{語氣},"><relation head="3" name="{modality}">似</relation></semantic></syntax><syntax class="D"><semantic class="{語氣},"><relation head="4" name="{modality}">應</relation></semantic></syntax><syntax class="P"><semantic class="{source},"><relation head="7" name="dummy">從</relation></semantic></syntax><syntax class="A"><semantic class="屬性值,等級,初,"><relation head="6" name="modifier">基本</relation></semantic></syntax><syntax class="Na"><semantic class="事務,教育,"><relation head="4" name="source">教育</relation></semantic></syntax><syntax class="VL"><semantic class="開始,"><relation head="8" name="dummy">開始</relation></semantic></syntax><syntax class="PERIODCATEGORY"><semantic class="{標點},"><relation head="0" name="主">。</relation></semantic></syntax></2936>
</doc>

Appendix B: Passage Format of the SenPrune System

他 屬性值,歸屬,另,|{他},男,
同時 {遞進},|時間,|時間,似,|屬性值,時間,似,
也 {也},
指出 表示,
， {標點},
父母 人,家,眾,
、 {標點},
教育 事務,教育,|教,
行政 事務,管理,
機關 規劃,|規劃,狡,*欺騙,莠,|部件,%器具,|道理,|機構,統稱,|屬性值,類型,
與 {partner},|{target},|{和},
司法 實施,patient=律法,警,|屬性值,性能,*實施,#律法,警,|屬性值,歸屬,#警,
機關 規劃,|規劃,狡,*欺騙,莠,|部件,%器具,|道理,|機構,統稱,|屬性值,類型,
， {標點},
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
防制 阻止,
青少年 人,幼,眾,
犯罪 做,content=罪,
的 {構助},|位置,$射擊,
這 時間,|屬性值,類型,特,
件 名量,&萬物,|信件,|部件,%人工物,#製造,工,
事情 事情,
上 {range},|{scope},|{動趨,上},|{發端},|{達成},|上去,|交,|安裝,|位置,|供,|到達,LocationFin=設施,purpose=比賽,體育,|前往,|前進,|做,|勒緊,|塗抹,|經受,content=傳播,|攀登,|屬性值,次序,前,|屬性值,時間,近,|屬性值,時間,近,昔,|屬性值,等級,高等,良,
， {標點},
應 {語氣},|同意,|答,|對待,|遵循,
分別 數量值,多少,單,|屬性值,舉止,共同,|分辨,|離別,
負 欠,|有,|依靠,|承擔,|揹起,|違背,|遭受,|輸掉,|輸掉,軍,|屬性值,特性,
起 {發端},|名量,&事物,|上升,|出現,|名量,&事件,|收,|拔出,|建立,|起身,|做,|發生,|開始,|編輯,
一 數量值,多少,基,|數量值,多少,單,|屬性值,幅度,全,|屬性值,頻率,|數量值,次序,第,
、 {標點},
二 數量值,多少,基,眾,|數量值,次序,第,
、 {標點},
三 數量值,多少,基,眾,|數量值,次序,第,
線 用具,線,*拴連,|位置,|物形,線,|規矩,政,|設施,路,|部件,%現象,|標誌,線,
的 {構助},|位置,$射擊,
責任 責任,
。 {標點},
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
父母 人,家,眾,
方面 部件,%實體,側,
， {標點},
他們 {他},眾,
對於 {concerning},|{concerning},|{concerning},
子女 人,家,眾,幼,
應該 {語氣},
做到 實現,
不 {否},

溺愛 愛惜,
及 {和},|到達,|{達成},
不 {否},
放縱 放縱,|屬性值,舉止,莽,莠,
； {標點},
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
教育 事務,教育,|教,
方面 部件,%實體,側,
， {標點},
有關 有關,|屬性值,特性,有關,
單位 部件,%組織,統稱,|屬性,多少,&實體,
需要 需求,
加強 優化,
生活 行動,|活著,|屬性,境況,&人,
輔導 教,
， {標點},
實行 實施,
常態 屬性,境況,常態,&實體,
分班 分類,
， {標點},
並且 {和},
多多 屬性值,程度,很,
關愛 注意,愛惜,
學生 人,*學,|人,*學,教育,
； {標點},
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
司法 實施,patient=律法,警,|屬性值,性能,*實施,#律法,警,|屬性值,歸屬,#警,
方面 部件,%實體,側,
， {標點},
有關 有關,|屬性值,特性,有關,
機關 規劃,|規劃,狡,*欺騙,莠,|部件,%器具,|道理,|機構,統稱,|屬性值,類型,
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
矯治 改正,醫,
犯罪 做,content=罪,
上 {range},|{scope},|{動趨,上},|{發端},|{達成},|上去,|交,|安裝,|位置,|供,|到達,LocationFin=設施,purpose=比賽,體育,|前往,|前進,|做,|勒緊,|塗抹,|經受,content=傳播,|攀登,|屬性值,次序,前,|屬性值,時間,近,|屬性值,時間,近,昔,|屬性值,等級,高等,良,
， {標點},
則 {強調},|{遞進},|{轉折},|{條件},|{但},|名量,&語文,|律法,|屬性,標準,&實體,
應 {語氣},|同意,|答,|對待,|遵循,
考慮 思考,|思想,
青少年 人,幼,眾,
犯罪 做,content=罪,
{{ XX
行爲 事情,行動,|屬性,舉止,&人,
及 {和},|到達,|{達成},
心理 精神,
}} XX
， {標點},
從 {LocationIni},|{TimeIni},|{source},|{StateIni},|納入,|從事,|實施,|遵循,|屬性值,主次,次,
根本 屬性值,幅度,全,|屬性值,主次,主,
矯正 改正,
， {標點},
以 {target},|{AccordingTo},|{cause},|{means},|{purpose},|地方,國家,專,(以色列),

防　抗住,|防守,|阻止,|設施,*保護,空間,
再　屬性值,頻率,再,
犯　攻打,|做,|違背,|違背,罪,|罹患,醫,|譴責,
。　{標點},
從　{LocationIni},|{TimeIni},|{source},|{StateIni},|納入,|從事,|實施,|遵循,|屬性值,主次,次,
以上 數量值,多少,多,|屬性值,類型,特,
的　{構助},|位置,$射擊,
調查 調查,
分析 分析,|屬性值,性能,
報告 告訴,|語文,
，　{標點},
不　{否},
難　爲難,|現象,莠,#不幸,|屬性值,可能,難,莠,|屬性值,好壞,壞,莠,|屬性值,難易,難,莠,|譴責,
看出 感知,
，　{標點},
在　{condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
整　改正,|修理,|做,|處罰,|損害,|整理,|屬性值,內容,齊,良,|屬性值,幅度,全,
個　名量,&實體,|屬性值,類型,特,
青少年　人,幼,眾,
犯罪 做,content=罪,
的　{構助},|位置,$射擊,
問題 問題,|現象,不幸,莠,
上　{range},|{scope},|{動趨,上},|{發端},|{達成},|上去,|交,|安裝,|位置,|供,|到達,LocationFin=設施,purpose=比賽,體育,|前往,|前進,|做,|勒緊,|塗抹,|經受,content=傳播,|攀登,|屬性值,次序,前,|屬性值,時間,近,|屬性值,時間,近,昔,|屬性值,等級,高等,良,
，　{標點},
家庭 團體,家,
的確 {信度},
是　是,|屬性值,正誤,正確,|屬性值,正誤,正確,良,|屬性值,類型,特,
最　屬性值,程度,最,|屬性值,距離,最,
重要 屬性值,主次,主,
的　{構助},|位置,$射擊,
因素 部件,%實體,
。　{標點},
遺憾 懊悔,
的　{構助},|位置,$射擊,
是　是,|屬性值,正誤,正確,|屬性值,正誤,正確,良,|屬性值,類型,特,
，　{標點},
該　欠,|值得,|適合,|屬性值,類型,特,
新聞 新聞,
報導 告訴,|新聞,
中　{進展},|{延續},|{scope},|時間,間,|地方,國家,專,(中國),|位置,間,|時間,今,|實現,|遭受,|屬性值,效用,優,良,|屬性值,等級,可,
並　{強調},|{和},|屬性值,時間,似,
未　{否},昔,
詳細 屬性值,內容,詳,良,
說明 說明,
今日 時間,今,日,
家庭 團體,家,
功能 屬性,性能,&實體,
退化 衰敗,|屬性值,風氣,壞,莠,|屬性值,標準,壞,
、　{標點},
父母 人,家,眾,

權威 人,#勢力,能,良,|屬性,勢力,&人,&組織,&信息,
式微 衰敗,
及 {和},|到達,|{達成},
不能 禁止,|屬性值,標準,^可,|無能,
負 欠,|有,|依靠,|承擔,|揹起,|違背,|遭受,|輸掉,|輸掉,軍,|屬性值,特性,
起 {發端},|名量,&事物,|上升,|出現,|名量,&事件,|收,|拔出,|建立,|起身,|做,|發生,|開始,|編輯,
教化 教,
的 {構助},|位置,$射擊,
原因 原因,
何在 位置,疑問,
， {標點},
但 {但},
可以 {語氣},|屬性值,標準,可,良,
想像 屬性值,能力,能,猜測,
， {標點},
不外乎 數量值,多少,少,|屬性值,程度,欠,
是 是,|屬性值,正誤,正確,|屬性值,正誤,正確,良,|屬性值,類型,特,
來自 緣于,
貧窮 屬性,貧富,窮,&人,&組織,|屬性值,貧富,窮,莠,
、 {標點},
父母 人,家,眾,
文盲 人,莠,愚,
或 {或},
受 收受,|忍耐,|經受,|遭受,
極 部件,%大地,頭,尾,|部件,%電,頭,尾,|屬性,程度,極,&實體,|屬性值,程度,極,
少 失去,|缺少,|停做,|數量值,多少,少,|屬性值,年齡,幼,
教育 事務,教育,|教,
、 {標點},
或者 {或},|{信度},
均 屬性值,幅度,全,|屬性值,舉止,勻,良,
在 {condition},|{scope},|{進展},|{location},|{time},|存在,|依靠,|活著,|處于,|緣于,
上班 從事,content=事務,#責任,
或 {或},
做生意 從事,content=事務,商,
、 {標點},
或者 {或},|{信度},
已經 {時態},昔,
離婚 分離,
或 {或},
經常 屬性值,頻率,經常,
爭吵 爭吵,
等等 {和},|等待,
原因 原因,
， {標點},
而 {轉折},|{LocationFin},|{StateFin},|{TimeFin},|{但},|{事件結局},|{和},
根本 屬性值,幅度,全,|屬性值,主次,主,
解決 消滅,|處理,
之 {指代},|{它},|{構助},
道 方法,|名量,&信息,|名量,&無生物,|規矩,|設施,路,|團體,莠,宗教,|認爲,|說,
， {標點},
似 {語氣},|屬性值,舉止,隱,|相像,
應 {語氣},|同意,|答,|對待,|遵循,
從 {LocationIni},|{TimeIni},|{source},|{StateIni},|納入,|從事,|實施,|遵循,|屬性值,主次,次,

基本 部件,%實體,根,|屬性值,主次,主,|屬性值,必要,要,|屬性值,等級,初,
教育 事務,教育,|教,
開始 開始,|過程,早,
。 {標點},