Automatic Labeling of Hypernymy-Troponymy Relation for Chinese Verbs

中文動詞上下位關係自動標記法

國立台灣師範大學英語學系碩士論文

Master Thesis

Department of English, National Taiwan Normal University

指導教授: 謝舒凱博士

Advisor: Dr. Shu-Kai Hsieh

研究生: 羅巧珊

Student: Chiao-Shan Lo

中華民國九十八年七月

July, 2009

Abstract

近年來，詞彙網路(Wordnet)已成爲計算語言學相關領域中最爲普遍利用的資源之一，對於資訊檢索(Information Retrieval)或是自然語言處理(Natural Language Processing)的發展有相當大的幫助。詞彙網路是由同義詞集(Synset)以及詞彙語意關係(Lexical Semantic Relation)所建構而成，例如以英語爲主的普林斯頓詞網(Princeton WordNet)、以及結合多個歐洲語言的歐語詞網(EuroWordNet)等，建構皆已相當完善。然而，一個詞網的建構並非一時一人之力所能完成，其所需要的人力以及耗費的時間相當可觀。因此，如何有效率並有系統的建構一個詞網是近年來研究致力的目標。而詞彙間的語意關係是構成一個詞網的主要元素，因此，如何自動化的抽取詞彙語義關係是建構詞網的重要步驟之一。中研院語言所已建立一個以中頻詞爲主的中文詞彙網路(Chinese WordNet, CWN)，旨在提供完整的中文辭彙之詞義區分。然而，在目前中文詞彙網路系統中，同義詞集間相互的語意關係乃是採用人爲判定標記，且這些標記之數量尚未達成可行應用之一定規模。因此，本研究提出一套半自動化的方法來自動標記詞彙間的語意關係，本篇論文針對動詞之間的上下位詞彙語意關係(Hypernymy-troponymy relation)，提出一種自動標記的方法，並抽取具有中文上下位關係之中文動詞組對。

本篇論文提出兩種並行之方法，第一，藉由句法上特定的句型(lexical syntactic pattern)，自動抽取出中文詞彙網路中具有上下位關係之動詞組。第二，我們利用bootstrapping的方法，透過中研院建構的中英雙語詞網(Sinica Bow)大量將普林斯頓英語詞網中的語意關係對映至中文。實驗結果顯示，此系統能快速並大量地自動抽取出具有上下位語意關係之中文動詞組，本論文盼能將此方法應用於正在發展中的中文詞網自動語意關係標記，以及知識本體之自動建構，進而能有效率的建構完善的中文詞彙知識資源。

關鍵詞：語義關係自動標記、動詞詞彙語意、動詞上下位關係、中文詞網

Abstract

WordNet-like databases have become crucial sources for lexical semantic studies and computational linguistic applications such as Information Retrieval (IR) and Natural Language Processing (NLP). The fundamental elements of WordNet are synsets (the synonymous grouping of words) and semantic relations among synsets. However, creating such a lexical network is a time-consuming and labor-intensive project. In particular, for those languages with few resources such as Chinese, is even difficult. Chinese WordNet (CWN), which composed of middle frequency words, has been launched by Academia Sinica based on the similar paradigm as Princeton WordNet. The synset that each word sense locates in CWN is manually labeled. However, the lexical semantic relations among synsets in CWN are only partially constructed and lack of systematic labeling. Therefore, in this thesis, two independent approaches were proposed to automatically harvesting lexical semantic relations, especially focused on the hypernymy-troponymy relation of verbs.

This thesis describes two approaches for discovering hypernymy-troponymy relation among verbs. Syntactic pattern-based approach is used for that sentence structures can always denote relations and reveal information among lexical entries. Bootstrapping approach, on the other hand, aims at exploiting an already existing database and combining them within a common, standard framework. From a large scale of input data, our proposed approaches can greatly and rapidly extract verb pairs that are in hypernymy-troponymy relation in Chinese, aiding the construction of lexical database in a more effective way. In addition, it is hoped that these approaches will shed light on the task of automatic acquisition of other Chinese lexical semantic relations and ontology learning as well.

Key word: automatic extraction, lexical semantic relation, troponymy, Chinese WordNet

ACKNOWLEDGEMENTS

終於到了寫謝詞的這一刻，從開始撰寫論文沒多久我就一邊構思著謝詞的內容，因爲一路走來，實在有太多人要感謝了。寫論文是一段漫長又煎熬的過程，這段過程中常常遇到各種挑戰與瓶頸，總是讓我灰心不已。不過很幸運地，這一路上總是有許多人伸出援手，不論是學術上的見解，或是精神上的支持，都給予我莫大的幫助。在此，我要向這些人深深表達我由衷的感謝。

首先，我要感謝我的指導教授，謝舒凱老師。碩二時，我擔任老師的助理，並且修了老師在研究所開的每一堂課，對計算語言學的認識可說是受到老師的啓蒙，讓我接觸到了語言學另一個全新的領域。在寫論文的過程中，老師總是給我非常自由的空間去發揮，並且對我提出的疑問跟想法都給予解答與支持，而謝老師沉穩的個性也是最能安定人心的力量，每當我因爲遇到瓶頸而焦躁不安時，老師總是有辦法不疾不徐地協助我解決困難。

另外，我也要感謝我的兩位口試委員：台大外文系的高照明老師，以及政大英語系的鍾曉芳老師。高老師在我兩次口試時，總是提出許多精闢的見解，不論是在語言學方面或是計算程式方面，都給了我很多很實用的建議，口試結束後高老師更是熱心地提供我需要的資源並回答我的疑問。而鍾老師也是在百忙之中抽空前來擔任口試委員，儘管如此，鍾老師還是在我的論文裡，密密麻麻地寫下她的意見並點出論文的缺點。感謝兩位老師的幫忙，這篇論文才能完成。

我也要感謝師大每位優秀的老師及同學，謝謝每位曾經教過我的老師，在師大的每一門課都是既紮實又豐富，一點一滴累積我對語言學的知識以及撰寫論文的能力。還有班上優秀的同學們，Nancy, Caroline, David, Fu-Pin, Clara, and Jessica 等等，雖然到了後來大家因爲工作或論文，各自努力很少見面，但是我們仍然會在空閒時交換心得，給彼此鼓勵。

除此之外，我要由衷的感謝中研院語言所的程式設計師，李龍豪先生。如果沒有你熱心的

幫忙，這篇論文不可能完成。還要感謝中研院的怡頻、兪庭以及淳涵，感謝你們在百忙之中放下手邊的工作來協助我分析那上千筆的語料。我也要感謝我的戰友們：儷蓉、書瑋、徐瑜。寫論文有多痛苦，眞的要寫過才知道，好險那些難熬的日子裡有你們的陪伴，互吐苦水，相互勉勵。

最後，也是最重要的，我要感謝我的家人，感謝爸媽總是無條件的在背後支持我，不論是在精神上或是物質上都給我莫大的幫助，還有妹妹不時捎來的關心與問候，都讓我倍感窩心。謝謝你們一路陪我走過來，支持我所做的每一個決定，謹以這本論文獻給你們---- 我最愛的家人。

Contents

1 Introduction **1**

1.1 Background 1

1.2 Motivation 3

1.3 Organization of the Thesis 4

2 Related Works **5**

2.1 WordNet-like Resources 5

2.1.1 Princeton WordNet [31] 6

2.1.2 EuroWordNet [45] 7

2.1.3 Sinica Bow [23] 8

2.1.4 Chinese WordNet [1] 10

2.1.5 HowNet [14] 12

2.2 Semantic Relations of Verbs 13

2.2.1 Semantic Relations of Verbs in WordNet 13

2.2.2 Semantic Relations of Verbs in EuroWordNet 16

2.2.3 Other Relations of Verbs 20

2.3 Troponymy 22

2.3.1 Definition of Troponymy 24

2.3.2 Distinguishing Manner 26
2.4 Automatic Discovery of Lexical Semantic Relation 28
2.4.1 Lexico Syntactic Pattern–Based Approach 29
2.4.2 Clustering-Based Approach 32
2.4.3 Bootstrapping Approach 33
2.5 Summary 35

3 Methodology **37**
3.1 Syntactic Pattern-Based Approach 37
3.1.1 Database: Chinese WordNet 37
3.1.2 Data Pre-processing 39
3.1.3 Syntactic Patterns in Chinese 41
3.1.4 Procedure 42
3.2 Bootstrapping Approach 44
3.2.1 Data Source 46
3.2.2 Procedure 48
3.3 Evaluation and Scoring 49
3.3.1 Evaluation 50
3.3.2 Scoring 54
3.4 Summary 55

4 Results and Error Analyses **56**
4.1 Results from Syntactic Pattern- based Approach 56
4.1.1 Error Analyses 58
4.1.2 Interim Summary 68

4.2 Results from Bootstrapping Approach 69
4.2.1 Error Analyses 70
4.3 Discussion 81
4.3.1 Comparison of Two Approaches 81
4.3.2 Comparison of the Results 83
4.3.3 Comparison of the Error Types 86
4.3.4 General Discussion 89
4.4 Summary 91

5 Conclusion **92**

5.1 Summary of the Thesis 92
5.2 Contribution 94
5.3 Limitations of the Present Study and Suggestions for Future Work 95

Appendix:

A Programming Code **104**

B Results from Syntactic Pattern-based Approach **107**

C Results from Bootstrapping Approach **110**

List of Tables

2.1 A finer-grained semantic relation among verbs. [9] 21

2.2 Semantic relations of verbs in Wordnet, EuroWordNet and VerbOcean 23

2.3 Three different types of Troponymy 28

4.1 General results of syntactic pattern-based approach 57

4.2 Error types and percentage 59

4.3 Overall results from bootstrapping approach 70

4.4 Non hypernymy-troponymy verb pairs (Total number of returned verb pairs= 11289) 71

4.5 General comparison of syntactic pattern-based and bootstrapping approach . . 82

4.6 Comparison of error types from results in two approaches 86

4.7 General comparison of the two approaches 89

List of Figures

2.1 The first two senses returned by CWN of the verb 走 'zao3, walk' 11

2.2 Four kinds of entailments among English verbs [31] 15

2.3 Translation-mediated LSR Prediction (The complete model) 33

2.4 Translation-mediated LSR Prediction (when translation equivalents are synonymous) 34

3.1 Bootstrapping model 45

3.2 Overall procedure of bootstrapping approach 50

Chapter 1

Introduction

1.1 Background

In recent years, there has been an increasing focus on the construction of lexical knowledge resources in the field of Natural Language Processing (NLP), such as Thesaurus, WordNets [31], EuroWordNet [45], FrameNet [6], HowNet [13], etc. Among these resources, Princeton WordNet[1], an electronic English lexical database, was started as an implementation of a psycholinguistic model of the mental lexicon. In WordNet, English nouns, verbs, adjectives, and adverbs are organized into synonym sets, called synsets. Synsets in WordNet are connected with each other by various kinds of paradigmatic lexical semantic relations, such as Meronymy and Holonymy (between parts and wholes), Hypernymy and Hyponymy (between specific and more general synsets), etc. These relations act as pointers between synsets. Due to the semantic relation-based property, WordNet has been widely used to solve a variety of problems in the field of NLP and has sparked off most interest both in theoretical and applicational sides, such as Information Retrieval (IR), lexical acquisition, automatic extraction, Word Sense Disambiguation (WSD), and so on. WordNet's growing popularity has prompted the modeling and

[1] http://wordnet.princeton.edu

construction of wordnets in other languages and various domains as well. EuroWordNet [45], which aims to build a multilingual database for several European languages, is a successful example. To date, in the field of NLP applications, WordNet and EuroWordNet serve as very crucial sources and have become a standard norm in evaluating semantic relations. WordNet covers a large scale of sense-based English lexicons (206941 word-sense pairs [2]). The extensive coverage of WordNet took immense labors and time. Further, semantic relations are unlimited, it takes years and intensive labors to steadily develop the scope and content. Consequently, there has been significant recent interest in finding methods to build a WordNet-like database in other languages with less efforts and time [5] [7] [9] [20] [21] [24] [28] [30] [32] [38] [39].

Lexical semantic relations among synsets are the foundations of a semantic network, but manually constructing all the relations is time-consuming and error-prone. Therefore, one of the most important steps toward efficiently constructing a WordNet-like database is to automatically extract lexical semantic relations. It is clearly necessary to develop an automated approach to harvest semantic relations; hence motivates researchers working on automated methods paralleled with manual verification, in order to ease the work. To date, most research on semantic relation harvesting have focused on *is-a* (c.f. [21] [32]) and *part-of* (c.f. [7] [20]) relations of nouns. Verb relations, on the contrary, have been studied less often especially the *is-a* relation among verbs– the hypernymy-troponymy relation. Several approaches have been suggested for automatically extracting semantic relations and they mainly fall onto three categories: lexico syntactic pattern-based approach, clustering-based approach, and bootstrapping approach, all of which will be introduced in this thesis.

[2]The statistics is updated to WordNet3.0

1.2 Motivation

As mentioned, creating a lexical semantic knowledge resource like WordNet is a time-consuming and labor-intensive task. Languages other than English and some European languages are facing with the lack of long-term linguistic supports, let alone those languages without balanced corpus available. In Chinese, constructing a wordnet is comparatively difficult owing to the fuzzy definition and classification among words, morphemes, and characters. The Chinese WordNet (hereafter CWN)[3], created by Academia Sinica, aims to provide complete sense inventory for each word based on the theory of lexical semantics and ontology. The synsets of each word in CWN are manually labeled but the semantic relations among synsets are only left partially constructed. In English or other European languages, several approaches were proposed to find semantic relations automatically such as lexico syntactic pattern-based approach [21], clustering approach [27] and bootstrapping approach [39]. However, in Chinese, there is few research studying on this issue. Huang et al. [24] has proposed a bootstrapping method mapping English WordNet to Chinese WordNet but the experimental data is small (about 200 lemmas) and several translational idiosyncrasies were ignored. Hence, bootstrapping method alone is not accredited enough. What is more, to the best of our knowledge, there is not yet a study focusing on the extraction of hypernymy-troponymy relation in any languages. The above mentioned issues reveal the need and the lack of a semantic relation-based wordnet in Chinese. This motivates this thesis to propose a lexico syntactic pattern-based approach assisted with bootstrapping to automatically label the semantic relations. Although the target in this thesis focuses only on hypernymy-troponymy relation of verbs, it is hoped that the results can serve as a useful framework for the subsequent studies on the extraction of semantic relations.

[3] http://cwn.ling.sinica.edu.tw/

1.3 Organization of the Thesis

The reminder of this thesis is organized as follows. Chapter 2 describes the relevant issues about lexical semantic databases and approaches on automatically extracting lexical information. It starts with an introduction of WordNet-like data resources along with how verb relations are classified in each database. Next, a detailed review on the definition of troponymy will be given. Finally, previous approaches on automated extraction will be reviewed and compared. In Chapter 3, we will introduce the two approaches adopted in this thesis including data sources, algorithm design, and procedure. Evaluation standard and scoring will also be presented in this chapter. Chapter 4 reports the results returned by each approach along with statistical representation and in-depth error analyses. The comparison of these two approaches will be discussed as well. Finally, Chapter 5 summarized the thesis along with our contribution and suggestions for future work.

Chapter 2

Related Works

This chapter devoted to the previous relevant studies and the introduction of some existing databases. In Section 2.1, I will begin with the introduction of several influential lexical semantic networks including Princeton WordNet, EuroWordNet, Chinese WordNet and Sinica Bilingual Ontological Wordnet. Section 2.2 turns the attention to the semantic relations of verbs which have different ways of classification in different Wordnet databases. Section 2.3 focuses on the target relation of this thesis–troponymy, which is the hypernymy-hyponymy relation among verbs. Finally, in Section 2.4, different approaches of discovering semantic relations in previous works will be discussed.

2.1 WordNet-like Resources

Lexical semantic resource has been considered vital resources in Natural Language Processing (NLP). In recent years, there has been an increasing focus on the construction of lexical knowledge resources such as Princeton WordNets [31], EuroWordNet [45], Mindnet [42], HowNet [13], VerbNet [26] etc. In this section, lexical semantic networks which are closely related to the this thesis will be introduced. Princeton WordNet, EuroWordNet, Sinica Bilin-

gual Ontological Wordnet, Chinese WordNet and HowNet will be introduced respectively in the following sections.

2.1.1 Princeton WordNet [31]

Launched in the early 1980s at Princeton University, WordNet has become a popular and crucial semantic network of English lexicon. Comprising more than 200,000 word meaning pairs, WordNet has grown to become an extensive electronic dictionary of English language [16]. Basically, WordNet differs from traditional alphabetical dictionaries in that it describes words with other words in a highly systematic way, mapping languages to concepts; also in that it organizes lexical information in terms of word meanings, but not word forms. WordNet organizes words into groups of near synonyms called synsets and synsets together define a unique sense (concept). What is more, in WordNet, the synsets are interlinked by means of bi-directional semantic relations such as meronymy, hypernymy and entailment, etc. Synsets and semantic relations, therefore, together serve as foundation of WordNet and set up this database. The semantic relations not only hold between words, but also between words and synsets, and between synsets themselves. The truth that WordNet is formed by synsets implies that only open class words in English are covered in WordNet since words in different categories (function words) cannot form synsets, as they are never interchangeable within a linguistic context. Therefore, WordNet contains only content words: Nouns, Verbs, Adjectives, and Adverbs. The following is quoted from Miller et al. in [31] defining the structures of each category.

> "Nouns are organized in lexical memory as topical hierarchies, verbs are organized by a variety of entailment relations, and adjectives and adverbs are organized as N-dimensional hyperspaces."

Started as an implementation of a psycholinguistic model of the mental lexicon, WordNet produces a combination of dictionary and thesaurus that is more intuitively usable. By the same token, WordNet has sparked off most interest both in theoretical and applicational sides, especially in NLP and IR. There are various off springs mapping WordNet's achievements onto languages other than English. Wordnets of European languages and Chinese, for example, are launched under similar spirit and will be discussed in the following sections.

2.1.2 EuroWordNet [45]

Being structured along the same lines as Princeton WordNet (specifically WordNet version 1.5), EuroWordNet is characterized by its multi-lingual nature. EuroWordNet is an integrated system that contains several European languages in which every sub-wordnet is a monolingual database. Each monolingual sub-wordnet represents a language –internal system constructed by synsets and basic semantic relations such as antonym, hyponym, meronym, etc. Besides, the equivalence relations between synsets in different languages and Princeton WordNet will be connected via the Inter-Lingual-Index (ILI) [46]. Therefore, by expanding words in one language to related words in another language via ILI, EuroWordNet may allow easy accessing of retrieving information stored in various languages. Also, the availability of multilingual makes each sub-wordnet compatible and comparable. Hence, not only the language- independent data is shared but also the language-specific differences can be maintained as well. Besides EuroWordNet's multilinguality which makes it distinct from Princeton WordNet, there are still some changes and additions in the EuroWordNet. One of the most distinct features is the explicit semantic relations across parts-of-speech (POS). In Princeton WordNet, nouns and verbs are not interrelated with each other by semantic relations; consequently, sometimes very similar synsets tend to be totally unrelated only because they belong to different POS. The following is

an example given in [44]: the noun *adornment* and the verb *adorn* expresses the same concept but they are not connected in WordNet for the different POS. On the other hand, words with different POS in EuroWordNet can be inter-linked with explicit synonymy and hyponymy relations. Therefore, the semantic relation between the noun *adornment* and the verb *adorn* can be described explicitly in EuroWordNet as follows where XPOS represents cross parts-of-speech:

```
{adore, V}  XPOS_NEAR_SYNONYM  {adornment, N}
```

Semantic relations used in Princeton WordNet were also adjusted in EuroWordNet. The most important relations of the Princeton WordNet have been maintained but extended in some ways [45]. First, for example, relations can have features; therefore, labels such as *conjunction*, *factive*, *reversed* and *negation* can be added to relations. Second, some existing relations have been broadened, like a more global near-synonym relation was proposed in EuroWordNet. Third, new semantic relations have been added as well; role-relations were used between entities and events. More detailed description of semantic relations in EuroWordNet will be introduced in the following section when discussing semantic relations among verbs.

Translingual knowledge is very important in the network era since uncountable information is disseminated across languages and boundaries via Internet; therefore, linguistic resources such as a multilingual thesaurus are indispensable for Information Retrieval. EuroWordNet is a successful example in following the paradigm of Princeton WordNet and also extend the design idea to the multilingual nature.

2.1.3 Sinica Bow [23]

Sinica Bow represents the Academia Sinica Bilingual Ontological Wordnet. It is a bilingual thesaurus integrated three resources: WordNet, English-Chinese Translation Equivalents Database

(ECTED), and Suggested Upper Merged Ontology (SUMO) [23]. Sinica Bow adopted WordNet version 1.6 and 1.7.1 which are used by most application so far [2]. And through ECTED, which is a crucial equivalence translation tool, a Chinese WordNet is bootstrapping from English WordNet. Constructed at Academia Sinica, the translation equivalence database was hand-crafted by a group of people. All Possible translations of an English synset word were extracted from several online bilingual (CE or EC) dictionaries. Afterwards, all the translated candidates will be checked by linguists who have near native-like ability in both languages. Finally, the translator selected three most appropriate translations whenever possible and tried to possess each translated entries into lexicalized words rather than descriptive phrases [25]. Another resource of Sinica Bow is SUMO, which represents an upper ontology constructed by Pease et al. [35] [34]. SUMO covers very general concepts such as time, spatial relations, physical objects, events and processes; the aim of SUMO is to link categories and relations coming from different top-level ontologies and to formalize different domains into a set of concepts, relations and axioms. SUMO can be applied onto the natural language processing, information retrieval, automated reasoning, and also the inter-operability in E-commerce, helping computers manage more details and structures. The three above resources together comprise the Sinica Bilingual Ontological Wordnet, which allow a versatile access of lexicon information combining lexical, semantic, and ontological information. Therefore, searching on Sinica Bow can return the following information: sense-based English-Chinese translation, semantic relations, English word-sense based ontology and inference, Chinese word-based ontology and inference.

Under the assumption that lexical semantic relations are universal [45] [16], they can be transported cross-lingually through accurate translation and mapping [24]. Therefore, with the integration of three key resources: ECTED, WordNet, and SUMO, Sinica Bow can plausibly

function as a medium for mapping WordNet to a Chinese one.

2.1.4 Chinese WordNet [1]

Creating a semantic relation database is a time-consuming and labor-demanding task. Creating a semantic relation-based wordnet in Chinese is even more difficult owing to the fuzzy definition and classification among words, morphemes, and characters. The Chinese WordNet (CWN) has been launched by Academia Sinica in 2006 and is continuously broadened its scope so far. Unlike some bilingual semantic databases such as Sinica Bow or HowNet, Chinese WordNet starts from Chinese per se but not through translating from another language. Lemmas included in Chinese WordNet mainly fall on the medium frequency words which are totally composed of 5600 lemmas and 13160 word senses [1]. Unlike Princeton WordNet or EuroWordNet, CWN is not a robust database that constructed mainly on the synsets and semantic relations. Rather, it aims to provide complete meanings and exquisite senses for each lemma based on the theory of lexical semantics and ontology. Also, each lemma has a corresponding English translation and is mapped onto WordNet via Sinica BOW. Figure 2.1, for example, shows the result of the verb 走 'zou3, walk' in Chinese Wordnet.

As Figure 2.1 shows, it is apparent that synsets and lexical semantic relations are not systematically labeled in CWN. In fact, the synsets of each word and the semantic relations among lemmas in CWN were manually tagged, and unlike WordNet or EuroWordNet, they have not been completely constructed yet. However, employing only human labor like the original WordNet on the construction of CWN would be wasteful and cannot guarantee compatibility. Hence, how to systematically and automatically discover the semantic relations is a crucial step toward constructing a synset and semantic relation-based Chinese Wordnet. Since CWN is composed of explicit senses of medium frequency words and can be seen as a detailed dictionary, this

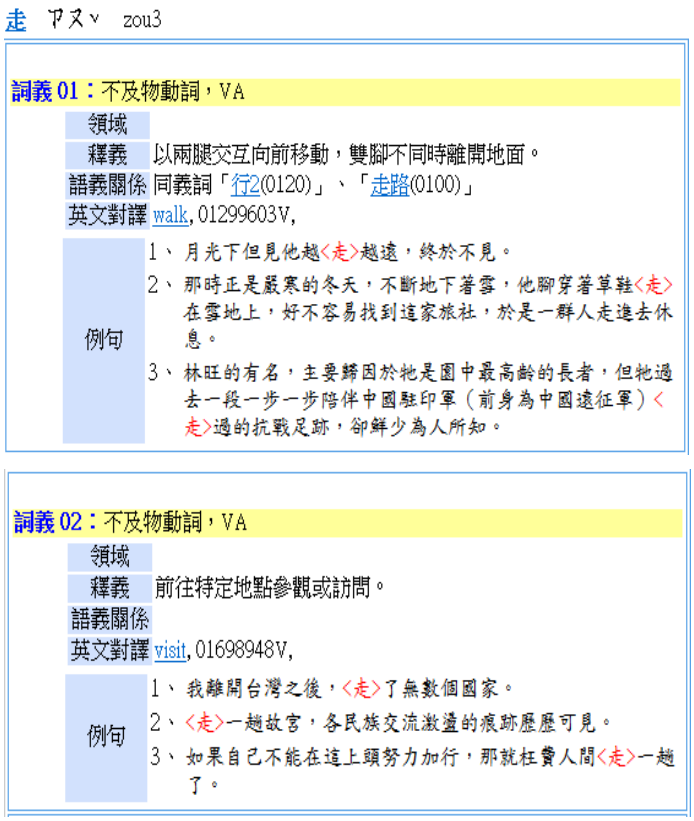


Figure 2.1: The first two senses returned by CWN of the verb 走 ‘zao3, walk’

present study tries to discover a pattern-based method of finding semantic relations (specifically the troponym relation) within CWN. Along with other approaches, such as bootstrapping from other existing WordNet [3] or bilingual dictionaries [25] (these will be introduced later), it is hoped that the results will be more direct and precise. While CWN is, until now, still devoted to extending its lemma from medium frequency words to high frequency words, our automated method on semantic relations extraction is hoped to enrich the completeness and speed up the construction of CWN.

2.1.5 HowNet [14]

Finally, another bilingual knowledge database is worth noting. HowNet is an on-line common-sense knowledge base which was constructed since 1988 by Dong [14] [13]. It is a large-scale bilingual lexical ontology for words and their meanings in both Chinese and English. HowNet, in many respect, is similar to Princeton WordNet but the biggest difference that characterizes HowNet and differentiates it from other WordNet-like databases is that HowNet relies less on hierarchical differentiation. In fact, HowNet unveils inter-concept relations and inter-attribute relations of each concepts [13]. It is proposed that the importance of constructing a knowledge system lies in the varied relations amongst concepts as well as those among the attributes of concepts but not taxonomic semantics. Therefore, it is obviously that HowNet and WordNet reflect a different view of semantic organization. As introduced, WordNet differentiates words by putting them under synsets and further differentiates synsets by assigning them to different taxonomy of semantics. On the other hand, HowNet does not provide glosses for each lexical concepts, rather, it combines sememes, which can be seen as the smallest basic semantic unit, from a less discriminating taxonomy to compose a semantic representation of meaning for each word sense. Under this fundamental, the knowledge structured by HowNet is a graph rather than a tree. How HowNet characterizes words can be better grasp by the following example [14]:

```
doctor|醫生 {human|人: HostOf={Occupation|職位}},domain= {medical|醫},
            {docotr|醫治:agent ={~}}}
```

As can be seen , the lexical concept describes in HowNet is not written in natural language but marked by basic sememes such as {Occupation— 職位} and {medical— 醫}. Despite the fact that HowNet is a powerful knowledge database, this thesis does not resort to it as

one of our data sources. The reason is that HowNet sketches lexical semantic relation in a syntagmatic perspective but not a taxonomic one. Therefore, the target relation in this thesis– the hypernymy-troponymy relation may exclude from this knowledge database.

2.2 Semantic Relations of Verbs

Although the target of this thesis is only on the hypernymy-troponymy relation among verbs which Fellbaum and Miller in [31] first dubbed as troponymy, yet a thorough review of semantic relations among verbs is needed. There is no consistency in classifying semantic relations of verbs. Therefore, in this section, I will introduce the semantic relations among verbs that defined in WordNet, EuroWordNet and other studies.

2.2.1 Semantic Relations of Verbs in WordNet

Verbs can not be classified into straightforward taxonomies like nouns do, rather, verbs are divided along semantic fields. Fellbaum in [15] proposed a detailed analysis of relations among verbs and indicates that **lexical entailment** lies under all verbal relation. WordNet identifies *Synonymy*, *Antonymy*, *Entailment*, *Troponymy*, and *Cause* among verb relations. This subsection is going to introduce and review semantic relations of verbs in WordNet. For the troponymy relation, which is the target relation in this thesis, a more detailed and complete review is needed. Therefore, this thesis left troponymy as an independent section and is given in Section 2.3.

Synonym

Basically, synset in WordNet indicates synonymous relation since synset is a group of words that together define a unique meaning. However, according to Miller and Fellbaum in [31],

there are few truly synonymous verbs can be found in the lexicon. Most synonymous verbs still have slight differences depending on the context and verbs also differed according to different speech register like in formal, technical, or colloquial environments. Examples are like *buy* vs. *purchase*, *sweat* vs. *perspire*. Owing to the truth that many apparently synonymous verbs still exhibit trivial differences, verb synsets in WordNet often expressed by periphrastic definitions rather than lexicalized synonyms and it reflects the fact that most verbs are manner elaboration of a more basic verb.

Antonym

Generally speaking, antonym is the concept which refers to a pair of words with opposite meaning, however, antonymous pairs of words are not simply ones with opposite meanings, but are suggested to be words with high relatedness both paradigmatically and syntagmatically. In WordNet, the antonym relation is also identified among verbs. Antonymous verb pairs cover the same semantic field and the same activity but differ only in the thematic role such as SOURCE or GOAL (for example *give* vs. *take*, *sell* vs. *buy*). Similar to synonymous verb pairs, there is few true antonym pairs which complementarily contrast with each other except static verbs and change verbs, such as *live* vs. *die* and *wake* vs. *sleep*.

Entailment

According to Miller [31], the entailment relation underlies all verbal relations:"the different relations that organize the verbs can be cast in terms of one overarching principle– lexical entailment." Entailment holds between two verbs V_1 and V_2 when the statement 'someone V_1s' entails the statement 'someone V_2s'. Lexical entailment is a unilateral relation, taking *sleep* and *snore* for examples: the proposition that *He is snoring* entails the proposition *He is sleeping*, as one cannot snore without sleeping. Therefore, owing to the property of unilateral relation,

the second proposition is necessarily to hold if the first one does but not the other way around. Lexical entailment also includes backward presupposition, such as *succeed* and *try*: If a person *succeeds* to do something, this entails that he has *tried* to do it.

The differences between verb pairs *snore/sleep* and *succeed/ try* lies in the **temporal inclusion**. *Succeed* entails *try* but neither verbs include the other in time, the proposition carries by these two verbs are in order : one *succeeds* after he *tries*. However, *snore* entails *sleep* and is properly included by it : the time span of *snore* is properly and entirely included in *sleep*. Figure 2.2 is a graphic representation of the relations among different kinds of entailment.

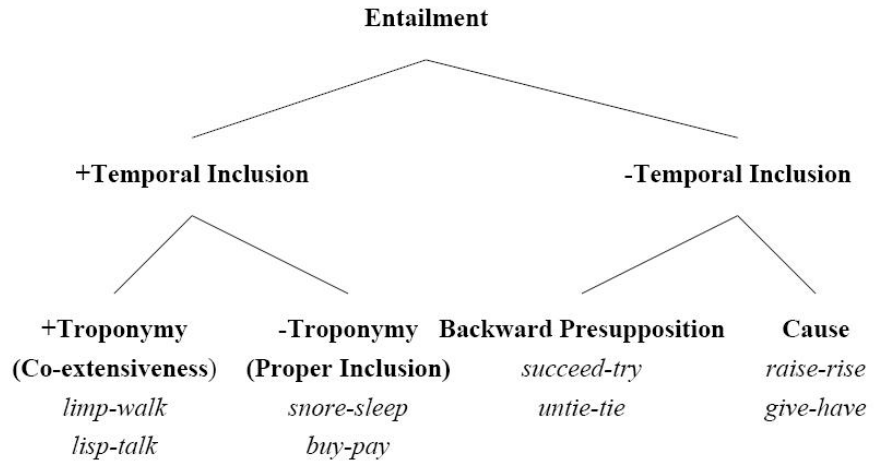


Figure 2.2: Four kinds of entailments among English verbs [31]

Cause

As Figure 2.2 shows, there are four kinds of entailments distinguished by two features {$\pm$ temporal inclusion, $\pm$ co-extensiveness}. WordNet identifies only *Cause* relation and *Troponymy* relation. The latter is the target of this thesis hence will be discussed in detail in

Section 2.3. *Cause* is a special form of entailment which only applies to temporally disjoint events. Two concepts are mentioned in WordNet, one is a causative verb concept like *give*, and the other is a resultative concept like *have*. Causative verbs carry the sense of *cause to be/become/happen/have* or *cause to do*. That is to say, they relate transitive verbs to either states or actions. For example, the synset *break, bust* in WordNet has a causative relation between *break, wear, wear out, fall apart* which indicate *cause to become pieces*.

2.2.2 Semantic Relations of Verbs in EuroWordNet

As mentioned, EuroWordNet is built along the same lines as Princeton WordNet in that synonymous meanings are joined in a synset and language-internal relations are expressed between synsets. However, semantic relations are being adjusted in EuroWordNet. The most distinct feature in EuroWordNet is the explicit cross-part-of-speech (POS) relations. The Princeton WordNet uses a rigid distinction between nouns and verbs, mainly because of their different syntactic role in English. However, this often leads to an undesirable consequence that very similar synsets are totally unrelated only because they belong to different syntactic categories. Therefore, in EuroWordNet, synsets can be inter-linked with one another across POS. That is, verbs do not related with verbs only, but also nouns or adjectives. The following is some important semantic relations of verbs in EuroWordNet [10] [4]:

NEAR_SYNONYMY and XPOS_NEAR_SYNONYMY

Unlike WordNet, EuroWordNet used the relation NEAR_SYNONYMY for semantically similar words. In many cases, there is a close relation between words but sometimes this relation is not sufficient to group the words into one synset. This is because the hyponyms linked to each of these words can not be exchanged. Therefore NEAR_SYNONYMY relation is used in

EuroWordNet to keep the hyponyms separate and still be able to express that the two words are close in meaning [4]. NEAR_SYNONYMY relation is used cross part-of-speech in EuroWordNet, therefore, morphologically very similar words can connect through semantic relations even when they belong to different POS. The followings are examples of verbs in synonymous relation:

(1) {move, V} XPOS_NEAR_SYNONYMY {movement, N}

(2) {sleep, V} XPOS_NEAR_SYNONYMY {sleep, N}

HYPERNYMY/HYPONYMY and XPOS_HYPERNYMY/HYPONYMY

A hyponymy relation implies that the hypernym may substitute the hyponym in a referential context but not the other way around. In EuroWordNet, HAS_ (XPOS) _HYPERNYMY/HYPONYMY relation is used to indicate the relation between a more general class and a more specific subtype either within or across POS. Examples for verbs are as follows:

(3) {move, V} HAS_HYPONYMY {walk, V}

(4) {love, V} HAS_XPOS_HYPERNYMY {emotion, N}

(5) {emotion, N} HAS_XPOS_HYPONYMY {love, V}

Note that the term 'troponymy' is not used in EuroWordNet. Verb pairs in hypernymy/ troponymy relation like *walk* and *move* are expressed by HAS_HYPERNYMY/HYPONYMY relation in EuroWordNet and it only occurs within one POS.

ANTONYM and XPOS_ANTONYM

In EuroWordNet, the Antonym relation referred to a more loosely defined notion which can be defined as the *opposition* of meaning in a context. As mentioned, antonym between different POS is allowed, as in the cases of synonymy and hypernymy/hyponymy. Antonyms typically form contrasting categories within the same dimension. This means that an antonym not only contrasts with another antonym in certain features but they have to share the same hypernym. This criterion prevents irrelevant pairs such as *love* and *car*. The following are examples of antonym relation in EuroWordNet:

(6) {open, V} ANTONYM {close, V}

(7) {live, V} XPOS_ANTONYM {dead, A}

CAUSE (XPOS)

The cause relation is used in WordNet to indicate the entailed relation between two verbs, one of which referring to an event causing a resulting event, process or state. Whereas in WordNet the causal relation only holds between temporally disjoint verbs, EuroWordNet, on the contrary, applies this relation across syntactic categories. Causal relation in EuroWordNet can be further distinguished with respect to the *factivity* of the effect. The *factive* feature denotes that a situation S_1 implies the causation of S_2, for example:

(8) {kill, V} XPOS_CAUSES {death, N} (factive)

(9) {death, N} IS_CAUSED_BY {kill, V}

On the other hand, *non-factive* indicates a situation S_1 probably or likely causes S_2 or S_1 is intended to cause some situation S_2, for example:

(10) {try, V} CAUSES {succeed, V} (Non-factive)

From the above examples, EuroWordNet does not further distinguish causal relation and backward presupposition like WordNet does, rather, only causal relation plus the feature of *factivity* are used in EuroWordNet.

SUBEVENT (XPOS)

Recall that Princeton WordNet identifies the relation *Entailment* on cases that cannot be expressed by the more specific hyponymy or cause relations. In this case the direction of the implication or entailment is indicated. For example, in the case of *snore/sleep*, the entailment direction is from *snore* to *sleep*: *snore* entails *sleep* but not the other way around. However, in EuroWordNet, the entailed relation with 'proper inclusion' can more adequately be described by means of the SUBEVENT relation which is very useful for many closely related verbs and appeals more directly to human-intuitions.

(11) {snore, V} IS_SUBEVENT_OF {sleep, V}

(12) {sleep, V} HAS_SUBEVENT {snore, V}

ROLE/INVOLVED

EuroWordNet identifies additional semantic relation ROLE and INVOLVED which indicates a link between a verb and a noun whose meaning is 'incorporated' in, or connected with the

meaning of the verb itself. Sometimes the most salient relation is not the hypernymy but the relation between the event and the involved participants. The subtype of ROLE/INVOLVED relation in verb-noun pairs also includes:

Involved-agent/ patient/ instrument/location/ direction

Involved-source-direction

Involved-target-direction

Examples like *write, V* and *pencil, N* are in INVOLVE-INSTRUMENT relation in that the meaning of the verb *write* incorporates the meaning of the noun *pencil*. The relations are expressed as follows:

(13) {write, V} INVOLVED_INSTRUMENT {pencil, N}

(14) {hammer, N} ROLE_INSTRUMENT {to hammer, V}

2.2.3 Other Relations of Verbs

From WordNet or EuroWordNet's classification, verb entries related to each other by their paradigmatic relations such as hyponymy, antonym and meronymy. However, this kind of classification may easily organized verbs into a flat or shallow hierarchy of classes [9]. Verbs can also relate with each other in a syntagmatic way. Syntagmatic relations constrain the contexts in which a word may be used, and can be seen as a complementary way of representing speaker's lexical knowledge [17] [11]. For example, Chklovski and Pantel in [9] proposed a finer-grained analysis of verb relations, which explicitly breaks out the temporal precedence between entities. Verbs which are related with one another syntagmatically together made up a broad-coverage semantic network called VerbOcean[1]. VerbOcean identified five semantic relations between

[1] http://demo.patrickpantel.com/Content/verbocean/

verbs. This can be summarized in Table 2.1 with a comparison to WordNet.

Semantic relation	**Example**	**Alignment with WordNet**	**Symmetric**
similarity	transform / integrate	synonyms or siblings	Y
strength	push / nudge	synonyms or siblings	N
antonym	open / close	antonymy	Y
enablement	wash / clean	cause	N
happens-before	buy / have; marry / divorce	cause; entailment	N

Table 2.1: A finer-grained semantic relation among verbs. [9]

Similarity

Very similar to the synonym in WordNet, verbs are often similar or related but not one hundred percent equals to each other. Similarity occurs between action verbs, for example, *transform/ integrate* or *produce/create*.

Strength

As mentioned, verbs can be very similar to one another but differ in their strength. When two verbs are similar, one of them may denote a stronger intention, more absolute in action like verb pairs *push* vs. *nudge*. Or in the case of change-of-state verbs, one may carry a more complete change than the other, such as *wound* vs. *kill*.

Antonym

Also similar to the classification in WordNet, antonymy is also known as semantic opposition. As Fellbaum described, it can arise from switching the thematic roles of verb pairs like *buy* vs. *sell*, *give* and *take*. Antonym can also occurred in stative verbs like *live* vs. *die*.

Enablement

Enablement can be seen as a type of causal relation hold between two verbs and satisfy the formula that V_1 is accomplished by V_2, examples are like *accomplish*/ *complete*, and *fight*/*win*.

Happens-before

The relation between two verbs may exhibit temporally disjoint intervals. The happens-before relation hold between two verbs corresponds to the *Cause* relation in WordNet which shows no temporal inclusion of two verbs but indicates two events happens in order such as *sell* happens before *buy*, and *marry* happens before *divorce*.

To sum up, semantic relations among verbs can be organized in terms of paradigmatic and syntagmatic perspectives. Paradigmatic organization relates words in hierarchy, that is, connect a more basic word to a more general concept or superordinate. Lexical resources like WordNet, EuroWordNet and many dictionaries identify paradigmatic relations among synsets or lexical entries. Some lexicons also supply syntagmatic relations between the target and other words by means of illustrative sentences. However, there is no consistency in the classification of semantic relations of verbs. In this section, we introduced semantic relations of verbs in WordNet, EuroWordNet and VerbOcean. The comparison of verb relations in these three databases can be summarized by Table 2.2.

2.3 Troponymy

Section 2.2 has introduced semantic relations of verbs, this section turns the focus back to the target relation in this thesis—the hypernymy/troponymy relation. Our literature survey reveals that, to the best of our knowledge, there is no study targeting at the troponymy extraction yet.

WordNet	**EuroWordNet**	**VerbOcean**
Synonym Synset (verb groups)	*NEAR_SYNONYM* *XPOS_NEAR_SYN* {Adore, V:: adornment, N}	*Similarity* Produce:: create
Hypernymy/Troponymy Walk:: limp	*HYPERNYM/HYPONYM* *XPOS_HYPERNYM/HYPONYM* {love, V}::{emotion, N}	*Strength* Wound:: kill
Antonym {close, shut}::{open}	*ANTONYM* *XPOS_ANTONYM* {live, V}::{dead, A}	*Antonym* Open:: close
Cause Kill::{die, decease, perish. . . }	*CAUSE (XPOS)* {kill, V}::{die, V} {kill, V}::{death, N}	*Enablement* Fight:: win
Entailment {divorce, split up}:: {marry, wed. . . }	*SUBEVENT(XPOS)* snore, V::sleep, V but, N::payment, N	*Happens-before* marry:: divorce
	ROLE/INVOLVED {write, V}::{pencil, N} {teach, V}::{teacher, N}	

Table 2.2: Semantic relations of verbs in Wordnet, EuroWordNet and VerbOcean

Troponymy relation is discussed less often than other semantic relations. In fact, this term was first dubbed and defined by Fellbaum and Miller in [31]. Therefore, in this section, I will try my best to cover a detailed introduction of troponymy basically based on Fellbaum et al.'s studies [15] [19] [16] [17] [31].

2.3.1 Definition of Troponymy

Basically, troponymy can be seen as the hypernymy-hyponymy relation of verbs. While **troponymy** is the term used to indicate the relation, **troponym**, on the other hand, is used to describe the subordinate word in this relation. For example, verb pair *move* and *walk* are in hypernymy/troponymy relation while *move* is the hypernym of *walk* and *walk* is the troponym of *move*. Hypernymy/troponymy relation, by its definition, may be corresponded to the *is-a* relation of nouns, but unlike nouns, verbs do not straightforward fit into the *is-a* relation nor related in a consistent mode, rather, verbs are connected by a **manner elaboration** from some superordinate verbs. As mentioned, entailment lies under all verbal relations thus the concept of entailment is important for troponymy to hold since troponymy is a special kind of entailment. Saying that troponymy is a kind of entailment indicates that a troponym inherits and entails its superordinate verb but is specified in different manner. Basing on this assumption, the definition of troponymy can be expressed by the formula: *to* V_1 *is to* V_2 *in some particular manner*. So, *walking* describes a manner of *moving*, elaborating on the speed and style. Therefore we say that *walk* is a troponym of *move* and *move* is the hypernym of *walk*. Similar to the *is-a* relation among nouns, troponymy also builds hierarchical structures with the semantically most inclusive verb at the root and increasingly specified subordinate verbs as the extending branches and leaves. However, unlike noun hierarchies which are tall and deep, verb hierarchies tend to be 'flatter' and 'bushy', most of them do not exceed more than three or four levels.

As mentioned, there are four kinds of lexical entailments of verb relations, demarcated by two features $\{\pm$ temporal inclusion, $\pm$ co-extensiveness$\}$. Temporal inclusion distinct verb pairs *succeed/try* from pairs like *snore/ sleep* and *move/walk* since the former pair demonstrates two temporally disjoint events while the latter indicate proper inclusion in time. But what criterion makes verb pair *snore/sleep* different from *move/walk* for that *snore* in not a troponym of *sleep* but *walk* is a troponym of *move*? According to Fellbaum and Miller [19], saying that troponymy is a particular kind of entailment involves **temporal co-extensiveness** for the two verbs. That is, the proposition of a troponym must and always entail its superordinate event occurs at the same time. For example, although *snore* entails *sleep* and is temporally included in *sleep*, we can not say that *snore* is a troponym of *sleep*, they are not in a hypernymy/ troponymy relation. This pair of verbs is related only by entailment and proper temporal inclusion. The important generalization here is that verbs related by entailment and proper temporal inclusion cannot be related by troponymy. On the contrary, for troponymy to hold—for we can say that *to* V_1 *is to* V_2 *in some specific manner*, the essential factor is the **co-extensiveness** in time: one can *sleep* before or after *snoring*, but not necessarily happened at the same time. Take another hypernymy/troponymy verb pair *limp* and *walk* for example, *limping* entails *walking* and *walking* can be said to be a part of *limping*, they are temporally coextensive, that is, *limp* and *walk* happen at the same time.

Beside the complicated distinction among verbs themselves, the troponymy relation is also different from the *is-a* relation among nouns in two ways [17]. First, the *is-a-kind-of* formula linking semantic related nouns may cause oddness when applying to verbs. For example, '(to) yodel is a kind of (to) sing.' sounds odd only when changing into gerund form 'yodeling is a

kind of singing' will make it acceptable. Second, in the case of nouns, *kind of* can be omitted without changing the truth statement, for instance, 'A donkey is a kind of animal.' equals 'A donkey is an animal.' By contrast, the same deletion makes verbs odd as the following sentences show: 'Murmuring is talking/ To murmur is to talk.' These differences indicate that there is more than just a *is-a* relation among concepts expressed by verbs and the way that used to distinct nouns and adjectives is not the same as the way we distinct verbs. To sum up, troponymy links verbs in a *manner* elaboration rather than *kind* relation.

2.3.2 Distinguishing Manner

The most important factors of troponymy are **manner elaboration** and **co-extensiveness** as introduced above. However, the so-called 'manner', in fact, can be further analyzed and distinguished. Fellbaum [31] examined a few areas of the verb lexicon where the encoding of a manner component has interesting consequences, and attempted to draw some distinctions among different kinds of troponymy.

Function vs. Manner

Many nouns that fit into the *is-a-kind-of* relation do not necessarily in hypernymy/hyponymy relation. Pustejovsky [40] discussed the semantic meaning of nouns like *pet* which he calls *roles*. Take *dog* and *pet* for example, *dog* can be a kind of *pet*, but the relation between *dog/animal* and *pet/animal* are different, the intrinsic meaning of these two words are not the same. There is no problem to state that *a dog is a kind of animal* but it is odd to say that *a pet is a kind of dog/animal*. The reason is that 'pet' is the *role* that an animal plays in a certain setting

or under certain condition which may be culturally and temporarily dependent. On the contrary, the *is-a* relation that links *dog* to its superordinate *animal* is biologically stable. The same phenomenon exists in the verb lexicon. Many verbs may have meaning aspects that depend on the context. For example, *run*, *walk*, and *swim* are all manners of motion verbs *move*. On the other hand, *run*, *walk*, *swim*, along with *bike*, *row* or *ski* are also manners of *exercise*. Similar to the role function of nouns, *exercise* is an unstable superordinate verb because the activities that can be defined as exercise may vary. But *move* is a stable hypernym of *run*, *walk*, *swim*, etc. because it inherits its meaning to the subordinate verbs. Therefore, the relation between *run*, *walk*, *swim*, etc. and *exercise* is being called the **function relation**. The distinction between genuine troponymy and the function relation is reflected in the fact that in cases like the above, *exercise* is not a part of these verbs' meanings in the same way that their superordinate *move* is.

Result vs. Manner

Besides verbs expressing the function or manner in which an action is carried out, English has many verbs that encode the *result* of an action but not the manner of achieving this result. For example, a result verb like *shut* does not carry the manner of the motion by itself, but depends on the noun object that denotes the entity that is acted upon. To make it clear, in the case of *shut*, the manner of shutting a window is definitely different from the manner of shutting a book. The manner is not specified by *shut* itself but by the noun object following *shut*. In contrast to result verbs, a manner of motion verb like *run* denotes a specific kind of traveling motion that does not vary depending on who or what does the running. For this reason, the result verbs are also called accomplishment verbs, and manner verbs are activity verbs.

Perhaps the most significant relation among lexicon semantics is the hypernymy/hyponymy

relation, which builds hierarchical structures. Troponymy, as this relation is called in the verb lexicon, apparently relates verbs in terms of a manner elaboration. The so-called 'manner', according to Fellbaum, can be further analyzed and distinguished and thus yields different types of troponymy. Table 2.3 summarized three different types of troponymy.

Manner verb	Function verb	Result verb
The event itself is carried out by the verb. e.g. move::run	The verb encodes the function that is context- or situation-dependent. e.g. exercise:: run	The verb encodes only the result of an action. e.g. shut:: snap

Table 2.3: Three different types of Troponymy

Fellbaum introduced a very detailed analysis of entailment in [16] [18] and also the complication of troponymy [15]. The clear distinction and definition of hypernymy/troponymy relation made by Fellbaum therefore motivates this thesis and also serves as the norm in describing and evaluating the result.

2.4 Automatic Discovery of Lexical Semantic Relation

Automated extraction of semantic resources has been widely studied by researchers. From Machine readable dictionaries [32] [41] to the unrestricted web corpus [22] [21] [9], Natural Language Processing researchers use these databases to develop algorithms for harvesting lexical semantic relations [21], observing rules [27], or extracting target words [20] [28]. Also, there has been a variety of studies focusing on the automatic acquisition of different lexical semantic relations, such as hypernymy/hyponymy [21], antonymy [28], meronymy [20] [7] and so on. Now I begin to discuss some approaches used in previous studies.

2.4.1 Lexico Syntactic Pattern–Based Approach

Automatically finding semantic relations among synsets by using pattern-based algorithm may be the most common approach [39]. It is generally agreed that the structure of a lexical entry in a dictionary sometimes reflects the relatedness of words and concepts [22]; also, certain structures or syntactic patterns usually define the semantic relation among each other. The following three subsections introduce some influential and novel studies on automated extraction.

Nakamura & Nagao (1988) and Hearst (1992)

Early studies analyzed and extracted information from Machine Readable Dictionaries (MRD) (cf. [5] [33] [32] [30]). By syntactic or phrasal patterns found in definitions, certain semantic relations can be accessed automatically. For example, Nakamura and Nagao [32] created a Machine Readable Database (MRD) from Longman Dictionary of Contemporary English (LDOCE), by extracting *key Verb/ key Noun* which express a 'key concepts' of the defined verb, the algorithm can automatically extract the taxonomic information and certain semantic relations. Basing on the similar spirit, Hearst [21] pioneered enlarging the scope to the unrestricted text as her database. Similar to previous pattern-based interpretation techniques, she proposed the LSPE (Lexico-Syntactic Pattern Extraction) method, upon which most of the related works are based, to automatically find hyponym relations among nouns. Manually constructed three patterns at first, Hearst discovered three more new patterns from seed instances by bootstrapping. After the pattern-discovery procedure, six lexico-syntactic patterns which denote the concept of 'including' or 'other than' were postulated. And basing on a text corpus, which contains terms and expressions that are not defined in Machine Readable Dictionaries; the six lexico-syntactic patterns successfully detected hypernymy-hyponymy relation and extracted these pairs from the sentences. The six syntactic patterns used in Hearst's algorithm

are as follows: (1) *X such as Y*; (2) *such X as Y*; (3) *Y, or other X*; (4) *Y, and other X*; (5) *X, including Y*; (6) *X, especially Y*. For terms that are present in the above patterns, this algorithm successfully captures the relation that Ys are hyponymy of Xs. Hearst's pattern-based approach successfully extracted nouns which are in hypernym-hyponym relation from large unrestricted text. Inspired by Hearst, it is hoped that other lexical relations can also be acquirable in the same way, or at least, from the similar paradigm.

Ramanand and Bhattachayya (2008)

Automated semantic relation discovering can be further used to evaluate the quality and the reliability of an existing thesaurus like Princeton WordNet. Ramanand and Bhattachayya [41] pioneered to propose an algorithm based on dictionary definitions to verify that synsets in WordNet are indeed synonymous. As mentioned in the previous section, WordNet resembles a thesaurus in that it represents word meanings primarily in terms of conceptual-semantic and lexical relation. Therefore, a synset, the foundation of WordNet, is constructed by assembling a set of synonyms that together define a unique sense. The basic idea of Ramanand and Bhattachayya's approach is that a lexical entry in the dictionary is usually defined in terms of its hypernyms or synonyms. Hence, they use the concept of synset to suggest that 'if a word *w* is present in a synset along with other words *w1,w2,...wk* , then there is a dictionary definition of *w* which refers to its hypernym or to its synonyms from the synset.' With this assumption, three groups of rules were applied in order in the algorithm for validating synonymy and hypernymy relation among Wordnet synsets. For example, for the rule which can denote hypernymy, the author defined that the definitions of words for particular senses often make references to the hypernym of the concept. Such as synset: { *brass, brass instrument*} and its hypernym: {*wind instrument*, *wind*}. The relevant definition of *brass instrument* is: a musical *wind instrument* of

brass or other metal with a cup-shaped mouthpiece, as the trombone, tuba, French horn, trumpet, or cornet. The positive results of Ramanand and Bhattachayya's approach proved that the basic idea behind the algorithm holds well: by using the definitions of each word, hypernymous or synonymous relations among words can be discovered.

Chklovski & Pantel (2004) and other related works

In [9], Chklovski and Pantel also used some simple lexico-syntactic patterns to automatically discover fine-grained verb semantics. The so-called 'fine-grained' semantics, as mentioned in section 2.2.3, relates words in a syntagmatic way. It is differed in focus from previous studies which related verbs to each other by organizing them into classes or identifying their frames or thematic role. It is also differed from WordNet which provides relations between verbs at a coarser level, such as hypernym, antonym, synonym, etc. Chklovski and Pantel related verbs in a broad-coverage semantics such as *strength*, *enablement*, and *temporal information*. Verbs which are related in a syntagmatical way together made up a semantic network called VerbOcean.

Syntactic pattern -based approach were then proposed by Chklovski and Pantel in order to automatically extract verb relations within VerOcean. The approach can be divided into two stages. First step was to extract highly associated verb pairs as experimental data which were output by a paraphrasing algorithm called DIRT proposed in [27]. Afterwards, syntactic patterns were used to detect the possible semantic relations among verbs. By examining pairs of verbs in known semantic relation, 35 syntactic patterns were selected manually. For example, one of the surface patterns used to identify the *strength* relation is: *X and ever Y*; *Yed or at least Xed*, and for patterns that denote enablement are: *Xed by Ying the* or *to X by Ying the*, etc. Since verbs are the primary core for describing events and events are exhibiting the rela-

tion of different entities, successfully identified fine-grained semantic relations is important in question-answering or summarization system.

2.4.2 Clustering-Based Approach

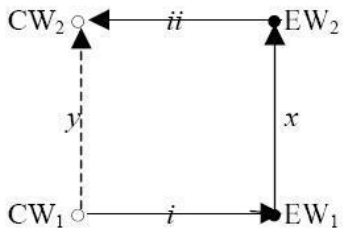
Finding semantic relations by clustering is another approach used in automated extraction. However, it is less common and has been applied only on *is-a* relation insofar [39]. Recently, Pantel and Ravichandran [38] proposed an algorithm labeling semantic classes and their concepts by clustering and then automatically harvest *is-a* relation using a top-down approach. Unlike a bottom-up approach which depends on surface syntactic patterns to discover hyponym relations (cf. [21]), Pantel and Ravichandran's is a top down approach. They made use of semantic classes which discovered by clustering algorithm called CBC (Clustering by Committee) proposed in [37] and output a ranked list of concept names for each semantic class. After knowing the general concept of a group of words, the extraction of hyponyms will be easy. For example, the following semantic class extracted automatically by CBC consists of words shown below:

(A) multiple sclerosis, diabetes, osteoporosis, cardiovascular, disease, Parkinson's, rheumatoid arthritis, heart disease, asthma, cancer, hypertension, lupus, high blood pressure, arthritis, emphysema, epilepsy, cystic fibrosis, leukemia, hemophilia, Alzheimer, myeloma, glaucoma, schizophrenia, ..

After the above semantic class (A) is labeled with the concept *disease* by using its member's lexical or syntactic dependencies, the *is-a* relationship can be extracted easily such as: *multiple sclerosis is a disease*, and *diabetes is a disease*, etc.

2.4.3 Bootstrapping Approach

The rich and structured semantic information in WordNet or EuroWordNet can be transported through accurate translation if the conceptual relations defined by lexical semantic relations (Hereafter, LSRs) remain constant in both languages. In Huang et al.'s paper [24] [25], cross-lingual LSR inferences were examined by bootstrapping a Chinese Wordnet with Princeton WordNet. As known, a semantic network is critical for knowledge processing, but it is not an easy task to create a semantic-relation based wordnet by manually construction. Therefore, by using an existing wordnet as a medium, it is hoped that a wordnet for a low-density language can be built through bootstrapping. The basic idea of bootstrapping lies in the universality of lexical semantic relations; through accurate translation between two languages, an existing database such as WordNet can serve as a potential common semantic network. In [24], a small set experiment data including 210 most frequently used Chinese words were served as input data and examined the validity of cross-lingual LSR inferences by bootstrapping a Chinese Wordnet from Princeton WordNet.



$x = EW_1 - EW_2$	LSR
$y = CW_1 - CW_2$	LSR
$i = CW_1 - EW_1$	Translation LSR
$ii = CW_2 - EW_2$	Translation LSR

The predicted LSR y = i + x + ii

Figure 2.3: Translation-mediated LSR Prediction (The complete model)

Figure 2.3 can demonstrate the bootstrapping method in Huang et al.'s study: CW_1 represents

the starting Chinese lemma which can be linked to EW_1 through the translation relation *i*, and therefore EW_1 can provide a set of LSRs based on WordNet. EW_2, under this situation, can be linked with EW_1 through relation *x*. LSR prediction is mapped back to Chinese when EW_2 is translated to CW_2. However, according to the authors, Figure 2.3 is only an ideal bootstrapping model since language translation involves more than semantic correspondences. Social and cultural factors also play a role in human choices of translation equivalents. Regardless this fact, the aim of Huang et al.'s study is to see how much lexical semantic information is inferable across different languages so the translational idiosyncrasies are ignored and the ideal model can be illustrated in Figure 2.4 where it is assumed that there is no translational discrepancy between CW_1/ EW_1 and CW_2/ EW_2.

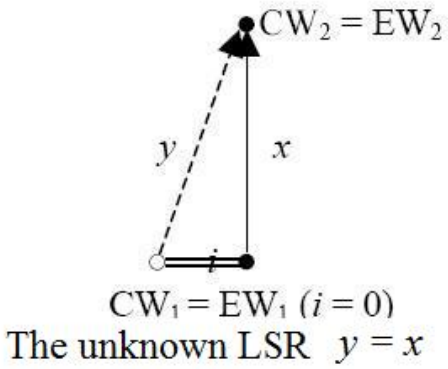


Figure 2.4: Translation-mediated LSR Prediction (when translation equivalents are synonymous)

Hypernymy, hyponymy and antonym relation are tested in [24] for these relations are transitive and allow clear logical predictions when combined. The results show that lexical semantic relation translations are indeed highly precise when they are logically inferable. Also, in terms of syntactic categories, it is observed that nominal semantic relations are more reliably inferred

cross-linguistically than verbal semantic relations. The fact that verbal relations are less reliable after bootstrapping is not a surprising result since polysemous lemmas have lower possibility of being synonymous to the corresponding English synset. That is to say, verb meanings are more mutable than noun meanings hence make it more difficult in cross lingual bootstrapping. In sum, the positive results in Huang et al.'s study set up a theoretical model for the following researches that LSRs do have the ability to transport cross-lingually, at least for hypernymy, hyponymy, and antonym relations.

2.5 Summary

In this chapter, the studies pertaining to the following issues are reviewed: semantic relation-based Wordnets, lexical semantic relations of verbs, troponymy, and automated approaches of extracting semantic relations.

Creating a semantic relation-based database is crucial for knowledge mining and natural language processing. Some semantic relations such as *is-a* relation, *part-of* relation have been successfully discovered by automated method. However, the literature survey reveals that there is no study targeting at hypernymy-troponymy relation in Chinese yet. Therefore, in this chapter, detailed definition and information of troponymy are being reviewed; also, three approaches on automatic extraction of semantic relations are introduced as well. To date, most researches on lexical relation discovering have been focused on *is-a* relation or *part-of* relation by using syntactic pattern-based or clustering-based approach. **Syntactic pattern-based** algorithm is a bottom-up approach and is the most commonly used one for that sentence structures can always denote relations and reveal information among lexical entries. On the other hand, the top-down **clustering-based** approach, which is used less often, has an advantage that it can identify hyponymous relations that do not explicitly appear in text or patterns. **Bootstrapping**

approach has been used for mapping English WordNet onto Chinese Wordnet but the experimental data is rather small and an important factor–translational idiosyncrasy– is being ignored. The literature review shows that none of these approaches were used to extract troponymy relation particularly in any languages: syntactic-based approach is used to extract *is-a* relation and *part-of* relation only among nouns, and although clustering-based approach has been used to extract relations of verbs, it focused on the syntagmatic relation but not hierarchical one. This thesis, therefore, is going to adopt the syntactic pattern-based and bootstrapping approach for extracting semantic relations. Clustering-based approach, unfortunately, is rather unachievable during this phase so will not be concerned as one of the approaches. It is hoped that the results can serve as a useful framework for the subsequent studies on the extraction of semantic relations.

Chapter 3

Methodology

This chapter presents two main approaches that this thesis adopts to automatically extract Chinese verb pairs in hypernymy-troponymy relation. These two approaches were implemented parallely and will be illustrated separately in the following sections. Section 3.1 describes how lexical syntactic pattern-based approach processes while Section 3.2 illustrates the procedure of bootstrapping approach. Finally, how to evaluate and score our results will be introduced in the final part of this chapter.

3.1 Syntactic Pattern-Based Approach

This section introduces the first approach used in the thesis, including the data source, two syntactic patterns in Chinese, the programming language and the procedure.

3.1.1 Database: Chinese WordNet

In this thesis, all verbs in Chinese WordNet (CWN) were used as the experimental data. Launched by Academia Sinica, CWN can be seen as a machine readable dictionary which covers 5600 medium frequency words and their explicit senses (until 2006). In recent years,

more and more studies on automatic extraction rely on large corpus or the unrestricted Web text as mining resources since the web serves as a strong application of data mining. The information on the web is immense, free, and easily available; it contains hundreds of billions of words or text, and can be used for all manners of language research. Despite the multi-faceted advantages of using the Web as corpus, it is more feasible for our approach to narrow down the scope to a dictionary-like database—- Chinese WordNet for the following considerations:

First, situation in Chinese is much more complicated in unrestricted text. Due to the varied forms and contents of the unrestricted text, language is easily metaphorized in free text. Take the *is-a* relation from web for example, we searched the corresponding pattern in Chinese *ShiYiZhong* 'is a kind of' from the search engine Google[1] and the following sentences are the first five results returned by it:

(1) 馬爺爺04：***筷子***是一種***槓桿*** (*Chopsticks* is a kind of *lever*.)

(2) 她說***寫作***是一種***治療***- WRETCH (*Writing* is a kind of *treatment*.)

(3) 博客來書籍館***跆拳***是一種***態度*** (*Tae-kwon-do* is a kind of *attitude*.)

(4) 「***勾心鬥角***」是一種***能力***? (*Intriguing against each other* is a kind of *ability*.)

(5) 天下文化書坊***亂***是一種***新商機*** (*Chaos* is a kind of *commercial possibility*.)

From the above example sentences, it is apparent that none of the noun pairs which fix the N_1 *is a kind of* N_2 pattern are in hypernymy-hyponymy relation. At least, none of these pairs indicate the *is-a* relation in WordNet; rather, all of them are being metaphorized to express some abstract

[1] http://www.google.com.tw/ (retrieved on 2008/11/5)

concepts. The first five results returned by the searching engine were not in accordance with our taxonomic aim, implying that they might have lowered the precision rate because of the large amount of indirect and unrelated data on the web. For this reason, unrestricted web texts are not considered to serve as the mining resource in this study. **Second**, in this thesis we proposed two syntactic patterns which can indicate hypernymy-troponymy relation. We assume that there might exist a lexical semantic relation between a defined entry and its dictionary definition, represented through certain syntactic patterns. In this thesis, it is postulated that the hypernym of the defined verb might appear in the definition. Therefore a dictionary-like database with explicit definition of each sense like CWN was chosen. **Third**, for a WordNet-like database in Chinese such as CWN, lexical semantic relations between senses or lemmas themselves have not been fully constructed yet. Hence, it is hoped that the results returned by our approaches can be applied directly onto the existing database, helping to enhance its integrity.

3.1.2 Data Pre-processing

Verbs with distinct senses were extracted from CWN word list, serving as the experimental data in the approach. This input verb list contains 10388 entries of verb senses in total and each verb entry is followed by its definition. As known, most verbs have more than one sense, and each sense of a given word might have their own hypernyms and troponyms. Note that the verb list extracted from CWN contains verbs with distinct senses, indicating that each entry represents one sense but not one verb; also, different senses of a same verb were marked by serial numbers. Next, for the sake of programming, we apply the following preprocessing methods to the source data:

- Tokenizing: Tokenization was applied onto the input text, aiming at separating punctua-

tion marks from words and identifying sentence boundaries.

- POS tagging: Assign word class to tokens through Chinese segmentator and POS tagger[2].

After pre-processing of the raw data, our input list then created. Example (6-a) to (6-e) show some example entries in this verb list:

(6) a. 安1_8(VH) @ 形容(Na) 以(P) 文字(Na) 的(DE) 方式(Na) 問候(VC) 他人(Nh) 的(DE) 問候語(Na)$

b. 按_1(VA) @ 以(P) 手(Na) 施力(VB) 使(VL) 人(Na) 或(Caa) 物(Na) 無法(D) 移動(VAC)$

c. 暗示_1(VA) @ 用(P) 間接(A) 的(DE) 方式(Na) 來(D) 表達(VC)$

d. 哀1_3(VH) @ 形容(Na) 會(D) 使(VL) 人(Na) 感到(VK) 悲傷(VH) 的(DE)$

e. 哀1_4(VC) @ 對(P) 他人(Nh) 產生(Nv) 關懷(Nv) 或(Caa) 同情心(Na)$

As example (6) shows, the @ symbol is used to separate the defined verb and its definition while $ is used for the ease of tokenization. From example (6-d) and (6-e), it can be noticed that verb entries are differentiated by senses but not word forms. The serial number is used to indicate the sense order of a lemma, for example, 哀1_3 in (6-d) represents the third sense of the first lemma of 哀.

[2]http://ckipsvr.iis.sinica.edu.tw

3.1.3 Syntactic Patterns in Chinese

Our first approach is basically based upon previous formulated ones that using syntactic patterns as indicators to find lexical semantic relations. As known, the structure of a lexical entry in a dictionary sometimes reflects the relatedness of words and concepts [21]. In this approach, two syntactic patterns were manually selected by examining pairs of verbs in known semantic relations:

Syntactic Pattern 1

Definitions of verbs for particular sense often refer to certain or specific manner of their hypernyms. Hence, the definition may appear in the lexical syntactic pattern that '*yi /yong /liyong... V_h... (by/with ... to V_h)*' where sentences or phrases followed by the target words 'yi/yong/liyong' are used to further illustrate certain manners. For example, in Chinese WordNet, the first sense definition of (VA) *zou*'to walk' is:

(7) 走 ：以 兩腿 交互 向前 移動
zou : Yi LiangTui JiaoHu XiangQian YiDong
To walk: moving forwards by two feet.

Our first pattern is, when a verb V_t's definition contains the pattern *'yi... V_h...'* or *'yong/liyong... V_h...'* in the sentence, we may take this verb V_h out or take out all verbs appearing after *yi* and *yong/liyong* if there is more than one. The verb(s) V_h could be labeled as the hypernym of V_t . Therefore, from the above example, we can label *yidong* 'to move' as the hypernym of the verb *zou* 'to walk'.

Syntactic Pattern 2

Deriving from Fellbaum's formula of troponymy [17]: *to V_1 is to V_2 in some particular manner*, we may assume that this 'particular manner' can be represented by an adverbial expression in Chinese. For example, the definition sentence of a defined verb V_t contains the pattern *'X di V_h (to V_h X-ly)'* where X-ly can be seen as a specific manner modifying V_h. Since adverbial phrase in Chinese often appeared with the suffix 'di', we then extract the verb that follows this adverb as the hypernym. For example, in Chinese WordNet, the first sense of (VA) *saomiau* 'to scan' is:

(8) 掃瞄 : 在 特定 範圍 內 快速 地 看 過 一次
saomiau : Zai TeDing FanWei Nei KuaiSu Di Kan Guo YiCi
To scan : to look fast within a certain scope.

Therefore, according to our second syntactic pattern, we will extract the verb *kan* 'to see or to look' which occurred following the adverb *kuaishudi* 'fast' in the definition sentence as the hypernym of *saumiao* 'to scan'.

3.1.4 Procedure

We implemented the first approach by using Python(2.5.2)[3] as the programming language (Programming Code is shown in Appendix A). As mentioned, in this approach we aim to find out possible verb pairs in hypernymy-troponymy relation by syntactic patterns: *'yi/yong/liyong... V_j... (by/with...to V_j)'* and *'X di V_j (to V_j X-ly).'* Therefore, *yi* 'by', *yong* 'with/by/to use' and *di* 'suffix for adverb' served as the target words for programming. First, verb entries containing

[3]http://www.python.org/

target words would be extracted out. Next, all verbs[4] occurring after the target words would be extracted again to serve as possible hypernyms. Therefore, taking example (6) as instant input, the program returned the results as shown in (9).

(9) a. 安1_8(VH) @ 問候(VC)

b. 按_1(VC) @ 施力(VB) 移動(VAC)

c. 暗示_1(VE) @ 表達(VC)

From input verb list (6-a) to (6-c), the program returned the results (9-a) to (9-c) where the first verbs are the troponyms and verbs after '@' symbol are their possible hypernyms. For example (6-d) and (6-e), the program will not returned any results for these entries do not contain the target words. It may be intuitively noticed that extracting *all verbs* that occurring after the target words would cause some errors. In fact, it seems an unavoidablely technical problem due to the syntactic structure of sentences. This type of error will be analyzed in detail in the following chapter. Lexical syntactic pattern-based approach can be summarized by the following algorithm:

[4]CWN subdivided verbs into 15 different subcategories including VA VAC VC VB VCL VD VE VF VG VH VHC VI VJ VK VL

Input: Verb sense list with definitions V_1, V_2, ... and V_n

Output: Predicted hypernyms of the input defined verbs

foreach *definition of verbs* V_{df} **do**
    Word segmentation and POS tagging via CKIP;

    foreach *definition of input verbs* V_i **do**
        check whether they contain the lexical syntactic pattern one;

        if *matched* **then**
            label the verb(s) V_j as a hypernym of V_i;

        end

    end

end

while *Unscheduled tasks remaining* **do**

    foreach *definition of verbs* V_{df} **do**
        check whether they contain the lexical syntactic pattern two;

        if *matched* **then**
            label the nominalized verb W_j as a hypernym of V_i ;

        end

    end

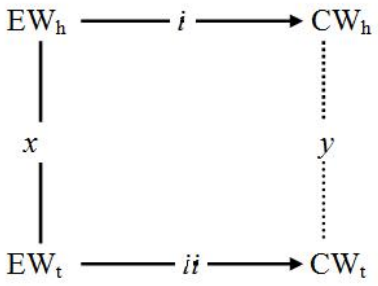
end

Algorithm 1: Algorithm to automatic labeling of hypernymy-troponymy relation

3.2 Bootstrapping Approach

It is tacit that Lexical Semantic Relations(LSRs) are universal [45] [16], that is, LSRs will hold true across different languages; hence, it can be plausibly postulated that LSRs can be transported and acquired by bootstrapping. In [24] [25], the validity of cross-lingual LSRs inferences was examined by bootstrapping a Chinese WordNet with Princeton WordNet. The results

verified that LSRs can be acquired by mapping from an existing WordNet through translational equivalent relation. Having positive outcomes in previous studies, we then adopted bootstrapping as an approach and expand the source data to a larger scope in order to find out Chinese hypernymy-troponymy verb pairs as more as possible. Unlike the bootstrapping approach used in [25] and [24] where LSRs are predicted by starting from a Chinese lemma linking to English WordNet synset, we started from targeting at certain LSRs in the excising WordNet (here, the hypernymy-troponymy relation), extracting English verb pairs in this relation and predicted that Chinese verb pairs can be acquired through translation equivalent relation. Figure 3.1 represents a general model of bootstrapping approach used in this thesis:



x = EW$_h$ - EW$_t$ Hypernymy-troponymy relation in English
y = CW$_h$ - CW$_t$ Inferred hypernymy-troponymy relation in Chinese
i = EW$_h$ - CW$_h$ Equivalent translation relation
ii= EW$_t$ - CW$_t$ Equivalent translation relation

Figure 3.1: Bootstrapping model

X represents our starting LSR in Princeton WordNet. By targeting at a certain relation, verb pairs which are known in the relation *x* can be extracted. Here, a verb pair EW$_h$-EW$_t$ is in hypernymy-troponymy relation where EW$_h$ stands for English hypernym and EW$_t$ for troponym. Assume that LSRs, including hypernymy-troponymy relation, can be transported cross-lingually, then *y* is the inferred relation transported from *x*, that is, LSR *y* is the inferred

hypernymy-troponymy relation in Chinese. Accordingly, when EW_h and EW_t are translated into Chinese with equivalent translation relations *i* and *ii*, a Chinese verb pair CW_h / CW_t which is in hypernymy-troponymy relation is then acquired.

3.2.1 Data Source

Data source used in this approach is different from the database (CWN) used in syntactic pattern-based approach. Since bootstrapping relies on transporting an existing wordnet database to another through equivalent translation, we then resorted to Princeton WordNet and the Sinica Bow as our data sources, especially depended on the hypernymy-troponymy relations in WordNet and English-Chinese Translation Equivalent Database(ECTED) in Sinica Bow.

Extracting WordNet relation

In WordNet, LSRs including hypernymy-troponymy relation are defined between synsets, that is, between senses of words. In this approach, we employed the latest version of WordNet (WordNet 3.0)[5] to extract verb synsets that are in hypernymy-troponymy relation. WordNet contains 13767 synsets of verbs in total, within which are 3315 synsets pairs having hypernymy-troponymy relation. A pair of verb synset which is in hypernymy-troponymy relation is shown as follow:

(10) Synset (‘exhale’%2:29:00::) ⇒ Synset (‘blow’%2:29:00::), Synset (‘snort’%2:29:00::)

As can be seen from example (10), synset to the left of the arrow is the hypernym while synsets to the right are the troponyms. Therefore, ‘exhale’ has two troponyms ‘blow’ and ‘snort’.

[5]http://wordnet.princeton.edu/

Note that hypernymy and troponymy are separated relations in WordNet; hence, we anchored hypernymous synsets and returned all the troponyms. The numerical expression follows the % symbol is the sense key of each synset. We extracted each sense key for the sake of further mapping with equivalent translation. Apparently, a synset will carry more than one troponyms and it is needed to split them into one-to-one correspondence as the following example shows:

(11) a. Synset ('exhale'%2:29:00::) ⇒ Synset ('blow'%2:29:00::)

b. Synset ('exhale'%2:29:00::) ⇒ Synset ('snort'%2:29:00::)

After all synset pairs were splited, a total of 13239 synset pairs that are in hypernymy-troponymy relation was yielded and served as our data source waiting for further mapping with Chinese equivalent translations.

ECTED

As mentioned in section 2.1.3, Sinica Bow is constructed by integrating WordNet, English-Chinese Translation Equivalence Database (ECTED), and SUMO. In this study, we extracted ECTED from Sinica Bow for it contains translation equivalent pairs based on a WordNet synset (WordNet version 1.6). Each translation was created manually. Most important of all, the translation was based on senses but not merely word forms which makes it different from any English-Chinese dictionary resources. The database includes synsets in WordNet, together with their synset id., and the equivalent translation in Chinese. We extracted out all equivalent translations of verb synsets which are 12127 in total. The following example shows the equivalent translation of verb synset 'exhale':

Synset id.	WordNet synset	Equivalent translation in Chinese
0003142V	exhale,expire,breathe_out	呼氣(HuChi), 吐氣(TuChi)

3.2.2 Procedure

Data source was processed by Python 2.5.2. With the help of Natural Language Toolkit(NLTK) version 0.9.8[6], WordNet relations can be easily processed, counted, and extracted (Programming Code are shown in Appendix A). Bootstrapping, in short, aims to transport an existing WordNet to another one. In this study, the main idea of this approach is to map English hypernymy-troponymy verb pairs to Chinese through equivalent translations. The mapping procedure first imports hypernymy-troponymy verb pairs extracted from WordNet, which are 13239 verb synset pairs in total, then performs a mapping process with ECTED. The two data sources were mapped by sense key, or synset id.[7], but not word forms. This prevents the errors of wrong mapping of polysemous verbs. Recall that in ECTED there are totally 12127 verb synset pairs translated from English into Chinese. The number is a bit lower than those extracted from WordNet. This discrepancy is mainly due to the different version of WordNet: we extracted LSR from WordNet 3.0[8] but the ECTED was created based on WordNet 1.6. The difference infers the truth that there must be some verb pairs failing to be mapped to Chinese. However, since the difference is rather small (1112) compare to overall synsets number (12127) and it is unworkable to find out the diversity in ECTED, this discrepancy will be ignored in the approach. Here, we limit our examination to the synsets that can find Chinese equivalent translation in ECTED.

[6]http://www.nltk.org/

[7]Synset id. is used in WordNet 1.6 while sense key is used in WordNet 3.0. Both of these representations refer to the specific identification of synset in WordNet. A pre-processing program was used to unify these two different expressions.

[8]The built-in WordNet version in NLTK(0.9.8) is 3.0

As mentioned, due to the different WordNet version used in data sources, verb synsets are not equal in these two data. Synsets pairs extracted from WordNet is more than those in ECTED, indicating that newly- added synsets in WordNet 3.0 were not included in ECTED. Under this situation, we skipped the synsets that can not find correspondent equivalent translation. After the mapping was done, a total of 11289 verb pairs were mapped to Chinese. Example (12) shows the results from bootstrapping; as shown, verbs to the left of the arrow is the hypernym, and verbs to the right are troponyms. And Figure 3.2 shows the overall procedure of bootstrapping approach:

(12) a. 呼氣(HuChi, 'exhale') ⇒ 吹(Chui, 'blow')

b. 呼氣(Huchi, 'exhale') ⇒ 噴鼻息(PenBiShi, 'snort')

3.3 Evaluation and Scoring

Both the two approaches return possible Chinese hypernymy-troponymy verb pairs and these results will be evaluated by a similar criterion. However, due to the different data sources (CWN in syntactic pattern-based approach and WordNet/ECTED in bootstrapping approach) used in these two approaches, the results should be calculated separately. Owing to the lack of ground truth data in the real world, that is to say, there is no database showing semantic relations of verbs available in the real world, we redefined the way in calculating the accuracy rate in this thesis. The following subsections introduce the criterion for evaluation and scoring.

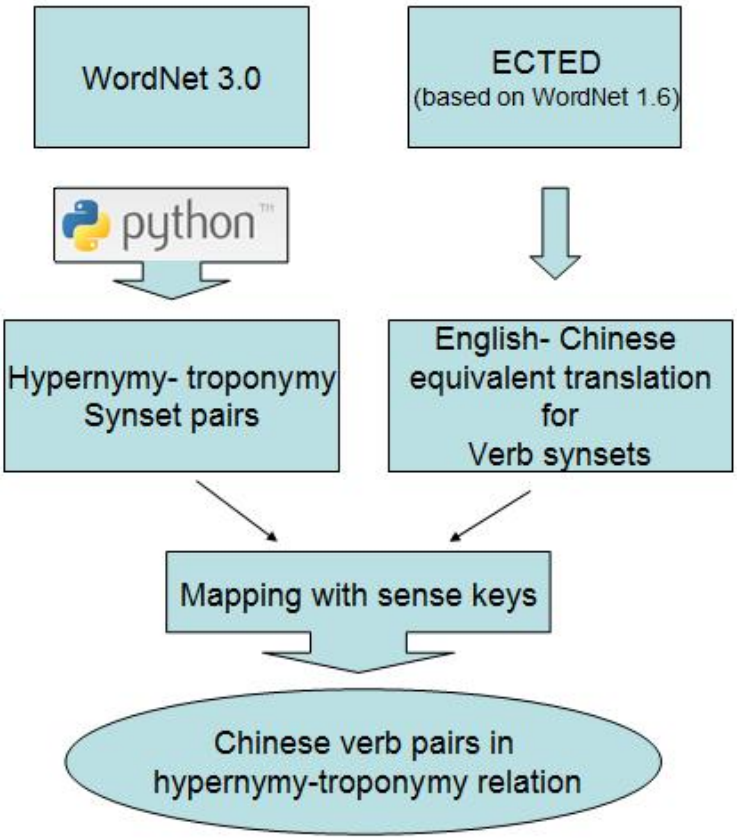


Figure 3.2: Overall procedure of bootstrapping approach

3.3.1 Evaluation

After the automatic extraction was done, human labor was used to inspect the correctness of the outcomes. There are totally four people who are all native to Chinese with linguistic background spending about three weeks evaluating the returned results from both approaches. Afterward, all the evaluated results will be double-checked again by the author. The criterion for evaluating and scoring the results will be illustrated in the following. Since there is not yet a Gold Standard of hypernymy-troponymy relation nor a WordNet-like database in Chinese, we adapted substitution test for evaluating the results. Substitution test is commonly used in linguistic literature [43]; EuroWordNet provided linguistic tests for each semantic relation to examine the validity [4]. In Tsai et.al [43], sentence formulae were created following the frame

in EuroWordNet to examine the validity of certain semantic relations in Chinese. Linguistic semantic tests help researcher check if two word meanings have a certain kind of semantic relation or not, and further ensure the quality and consistence of the database. Therefore, following the previous framework, a set of sentence formulae based on properties of troponymy was created to verify the correctness of hypernymy-troponymy verb pairs.

According to Fellbaum's detailed analysis of verb relations [17], it is known that troponymy is a kind of entailment and is characterized by the properties of **temporal inclusion** and **co-extensiveness**. **Temporal inclusion** distinguishes troponymy from other kinds of entailment like backward-presupposition and cause relation. **Co-extensiveness**, on the other hand, differentiates troponymy from verb pair such as *snore/sleep* which also involves temporal inclusion. Due to the above properties of troponymy, the following sentence formulae are proposed to test the validity of semantic relation and evaluate the results found by our algorithm.

Sentence Formulae for Hypernymy-Troponymy Relation

If a pair of verb (V_h, V_t) is in hypernymy-troponymy relation, then:

(i) V_h is the hypernym of V_t , V_h has the superordinate conception of V_t.

(ii) V_t is the troponym of V_h, V_t inherits properties of V_h, and is specified in some certain manners.

Sentence Formulae:

S_1(V_h, V_t) = V_t是一種V_h的方式。

V_t Shi Yi Zhong V_h De FanShi.

'V_t is a way of V_h.'

S_2(V_h, V_t) = 如果他(正在)V_t，那麼他同時一定[在/會]V_h。

RuGuo Ta ZhengZai V_t-ing, NaMa Ta TongShi YiDingZai V_h-ing

'If he is V_t-ing, then he must be V_h-ing at the same time.'

S_3(V_h, V_t) = 如果他(正在)V_h，那麼他同時一定[在/會]V_t。*

RuGuo Ta ZhengZai V_h-ing, NaMa Ta TongShi YiDingZai V_t-ing

'If he is V_h-ing, then he must be V_t-ing at the same time.'*

S_1 is derived from the formula of troponymy defined by Fellbaum in [31]: *to* V_1 *is to* V_2 *in some particular manner*. However, only S_1 is not enough to verify troponymy, for example, verb pair (*ChengGong* 'succeed'/*ChangShi* 'try') is not in hypernymy-troponymy relation but it seems grammatical when they fit into S_1: *ChangShi Shi YiZhong ChengGong De FangShi* 'Trying is a way of succeeding.' Therefore, testing sentences need to cover other properties of troponymy. S_2, hence, examine whether a verb pair has the properties of temporal inclusion and co-extensiveness. S_3 is counter to S_2; the reason of using S_3 is to avoid synonym pairs. Take two synonym verbs (*zou* 'to walk'/ *bushing* 'to go on foot') for example, there is no grammatical problems when substituting these verbs into S_2 and S_3: *RuGuo Ta ZhengZai Zou，NaMa Ta TongShi YiDing Zai BuShing* 'If he is walking, then he must be going on foot at the same time', *RuGuo Ta ZhengZai BuShin, NaMa Ta TongShi YiDing Zai Zou* 'If he is going on foot, then he must be walking at the same time'. Therefore, the final condition for troponymy to hold is that testing by S_3 must be WRONG. The rules of evaluating hypernymy-troponymy pairs can be summarized as following rules:

Rule

For each verb pair (V_h, V_t), if:

S_1(V_h, V_t) and S_2(V_h, V_t) are grammatical and meaningful.

And:

$S_3(V_h, V_t)$ is ungrammatical and makes no sense.

Then:

Verb pair (V_h, V_t) is in hypernym-troponymy relation.

Example 1:

(V_h, V_t) = (*yidong* 'to move', *zou* 'to walk')

$S_1(V_h, V_t)$ = Zou Shi YiZhong YiDong De Fang Shi

'Walking is a way of moving'

$S_2(V_h, V_t)$ = RuGuo Ta ZhengZai Zou, NaMa Ta TongShi YiDing Zai YiDong

'If he is walking, then he must be moving at the same time.'

$S_3(V_h, V_t)$ = RuGuo Ta ZhengZai YiDong, NaMa Ta TongShi YiDing Zai Zou*

'If he is moving, then he must be walking at the same time.'*

The above example shows that $S_1(V_h, V_t)$, $S_2(V_h, V_t)$ are grammatical and $S_3(V_h, V_t)$ is not because if someone is moving, he could also be running, jumping, sliding, etc. Verb pair (*YiDong*'to move', *Zou* 'to walk)' satisfies all the testing sentences therefore they are in hypernymy-troponymy relation.

In order to prove that the sentence formulae do have the validity of distinction, we take another verb pair which does not in hypernymy-troponymy relation for example.

Example 2:

(V_h, V_t) = (*ChengKung* 'succeed', *ChangShi* 'try')

$S_1(V_h, V_t)$ = ChangShi Shi YiZhong ChenKung De FangFa

'Trying is a way of succeeding.'

$S_2(V_h, V_t)$ = RuGuo Ta ZhengZai ChangShi, NaMa Ta TongShi YiDing Hui ChengKung*

'If he is trying, then he must succeed at the same time.'*

$S_3(V_h, V_t)$ = RuGuo Ta ChengKung，NaMa Ta TongShi YiDing Zai ChangShi*

'If he succeeds, then he is trying at the same time.'*

As example sentences show, verb pair (*ChengKung* 'to succeed, *ChangShi* 'to try') does not satisfy the rule for that $S_2(V_h, V_t)$ is non-factive and thus ungrammatical.

3.3.2 Scoring

Results returned by each approach were analyzed separately. From syntactic pattern-based approach, the program returned a list of verb pairs which are possible in hypernymy-troponymy relation. For scoring the results, the correct ones were calculated by filtering through testing sentences. The input verb and its correct hypernyms together make up a hypernymy-troponymy pair. As mentioned, one verb can have more than one hypernym, under this situation, it is scored as different hypernymy-troponymy pairs. For example, when 束1_1 (Shu, 'to gird') carries the sense of "encircle or bind", two hypernyms were returned by the program: 纏繞 (ChanRao, 'to encircle') and 固定 (GuDing, 'to fix'). Therefore, the results were calculated as two separate hypernymy-troponymy pairs: 束1_1 (Shu, 'to gird') / 纏繞 (ChanRao, 'to encircle') and 束1_1 (Shu, 'to gird') / 固定 (GuDing, 'to fix'). As for the results from bootstrapping approach, there is only one-to-one correspondence will be returned. Therefore, each possible hypernymy-troponymy pairs will be evaluated and calculated.

After all the results were evaluated, the accuracy was calculated. Instead of computing the precision and recall rate, we reckoned the accuracy rate since there is not yet a ground truth database existing in the real world. That is, there is not yet an available database showing all the hypernymy-troponymy verb pairs. Note that our approaches relied on two different

data sources–CWN for syntactic pattern-based approach and Princeton WordNet for bootstrapping. For syntactic pattern-based approach, we aim to semi-automatically find out hypernymy-troponymy verb pairs as more as possible from CWN. Similarly, bootstrapping approach extracts *all* synsets pairs in our target relation from WordNet, which can be seen as a standard database. Owing to the lack of Gold Standard in Chinese, we can only calculate proportion of correct results retrieved from our program. In order to make a distinction from *precision/recall rate*, we termed our statistical representation as *accuracy rate* which is redefined as follow:

$$\text{Accuracy rate} = \frac{Numberofcorrectwordsretrievedbyprogram}{Numberoftotalretrievedwords}$$

With regard to the returned results that are not in our target relation, a qualitative analysis was conducted so as to investigate the causations lead to the errors.

3.4 Summary

Chapter Three presents two approaches adopted in this thesis aiming to extract verb pairs in hypernymy-troponymy relation. Lexical syntactic pattern-based method is commonly used in finding semantic relations because the structure of a sentence can reflects the relatedness of words and concepts. Bootstrapping approach, on the other hand, aims to exploit an already existing resource and combining them within a common, standard framework. These two proposed approaches were processed parallely for that there is methodologically and experimentally no direct connection between them. The results will be analyzed and calculated separately as well. In-depth results will be analyzed in the following chapter.

Chapter 4

Results and Error Analyses

This chapter reports the results of syntactic pattern-based and bootstrapping approach, along with quantitative evaluation of the results and detailed error analyses. Section 4.1 displays the results returned by the first approach and discusses some types of semantic and syntactic problems that lead to the errors. Section 4.2 examines the results of bootstrapping approach and discusses some problems observed from the results. Although the thesis presents two parallel approaches in finding hypernymy-troponymy verb pairs in Chinese, an integrated comparison and discussion is needed. Thus, Section 4.3 dedicated to the discussion of the two approaches used in the study. Finally, a general summary of this chapter is given in the last section.

4.1 Results from Syntactic Pattern- based Approach

This section displays the results returned by syntactic pattern -based approach. As mentioned, two syntactic patterns in Chinese were manually selected to extract hypernymy-troponymy relation. Table 4.1 shows a general results and accuracy rate of the approach (correct returned pairs are shown in Appendix B).

As table 4.1 shows, the first column displays the number of verb entries that are in ac-

	Verb entry	Correct troponymy pairs retrieved	Incorrect troponymy pairs retrieved	Total troponymy pairs retrieved	Accuracy Rate
Pattern 1: *yi/yong/liyong/...V*	1058	828	590	1418	58.39 %
Pattern 2: *...Di V*	174	161	50	211	76.30 %
Total	1232	989	640	1629	60.71 %

Table 4.1: General results of syntactic pattern-based approach

cordance with the syntactic patterns. Next, from these verb entries, the program extracted possible hypernymy-troponymy pairs, of which correct and incorrect ones were calculate and shown in the second and the third column respectively. For syntactic pattern 1, there were totally 1058 verb entries being extracted from the input verb list (10388 entries in total). Within the 1058 verb entries, which contain target words *yi*, *yong* or *liyong*, a total of 1418 possible hypernymy-troponymy verb pairs was found, of which were 828 correct pairs and 590 incorrect ones, leading to a 58.39% accuracy rate. In terms of syntactic pattern 2, there were only 174 verb entries out of 10388 fitting in with the pattern, and from this 174 verb entries, a total of 211 possible hypernymy-troponymy pairs were extracted, of which included 161 correct ones and 50 incorrect ones resulting in 76.30% accuracy rate. As for the overall results, there were totally 1629 possible hypernymy-troponymy pairs being extracted from 1232 verb entries, averaging 1.32 hypernyms per input verbs, indicating that for each input verb entry, the program extracted more than one possible hypernyms. Nevertheless, the precision rate shows how the program performs in this experiment. From table 4.1 we see that there were totally 989 correct hypernymy-troponymy pairs and 640 incorrect pairs being extracted, leading to a 60.71% ac-

curacy rate.

A number of facts can be deduced from the foregoing statistic. First is that verb entries conforming to the lexical syntactic patterns are comparatively few. We use only two syntactic patterns to extract verb relation, this will first filter out most entries in the input list. Only 1232 verb entries out of 10388 input entries were found to comply with syntactic pattern 1 and 2, indicating that there is only a few proportion of verb entries from which the program can extract possible hypernymy-troponymy relation. Another generalization relates to the performance of each syntactic patterns. As can be seen in Table 4.1, syntactic pattern 1 identified a lot more verb entries than pattern 2, however, in terms of the accuracy rate, syntactic pattern 2 performs better (76.30%) than pattern 1 (58.39%). Although there is only a small portion of possible hypernymy-troponymy verb pairs extracted by syntactic pattern 2 (compared to 1418 possible pairs extracted by pattern 1), the accuracy rate is rather high. This presents that syntactic pattern 2 is more stable and reliable in extracting hypernymy-troponymy relation. This may due to the syntactic and semantic problems of pattern 1, for example, the ambiguity of one of the target words *yi*. More detailed analyses will be described in the following section.

4.1.1 Error Analyses

By observing the results returned by our program, unwanted or incorrect answers were found to cause the decrease of the performance. A total of 640 incorrect hypernymy-troponymy verb pairs (namely, 640 incorrect hypernyms) were found within the results returned by syntactic pattern 1 (590 incorrect hypernyms) and pattern 2 (50 incorrect hypernyms). This subsection probes into different types of errors which are generalized as follows:

- Type 1. The ambiguity of *yi*
- Type 2. Errors caused by syntactic structure
- Type 3. Verb and verb phrase
- Type 4. The problem of synonym or near-synonym
- Type 5. Abstract concept verbs
- Type 6. Wrong tagging of POS

Each error type will be discussed in detail in the following subsections and the percentage of each error type was listed in Table 4.2:

Error type	Incorrect Hypernyms from Pattern 1	Incorrect Hypernyms from Pattern 2	Total Incorrect Hypernyms	Error Rate
Type 1	194	0	194	30.31%
Type 2	215	22	237	37.03%
Type 3	71	8	79	12.34%
Type 4	56	17	73	11.40%
Type 5	38	1	39	6.09%
Type 6	16	2	18	2.81%
total	590	50	640	100%

Table 4.2: Error types and percentage

Type 1. The ambiguity of *yi*

Recall that the first syntactic pattern used in the first approach is *'yi/yong/liyong... V_h... '(by/with...to V_h)* for it satisfied the definition of troponym— to V_t is to V_h by a specific way or with a particular

manner where V_t is the troponym of V_h. In Chinese, it could be expressed by the target words *yi* or *yong*. Therefore, hypernymy-troponymy pairs were expected to be found by first extracting sentences containing the target words. Unfortunately, errors occurred when extracting the target *yi* because not every *yi* extracted by our program is the correct ones. This is due to the inability of WSD (Word Sense Disambiguation). According to the classification and definition in CWN, *yi* serves as both preposition (P) and conjunction (Cbb). In this approach, only the preposition *yi* will be extracted. Nevertheless, *yi* is still ambiguous due to its multi-senseness. When *yi* serves as a preposition, there are about ten different senses classified by CWN including *by*, *in order to*, *dependant on*, *according to*, etc. In this approach, it is expected that the preposition *yi* is used as 'by...manner' or 'with...manner' as shown in the following example:

(1) 保存_3(VC) @ 以(P) 原本(A) 狀態(Na) 收藏(VC)
BaoChun @ Yi YuanBen ZhuangTai ShouCang
to keep @ To store (something) by (their) original state.

(2) 走_3(VA) @ 以(P) 兩(Neu) 脚(Nf) 交互(D) 向(P) 前(Ncd) 移動(VAC)
Zou @ Yi Liang Jiao JiaoHu Xiang Qian YiDong
to walk @ moving forwards by two feet.

In example (1), *BaoChun* 'to keep' can be seen as a way of *ShouChang* 'to store/to stock', elaborating by a certain manner– to store by something's original state. Therefore, we obtain a hypernymy-troponymy pair: (*ShouChang* 'to store'/ *BaoChun* 'to keep'). By the same token, we get another pair (*YiDong* 'to move'/ *Zou* 'to walk') from example (2).

However, not every *yi* that found by the program carries the meaning we want. As mentioned, *yi* is a multi-sense preposition which carries about ten different senses. From the results, it is observed that *yi* is also wildly used as 'in order to (V_j)'. Under this situation, *yi* is found to be failed in indicating hypernymy-troponymy relation. The following examples contain the

preposition *yi* with the sense of 'in order to' within the definition of each verb entry:

(3) 測_3(VC) @ 考驗(Na) 後述(Na) 對像(Na) 以(P) *發現(VE) 事實(Na)
Ce @ KaoYan HouShu DuiXiang Yi *FaXian ShiShi.
to surmise @ testing someone in order to discover the truth.

(4) 施肥_1 @ 給予(VD) 肥料(Na) 以(P) *促進(VK) 植物(Na) *生長(VH)
ShiFei @ GeiYu FeiLiao Yi *CuJin ZhiWu *ShengChang
to fertilize @ fertilize plants in order to promote the growth.

In (3) and (4), none of the verbs that occur after *yi* is the correct hypernyms of the defined verb, namely, *FaShien* 'to discover' and *Ce* 'to surmise' is not a hypernymy-troponymy verb pair and either is (*ChuJing* 'to promote'/*ShiFei* 'to grow') nor (*ShengZhang* 'to grow'/ *ShiFei* 'to fertilize'). Therefore, When *yi* carries the meaning 'in order to V_j', it is failed to extract V_j as the hypernym. As Table 4.2 displays, errors caused by the ambiguity of *yi* yielded 194 tokens out of total 640 incorrect results, result in 30.31% error rate which is comparatively high. Also, ambiguity does not occur in other target words such as *yong* 'by/with' in pattern 1 and *di* 'adverbial suffix' from syntactic pattern 2 for these words carries only one sense under certain POS, that is, there is only one sense of *yong* when it serves as a preposition and so is *di* when it serves as an adverb.

Word Sense Disambiguation has long been an open central problem at the lexical level in NLP, ambiguous words can lead to irrelevant or unwanted information retrieval just as the ambiguous preposition *yi* shows here. Many approaches or disambiguators have been proposed to solve problems caused by ambiguity [36] , while many approaches were used to disambiguate words in English, it is also found by researchers that features that are important for disambiguation in English is not the same for that in Chinese [12]. For example, parse, predicate-argument

information and selectional restriction play important role for disambiguation in English but rather minor in Chinese [12] [47]. When the ambiguity occurs at the very basic level—the lexical ambiguity, it seems even harder to disambiguate the polysemous senses automatically, at least in the present approach, sense determination has to be done manually.

Type 2. Errors caused by syntactic structure

The second error type is caused by problems related to syntactic structures. As shown in Table 4.2, a total of 237 out of 640 incorrect hypernyms belong to this type, ranking the highest with a percentage of 37.03%. Both Pattern 1 and Pattern 2 involved in this type of error. As mentioned, all verbs occurring after the target words will be extracted as the possible hypernyms; however, a lot more unwanted results were extracted as well and consequently lowered the accuracy rate. Extracting *all* verbs implies the regardlessness of the syntactic structure, or to be more precise, the constituency of the sentences. This can be illustrated by the following examples:

(5) 打1_25(VC) @ 以(P) *彈(VC) 棉花(Na) 的(DE) 方式(Na) 製造出(VC) 棉被(Na)
Da @ Yi *Tan MianHua De FangShi ZhiZaoChu MianBei
to pluck @ producing a bed quilt by flipping the cotton.

(6) 釣1_1(VC) @ 用(P) *連結(VC) 著(Di) 鉤子(Na) 的(DE) 竿子(Na) 或(Caa) 長線(Na)
Diao @ Yong *LianJie Zhao GouZi De GanZi Huo ChangXian
捕捉(VC) 水(Na) 中(Ng) 生物(Na)
BuZhuo Shui Zhong ShengWu
to fish @ To catch creatures in the water by using a pole with hook and fishing line.

Taking example (5) and (6) as inputs, the program extracted all verbs occurring after target words as candidates and hence returned possible hypernymy-troponymy pairs as follows:

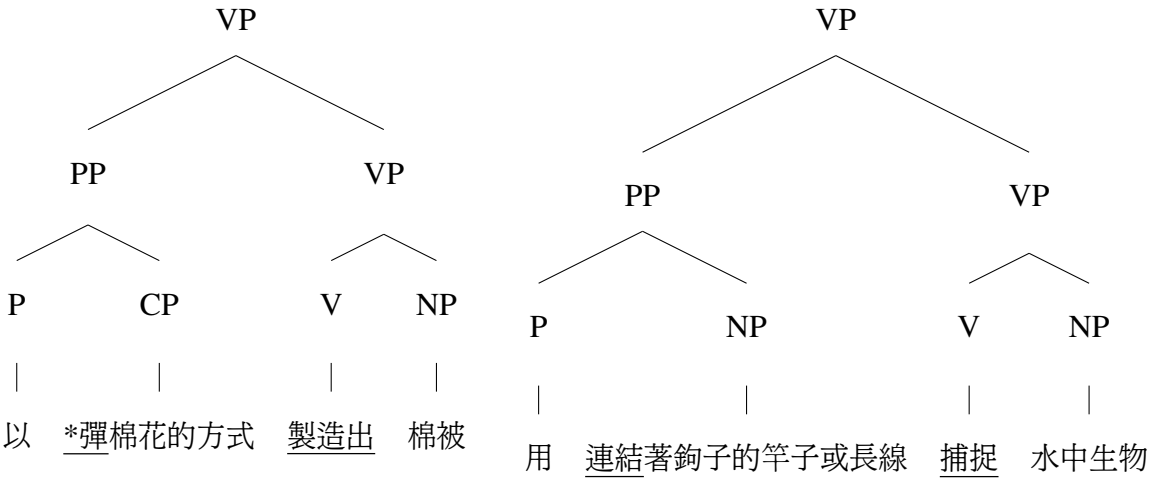
(i) *打1_25(VC)/ 彈(VC) *(*Da* 'to pluck'/*Tan* 'to flip')

(ii) 打_25(VC)/ 製造出(VC)　　(*Da* ‘to pluck’/*ZhiZaoChu* ‘to produce’)

(iii) *釣1_1(VC)/ 連結(VC)　　*(*Diao* ‘to fish’/*LianJei* ‘to connect’)

(iv) 釣1_1(VC)/ 捕捉(VC)　　(*Diao* ‘to fish’/ *BuZhou* ‘to catch’)

The wrongly extracted hypernyms (labelled with *) were caused by the problem of syntactic structure. The following tree diagrams illustrate the boundaries of each phrase. The verb *Tan* ‘to flip’ belongs to the preposition phrase that headed by *yi*(P), and the other verb *ZhiZaoChu* ‘to produce’ belongs to the head of the entire verb phrase which is the proper slot for the predicted hypernym to appear. By the same token, one of the verbs in example (6), *LienJie* ‘to connect’, is syntactically embedded into a noun phrase and consequently leads to an incorrect prediction.



This problem involves in sentence parsing during the pre-processing stage. In this thesis, input data were processed only by segmentation and POS tagging, the parsing part was left undone in order to limit the required time and resources for processing the input data. What is more, although syntactic structure is known to raise errors or problems in automatic extraction of semantic relations, it is not rigid or regular enough to generate rules without human decision. However, this issue can be left for future works by using parsers in Chinese such as

Sinica Treebank[1] which can do sentence parsing and automatic semantic role assignment for structured trees [8].

Type 3. Verb and verb phrase

A total of 73 (12.34%) out of 640 errors belongs to this error type. This type of error shows that hypernymy-troponymy relation sometimes lies between verbs and verb phrases but not between verbs themselves. To be more precise, the extracted verb alone can not indicate the existence of hypernymy-troponymy relation, rather, it is the whole verb phrase that demonstrate the target relation. The following are two examples:

(7) 標示_1(VC) @ 以(P) 文字(Na) 或(Caa) 記號(Na) *提供(VC) 訊息(Na)
BiaoShi @ Yi WeiZi Huo JiHao *TiGong XunXi
to label @ Provide messages by words or symbols.

(8) 敬禮_1(VB) @ 以(P) 動作(Na) *表示(VE) 敬意(Na), 降低(Nv) 頭部(Nc) 高度(Na)
JinLi @ Yi DongZuo *BiaoShi JingYi JiangDi TouBu GaoDu
爲(VG) 典型(Na)
Wei DianShing
to salute @ showing respect by movement, typically by lower one's head.

In example (7) and (8), both verbs after *yi* are not the hypernyms of the input verbs for they can not grammaticize or make sense the evaluating sentences; therefore, *TiGon* 'to provide' is not the hypernym of *BiaoShi* 'to label' and *BiaoShi* 'to show' is not the hypernym of *JinLi* 'to salute', either. If we scrutinize the results and their sentences, we can find that the incorrect hypernyms found by the program are not totally irrelevant with the input verb. Rather, the whole verb phrase is more reasonable to indicate a hypernymy-troponymy relation with the

[1] http://rocling.iis.sinica.edu.tw/CKIP/treebank.htm

input verb. Therefore, although *TiGon* 'to provide' alone is not valid enough to indicate a troponym relation with *BiaoShi* 'to label', yet the whole verb phrase *TiGon*(VE)+ *ShunShi*(Na) 'to provide message' can be seen as the hypernym of *BiaoShi* 'to label' for that labeling is a way of providing messages. And this is also true for *JinLi* 'to salute' and *BiaoShi*(VE)+ *JinYi*(Na) 'to show respect' in example (8). Despite that there might exist a semantic relation between verbs and verb phrases, it is not considered as hypernymy-troponymy pairs in our experiment, since lexical semantic relations arise within lexemes but not phrases. At least, we limit the semantic relation to *lexicalized* verbs which are segmented and tagged by a single POS in Chinese.

Type 4. The problem of synonym or near-synonym

It is common to define an unknown word by other synonymous words. By observing the results, it is found that synonymous verbs are used to define other verbs even when the definition sentences are in accordance with our target syntactic patterns. As example (9) shows, the definition sentence complies with our first syntactic pattern: '*yi*...to V_j', but the verb *HuaChu* 'to paint', which occurs after *yi*, is not the hypernym of *HuaTu* 'to paint'. Rather, *HuaChu* 'to paint' and *HuaTu* 'to paint' demonstrate a synonymous relation.

(9) 畫圖_1(VA) @ 以(P) 筆(Na) 或(Caa) 其他(Neqa) 工具(Na) *畫出(VC) 圖形(Na)
HuaTu @ Yi Bi Huo QiTa GongJu *HuaChu TuShing
或(Caa) 圖案(Na)
Huo TuAn
to paint @ Drawing figures or patterns by pens or other equipment.

(10) 拍3_1(VC) @ 用(P) 攝影(Na) 器材(Na) *拍照片(VB)
pai @ Yong SheYing QiCai *PaiZhaoPian
to photograph @ Taking photos by photographic apparatus.

Another example is shown in (10), one of the sense of *Pai* 'to photograph' is to take photos. The program extracted *PaiZhaoPien* 'to take photos' as the possible hypernym of *Pai* 'to photograph', notice that we ignore the problem of verb phrase here since *PaiZaoPien* 'to take photos' or *PaiZao* 'to take photos' have already been lexicalized in CWN or CKIP. Nevertheless, *Pai* 'to photograph' and *PaiZhaoPien* 'to take photos', under this sense, are in a synonymous relation rather than a troponymy relation. The problem of synonym also occurred in syntactic pattern 2 *'X Di V_j (to V_j X-ly)'* as shown in example (11) where *YaoDong* 'to shake' and *YaoHuang* 'to rock' may be intuitively judged as synonymous rather than in a hypernym-troponymy relation.

(11) 搖晃_1(VAC) @ 物體(Na) 不(D) 規律(Na) 地(DE) 來回(VCL) *搖動(VC)
YaoHuang @ WuTi Bu GuiLu Di LaiHui *YaoDong
to rock @ Objects shake back and force irregularly.

As table 4.1 displays, a total of 73 (11.40%) synonymous errors was found in the results. Although this type of error does not show a high percentage of error rate, yet this problem may not be easily solved for it involves in how definitions were made and described by lexicographers, the synonymy can only be filtered out manually.

Type 5. Abstract concept verbs

By observing the incorrect returned results, it is found that the appearance of certain verbs always leads to the error. These verbs are ***成/成爲***, *Cheng/ChengWei 'become'*, ***作/作爲***, *Zuo/ZuoWei 'be'* as the following examples show:

(12) 編_9(VA) @ 用(P) 已(D) 有的(Nh) 材料(Na) 或(Caa) 想法(Na) 創作(VA)
Bien @ Yong Yi YiuDe ChaiLiao Huo ShiangFa ChungZuo
*成(VG) 後述(Na) 作品(Na)
*Cheng HouShu ZuoPing

to arrange @ To create the following works by using already existing materials or thoughts.

(13) 設_2(VC) @ 以(P) 後述(Na) 陳述(Na) *爲(VG) 推測(VE) 依據(Na)
She @ Yi HouShu ChenShu ZuoWei TuiCe YiJu
to assume @ To make an assumption by the following statement.

In Chinese, verbs such as *Cheng 'become'* or *Wei 'be'*as shown in example (12) (13) function as copula-like verb where nouns or predicates are attached after them. Despite their grammatical function in Chinese, these verbs indeed, semantically carry very abstract concepts. To be more precise, these verbs are hard to be associated with concrete events, actions, or images. At least, it makes no sense when we substitute these pairs into the very basic definition of troponym: V_t *Shi YiZong* V_h *De FanShi* 'To V_t is to V_h in a particular manner'. Take (12) for example, '*She Shi Yi Zong Wei De FanShi' (? to assume is a way to be) makes no sense in Chinese. There are totally 39 returned verbs belong to this type of errors, accounted for 6.09% error rate.

Type 6. Wrong tagging of POS

As can be seen in Table 4.1, there are 18 errors out of 640 belong to the wrong tagging of POS, ranking the lowest with a percentage of 2.81% error rate. POS tagging is the process of assigning a part of speech or other lexical class marker to each word in corpus. Taggers play an increasing important role in speech recognition, natural language parsing and information retrieval. In the experiment, we use CKIP Word Segmentation tool to segment sentences and to tag POS. However, Chinese part-of-speech tagging is more difficult than its English counterpart because it needs to be solved together with the problem of unknown words and word segmentation. Also, Chinese part-of-speech classes are very ambiguous; many words can be

both adjective or noun, noun or verb without any change in morphology. There is no taggers or segmentators being proved one hundred percent accurate in Chinese owing to the above reasons [29]. Therefore, incorrect tagging is sometimes found in tagger:

(14) 拍攝_1(VC) @ 用(P) 攝影(Na) 器材(Na) *記錄(Na) 影像(Na)
PaiShe @ Yong SheYing QiCai *JiLu YingXiang
to photograph @ To record images with photographic apparatus.

(15) 確保_1(VE) @ 確實(VH) 地(DE) *保障(Na) 後述(Na) 對像(Na) 的(DE) 安全(Na) 或(Caa) 存在(Na)
QueBao @ QueShi Di *BaoZhang HouShu DuiXiang De AnQuan Huo ChunZai
to guarantee @ accurately indemnify the following subject's safety or existence.

In example (14), *JiLu* 'to record' can be served as both noun and verb according to CWN, obviously, *JiLu* in example (14) functions as a verb but is wrongly tagged as a noun in CKIP. Similarly, *BaoZhang* 'to indemnify' can function as a verb or as a noun (security) according to CWN. Apparently, it is the verb usage in example (15) but is wrongly tagged as noun (Na) through CKIP.

4.1.2 Interim Summary

Section 4.1 reports the results returned by lexico syntactic pattern-based approach. In general, this approach reached a 60.7% accuracy rate. For those incorrect results that decrease the performance, an in-depth investigation was made. In terms of the performance of each syntactic pattern, Pattern 2 was found to be more stable and reliable than Pattern 1. In total, six types of errors were generated by scrutinizing the incorrect results, with examples given above.

4.2 Results from Bootstrapping Approach

Bootstrapping approach returned totally 11289 Chinese verb pairs mapped from WordNet. The large amount of results were manually evaluated through testing sentences by a group of people[2] who are all native to Chinese with linguistic background. Not only correct hypernymy-troponymy verb pairs were calculated, several tags were also assigned to mark incorrect verb pairs. Evaluative tags are shown as follows:

- Hypernymy- troponymy: Returned Chinese verb pairs are in hypernymy-troponymy relation.
- Non-lexicalized verb: One or both of the verbs in returned pairs are not lexicalized in Chinese; that is, they may be expressed by phrases or sentences. But conceptually or semantically, the hypernymy- troponymy relation holds true.
- Other relation: Returned verb pairs stand in other lexical semantic relations.
- Unrelated: Returned pairs are unrelated to each other.

Table 4.3 displays a general result from bootstrapping approach. As can be seen, a total of 8305 verb pairs out of 11289 were marked hypernymy- troponymy relation, resulting in a 73.56% accuracy rate (correct returned pairs are shown in Appendix C). In Huang et al.' studies [24][25], a small portion of experimental data in Chinese (210 lemmas including N, V, Adv. and Adj.) was used to investigate how LSRs can be inferred by bootstrapping. In terms of LSRs inference for verbs, the precision for English-to-Chinese semantic relation inference is around 80%, within which 70% for hypernym inference and 82.4% for hyponymy inference. Much similar to previous studies, our results reach a 73.56% accuracy rate. With a much larger

[2]There are totally four people spending about two weeks on evaluating and tagging the returned results.

Tags	Number of verb pairs	Percentage
Hypernymy- troponymy	8305	73.56%
Non- lexicalized verb	893	7.91%
Other relation	700	6.20%
Unrelated	1391	12.32%
Total	11289	100%

Table 4.3: Overall results from bootstrapping approach

data source, the precision rate is a bit lower than the results in [24]. Previous studies focused on how and which LSRs can be inferred through bootstrapping. Only correct results and the feasibility of the approach were discussed in the studies. Without more in-depth analyses, some problems that cause the error were merely concluded by translational idiosyncrasies. In this thesis, incorrect results receive a scrutinized discussion. For returned verb pairs which can not be assigned a hypernymy- troponymy relation, they are roughly classified into three types of errors. Surprisingly, as Table 4.3 shows, 12.32% returned pairs were marked *Unrelated* which is the highest one within the three error types. Unlike *Non-lexicalized verb* and *Other relation* in which returned pairs are still stand in certain LSRs, *Unrelated* verb pairs show no semantic connections between two verbs. These results contradict our assumption and deserve an in-depth investigation. Therefore, in the following subsections, several causations that lead to the errors will be discussed.

4.2.1 Error Analyses

Evaluative tags including *Non-lexicalized verbs*, *Incorrect*, and *Other relation* were assigned to returned pairs which are not in hypernymy- troponymy relation.

Tags	Number of verb pairs	Percentage
Non- lexicalized verb	893	7.91%
Other relation	700	6.20%
Unrelated	1391	12.32%
Total	2984	26.43%

Table 4.4: Non hypernymy-troponymy verb pairs (Total number of returned verb pairs= 11289)

As Table 4.4 shows, there are totally 2984 verb pairs being divided into three types of errors. These three types of errors were further scrutinized in this subsection to investigate the reasons that decrease the performance, that is, what makes our returned verb pairs incorrect and what other relations could a returned pair stands in if they are not in our target relation. Generally speaking, errors arise from translational idiosyncrasies along with inaccuracy translations and conceptual differences cross languages. They will be illustrated as follows:

Non-lexicalized verbs

Without further illustration, a single word in one language often has meanings that require several words in another language to explain. By observing our results, it is noticed that many verbs, especially for troponyms, could not be described by a single lexeme in Chinese. Consider the following examples:

(16) a. anesthetize ⇒ cocainize

b. 麻醉 (MaZue) ⇒ 用古柯鹼麻醉 (Yong GuKeJian MaZue)

(17) a. laugh ⇒ chuckle

b. 笑 (Shiao) ⇒ 咯咯地笑 (GeGe Di Shiao)

Example (16-a) is a synset pair in WordNet standing in hypernymy-troponymy relation. This thesis hereafter uses the arrow to indicate troponymous relation of which the word to the left of the arrow is the hypernym and the word to the right is the troponym. Hence, under a certain sense in WordNet, 'anesthetize' is the hypernym of 'cocainize' or 'to cocainize' is a way to 'anesthetize'. After mapping through ECTED, the program returned Chinese verb pairs as (16-b) shows. Apparently, there is no single lexeme in Chinese can properly describe the troponym 'cocainize'; rather, it is expressed by phrase in which more specific manner (*Yong GuKeJian* 'use cocaine') is added to a general verb (*MaZue* 'to anesthetize'). Similarly, the troponym 'chuckle' in (17-a) can not be properly translated into one single lexeme in Chinese but has to be expressed by an adverbial phrase *GeGe Di Shiao* 'to laugh (chucklely)'. By observing the results which were tagged *Non-lexicalized verbs*, they are wildly found to be expressed by lexico-syntactic patterns *yi/yong*...V_h 'by/with (manner) to V_h' or ...*Di* V_h 'to V_h X-ly'. Interestingly, this conforms to our lexico-syntactic patterns used in previous approach for we assumed that troponyms are expressed by adding specified manners to more general verbs. This also explains why non-lexicalized verbs appear more often in the troponyms but not hypernyms. Translational discrepancy also appears in verbalized words, here is an example:

(18) a. prepare ⇒ summerize

b. 準備 (ZuenBei) ⇒ 爲夏天做準備 (Wei ShaTien Zuo ZuenBei)

English is full of inflectional information; syntactic categories can be flexibly changed and created by altering in morphology. As (18-a) shows, the troponym 'summerize' is verbalized from its noun 'summer', meaning to prepare for summer. Although verbalization is not restricted to English, yet in Chinese, verbalization requires several words to express as (18-b) shows. Foreign words also affect the preciseness in equivalent translation, especially revealing in terminology, jargon, or vernacular. Consider the following example:

(19) a. score ⇒ eagle

b. 得分 (DeFen) ⇒ 在打出低於標準桿兩桿的桿數 (Zai DaChu DiYu BiaoZuenGan Liang Gan De GanShu)

The troponym 'eagle' in (19-a) is used as a terminology in golf, meaning to shoot in two strokes under par, thus is a way to score. However, with no precise translation can be found in Chinese, it can not but to be expressed following the definition of its origin as can be seen in (19-b).
From a total number of 11289 verb pairs acquiring from bootstrapping approach, 893 pairs involve in this type of error, account for 7.91% of total as shown in Table 4.4. Comparatively speaking, the proportion is not the highest one; nevertheless, the problem of non-lexicalization might not be easily solved. This problem can be traced to the very beginning issue that there is no word delimiters in Chinese. Without blank to mark word boundaries, the distinction between words, morphemes, and lexicons are not easily defined. Note that all verb pairs marked by *Non-lexicalized verb* are conceptually or semantically standing in a hypernymy-troponymy relation except for the concepts are described by phrases or sentences. Lexicalization certainly yields problems especially in creating a WordNet-like database. Although in the construction of ECTED, each translated entries were expressed by lexicalized words rather than descrip-

tive phrases as possible as they can [25], yet our results suggest that there is still room for modification.

Other relations

Returned results sometimes were found to stand in other relations. We grouped these wrong pairs together because they are not unrelated to each other nor in our target relation. The most common semantic relation found within wrongly-acquired Chinese verb pairs is near-synonymy or synonymous relation. These can be illustrated by the following examples:

(20) a. disconnect ⇒ detach

b. 使分離 (ShiFenLi) ⇒ 使分開 (ShiFenKai)

(21) a. end ⇒ lapse

b. 結束 (JienShu)⇒ 終止 (ZhongZhi)

Under reasonable postulation, returned verb pairs should stand in hypernymy-troponymy relation since LSRs can be inferred by bootstrapping. However, the returned pairs such as (20-b) and (21-b) exhibit unexpected results in that both of these pairs are indeed in a near synonymous relation. If we look back to the glosses of their English counterparts in WordNet, we might discover the differences in conceptualizing events across languages. To be more specific, 'disconnect' in (20-a), according to its gloss in WordNet, means 'to make disconnected, disjoin or unfasten' while its troponym 'detach' carries the meaning that 'cause to become detached or separated'. Apparently, in English, 'to detach' can be seen as a way of making things disconnected although the distinction might be slight. However, this distinction disappeared when they are translated into Chinese. As (20-b) shows, both *ShiFenLi* and *ShiFenKai*

carry nearly the same meaning that ‘to cause something to separate’. Similarly, example (21-a) and (21-b) reveal the same problem. Under certain sense in English, ‘to lapse’ is ‘to end’ at least for a long time, thus can be seen as a troponym of ‘end’. Nevertheless, the differentiation vanished after translated into Chinese. As (21-b) shows, *Jieshu* and *ZhongZhi* are intuitively judged as near synonymous. The difference between ‘to end’ and ‘to lapse’ can not be found in their Chinese counterpart, at least, this difference has not be transported in ECTED. Moreover, many returned pairs are found to be exactly the same. For example, a pair of hypernymy-troponymy verb in (22-a) were mapped to two identical verbs in Chinese through ECTED as shown in (22-b).

(22) a. change ⇒ convert

b. 改變 (Gaibien) ⇒ 改變 (Gaibien)

Similar to examples (20) and (21), example (22) reveals the same problem in that the concept which makes a pair of verb synset stand in a hierarchical relation disappears after transported into Chinese. This time, as (22-b) shows, bootstrapping returned two identical verbs which are certainly impossible to have a hypernymy-troponymy relation. This indicates the difference in conceptualizing the manners of verbs. In English, the concept of ‘change’ is broader than ‘convert’ and thus includes ‘convert’. Despite the slight difference in English, there is no such distinction in Chinese. Therefore, both ‘change’ and ‘convert’ were translated into *GaiBien* in Chinese.

In addition to near- synonymy or synonymy relation found in returned pairs, reversed relation is also observed within our results. The reversed relation here indicates that for an English

hypernymy-troponymy synset pair E_h and E_t, each were mapped to the equivalent translations in Chinese C_h and C_t. However, the expected $C_h \Rightarrow C_t$ relation did not exist, rather, there is a reversed relation between C_h and C_t in which C_t is more reasonable to be the hypernym of C_h. More concrete example is shown as follow:

(23) a. expectorate ⇒ spit

b. 吐痰 (TuTan) ⇒ 吐(Tu)

As example (23-b) shows, *Tutan* 'to spit (phlegm)' and *Tu* 'to throw up (anything from the stomach or lung)' were returned by mapping from (23-a). However, the returned verb pair does not stand in a hypernymy-troponymy relation, that is, *TuTan* is not the hypernym of *Tu* for they can not pass our evaluating sentences. With a further observation, it is more reasonable to judge that *Tu* is actually the hypernym of *TuTan*. Once again, the reason that leads to this kind of error is due to how languages conceptualize verbs and how manners are distincted by verbs across languages. According to the gloss in WordNet, 'expectorate' is defined as 'to discharge (phlegm or sputum) from the lungs and out of the mouth' and 'spit' is defined as 'expel or eject (saliva or phlegm or sputum) from the mouth'. These two verbs are districted by manner. However, when these two verbs transported into Chinese, the distinction of manner vanished.

Above examples validate the truth that linguistic ontologies vary, the way to describe a concept may not be exactly the same across languages. Especially, it is found that English conceptualizes verbs in a more detailed way in which manners are distinguished by different lexeme. On the contrary, Chinese conceptualizes verb meanings in a broader or more general mode. Within a total of 11289 returned verb pairs in Chinese, there are 700 pairs being marked

to stand in semantic relations other than hypernymy-troponymy relation, accounted for 6.20% of total. Whether these incorrect verb pairs stand in near-synonymy relation or reversed relation are not further divided because they come up from similar causations as described in this subsection.

Unrelated

There are totally 1391 returned pairs being evaluated as unrelated, accounted for 12.32% proportion which is the highest within the three error types. Unrelated pairs, unlike the above mentioned error types, do not stand in hypernymy-tropoymy relation nor any lexical semantic relations. Having so many incorrect returned pairs from bootstrapping is somehow surprising since it is unexpected that a hypernymy-troponymy synset pair in WordNet turns out to be unrelated in Chinese after bootstrapping. Therefore, the causations which decrease the performance deserve a scrutinized investigation. By observing the results, unrelated verb pairs in Chinese can be ascribed to two main problems: inaccurate translation in ECTED and abstract concept verbs such as 'make', 'be', and 'put' in English.

The first problem leads to unrelated verb pairs is due to inaccurate translation in ECTED. Consider the following examples:

(24) a. change_ integrity ⇒ condense

b. 整型 (ZenShing) ⇒ 濃縮 (NongSuo)

(25) a. dance ⇒ folk_ dance

b. 舞蹈 (WuDao) ⇒ 民間舞曲 (MinJenWuChu)

Intuitively, example (24-b) could not be judged as a hypernymy-troponymy pair for that *Nong-Suo* 'to condense' is hardly being associated with *ZenShing* 'to do practice surgery'. If we compare (24-b) with its English counterpart (24-a), it is apparently that there is a translational discrepancy between 'change integrity' and 'ZenShing'. According to the definition in WordNet, 'change_ integrity' describes something 'changes in physical make-up' which is wrongly translated into '*ZenShing*' in Chinese. Established by common usage, *ZenShing* in Chinese almost exclusively means 'to do the practice surgery'. The incorrect translation in ECTED consequently leads to the unrelated verb pair returned by bootstrapping. Similarly, (25-b) displays an example of inaccurate translation in ECTED for that 'folk_ dance' was wrongly translated into a noun *MingJenWuChu* 'a dance music for folk_ dance'. Inaccurate translation sometimes involves in imprecise translation as well, that is, the translation itself is not totally unrelated to its English counterpart. Rather, the Chinese translation is indistinct and does not match the proper sense. Consider the following example:

(26) a. trouble ⇒ erupt

b. 打擾 (DaRao) ⇒ 發疹 (FaZhen)

As example (26-a) shows, when 'trouble' serves as a hypernym of 'erupt', it carries the meaning that 'to cause bodily suffering and make sick or indisposed'. Under this sense, 'trouble' is imprecisely translated into *DaRao* 'to bother' in Chinese which makes it intuitively hard to be connected with *FaZen*.

Another problem causing the returned pairs to be unrelated is that verbs with abstract concepts generate problems when transported into another languages. In English, the concepts of verbs such as 'be', 'make', or 'seem' which does not exhibit any literal spatial properties are

hardly conveyed to Chinese equivalently. Following examples give some illustrations:

(27) a. be ⇒ gape

b. 是 (Shi) ⇒ 張開 (ZhangKai)

(28) a. be ⇒ sit

b. 在(Zai) ⇒ 坐 (Zuo)

The copula verb 'be' is used widely in English. However, the concept of 'be' is rather abstract. For example, one of the sense of copula 'be' in WordNet is defined as 'having the quality of being'. Under this sense, there are more than one hundred direct troponyms can be found in WordNet. The equivalent translation of copula 'be' in Chinese is *Shi* which is used to connect with an adjective or a predicate noun. However, the copula *Shi* in Chinese does not carry any concrete meaning nor showing the quality of being as its English counterpart does. Therefore, when a hypernymy-troponymy pair containing copula 'be' is mapped to Chinese, none of them indicates hypernymy-troponymy relation as can be seen in (27-b). Similarly, another sense of copula 'be', according to WordNet, is 'occupy a certain position or area; be somewhere' and carries a troponym such as 'sit' (be located or situated somewhere) as example (28-a) shows. Again, copula 'be' under this sense, is translated into monomorphemic word *Zai* in ETCED. In Chinese, the verb *Zai* indicates 'being, to be located at' which is rather a morpheme that shows no concrete concept unless it is attached to other morphemes. Consequently, *Zuo* and *Zai* are not considered in hypernymy-troponymy relation.

Abstract concept verb reveals in Chinese too. Throughout our returned results tagged *Unrelated*, it is found that many verbs exhibit rather abstract concepts such as 使 *Shi 'make, let'*,

成爲 *ChenWei 'become'* or ***似乎*** *SiHu 'seem'* which are hard to be linked to any troponyms. Consider the following examples:

(29) a. make ⇒ prepare

b. 使 (Shi) ⇒ 準備 (ZuenBei)

(30) a. become ⇒ reduce

b. 成爲 (ChenWei) ⇒ 精簡 (JinJen)

(31) a. seem ⇒ glitter

b. 似乎 (SiHu) ⇒ 閃耀 (ShanYao)

The above returned Chinese verbs *Shi*, *ChenWei*, and *Sihu* come from their English correspondents 'make', 'become', and 'seem' which exhibit no translational inconsistencies. However, none of the Chinese pairs passed through our testing sentences, at least, it makes no sense when we substitute these pairs into the very basic definition of troponym: V_t *Shi YiZong* V_h *De FanShi* 'To V_t is to V_h in a particular manner'. Take (29-b) for example, '*ZuenBei Shi Yi Zong Shi De FanShi' (? to prepare is a way to make) makes no sense in Chinese. Hence, *Zuenbei 'to prepare'* and *Shi 'to make(causative)'* were not regarded as in hypernymy-troponymy relation.

Most of the unrelated verb pairs arise from inaccurate or imprecise translation in ECTED as shown by above examples. Conceptual idiosyncrasies aggravate translational problems as well. As mentioned, the equivalent translation database was manually created basing on WordNet synsets. Nevertheless, due to a large amount of data (12127 verb synset pairs) and the mutability of polysemous words, it is hard to reach a hundred-percent correspondence. While

the inaccurate translations can be corrected by double-check, conceptual idiosyncrasies, on the contrary, are hard to prevent and consequently become the toughest problem in bootstrapping.

Interim Summary

Section 4.2 reports the results returned by bootstrapping approach. Bootstrapping reaches a 73.56% accuracy, which is higher than syntactic pattern-based approach. For those incorrect results that decrease the performance, an in-depth investigation was made. In general, incorrect results can be roughly divided into three error types which are mainly caused by translational idiosyncrasies and conceptual inconsistencies across different languages.

4.3 Discussion

This section aims to provide some general discussion about syntactic pattern-based approach and bootstrapping approach. In this study, the two proposed approaches were processed separately for that there is methodologically and experimentally no direct connection between them. However, a general comparison and discussion on the performance of each approach is needed. Databases, results, and error types of these two approaches will be compared in the following subsections.

4.3.1 Comparison of Two Approaches

The most evident distinction between our two proposed approaches is that they applied on different data sources. For lexico syntactic-pattern based approach, we relied on a dictionary-like database where semantic relations may lie between the defined verb and its definition, represented through certain syntactic patterns. Chinese WordNet, for this reason, was chosen to serve as the data source for syntactic pattern-based approach. Unlike any other on-line dic-

tionaries, CWN was created basing on sense-splited entries along with explicit definitions and examples. Therefore, semantic relations can be built among each sense-splited set. On the other hand, bootstrapping approach exploits already existing resources and combining them within a common, standardized framework. In this thesis, bootstrapping approach resorted to Princeton WordNet, which has long been regarded as the gold standard for evaluating semantic frame induction, and ECTED in Sinica Bow since it provides us with a sense-based translation but not merely word-to-word correspondence. Different data sources were chosen according to different requirements in each approach. Despite the differences of the database, these two approaches point to the same goal: to extract possible hypernymy-troponymy verb pairs in Chinese. The performance of each approach was evaluated with the same metrics. Returned results from both approaches were evaluated and computed to see how precise these retrieved results are. For results retrieved from both approaches, accuracy rate was calculated in which the number of correct results retrieved from the approach was divided by total number of retrieved results. For the ease of comparison of the two approaches, Table 4.5 was given.

	Syntactic pattern-based approach	Bootstrapping approach
Database	CWN (10,388 verb senses)	WN (13,767 verb synsets) ECTED (12,127 sense-based ETs)
Total hypernymy-troponymy pairs retrieved	1619	11,289
Correct hypernymy-troponymy pairs retrieved	989	8305
Accuracy rate	60.71%	73.58%

Table 4.5: General comparison of syntactic pattern-based and bootstrapping approach

The comparison of syntactic pattern-based and bootstrapping approach can be interpreted

from two fold: **quantity** and **quality**. In terms of quantity, we need to compare the total number of verb pairs retrieved by both approaches. Input verb data was extracted from each database, the total number of verb in each database shows no significant discrepancy as can be seen from Table 4.5. However, for total results retrieved by each approach, there is a great disparity in number. From a total of 10388 verb senses in CWN, only 1619 verb pairs were returned by syntactic pattern-based approach. Compare to verb pairs retrieved by bootstrapping (11289 verb pairs), syntactic pattern-based approach extracted much fewer in return. The large discrepancy in scale is reasonable since only two syntactic patterns were used in the approach. Input entries conforming to our two lexical syntactic patterns are comparatively few but all of them can be retrieved by syntactic pattern-based approach. On the contrary, bootstrapping aims at retrieved all lexical information from an already existing database, here, the Princeton WordNet. Therefore, in this approach, all verb synsets pairs in WordNet which stand in hypernymy-troponymy relation will be returned as candidates as long as correspondent equivalent translations can be found in ECTED. From a qualitative perspective, the accuracy rate of each approach should be compared. As can be seen from Table 4.5, bootstrapping performs better than syntactic pattern-based approach (about 13% higher in accuracy rate). The difference between two approaches in performance should not be dismissed. However, only comparing the statistics is not valid enough to make any implications; comparisons on the returned results should be looked into as well. Therefore, a discussion on the results returned by each approach is given in the following subsection.

4.3.2 Comparison of the Results

By looking into the correct results, some characteristics of each approach can be observed. Syntactic pattern-based approach returned totally 989 correct hypernymy-troponymy pairs while

bootstrapping returned 8305 correct hypernym-troponymy pairs. After cross-comparing the results, only 38 hypernymy-troponymy pairs were found overlapped (These overlapping pairs were shown in Appendix B). The number of overlapped verb pairs was few. This may due to one of the characteristics of Chinese WordNet— the database on which syntactic pattern-based approach applied. Monomorphemic words (單字詞, *Dan ZhiCi*) are the foundation of Chinese WordNet [1]. That is to say, the definition of each sense tends to be made on the minimum morpheme in Chinese. For example, the monomorphemic word 沖 *Chong* carries totally different senses in 沖洗 (ChongShi, 'to wash') and 沖咖啡 (ChongKaFei, 'to make a coffee'). Under this situation, Chinese WordNet defines the monomorphemic word *Chong* and disambiguates the senses by splitting *Chong* into different entries. Example (32) and (33) show the definitions of *Chong* in different senses:

(32) 沖1_2 @ 用(P) 大量(D) 且(Cbb) 較(Dfa) 強(VH) 的(DE) 水(Na) 去除(VC) 特定(A) 對像(Na) 上(Ncd) 的(DE) 附著物(Na)
Chong @ Yong DaLiang Qie Jiao Jiang De Shui QuChu TeDing DuiXiang Shang De FuZuoWu
to wash@ Using a large amount of water to remove dirt.

(33) 沖1_3 @ 以(P) 水(Na) 調製(VC) 後述(Na) 飲料(Na) 或(Caa) 食品(Na)
Chong @ Yi Shui TiaoZhi HouShu YinLiao Huo ShiPin
to make @ To make drinks or food by water.

As above examples show, Chinese WordNet intends to give explicit senses of the minimum units in Chinese [1]. Consequently, syntactic pattern-based approach returned verb pairs which are sometimes hard to be intuitively judged as hypernymy-troponymy pairs due to the problem of polysemy. From input entries (32) and (33), the program returned verb pairs as example (34) show:

(34) a. 去除 (QuChu,‘to remove’) ⇒ 沖1_2 (Chong, ‘to wash’)

b. 調製(TiaoZhi, ‘to blend’) ⇒ 沖1_3 (Chong, ‘to make’)

Intuitively, example (34-a) and (34-b) may raise controversy when being evaluated as hypernymy-troponymy pairs. However, by looking back to the definition of the specific sense hold by the verb, the uncertainty can be eliminated. Therefore, for *Chong1_2, ‘to wash’*, which means the second sense of the first lemma, a hypernym *QuChu, ‘to remove’* was returned. Owing to the same reason, there are few hypernymy-troponymy pairs returned from both approaches being found overlapping with each other. For example, syntactic pattern-based approach tends to returned monomorphemic word as (34-a) shows; on the other hand, bootstrapping returned verb pairs through translation which did not limited to monomorphemic word as (35) shows:

(35) 除去 (ChuQu, ‘to remove’) ⇒ 沖洗(ChongSi, ‘to wash’)

Another characteristic related to the returned results was also observed. It is found that syntactic pattern-based approach returned more *direct* results in Chinese, that is, verbs are easily inferred in concept because syntactic pattern-based approach starts and bases on a Chinese database per se but not through translating from another languages. On the contrary, verb pairs retrieved by bootstrapping sometimes seems circuitous and indirect since it involves translations between two languages. The following examples were verb pairs returned by bootstrapping approach which are judged as hypernymy-troponymy pairs:

(36) a. embroider ⇒ purl

b. 刺繡 (CiShow) ⇒ 加繡邊 (JiaShowBien)

(37) a. look ⇒ ogle

b. 注意(ZuYi)⇒ 眼帶挑逗 (YenDaiTiaoDou)

Although we evaluated example (36-b) and (37-b) as hypernymy-troponymy pairs, it can be found that *JiaShowBien* and *YenDaiTiaoDou* are rather indirect verbs in Chinese, at least, both these verbs can not be queried either in CWN or the online revised Chinese dictionary[3].

4.3.3 Comparison of the Error Types

Knowing what lowers the accuracy rate can help us understand the advantages and limitations of both approaches. Table 4.6 gives a comparison of different error types by looking into incorrect returned results from both approaches.

Error type	
Syntactic pattern-based approach	**Bootstrapping approach**
Synonym or near synonym	Synonym or near-synonym
Verb and verb phrase	Non-lexicalized verb
Abstract concept verbs	Abstract concept verbs
The ambiguity of *yi*	Translational idiosyncrasy
Errors caused by syntactic structure	
Wrong tagging of POS	

Table 4.6: Comparison of error types from results in two approaches

[3]http://dict.revised.moe.edu.tw/

In general, more error types were found in syntactic pattern-based approach. This is inferable since its accuracy rate is lower than bootstrapping. When these error types were put together, we may find that the problems of near-synonym, non-lexicalization and abstract concept verbs cause the errors in both approaches (Error types found in both approaches were shown in shadow box). The problem of non-lexicalization arises from the linguistic characteristic that there is no word delimiters in Chinese. Unlike English, where spaces or blanks are used to delimit word boundaries, Chinese encounters difficulties in clearly defining a word or lexeme. Among the returned results, some are found that hypernymy-troponymy relation occurs between verbs and verb phrases but not within verbs themselves, at least, they have not been lexicalized nor can be POS tagged within one single part of speech in CKIP segmentator and tagger. On the other hand, the problem of near-synonym comes from different causations. In syntactic pattern-based approach, synonyms were sometimes retrieved despite the definition sentences comply with our syntactic patterns as the following example shows:

(38) 畫圖_1(VA) @ 以(P) 筆(Na) 或(Caa) 其他(Neqa) 工具(Na) *畫出(VC) 圖形(Na)
HuaTu @ Yi Bi Huo QiTa GongJu *HuaChu TuShing
或(Caa) 圖案(Na)
Huo TuAn
to paint @ Drawing figures or patterns by pens or other equipment.

However, the problem of synonym from bootstrapping approach reflects the discrepancy of how different languages conceptualize verbs. For example, a hypernymy- troponymy synset pair 'change ⇒ convert' was translated into two identical verbs in Chinese '*Gaibien* ⇒ *Gaibien*'. This shows that a slight conceptual difference in manner of 'change' and 'convert' does not exist in Chinese. In other words, Chinese conceptualizes verb meanings in a broader or more general mode. Also, as Table 4.6 shows, abstract concept verbs in Chinese were returned

by both approaches. Recall that verbs such as *成(爲), ChengWei 'become'*, *作(爲), ZuoWei 'be, serve as'* or *是, Shi 'be'*, *在, Zai 'be'*, *似乎, Sihu 'seem'* are hard to convey concrete concepts. These verbs, rather, grammatically function as copulative verbs to which nouns, adjectives or predicates are attached. Therefore, abstract concept verbs then failed to indicate any hypernymy-troponymy relation.

By looking into the rest of the error types made by each approach, we may deduce that bootstrapping algorithm is more stable than syntactic pattern-based approach. This is proved by the errors found within incorrect results. The main problem in bootstrapping approach, as mentioned, lies in the translational and conceptual discrepancies across languages. Since we relied on ECTED as a translational medium to map between English and Chinese, errors became unavoidable as long as the equivalent translations were inaccurate or imprecise in ECTED. It is also found that English conceptualizes verbs in a more detailed way in which manners are distincted by different lexemes. However, except for the problems related to translational idiosyncrasies, the mapping algorithm itself shows no mismatch nor neglect. On the contrary, for syntactic pattern-based approach, more error types are related to algorithmic process itself. For example, errors caused by syntactic structures were due to the insufficiency in pre-proseccing such as parsing as the following example shows:

(39) 打1_25(VC) @ 以(P) *彈(VC) 棉花(Na) 的(DE) 方式(Na) 製造出(VC) 棉被(Na)
Da @ Yi *Tan MianHua De FangShi ZhiZaoChu MianBei
to pluck @ producing a bed quilt by flipping the cotton.

Syntactic pattern-based algorithm returned *all* verbs occurring after target words. In example (39), both *Tan* and *ZhiZaoChu* were extracted to be the hypernyms of *Da* since they appear

after the target words *yi*. This problem involves in sentence parsing during the pre-processing stage. In this thesis, input data were processed only by segmentation and POS tagging through CKIP segmentator and tagger, the parsing part was left undone in order to limit the required time and resources for processing the input data.

4.3.4 General Discussion

Overall, bootstrapping approach exhibits a better execution than syntactic pattern-based approach both in quantity and quality. However, it might seem arbitrary to jump to the conclusion that bootstrapping is **better** than syntactic pattern-based approach. From the above comparisons on the databases, results, and error types of each approach, we know that there are advantages and limitations in syntactic pattern-based and bootstrapping approach. Table 4.7 provides a clear comparison of the two approaches on the limitations and advantages.

Syntactic pattern-based approach		**Bootstrapping approach**	
Advantage	Limitation	Advantage	Limitation
Returned more direct results for it started and based on a Chinese database.	Error types were related to algorithmic process itself.	More stable and reliable. The mapping algorithm itself showed no mismatch nor neglect.	Translational and conceptual discrepancies across languages.
	The number of retrieved verb pairs was comparatively few.	Immensely extracted LSRs from an existing WordNet database.	

Table 4.7: General comparison of the two approaches

As shown in Table 4.7, syntactic pattern-based approach mainly limited to the small scale of retrieved results since there were only two syntactic patterns used in this approach. Also, more error types were related to algorithmic process itself because sentence parsing process and word sense disambiguation process were left undone in the approach. On the other hand, translational and conceptual idiosyncrasies became the major barriers for bootstrapping approach. In Fellbaum's [16] and Vossen et al.'s studies [44], it was clarified that LSRs are 'universal'. That is, LSRs will hold true across different languages. In Huang et al.'s studies [24] [25], certain LSRs were also proved transportable across different languages through bootstrapping. However, little research put attention on the issue of language idiosyncrasies. In this thesis, our results from bootstrapping exhibit that language idiosyncrasies do play important roles in bootstrapping and deserved notices.

Despites the limitations, both of the proposed approaches do semi-automatically extract a large amount of Chinese verb pairs which are in hypernymy-troponymy relation. Terefore, instead of comparing these two approaches in a competing way, they should be implemented complementarily for that syntactic pattern-based approach returned direct results (not through translation from other languages) while bootstrapping can immensely extract LSRs from an existing WordNet database. In this thesis, a total of 9294 hypernymy-troponymy pairs were extracted by syntactic pattern-based and bootstrapping approach. A Chinese lexical database such as CWN contains only 143 verbs which have hypernymous linkage and only 27 verbs having hyponymous connections[4]. Despite that there is still limitations of the two approaches, a large amount of returned hypernymy-troponymy verb pairs can definitely enhance and help the contribution of a WordNet-like database.

[4]The statistic is up to 2006

4.4 Summary

This chapter has presented elaborate results returned from syntactic pattern-based approach and bootstrapping approach. In addition to the statistical representation of the results, scrutinized error analyses were given in this chapter as well. Knowing what causes the errors helps us understand the shortages and advantages of each approach. Section 4.1 reports the results of syntactic pattern-based approach in which the accuracy rate reaches 60.71%. Within the returned results which are not in hypernymy-troponymy relation, there are totally six different error types being observed. Most of the errors generated from syntactic pattern-based approach are related to insufficient pre-processing such as inaccurate tagging of POS and the lack of sentence parsing in syntactic structures. In Section 4.2, results of bootstrapping approach along with error analyses are given. Bootstrapping reaches a 73.56% accuracy rate, which is about 13% higher than syntactic pattern-based approach. Errors that lead to the incorrect results in bootstrapping can be roughly divided into three types which relates much to the problem of translational idiosyncrasies and conceptual inconsistencies across languages. Finally, a general discussion and comparison of these two approaches was conducted in Section 4.3. Both approaches carry their own benefits and limitations. Syntactic pattern-based approach returned more direct results for it started and based on a Chinese database per se but the number of retrieved verb pairs were comparatively few. On the other hand, bootstrapping resorted to an existing WordNet and immensely transported lexical information into Chinese. Despite the advantages and limitations of each approach, both of them do semi-automatically extract a large amount of Chinese verb pairs which are in hypernymy-troponymy relation, contributing to the enhancement of WordNet-like database such as CWN.

Chapter 5

Conclusion

In this chapter, the summary of this thesis will be given in Section 5.1. Section 5.2 discusses the contributions and implications of this thesis, followed by the limitations of the study and suggestions for future works in Section 5.3.

5.1 Summary of the Thesis

In this thesis, we proposed two independent approaches to semi-automatically extract lexical semantic relations. Given the fact that a sense-based and LSRs-connected database such as Princeton WordNet has become a key source for Natural Language Processing, automatically extracting lexical semantic relations, hence, becomes an important issue in recent years. In this study, we target at hypernymy-troponymy relation among verbs and proposed syntactic pattern-based approach and bootstrapping approach to automatically label verb pairs which stand in our target relation. **Syntactic pattern-based approach** is used for that sentence structures can always denote relations and reveal information among lexical entries. We assumed that LSRs may lie in a large dictionary-like database, represented through certain syntactic patterns. Therefore, we then used two manually selected syntactic patterns which can possibly exhibit

hypernymy-troponymy relation between a defined verb and its definition sentence in this approach. The results shows that with a 60.71% accuracy rate, syntactic pattern-based approach successfully semi-automatically extract direct and lexicalized verb pairs in Chinese. On the other hand, **bootstrapping approach** aims at exploiting an already existing database and transported lexical and semantic information into another language. In this approach, we resorted to Princeton WordNet as our database and transported hypernymy-troponymy synset pairs into Chinese through mapping with an equivalent translation database ECTED. The results shows that bootstrapping approach can immensely transport lexical information from Princeton WordNet into Chinese with a 73.56% accuracy rate.

In addition to analyzing the accuracy and errors of both approaches, we also compared syntactic pattern-based approach and bootstrapping approach on the databases that they applied on, the returned results and error types. Both approaches carry their own benefits and limitations. Bootstrapping approach has the advantages in returning a large amount of results in a short time but is obstructed by the cross-lingual differences in translating and conceptualizing. On the other hand, syntactic pattern-based approach returned direct results for it started and based on a Chinese database per se but not through translating from another lanagues. Nevertheless, the number of retrieved verb pairs was comparatively few since only two syntactic patterns were manually selected. Given the fact that each approach does successfully extract hypernymy-troponymy verb pairs despites their limitations, they should be implemented complementarily but not compared in a competing way.

5.2 Contribution

Creating a WordNet-like database requires intensive human labors and time. Neither the WordNet nor the EuroWordNet construction methodology seems to be efficient for additional languages. Researchers have long been studied on the combination of computational technology and natural language processing or knowledge extraction, aiming at automatically learning lexical semantic relations and constructing semantic hierarchies [7] [9] [20] [21] [28] [30] [32] [41]. Despite the fact that much work has be done on automatically extracting lexical information, this thesis differs in two ways. In terms of language, previous studies deal with English or other European languages; little work has been done for Chinese. In terms of lexical semantic relations, our literature survey reveals that hypernymous relation was studied and extracted within nouns; the hypernymy-hyponymy relation between verbs is rarely studied so far. To our knowledge, there has not yet been a completely constructed WordNet-like database in Chinese. A sense-based taxonomic database—Chinese WordNet is impoverished in labeling semantic relations. In total, our semi-automatic approaches extracted 9256 Chinese verb pairs which are in hypernymy-troponymy relation. Compare to Chinese WordNet, where only 143 hypernymous relation and 27 hyponymous relation were labeled to verbs, our returned results can definitely enhance and enrich the integrity of WordNet-like database in Chinese. While the fully automated labeling is nearly impossible, our approaches generates the labeling semi-automatically, means that the final evaluating part would be refined by linguists or domain experts. Nevertheless, the approaches proposed in this thesis can greatly and rapidly extract lexical information at the first stage, aiding the construction of lexical database in a more effective way.

5.3 Limitations of the Present Study and Suggestions for Future Work

This thesis proposed two approaches on semi-automatic extraction of semantic relations. In addition to evaluating the results extracted from our approaches, an investigation into errors that decrease the performance was also provided. In doing so, limitations of this present study can be discovered. Also, modifications and suggestions can be provided for future work:

- For syntactic pattern-based approach, some limitations were found by analyzing errors and results. First of all, there were only two manually selected patterns being used. Most input entries were then filtered out for they do not comply with our syntactic patterns. For further research, a pattern discovery procedure is suggested before executing this approach.

- Second, by scrutinizing the incorrect returned results, we have learned that Word Sense Disambiguation (WSD) is needed. Also, the syntactic structure of a sentence will influence the accuracy of extracted verbs. A possible direction for future research is to do the sense disambiguation and sentence parsing in a pre-processing procedure. In this study, we left the parsing undone in order to limit the required time and resources for processing the input data. Therefore, sentence parsing can be left for future works by using parsers in Chinese such as Sinica Treebank[1].

- Third, for all the Chinese verbs, they are subdivided into 15 different subcategories in CWN including VA, VAC, VB, VC, VD, VE, VF, VG, VH, VI, VJ, VK, VK, VCL, VHC. With such a diverse part-of-speech in Chinese verbs, it might become an issue in studying what roles could POS play in finding lexical semantic relations. For example,

[1] http://rocling.iis.sinica.edu.tw/CKIP/treebank.htm

do subcategories of verbs matter? If so, the information of POS may serve as a filter during the process of automatic extraction.

- Fourth, in syntactic pattern-based approach, we resorted to Chinese WordNet as our data source. Despite that CWN is a sense-split database, it contains only medium frequency words including 5600 lemmas and 13160 word senses in total. For future work, data source can be extend to a more general dictionary such as The Revised Chinese Dictionary released by Ministry of Education(教育部重編國語辭典修訂本)[2]. Future work can even extend the database to the limitless free text on the Internet– the Web As Corpus. In this thesis, we resorted to dictionary-like database because a preliminary research on the web showed that language is easily being metaphorized and hence becomes unstable and unreliable. However, the immenseness of the free text can not be neglected. A possible direction for future work is to incorporate web text into database, assisting with the information of POS, collocation, or other filters.

- For bootstrapping approach, it has been found that imprecise translations and conceptual discrepancy will certainly lower the accuracy. While the conceptual differences are hard to prevent, a modification on the equivalent translations is somehow doable.

- Last but not least, extending the approaches to other semantic relations is also suggested. As mentioned, WordNet-like semantic networks were built on the paradigmatic semantic relations. This thesis narrows down the scope to the hypernymy-troponymy relation of verbs only. It is hoped that the proposed approaches, especially bootstrapping can be applied onto other paradigmatic relations such as antonym, synonymy, and meronymy, etc. What is more, lexical semantic relations can also be extended from a syntagmatic per-

[2]http://dict.revised.moe.edu.tw

spective, such as thematic relations. More lexical semantic databases, such as HowNet[3], VerbNet[4], VerbOcean[5], etc. can be probed into for future work.

[3]http://www.keenage.com/

[4]http://verbs.colorado.edu/ mpalmer/projects/verbnet.html

[5]http://demo.patrickpantel.com/Content/verbocean/

Bibliography

[1] 黃 居 仁. ***中 文 詞 彙 意 義 的 區 辨 與 操 作 原 則***. available at http://cwn.ling.sinica.edu.tw/.

[2] 張如瑩 and 黃居仁. 中央研究院中英雙語知識本體詞網 (sinica bow)：結合詞網，知識本體，與領域標記的詞彙知識庫. In ***第十六屆自然語言與語音處理研討會（ROCLING XVI）***, Greenbay, Taipei., 2004.

[3] 謝舒凱, Petr Šimon, and 黃居仁. 大規模詞彙語意關係自動標記之初步研究：以中文詞網 (chinese wordnet) 爲例. In ***中華民國計算語言學國際會議***, 交通大學, 2006.

[4] A. Alonge. *Definition of the links and subsets for verbs.* EuroWordnet deliverable D006 at http//www.hum.uva.nl/ ewn/docs.htm, 1996.

[5] H. Alshawi. Processing dictionary definitions with phrasal pattern hierarchies. *American Journal of Computational Linguistics*, 13.3:195–202, 1987.

[6] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The berkeley framenet project. In *Proceedings of the COLING-ACL*, 1998.

[7] A. Berland and E. Charniak. Finding parts in very large corpora. In *proccedings of ACL-1999*, pages 57–64, College park, MD, 1999.

[8] K.J Chen and Y.M. Hsieh. Chinese treebanks and grammar extraction. In *Proceedings of the first International Joint Conference on Natural Language Processing.*,

2004.

[9] T. Chklovski and P. Pantel. Large-scale extraction of fine-grained semantic relations between verbs. In *Proceedings of KDD Workshop on Mining for and from the Semantic Web (MSW-04)*, pages 12–23, Seattle, WA, 2004.

[10] S. Climent, H. Rodriguez, and J. Gonzalo. Definition of the links and subsets for nouns. In *EuroWordNet deliverable D005, http://www.hum.uva.nl/ ewn/docs.htm*, 1996.

[11] D.A. Cruse. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.

[12] H. Dang, Y. Ching, M. Palmer, and F. Chiou. Simple features for chinese word sense disambiguation. In *proccedings of COLING02*, pages 133–138, Taipei, Taiwan, 2002.

[13] Z. Dong and Q. Dong. *An Introduction to HowNet*. available from http://www.keenage.com.

[14] Z. Dong and Q. Dong. *HowNet and the computation of Meaning*. N.J: World Scientific Publishing Co., 2006.

[15] C. Fellbaum. English verb as a semantic net. *International Journal of Lexicography*, 3:181–303, 1990.

[16] C. Fellbaum. *WordNet*. The MIT press, 1998.

[17] C. Fellbaum. On the semantics of troponymy. In Rebecca Green, A.Carol Bean, and H.M.Sung, editors, *The semantics of relationships: an interdisciplinary perspective*, pages 23–24, 2002.

[18] C. Fellbaum. *On the semantics of Troponymy*. Cognitive Science Laboratory, Princeton University, 2002.

[19] C. Fellbaum and G. Miller. Folk psychology or semantic entailment?– a reply to rips and conrad. *The Psychological Review*, 97:565–570, 1990.

[20] R. Girju, A. Badulescu, and D. Moldovan. Automatic discovery of part-whole relations. *Computational Linguistics*, 31(1):12–24, 2006.

[21] M. A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *proceedings of the Fourth International Conference on Computaional Linguistics (COLING)*, pages 539–545. Nantes, France, 1992.

[22] M. A. Hearst. Automatic discovery of wordnet relations. In C. Fellcaum, editor, *WordNet: An Electronic Lexical Database and Some of its Applications*. MIT press, 1998.

[23] C.R. Huang, F.J. Lo, R.Y. Chang, and S.M. Chang. Reconstructing the ontology of the tang dynasty: A pilot study of the shakespearean-garden approach. In *The OntoLex 2004 Workshop*, Lisbon, 2004.

[24] C.R. Huang, E. Tseng, and B. S. Tsai. Translating lexical semantic relations: The first step towards multilingual wordnets. In Grace Ngai, Pascale Fung, and Kenneth W. Church, editors, *Proceedings of the COLING 2002 Workshop "SemaNet: Building and Using Semantic Networks"*, pages 2–8, 2002.

[25] C.R. Huang, I. J. Tseng, B. S. Tsai, and B. Murphy. Cross-lingual portability of semantic relations: Bootstrapping chinese wordnet with english wordnet relations. *Language and linguistics*, 4.3:509–532, 2003.

[26] K. Kipper-Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon.* PhD thesis, University of Pennsylvania, 2005.

[27] D.K Lin and P. Pantel. Dirt - discovery of inference rules from text. In *Proceedings of ACM Conference on Knowledge Discovery and Data Mining (KDD-01)*, pages 323–328, San Francisco, CA., 2001.

[28] D.K. Lin, S.J. Zhao, L.J. Qin, and M. Zhou. Identifying synonyms among distributionally similar words. In *IJCAI-03*, 2003.

[29] W.Y. Ma and K.J. Chen. Introduction to ckip chinese word segmentation system for the first international chinese word segmentation bakeoff. In *Proceedings of ACL 2nd SIGHAN Workshop on Chinese Language Processing*, Seattle, WA, 2003.

[30] J. Markowitz, T. Ahlswede, and M. Evens. Semantically significant patterns in dictionary definitions. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 112–119, 1986.

[31] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 3.4:235–244, 1990.

[32] J. Nakamura and M. Nagao. Extraction of semantic information from an ordinary english dictionary and its evaluation. In *Proceedings of the Twelfth International Conference on Computational Linguistics*, pages 459–464,, Budapest., 1988.

[33] J. Nakamura, K. Sakai, and M. Nagao. Automatic analysis of semantical relation between english nouns by an ordinary english dictionary. In *Institute of Electronics, Information and Communication Engineers of Japan, WGNLC*, Japan, 1987.

[34] I. Niles and A. Pease. Toward a standard upper ontology. In *In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001).*, Ogunquit, Maine., 2001.

[35] I. Niles and A. Pease. Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering. (IKE 2003)*, Las Vegas, Nevada., 2003.

[36] Martha Palmer and Zhibao Wu. Verb semantics for english-chinese translation. *Machine Translation*, 10:59–92, 1995.

[37] P. Pantel and D.K Lin. Automatically discovering word senses. In *Proceedings of Human Language Technology / North American Association for Computational Linguistics (HLT/NAACL-03)*, pages 21–22, Edmonton, Canada, 2003.

[38] P. Pantel and D. Ravichandran. Automatically labeling semantic classes. In *Proceedings of HLT/NAACL-2004.*, Boston, MA, 2004.

[39] M. Pennacchiotti and P. Pantel. A bootstrapping algorithm for automatic harvesting semantic relations. In *Proceedings of Inference in Computational Semantics (ICOS-06)*, pages 87–96, Buxton, England, 2006.

[40] J. Pustejovsky. *The Generative Lexicon*. MA:MIT Press, 1995.

[41] J. Ramanand and P. Bhattacharyya. Towards automatic evaluation of wordnet synsets. In *Global Wordnet Conference (GWC08)*, 2008.

[42] S. Richardson, W. Dolan, and L. Vanderwende. Mindnet: acquiring and structuring semantic information from text. In *36th Annual meeting of the Association for Computational Linguistics*, volume 2, pages 1098–1102, 1998.

[43] B.S. Tsai, C.R. Huang, S.C. Tseng, J.Y. Lin, K.J. Chen, and Y.S. Chuang. 中文詞義的定義與判定原則. ***中文信息學報*** *(Journal of Chinese Information Processing)*, 16.4:21–31, 2002.

[44] P. Vossen. Eurowordnet: a multilingual database for information retrieval. In *Proceedings of the DELOS workshop on Cross-language Information Retrieval*, Zurich, 1997.

[45] P. Vossen. *EuroWordNet: a multilingual database with lexical semantic networks.* Kluwer Academic Publishers, 1998.

[46] P. Vossen, P. Diez-Orzas, and W. Peters. The multilingual design of eurowordnet. In P. Vossen, N. Calzolari, G. Adriaens, A. Sanfilippo, and Y. Wilks, editors, *Proceedings of the ACL/EACL-97 workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 1–8, Madrid, 1997.

[47] Yun Xin. Srcb-wsd: Supervised chinese word sense disambiguation with key features. In *proccedings of SemEval-2007*, pages 300–303, Prague, Czech Republic, 2007.

Appendix A

Programming Code

Programming language: Python

Version: 2.5.2

Syntactic pattern-based approach-1

```
#-----------------------------------------------------

#!/usr/bin/env python

# -*- coding: big5 -*-

import nltk, re, pprint

f = open('First_result.txt')

s = f.read()

f.close()

uni_string = unicode(s,'big5')

aa = re.findall(r'.+@|...?\(V[A-GI-KM-Z]+\)|...?\(Nv\)', uni_string)

for word in aa:
```

```
    print word
```

Syntactic pattern-based approach-2

```
import nltk, re, pprint
f = open('cwnsenselist.txt')
s = f.read()
f.close()
uni_string = unicode(s,'big5')
a = nltk.tokenize.regexp_tokenize
(uni_string,pattern=r'\s\$', gaps=True
target1 = u'以(P)'
target2 = u'地(DE)'
target3 = u'用(P)'
for word in a:
    if target1 in word :
        print word
    elif target2 in word:
        print word
    elif target3 in word:
        print word
```

Bootstrapping approach-1

```
import nltk
from nltk.corpus import wordnet as wn
#for synset in list(wn.all_synsets('v')):
```

```
from itertools import islice
for synset in islice(wn.all_synsets('v'), 13767):
    if len(synset.hyponyms())!=0:
        print synset, synset.hyponyms()
```

Bootstrapping approach-2

```
import nltk
from nltk.corpus import wordnet as wn
#for synset in list(wn.all_synsets('v')):
from itertools import islice
for synset in islice(wn.all_synsets('v'), 13767):
    verb=synset.lemmas[0].key
    if len(synset.hyponyms())!=0:
        for s in synset.hyponyms():
            hypon=s.lemmas[0].key
            print verb+"=>"+hypon
            f=open("troponym.txt","a+")
            f.write(verb+"=>"+hypon+"\n")
            f.close()
```

Appendix B

Results from Syntactic Pattern-based Approach

Syntactic pattern-based approach returned totally 989 hypernymy-troponymy pairs. As space is limited, we only list 216 pairs in the Addendix. There are totally 38 verb pairs being overlapped with the correct results from Bootstrapping; these overlapped pairs were shown in shadow box:

Hypernym ⇒ Troponym			
增加VHC⇒ 倍1_1	指示VE⇒ 比1_6	傳達VD⇒ 講_1	傳達VD⇒ 講_14
傳達VD⇒ 說1_1	傳達VD⇒ 說1_16	摩擦VC⇒ 擦_3	製作VC⇒ 擦_5
呼氣$VA \Rightarrow$ 吹_1	接合$VB \Rightarrow$ 縫合_1	改變$VC \Rightarrow$ 革新_1	離開 VC⇒ 滾_1
修飾VC⇒ 化妝_1	移動 VAC⇒ 揮_1	傳送$VD \Rightarrow$ 寄_1	增加$VHC \Rightarrow$ 加倍_1
發聲$VA \Rightarrow$ 叫1_1	纏繞VC⇒ 捲_6	脅迫VF⇒ 勒索_1	摩擦$VC \Rightarrow$ 磨1_7
摩擦$VC \Rightarrow$ 抹1_2	對待$VC \Rightarrow$ 虐待_1	移動$VAC \Rightarrow$ 爬_1	打$VC \Rightarrow$ 拍2_1
移動$VAC \Rightarrow$ 跑_1	批評VC⇒ 抨擊_1	要求$VF \Rightarrow$ 請1_1	離開$VC \Rightarrow$ 去1_11

Hypernym ⇒ Troponym			
改變VC ⇒ 染1_2	移動VAC ⇒ 抬_2	接收VC ⇒ 聽1_1	控制VC ⇒ 限制_1
決定VE ⇒ 確定_2	行走VA ⇒ 散步_1	移動VAC ⇒ 走_1	離開VC ⇒ 走_5
呼吸VC ⇒ 喘_1	誘捕VC⇒ 釣魚_1	戴VC⇒ 頂_4	說話VA ⇒ 侃侃而談_1
製作VG ⇒ 吹_3	去除VC⇒ 吹_6	釣魚VA ⇒ 垂釣_1	對待VC ⇒ 慈1_4
穿入VCL⇒ 刺1_1	搭配VC ⇒ 搭1_9	進行VC ⇒ 搭2_4	回應VC ⇒ 答2_3
施力VB ⇒ 打1_1	攪拌VG ⇒ 打1_9	擊VC ⇒ 打1_23	製造VC ⇒ 打1_25
攻擊VC ⇒ 打1_29	取得VC ⇒ 打1_30	輸入VC ⇒ 打1_39	計算VC ⇒ 打1_41
傳送VD⇒ 打1_67	撞擊VC⇒ 打破_1	問候VC⇒ 打招呼_1	注射VC⇒ 打針_1
互動VA⇒ 待2_2	押到VC⇒ 帶2_10	支撐VC⇒ 擔1_1	抵押VC⇒ 當4_1
傳達VD⇒ 道5_6	引述VC⇒ 道5_2	倚靠VC⇒ 抵1_1	換取VC⇒ 抵1_6
抵押VC⇒ 典2_1	劃上VC⇒ 點1_4	指示VE⇒ 點1_21	動VAC⇒ 點頭_2
捕捉VC⇒ 釣1_1	維持VJ⇒ 撐_7	抵住VC ⇒ 撐竿_1	倚靠VC⇒ 頂_6
顯露出VJ⇒ 表現_8	固定VHC⇒ 別3_1	固定VHC ⇒ 別3_2	取出VC ⇒ 抽_1
輕觸VC ⇒ 啵_4	施力VB ⇒ 撥_1	施力VB ⇒ 撥_2	顯影VA ⇒ 沖1_7
計算VC ⇒ 撥_4	施力VB⇒ 撥_5	演奏VC ⇒ 撥_6	調製VC ⇒ 沖1_3
分開VHC⇒ 剝1_2	得到VJ ⇒ 博2_1	增加VHC ⇒ 補1_13	去除VC ⇒ 沖1_2
幫助VC⇒ 補助_1	移動VAC ⇒ 步_1	取出VC ⇒ 採_2	噴VC ⇒ 沖1_1
操控VC ⇒ 踩_2	行走VA ⇒ 踩_4	對待VC ⇒ 踩_5	固定VHC⇒ 持_1
搜集VC ⇒ 採訪_1	告狀VB⇒ 參3_1	存在VA ⇒ 藏1_3	攝入VC ⇒ 吃_5
存放VC ⇒ 藏1_6	確定VK ⇒ 測_2	進入VCL ⇒ 插_3	取得VC ⇒ 吃_13
附著VA ⇒ 纏1_1	相鬥VA⇒ 吵架_1	發聲VA ⇒ 扯_3	吸收VC ⇒ 吃_3
拉動VC ⇒ 扯_8	拉動VC ⇒ 扯鈴_2	呈現VJ ⇒ 陳1_2	測量VC ⇒ 稱2_1
頂住VC ⇒ 撐_2	阻擋VC⇒ 撐_3	抵住VC ⇒ 撐_4	表現出VC⇒ 呈現_3

牴觸VJ⇒ 頂_7	牴觸VJ⇒ 頂嘴_1	制定VC⇒ 訂_1	定VHC⇒ 訂_5
穿過VCL⇒ 釘2_1	約束VC⇒ 釘2_2	施力VB⇒ 丟_1	磨練VC⇒ 動心忍性_1
晃動VAC⇒ 抖_3	攻擊VC⇒ 鬥1_6	刺激VC⇒ 逗1_1	發笑VA⇒ 逗1_4
拿VC ⇒ 端2_1	分離VHC⇒ 斷_2	互動VA⇒ 對_11	取得VC⇒ 奪_1
傳達VD⇒ 發2_7	給VD ⇒ 發2_8	傳送VD⇒ 發送_3	處罰VC⇒ 罰款_1
處罰VC⇒ 罰錢_1	射門VA⇒ 罰球_2	尋找VC⇒ 翻1_8	增加VHC⇒ 翻1_13
檢視VC⇒ 反觀_1	蒐集VC⇒ 訪1_5	注入VC⇒ 放_7	破壞VC⇒ 放_32
獎賞VD⇒ 分封1_1	覆蓋VC⇒ 封1_3	獎賞VD⇒ 封2_1	接補VC⇒ 縫1_1
進入VCL ⇒ 鑽1_2	保全VC ⇒ 自保_1	支撐VC⇒ 扶1_1	靠VJ⇒ 扶1_2
食用VC⇒ 服_2	潛水VA⇒ 浮潛_1	協助VC⇒ 輔選_1	承載VC⇒ 負1_1
評定VE⇒ 改1_4	修飾VC⇒ 改1_6	拍擊VC⇒ 蓋1_7	駕駛VA⇒ 趕_3
碾VC⇒ 桿_3	逼迫VF⇒ 趕盡殺絕_2	分開VHC⇒ 割_2	去除VC⇒ 割_6
翻VC ⇒ 耕1_1	付出VC⇒ 耕1_6	移動VAC⇒ 游1_3	取代VC⇒ 更新_1
批評VC⇒ 攻1_4	支撐VC⇒ 鉤1_8	履行VC⇒ 估價_1	擔任VG⇒ 雇用_1
察覺VK⇒ 觀望_1	執行VC⇒ 貫1_5	清洗VC⇒ 灌1_6	扣入VC⇒ 灌1_9
製作VC⇒ 灌1_13	宣告VE⇒ 廣播_1	播放VC⇒ 廣播_2	開導VC⇒ 規勸_1
通知VC⇒ 函請_1	說話VA⇒ 喊_1	叫喊VA⇒ 喝2_1	取得VC⇒ 黑1_9
侵入VCL⇒ 黑1_10	檢測VC⇒ 衡量_1	修飾VC⇒ 畫1_9	取VC ⇒ 啄_1
遮蓋VC⇒ 化妝_2	發聲VA⇒ 喚_1	替代VJ⇒ 換成_1	得到VJ⇒ 換取_1
補償VD⇒ 恢復_4	回答VE⇒ 回嘴_1	施力VB⇒ 擊_1	拍打VC⇒ 擊_5
犧牲VJ⇒ 祭1_1	固定VHC ⇒ 抓_1	送VD ⇒ 寄回_2	固定VHC⇒ 夾_5
推測VE⇒ 假設_1	攜帶VC⇒ 肩1_4	修飾VC⇒ 剪_2	輸入VC⇒ 鍵入_1
引述VC⇒ 講_2	彰顯VC⇒ 志2_4	談論VE⇒ 講_13	存在VA⇒ 降1_12
交換VC⇒ 交談_1	磨VC ⇒ 嚼_1	取得VC⇒ 殖民_1	切割VC⇒ 鋸_2

Appendix C

Results from Bootstrapping Approach

Bootstrapping approach returned totally 8305 hypernymy-troponymy pairs. As space is limited, we only list 236 pairs in the Addendix:

Hypernym ⇒ Troponym			
讚美 ⇒ 頌揚	讚美 ⇒ 稱賀	讚美 ⇒ 諂媚	讚美 ⇒ 吹捧
觀察 ⇒ 偵察	觀察⇒ 追蹤	觀察 ⇒監聽	觀察 ⇒ 賞鳥
觀光 ⇒ 參觀	讓步 ⇒服從	變質 ⇒玷污	變熱 ⇒ 使過熱
變熱 ⇒ 灼傷	變緊 ⇒ 轉動	變緊 ⇒ 旋轉	變緊 ⇒使堅固
變緊 ⇒ 綁緊	體會 ⇒回味	變甜 ⇒加糖	變甜 ⇒ 覆上糖衣
變乾 ⇒ 乾涸	變窄⇒ 壓縮	變弱 ⇒ 模糊	變弱 ⇒ 鬆弛
變弱 ⇒放鬆	變弱 ⇒ 瓦解	變弱 ⇒ 使放鬆	變污 ⇒ 生斑點
變污 ⇒ 變色	變白 ⇒ 漂白	變平 ⇒ 壓	變平 ⇒拍平
攪亂 ⇒ 戳	攪 ⇒ 起泡沫	鑄造 ⇒ 沙模鑄造	鑄造 ⇒ 製版
學習 ⇒ 訓練	離開 ⇒出門	離開 ⇒ 逃	醫治 ⇒ 診療
轉讓 ⇒ 交換	轉讓 ⇒ 發行	轉讓 ⇒ 交付	轉變 ⇒ 減低

鑄造 ⇒ 改造	鑑定 ⇒ 查驗	襲擊 ⇒ 包圍	襲擊 ⇒ 轟炸
襲擊 ⇒ 反攻	襲擊 ⇒ 閃擊	襲擊 ⇒ 放毒氣	襲擊 ⇒ 攻擊
襲擊 ⇒ 逮捕	襲擊 ⇒ 炮轟	襲擊 ⇒ 打擊	襲擊 ⇒ 砲擊
聽從⇒ 遵從	聽從 ⇒ 討好	聽 ⇒ 偷聽	歡迎 ⇒ 行額手禮
彎曲 ⇒ 成弧形	彎下 ⇒ 蜷縮	彎下 ⇒ 擠緊	灑 ⇒ 滋潤
彎 ⇒ 卑躬屈膝	彎 ⇒ 彎腰	彎 ⇒ 斜倚	彎 ⇒ 傾斜
驅散 ⇒ 解散	辯論 ⇒質疑	辯解 ⇒ 掩飾	護送 ⇒ 隨侍
護送 ⇒ 防衛	露出 ⇒ 揭幕	露出 ⇒ 去除	露出 ⇒ 打開
轟炸 ⇒ 低空水平轟炸	屬於 ⇒ 附屬	闡明 ⇒ 澄清	釋放 ⇒ 保釋
轟炸 ⇒ 定形轟炸	釋放 ⇒ 假釋	釋放 ⇒ 解咒	觸怒 ⇒ 激怒
轟炸 ⇒ 俯衝轟炸	觸怒 ⇒ 冒犯	觸怒 ⇒ 激憤	警告 ⇒ 威脅
屬於 ⇒ 原來屬有	競選 ⇒ 重新競選	競選 ⇒ 登記	籌備 ⇒ 鋪設
籌備 ⇒ 架置	競爭 ⇒ 賽跑	競爭 ⇒ 玩	競爭 ⇒ 匹敵
競爭 ⇒ 爭取	獻給 ⇒書寫	懸掛 ⇒下垂	勸告⇒ 提醒
勸告 ⇒ 遏阻	勸告 ⇒ 當顧問	勸告 ⇒提出	關緊 ⇒ 鎖上
關緊 ⇒ 鞏固	關緊⇒ 繫住	關緊 ⇒ 使接合	關緊 ⇒ 夾住
關緊 ⇒ 拴緊	關緊 ⇒掛	關緊 ⇒ 鏈住	關緊 ⇒ 結成繩環
關緊 ⇒ 鉚接	關緊 ⇒ 釘	關緊 ⇒ 扣緊	關緊 ⇒ 鉤住
關緊 ⇒ 閂上	關閉 ⇒ 放下	關閉 ⇒密封	關閉 ⇒ 塞住
關閉 ⇒ 暫時休會	關心 ⇒ 看	證實 ⇒ 清查	證實 ⇒ 留意
證實 ⇒ 檢查	繪畫 ⇒ 人體彩繪	繪畫 ⇒ 畫壁畫	繪畫 ⇒ 劃線
繪畫 ⇒ 畫漫畫	繪畫 ⇒ 描繪	繪畫 ⇒ 畫輪廓	繪畫 ⇒ 亂畫
繪畫⇒ 圖示	穩定 ⇒平靜	穩定 ⇒ 定居	穩定 ⇒使堅定
離職 ⇒ 辭職	離職 ⇒ 退休	鎮靜 ⇒ 安撫	鎮靜 ⇒使平靜

轉變 ⇒ 使平行	轉變 ⇒ 編輯	轉變 ⇒ 改過自新	轉變 ⇒ 濃縮
轉變 ⇒ 鹼化	轉變 ⇒ 倒空	轉變 ⇒ 擴張	轉變 ⇒ 膠化
轉變 ⇒ 緩和	轉變 ⇒ 模糊	轉變 ⇒ 結束	轉變 ⇒ 影響
轉變⇒ 氧化	轉變 ⇒ 冷凍	轉變 ⇒ 變直	轉變 ⇒ 上弦
轉變⇒溶解	轉變 ⇒ 複雜化	轉變 ⇒ 軍事化	轉變 ⇒ 自然化
轉變 ⇒ 皂化	繞 ⇒ 解開	擺蕩 ⇒ 波動	斷裂 ⇒ 碎開
斷裂 ⇒ 起皺	斷裂 ⇒ 倒下	隱藏 ⇒ 隱藏	隱匿 ⇒ 逃避追緝
隱匿 ⇒ 藏匿	隱匿 ⇒ 埋入土中	隱匿 ⇒ 採低姿態	鞠躬 ⇒ 屈膝
演講 ⇒講道	演戲 ⇒比手劃脚	熄滅 ⇒斷電	漠視 ⇒不理睬
漂浮 ⇒再浮起	漂浮 ⇒順潮漂流	漫步 ⇒閒逛	移動 ⇒ 急馳
漫步 ⇒徘徊	演 ⇒充當配角	演 ⇒模仿	演 ⇒扮演
演 ⇒嘲諷模仿	演 ⇒激動表達	漏 ⇒船底漏水	歌頌 ⇒報佳音
歌頌 ⇒唱讚美詩	歌頌 ⇒唱小調	滾動 ⇒翻觔斗	滿足 ⇒解渴
滿足 ⇒適合	滿足 ⇒賴以維生	滿足 ⇒取悅	滲出 ⇒分泌
滲出 ⇒噴出	摺疊 ⇒弄皺	摺疊 ⇒擠壓	摺疊 ⇒打褶
摺疊 ⇒摺疊	摺疊 ⇒交叉	摺疊 ⇒起皺	摺疊 ⇒使癟掉
撤職 ⇒開除	撤職 ⇒解僱	撤職 ⇒逐出	撤職 ⇒強迫退休
敲詐 ⇒勒索	撤退 ⇒退潮	摸 ⇒脚尖觸摸	摸 ⇒嚙合
摸 ⇒飛速掠過	摸 ⇒按	摸 ⇒(棒球)觸殺	移動 ⇒飛
移動 ⇒環行	移動 ⇒前進	移動 ⇒繞行	移動 ⇒超速
移動 ⇒踩	移動 ⇒乘船	對待 ⇒勢利眼	對待 ⇒照料
對待 ⇒不予理會	對待 ⇒寵壞	對待 ⇒嘲弄	創作 ⇒繪畫
喝 ⇒啜飲	喝 ⇒吸	喝 ⇒牛飲	喝 ⇒舔
陳列 ⇒揮舞	陪 ⇒護送	陪 ⇒尾隨	陪 ⇒散步