Automatic labeling of troponymy for Chinese verbs

羅巧珊　Chiao-Shan Lo*+　　陳怡蓉　Yi-Rung Chen+
jcland0407@hotmail.com　　　yrsmile117@gmail.com

林芝佑　Chih-Yu Lin+　　　謝舒凱　Shu-Kai Hsieh*+
jyyow@msn.com　　　　　　shukai@gmail.com

*Lab of Linguistic Ontology, Language Processing and e-Humanities,
+Graduate School of English/Linguistics,
National Taiwan Normal University

Abstract

以同義詞集與詞彙語意關係架構而成的詞彙知識庫，如英語詞網 (Wordnet)、歐語詞網 (EuroWordnet)等，已有充分的研究，詞網的建構也已相當完善。基於相同的目的，中研院語言所亦已建立大規模之中文詞彙網路 (Chinese Wordnet,CWN)，旨在提供完整的中文辭彙之詞義區分。然而，在目前之中文詞彙網路系統中，由於目前主要是採用人為判定來標記同義詞集之間的語意關係，因此這些標記之數量尚未達成可行應用之一定規模。因此，本篇文章特別針對動詞之間的上下位詞彙語意關係 (Troponymy)，提出一種自動標記的方法。我們希望藉由句法上特定的句型 (lexical syntactic pattern)，建立一個能夠自動抽取出動詞上下位的系統。透過詞義意判定原則的評估，結果顯示，此系統自動抽取出的動詞上位詞，正確率將近百分之七十。本研究盼能將本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：中文詞彙網路、語義關係自動標記、動詞詞彙語義

Abstract

Synset and semantic relation based lexical knowledge base such as wordnet, have been well-studied and constructed in English and other European languages (EuroWordnet). The Chinese wordnet (CWN) has been launched by Academia Sinica basing on the similar paradigm. The synset that each word sense locates in CWN are manually labeled, however, the lexical semantic relations among synsets are not fully constructed yet. In this present paper, we try to propose a lexical pattern-based algorithm which can automatically discover the semantic relations among verbs, especially the troponymy relation. There are many ways that the structure of a language can indicate the meaning of lexical items. For Chinese verbs, we identify two sets of lexical syntactic patterns denoting the concept of hypernymy-troponymy relation. We describe a method for discovering these syntactic patterns and automatically extracting the target verbs and their corresponding hypernyms. Our system achieves satisfactory results and we beleive it will shed light on the task of automatic acquisition of Chinese lexical semantic relations and ontology learning as well.

Key word: troponymy, automatic labeling, lexical syntactic pattern

1   Introduction

In recent years, there has been an increasing focus on the construction of lexical knowledge resources in the field of Natural Language Processing, such as Thesaurus, Wordnets, Mindnet, Hownet, VerbNet, etc. Among these resources, Princeton WordNet[1], started as an implementation of a psycholinguistic model of the mental lexicon, has sparked off most interest both in theoretical and applicational sides. WordNet's growing popularity has prompted the modeling and construction of wordnet in other languages and various domains as well. However, 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. This has motivated researches into automatic methods paralleled with manual verification, in order to ease the work.

In Chinese, constructing a semantic relation-based wordnet is comparatively difficult owing to the fuzzy definition and classification among words, morphemes, and characters. The Chinese Wordnet (CWN)[2], created by Academia Sinica, aims to provide complete senses for each word based on the theory of lexical semantics and ontology. However, the synsets of each word in CWN are manually labeled and the semantic relations among synsets are not fully constructed. In this present paper, we try to propose an algorithm which can automatically label the semantic relations among verbs, especially focused on the hypernymy-troponymy relation. According to Fellbaum [2], lexical entries in a dictionary can reflect the relatedness of words and concepts. Such relations reflect the paradigmatic organization of the lexicon. Also, there are many ways that the structure of a language can indicate the meaning of lexical items.

This paper is organized as follows: In the next section we briefly outline the main research on the automatic discovery of lexical semantic relations, which motivates the present study. Then we discuss the concept of troponymy between verbs. Section 3 introduces our proposal and experiments. Section 4 shows the results and discussion of this method; Section 5 concludes this paper with future directions.

2   Literature Review

There has been a variety of studies on the automatic acquisition of lexical semantic relations, such as hypernymy/hyponymy [6], antonymy [7], meronymy [5] and so on. In Section 2.1 we will review Hearst's approach, which most of the works on automatic labeling of word sense relations are based upon. To the best of our knowledge, there is no study targeting at troponymy extraction yet, so in Section 2.2, we first define what troponymy is, the complexity of troponymy, and discuss how we can infer troponymy motivated by Hearst's approach.

2.1   Syntactic patterns and semantic relation

The structure of a lexical entry in a dictionary reflects the relatedness of words and concepts; also, certain structures or syntactic patterns usually define the semantic relation among each other. Hearst [6] proposed a **lexico-syntactic pattern** based method for automatic acquisition of hyponymy from unrestricted texts. Basing on a text corpus, which contains terms and expressions that are not defined in Machine Readable Dictionaries, she postulates six lexico-syntactic

---

[1]http://wordnet.princeton.edu
[2]http://cwn.ling.sinica.edu.tw/

patterns to automatically detect hypernymy-hyponymy relation and extract these pairs from the sentences. Lexico- syntactic patterns which denote the concept of "including" or "other than" may often possibly reveal the hypernymy-hyponymy relation. 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.

According to Miller [4] and Fellbaum [2], the lexical database WordNet resembles a thesaurus in that it represents word meanings primarily in terms of conceptual-semantic and lexical relation. A synset, therefore, is constructed by 'assembling a set of synonyms that together define a unique sense.' If one sense of a word is the same to another word, they share the same synsets and they are synonyms (at least partial synonyms). Ramanand and Bhattacharyya [8] hence use the concept of synset to suggests that 'if a word w is present in a synset along with other words $w_1, w_2, \dots w_k$, then there is a dictionary definition of w which refers to one or more of $w_1, w_2, \dots w_k$, and/or to the words in the hypernymy of the synset.' With this assumption, he applies groups of rules to validate synonymy and hypernymy relation among corpus. 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. Also, another rule detected partial hypernymy: many words in the wordnet are made up of more than one word, which are called 'multiwords'. In many cases, hypernyms or synonyms of such words are not entirely present in the definitions of words, but parts of them can be found in the definition.

2.2 Troponymy

As known, synsets in WordNet are connected with each other by various kinds of lexical semantic relations, such as Meronymy and Holonymy (between parts and wholes), Hypernymy and Hyponymy (between specific and more general synsets) and so on. Among them, the most important semantic relation in Wordnet is the hypernymy/hyponymy relation which links general and more specific concepts in both directions [7]. In Fellbaum's study [3], she defined the hyponymy-hypernymy relation among verbs as troponymy. Basing on this definition, troponymy, may at first sight appear like the relations of hypernymy/hyponymy among nouns: The subordinate concept contains the superordinate, but adds some additional semantic specification of its own. However, the semantic organization of verbs is more complicated than that of nouns and "the semantic relations so important for the organization of the noun lexicon are not the relations that hold among verbs" [4]. Hence, not all verbs can be placed under a single top node and verbs do not seem obviously related in a consistent manner like nouns do. According to Fellbaum and Miller [4], saying that troponymy is a particular kind of entailment involves temporal co-extensiveness for the two verbs. As known, entailment is a unilateral relation, taking *snore* and *sleep* for example, *snoring* entails *sleeping* but not the other way around. Although *snore* entails *sleep* and is included in *sleep*, we can not say that *snore* is a troponym of *sleep*; these two verbs are not in a hypernymy/ troponymy relation. Hence, for troponymy to hold, the essential factor is the co-extensiveness in time: one can sleep before or after snoring, but not necessarily happened at the same time. On the other hand, the activities denoted by the hypernym/ troponymy relation verbs must be coextensive in time. The following is an example from Lin et al [7]. *Reason* is a troponym of *think* because to *reason* is to *think* in a particular manner (logically). Therefore, the definition of *reason* naturally includes *to think "at the same time"* and thus inherits the property of *think*.

Beside the complicated distinction among verbs themselves, the troponymy relation is also different from the *is-a* relation among nouns in two ways [1]. 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. Rather than *kind*, troponymy seems to link verbs in a *manner* elaboration. Basing on the above properties of troponym, we postulate two syntactic patterns as the possible environments for discovering troponyms. More details will be discussed in the following section.

Our literature survey revealed that although some work had been done in automatically detection of hypernymy-hyponymy relation, none of them focus on hypernymy-troponymy relation of verbs. Therefore, in this paper, we attempt to propose a lexical pattern-based algorithm to tackle with this issue.

3   Algorithm

To automatically label the troponymy relation among verb senses, in the following, we propose an algorithm which applies three main steps and two rules.

3.1   First step: finding word senses

Most of words have more than one sense, and each sense of a given word might have their different hypernyms and troponyms. Therefore, to find semantic relations among verbs, our first step is to extract the definitions of each verb, by using web search. The input data used here composed 168 verbs which were extracted from *Sinica Corpus*. Although the 168 input verbs were randomly chosen by the authors, we firstly delimited out inputs labeled with syntactic categories `VA,` `VC` and `VAC,` respectively[3] for they contain most verbs that are commonly used. We then do search queries of each verb on Chinese Wordnet. If the result for a given word cannot be found here, then turn to the online version of the MOE Revised Chinese Dictionary[4] to find each sense of a given verb.

3.2   Second step: word segmentation

For later rule application, our next step is to do segmentation and POS (part-of-speech) tagging in each of the verbs' definitions via the online CKIP Chinese word segmentation system[5]. The following example shows the segmented result of one target verb and its definition:

買 (VAC) : 以 (P)   金錢 (Na)   購進 (VC)   物產 (Na)
　mai :　 yi　 jingqien　 gojin　 wuchan
　to buy :　with　 money　 purchase　 products
　　　　to buy : to purchase products with money.

By doing this, each word in this definition is segmented and POS tagged. After this step is done, our input data are established, which includes different entries[6] of each words' senses and all of the words are segmented and POS tagged.

3.3 Third step: Apply Rule 1 and Rule 2

After getting each of the verb's definitions, we propose two rules to find a given verb sense's hypernym.

3.3.1 Rule 1 Application

Rule 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 of '以 *(yi)* / 用 *(yong)* … V_j… *(by/with …. to V_j)*'.

Our first rule is, when a verb (V_i)'s definition contains the pattern '以 *(yi)*… V_j …' or '用 *(yong)* … V_j … ' in a sentence, we could take this verb V_j or take all these verbs in this sentence out if there is more than one verb in this sentence. The verb(s) V_j could be labeled as a hypernym of V_i. For example, the definition of the verb 走 (zou, 'to walk') correspond to this pattern:

走 : 以(P) 兩(Neu) 腿(Na) 交互(D) 向(P) 前(Ncd) 移動(VAC)
zou : yi   lian   tuei   jiaohu   xian qien   yidong
to walk: by two   feet   mutually   towards front   to move
to walk: moving forwards by two feet.

Thus, 移動 (yidong, 'to move') could be labeled as the *hypernym* of the verbs 走 (zou, 'to walk').

3.3.2 Rule 2 Application

Rule 2: Deriving from the *is-a* relation of noun phrases, we may assume that, for verbs, a troponym is a certain way or a specific manner of its hypernym. Hence, the definition might appear in the pattern of '一種 *(yizhong)* … W_j 方式 *(fangshi)* (a way of W_j).

The second rule is, if a verb (V_i)'s definition contains the pattern '一種 *(yizhong)* … W_j 方式 *(fangshi)* (a way of W_j)*, then we could label this nominalized verb W_j as a hypernym of V_i. For example, the definitions of the verb 煎 (jian, 'to fry') is:

煎: 一(Neu) 種(Nf) 烹飪(VC) 方式(Na)
jian: yi   zhong   pengren   fangshi
to fry: one   kind   to cook   way
to fry: a way of cooking.

Since they follow the Rule 2 pattern, the verb 烹飪 (penren, 'to cook') is the hypernym of the verb 煎 (jian, 'to fry'), standing for a specific manner of cooking.

---

[6]Different entries in the data are separated by the step of tokenization.
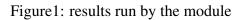
3.4 Algorithmic Representation

The proposed algorithm outlined above could be summarized as the following:

Input: Verbs V_1, V_2, ... and V_n, web search for each definition of them.
Output: Predicted hyponyms between 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**
      $\mid$ 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**
      $\mid$ label the nominalized verb W_j as a hypernym of V_i ;
    end
  end
end

Algorithm 1: Algorithm to automatic labelling of troponomy

4 Experiment and Result

We implement our proposed method in Python (2.5.2). The module at first aims to extract entries containing our targets: '以 (yi) (P) / 用 (yong) (P) / 一種 (yi zhong) (Nf)' in the definition. Afterwards, all the verbs that occur after the targets will be extracted as the possible candidates. Figure 1 illustrates some of the results of a run of the labeling algorithm on Python, where the verbs occurred before the symbol '@' are the input verbs and other verbs occurred after '@' are their possible hypernyms.

```
IDLE 1.2.2      \verb==== No Subprocess \verb====

抖3(VAC)@
  重複(VC)
  晃動(VAC)
搖晃2(VAC)@
  搖動(VC)
震2(VAC)@
  震動(VAC)
  搖撼(VC)
...
```

Figure1: results run by the module

5   Evaluation and Discussion

To evaluate the system, we adopt the **substitution tests** [9] to examine whether a predicted pair of verbs has the relationship of troponymy. The specific substitution test that we implement is introduced in section 5.1. The acceptability of the sentences which contain the pairs of verbs we are examining is decided manually by 3 linguists. In section 5.2, we calculate the precision rate to properly evaluate the system. Some problems and further directions of our work are discussed in section 5.3.

5.1   The Substitution Test

Substitution test is commonly used in linguistic literature [9]. We apply the sentence pattern to the possible hypernyms found by our module: "如果他在 V_1 -ing, 那麼他便是在 V_2 -ing" (If he is V_1 -ing, then he is V_2 -ing). If the sentence "If he is V_1 -ing, then he is V_2 -ing" is always true but the other way around "If he is V_2 -ing, then he is V_1 -ing" is not, we say that V_2 is V_1's hypernym and V_1, on the other hand, is V_2's troponym. For example, we place the two verbs 走 (zou, 'to walk') and 移動 (yidong, 'to move') into this sentence. The result is that "If he is walking, then he is moving" is true, and that "If he is moving, then he is walking" is not necessarily true because if he is moving, he can also be running. We can therefore describe that 走 (zou, 'to walk') and 移動 (yidong, 'to move') are in a trponymy relation.

   The reason for testing both these two sentences is to avoid the synonym pairs. We take two synonymous verb 立 (li, 'to stand up') and 站 (zhan, 'to stand') for example. "If he is standing up, then he is standing" is true, but (to stand up) is not (to stand)'s troponym. On the contrary, if we test by two sentences, this situation can be avoided. The second sentence "If he is standing, then he is standing up" is also true. Thus they are not in the relation of troponymy. By placing the verbs in the sentence pattern, we manually evaluated all the possible hypernyms found by our system and then we calculated the precision rate of the system.

5.2   Evaluation

The precision rate of our system is calculated as `# of correct answers given by system` / `# of answers given by system` $= \dfrac{93}{133} = 69.9\%$. There are totally 133 possible hypernyms returned by our system, and after manually filtered with the substitution test, there are 93 verbs left. The system shows that there are still some unwanted or incorrect hypernyms returned by our system. This will be further discussed in the following section. Although our system could have further improvement, but the high precision rate shows that our system has certain quality of performance and the design of our system is in the right direction.

5.3   Problems

There are two major problems in our system. First is the problem of synonym. In our input, the possible hypernyms returned by our system may be the synonym to the input verb. For example: for the input 揮 (huei, 'to wave'), two possible hypernyms were returned by the system, including 揮 (VC) (huei, 'to wave'), and 移動 (VAC) (yidong, 'to move').The underlined is an undesired synonym. Such instances will lower the precision rate. The second problem, and might be the most difficult one, is the ambiguity of the preposition 以 (yi, 'by or woth') and 用 (yong, 'by or with'). In CWN, when 以 (yi, 'by or with') serves as a preposition (it could

also serve as conjunction or modal), there are over 20 different polysemous meanings including *'because of'*, *'according to'*, *'in order to'*, *'with'* etc. In this case, we have to determine which meaning the preposition '以 (yi) / 用 (yong)' belongs to and do the sense determination manually. This could remain a main problem for computational linguistics since so far, the disambiguation of these polysemy could not be fully solved using any system or algorithm.

Yet it is also very likely that hypernyms exists in inputs other than the patterns that we suggest. How to include those instances based on other methods, for example, by taking suffix-like forms as an indicator of troponymy (想 / 細想、回想 etc), is what we will consider in the future research.

6 Conclusion

In this study, we have suggested an automatic labeling method of troponymy for Chinese verbs. We identify the lexical-syntactic pattern '以 (yi) / 用 (yong)... V_j ... (by/with ....to V_j)' or '一種 (yizhong) ... W_j 方式 (fangshi) (a way of W_j)' that occur in the definition for V_i to see whether V_j and W_j indicate the troponymy relation of V_i. Though the range of the data is not exceedingly extensive, our approach has the advantages of low-cost and less-effort over other methods for automatic acquisition of lexical semantic relations from unrestricted text. Possible future works could be test the similar methods on the hypernymy/hyponymy relations among nouns. We beleive the comparison of the results will speed up the construction of Chinese lexical semantic relations knowledge base and ontology as well.

References

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

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

[3] 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.

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

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

[6] Marti 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.

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

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

[9] Bo-Sheng Tsai, Chu-Ren Huang, Shu-Chuan Tseng, Jen-Yi Lin, Keh-Jiann Chen, and Yuan-Shun 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.