A Rule-Based and Corpus-Oriented Approach to Prepositional Phrases Attachment

Kuang-hua Chen and Hsin-Hsi Chen

Department of Computer Science and Information Engineering

National Taiwan University

Taipei, Taiwan, R.O.C.

khchen@nlg.csie.ntu.edu.tw, hh_chen@csie.ntu.edu.tw

Abstract

Prepositional Phrase is the key issue in structrual ambiguity. Recently, researches in corpora provide the lexical cue of association among prepositions and other words and these information could be used to resolve partly ambiguity resulted from prepositional phrases. Two possible attachments are considered in the existing approaches: either noun attachment or verb attachment. In this paper, four different attachment are told out according to their functionalities: noun attachment, verb attachment, sentence-level attachment, and predicate-level attachment. Both lexical knowledge and semantic knowledge are envolved resolving attachment in the proposed mechanism. Experimental results show that considering more types of presitional phrases is useful in machine translation.

1 Introduction

Prepositonal phrases are usually ambiguous. The well- known sentence shown in the following is a good example.

Kevin looked at the girl with telescope.

Whether the prepostional phrase **with telescope** modifies the head noun **girl** or the verb **look** are not resolved by using only one knowledge source. Many researcheres observe text corpora and learn some knowledge based on language model to determine the plausible attachment. For example, we could expect that the aforementioned prepositional phrase is usually attached to verb according to text corpora. However, the correct attachment is dependent of world knowledge sometimes.

Some approaches to determination of PPs are reported in literature [Kimball, 1973; Frazier, 1978; Ford *et al.*, 1981; Shieberk, 1983; Wilk *et al.*, 1987; Liu, *et al.*, 1990; Chen and Chen, 1992; Hindle and Rooth, 1993; Brill and Resnik, 1994]. These resolutions fall into three categories: syntax-based, semantics-based and corpus-based approaches. The brief discussion are described in the following:

1. Syntax-based

- Right Assocation [Kimball, 1973]

  The PPs always modifies the nearest component preceding it.

- Minimal Attachment [Frazier, 1978; Shieber, 1983]

  The correct attaching point of a PP in a sentence is determined by the number of nodes in parsing tree.

2. Semantics-based

- Lexical Preference [Ford *et al.*,1982]

  The real attaching point must satisfy some constraints, e.g., verb features. Different verb accompany with the same PPs may have the different attaching points. The following sentences show the different attaching points.

  I eat an apple <u>on the table</u>.

  I put an apple <u>on the table</u>.

- Preference Semantics [Wilk *et al.*, 1987]

  The method tell us the real attaching point must be determined by the preference of verbs and prepositions.

3. Corpus-based

- [Liu *et al.*, 1990]

They use semantic score and syntactic score to determine the attaching point. These scores are trained from text corpora.

- Lexical Association [Hindle and Rooth, 1993]

  This method applies statistical techniques to discover the lexical association from text corpora. Thus, the attchment of PPs is determined.

- Model Refinement [Brill and Resnik, 1994]

  Their approach assumes every PP modifies the immediately previous noun and uses rules trained from text corpora to change the errone attachments.

In the sections what follows, we will first present our perspective from machines translation to this problem. Section 3 will discuss the detail resolution to PPs attachment, which considers more different attachments. Section 4 will conduct a series of expeiments to investigate our approach. Finally, a few concluding remarks bring this paper to an end.

2 Our Perspective

From the viewpoint of machine translation, in particular, English-Chines machine translaton [Chen and Chen, 1995], the main shortcoming of the approaches mentioned in previous section is that they all consider either PPs modify nouns or PPs modify verbs. Although, PPs usually modify nouns or verbs, there are some counter examples even in the simple sentences like "there is a book on the table" and "The apple has worm in it". In the first example, the PP "on the table" is neither used to modify the copula verb nor the noun phrase "a book". It describes the situation of the whole sentence. The second example shows that the PP "in it" is also not a modifier, but a complement to the preceding noun phrase. That is, the PP has a nonrestrictive usage. To transfer PPs among different languages, we must capture the correct interpretation. Therefore, we distinguish four different prepositional phrases.

- Predicative Prepositional Phrases (PPP): PPs that serve as predicates.

  He is <u>at home</u>.

  他<u>在家</u>。

  He found a lion <u>in the net</u>.

  他發現獅子<u>在網子裡</u>。

- Sentential Prepositional Phrases (SPP): PPs that serve functions of time and location.

  There is no parking <u>along the street</u>.

  <u>這條街上</u>禁止停車。

  We had a good time <u>in Paris</u>.

  <u>在巴黎</u>我們有一段美好的時光。

- Prepositional Phrases Modifying Verbs (VPP)

  I went to a movie <u>with Mary</u>.

  我<u>和瑪莉</u>去看電影。

  I bought a book <u>for Mary</u>.

  我<u>為瑪莉</u>買了一本書。

- Prepositional Phrases Modifying Nouns (NPP)

  The man <u>with a hat</u> is my brother.

  <u>戴帽子</u>的人是我哥哥。

  Give me the book <u>on the desk</u>.

  把<u>桌上</u>的書給我。

It is obvious that these four different prepositional phrases have their own appropriate positions in Chinese. That is if we determine the type of a prepositional phrase, its position in Chinese is also determined.

3 The Resolution to PP-Attachment

In the previous section, four different types of PPs are defined according to their functionalities. Thus, the resolution to this problem is to determine which type the PPs belong to. The basic steps are:

- Check if it is a PPP .
- Check if it is an SPP.
- Check if it is a VPP.
- Otherwise, it is an NPP.

Now, the problem is what constitutes the mechanism of each step.

Oxford Advanced Learner's Dictionary (OALD) [Hornby, 1989] uses 32 verb patterns to describe the usage of verbs. Table 1 lists these verb patterns. Since some frames defined by OALD cannot be distinguished from each other by the contexts, we define the hyper types for these frames. These definitions are shown in Table 2. That is to say, Ln and Tn will be recognized as Vn and the real type is decided by the verb itself and the type. If a verb is a linking verb and its type is Vn, its real type would be Ln. In addition, the subcategorization frame In/pr is regarded as In and Ipr. [Chen and Chen, 1994] has proposed a method to determinate the predicate argument structure of an sentence. Once the predicate-argument structure is judged as Vnpr, Ipr, Dprf, Dprt, or Dprw, the underlying prepositional phrase is PPP.

As for SPP, VPP, and NPP, the rules are dependent of the lexical knowledge and semantic usage. That is to say, the semantic tag should be assigned to each word.. [Chen and Chen, 1992] describes the semantic hierarchy for noun and verb (shown in Figure 1 and Figure 2). However, manually building a lexicon with semantic tag information is a time-consuming and human-intensive work. Fortunately, an on-line thesaurus provides this information. Roget's thesaurus defines a semantic hierarchy with 1000 leaf nodes shown in Table 3. Each leaf node contain words with this semantic usage, that is, these words have the semantic tags represented by these leaf nodes. We just map these leaf nodes to the semantic defintions listed in Figure 1 and Figure 2. Therefore, nouns and verbs in running texts could be easily assigned semantic tags in our semantic definitions.

In general, four factors contribute the determination of PP-attachment: 1) verbs; 2) accusative nouns; 3) prepositions; and 4) oblique nouns. We use a 4-tuple <*V*, *N*1, *P*, *N*2> to denote the relationship of PP attachment, where *V* denotes semantic tag of verbs, *N*1 denotes the semantic tag of accusative noun, *P* denotes the preposition and *N*2 denotes the semantic tag of oblique noun. For example, the following sentece has the 4-tuple <non_speech_act, human, with, object>.

Kevin watched the girl with telescope.

Having the 4-tuple in advance, we could apply 65 rule-templates listed in Appendix to determine what the PP type is by aforementioned steps. That is, apply SPP rule-template first, and then VPP rule-template. If none succeeds, the PP should be an NPP. We summarize the algorithm as follows:

Algorithm 1:
Resolution to PP-Attachment

(1) Check if it is a PPP according to the predicate-argument structure.

(2) Check if it is an SPP according to 21 rule-templates for SPP.

(3) Check if it is a VPP according to 44 rule-templates for VPP.

(4) Otherwise, it is an NPP.

Table 1 The Predicate-Argument Structures Defined in OALD

Types	Subcategorization Frames	Types	Subcategorization Frames
La	Linking verb + adj.	Tng	Transitive verb + noun + -ing form
Ln	Linking verb + noun	Tni	Transitive verb + noun + infinitive
I	Intransitive verb	Cna	Complex-transitive verb + noun + adj.
Ipr	Intransitive verb + prep.	Cnn	Complex-transitive verb + noun + noun
Ip	Intransitive verb + particle	Cnn/a	Complex-transitive verb + noun + as + noun (adj.)
In/pr	Intransitive verb + noun or prep.	Cnt	Complex-transitive verb + noun + to-infinitive
It	Intransitive verb + to-infinitive	Cng	Complex-transitive verb + noun + -ing form
Tn	Transitive verb + noun	Cni	Complex-transitive verb + noun + infinitive
Tnpr	Transitive verb + noun + prep.	Dnn	Double-transitive verb + noun + noun
Tnp	Transitive verb + noun + particle	Dnpr	Double-transitive verb + noun + prep.
Tf	Transitive verb + finite that-clause	Dnf	Double-transitive verb + noun + finite that-clause
Tw	Transitive verb + wh-clause	Dprf	Double-transitive verb + prep. + finite that-clause
Tt	Transitive verb + to-infinitive	Dnw	Double-transitive verb + noun + wh-clause
Tnt	Transitive verb + noun + to-infinitive	Dprw	Double-transitive verb + prep. + wh-clause
Tg	Transitive verb + -ing form	Dnt	Double-transitive verb + noun + to-infinitive
Tsg	Transitive verb + noun's + -ing form	Dprt	Double-transitive verb + prep. + to-infinitive

Table 2 Hypertypes for Predicate-Argument Structures

Hypertypes	Types	Hypertypes	Types	Hypertypes	Types
Vnpr	Dnpr, Tnpr	I	I	Cna	Cna
Vn	Ln, Tn	Ipr	Ipr	Cnn/a	Cnn/a
Vnt	Cnt, Dnt, Tnt	Ip	Ip	Dnf	Dnf
Vni	Cni, Tni	Tnp	Tnp	Dprf	Dprf
Vng	Cng, Tng	Tf	Tf	Dnw	Dnw
Vnn	Cnn, Dnn	Tw	Tw	Dprw	Dprw
Vt	It, Tt	Tg	Tg	Dprt	Dprt
La	La	Tsg	Tsg		

Table 3 Classification of Roget's Thesaurus

CLASS	SECTION	TAG	CLASS	SECTION	TAG
ABSTRACT RELATIONS	Existence	1 - 8	SPACE	In General	180 - 191
	Relation	9 - 24		Dimensions	192 - 239
	Quantity	25 - 57		Form	240 - 263
	Order	58 - 83		Motion	264 - 315
	Number	84 - 105	MATTER	In General	316 - 320
	Time	106 - 139		Inorganic	321-356
	Change	140 - 152		Organic	357 - 449
	Causation	153 - 179	AFFECTIONS	In General	820 - 826
INTELLECT	Formation of Ideas	450 - 515		Personal	827 - 887
	Communication of Ideas	516 - 599		Sympathetic	888 - 921
VOLITION	Individual	600 - 736		Moral	922 - 975
	Intersocial	737 - 819		Religious	975 - 1000

- entity
  - + concrete
    - + animate
      - + human (e.g. boy)
      - – human (e.g. cat)
    - -animate
      - object (e.g. card)
      - vehicle (e.g. car)
  - -concrete
    - + adv
      - frequent (e.g. time)
      - manner (e.g. way)
      - space
        - location (e.g. bookstore)
        - direction (e.g. south)
        - dimension (e.g. width)
      - time (e.g. April)
      - order (e.g. regularity)
    - -adv
      - event (e.g. earthquake)
      - motion (e.g. transfer)
      - number (e.g. dozen)
      - abstract (e.g. fact)
      - form (e.g. frame)
      - product (e.g. writing)
      - religion (e.g. heaven)
      - sensation (e.g. pain)
      - volition (e.g. will)

Figure 2 Semantic Tags for Nouns

- happening
  - state
    - -mental (e.g. cost, have, own)
    - +mental (e.g. know, think, like)
  - linking (e.g. become, grow, look)
  - perception (e.g. see, taste, feel)
  - action
    - act
      - +mental (e.g. realize, understand, recognize)
      - -mental
        - -speech act (e.g. catch, hit, kill)
        - +speech act (e.g. say, tell, state)
    - activity
      - -mental
        - - montion (e.g. work, drive, draw)
        - +montion (e.g. come, fall, go)
      - +mental (e.g. remember, learn, read)

Figure 1 Semantic Tags for Verbs

5 Experiments

The Penn Treebank [Marcus *et al.*, 1993] is used as the testing corpus, which consists of over 4.5 million words of American English. The following is a real example extracted from this treebank.

```
(
    (S  (ADVP (NP Next week) )
        (S
        (NP (NP some inmates)
            (VP released
                (ADVP early)
                (PP from
                    (NP the Hampton County
jail
                        (PP in
                            (NP
Springfield))))))
        will be
        (VP wearing
            (NP (NP a wristband)
                (SBARQ
                 (WHNP that)
                 (S (NP T)
                    (VP hooks up
                        (pp with
                            (NP a special
jack
                                (PP on

(NP their home phones)))))))))))
        .)
```

The PPs contained in Penn Treebank are collected and associated with one label of PPP, SPP, VPP, or NPP. For example, the aforementioned PPs are extracted as follows.

<from the Hampton County jail, VPP>
<in Springfield, NPP>
<with a speical jack, VPP>
<on their home phones, NPP>

These extracted PPs constitute the standard set and then the attachment algorithm shown in previous section are applied to attaching the PPs. Finally, the attached PPs are compared to the standard set for performance evaluation. The results are shown in Table 8.

Table 8 Experimental Results

	Total	Correct
SPP	750	750
VPP	6392	4923
NPP	7230	7230
PPP	387	387
Total	14750	13290

6 Concluding Remarks

Many approaches are proposed to build predicate-argument structures in dictionaries automatically. The statistics-based approach and linguistic theory are integrated to determine the predicate-argument structures in the chapter. These two kinds of methods are complementary. Statistics-based approach is robust. It provides simple language models to analyze unrestricted texts. However, it may need large completely-annotated corpus to treat complex linguistic phenomena. Linguistic theory gives such a supplement. Well-formed patterns can be explained properly by universal principles, so that they can be formulated in terms of rules easily. The experimental results show that the integrated mechanisms are useful for further researches on large volume of real texts.

Reference

Aone C. and McKee, D. (1993), "Acquiring Predicate-Argument Mapping Information from Multilingual Texts," *Proceedings of Workshop on Acquisition of Lexical Knowledge from Text*, 1993, pp. 107-116.

Brent, M. (1991), "Automatic Acquisition of Subcategorization Frame from Untagged Text," *Proceedings of the 29th Annual Meeting of ACL*, 1991, pp. 209-214.

Brill, E. and Resnik, P. (1994), "A Rule-Based Approach to Automated Prepositional Phrase Attachment Disambiguation," *Proceedings of COLING-94*, 1994, pp. 1198-1204.

Chen, K.H. and Chen, H.H. (1992b), "Attachment and Transfer of Prepositional Phrases with Constraint Propogation," *Computer Processing of Chinese and Oriental Languages: An International Journal of the Chinese language Computer Society*, 6(2), 1992, pp. 123-142.

Ford, M.; Bresnan, J and Kaplan, R. (1982), "A Competence-Based Theory of Syntactic Closure," The Mental Representation of Grammatical Relations (Bresnan, J. Eds.), MIT Press, 1982, pp. 727-796.

Frazier, L. (1978), *On Comprehending Sentences: Suntactic Parsing Strategies*, Doctoral Dissertation, University of Connecticut, 1978.

Hindle, D. and Rooth, M.. (1993), "Structural Ambiguity and Lexical Relations," *Computational Linguistics*, 19(1), 1993, pp. 103-120.

Hornby, A.S. (1989), *Oxford Advanced Learner's Dictionary*, Oxford University Press, 1989.

Kimball, J. (1973), "Seven Principles of Surface Structure Parsing in Natural Language," *Cognition*, 2, 1973, pp. 15-47.

Liu, C.L.; Chang, J.S. and Su, K.Y. (1990), "The Semantic Score Approach to the Disambiguation of PP Attachment Problem," *Proceedings of ROCLING-90*, Taiwan, R.O.C., 1994, pp. 253-270.

Manning, C. (1993), "Automatic Acquisition of a Large Subcategorization Dictionary from Corpora," *Proceedings of the 31st Annual Meeting of ACL*, 1993, pp. 235-242.

Shieber, S (1983), "Sentence Disambiguation by a Shift-Reduced Parsing Technique," *Proceedings of IJCAI-83*, Kahlsruhe, Germany, 1983, pp. 699-703.

Ushioda, A.; Evans, D.; Gibson, T. and Waibel, A. (1993), "The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora," *Proceedings of Workshop on Acquisition of Lexical Knowledge from Text,* 1993, pp. 95-106.

Wilks, Y; Huang, X.H. and Fass, D. (1985), "Syntax, Preference and Right Attachment," *Proceedings of IJCAI-85*, Los Angeles, CA, 1985, pp. 779-784.

Appendix

The following lists rule-templates for PP-attachment. Every template consists of four elements <*V*, *N*1, *P*, *N*2>. The curl bracket pair denotes *OR*, the underline denotes *DON'T CARE* and ~ denotes *NOT*.

I. Rule-template for SPP

1. <___, ___, about, time>
2. <___, ___, across, location>
3. <___, ___, after, time>
4. <___, ___, along, location>
5. <___, ___, among, location>
6. <___, ___, at, {location, time}>
7. <___, ___, before, time>
8. <___, ___, between, {location, time}>
9. <___, ___, by, time>
10. <___, ___, during, time>
11. <___, ___, in, {location, time}>
12. <___, ___, in_front_of, location>
13. <___, ___, near, location>
14. <___, ___, next_to, location>
15. <___, ___, on, time>
16. <___, ___, out_of, {abstract, location}>
17. <___, ___, over, {location, time}>
18. <___, ___, through, {abstract, event, time}>
19. <___, ___, under, time>
20. <___, ___, with, abstract>
21. <___, ___, without, abstract>

II. Rule-template for VPP

1. <motion, ___, about, {object, location}>
2. <at_ment, ___, about, object>
3. <action, event, after, concrete>
4. <at_ment, {abstract, event}, after, {event, no, time}>
5. <motion, ___, across, {location, object}>
6. <{at_nonmen, ai_nonmen}, ___, along, {location, object}>
7. <~motion, ~{concrete, location}, among, {concrete, location}>
8. <~{at_nonmen, ai_nonmen}, ___, at, {animate, object}>
9. <{at_nonmen, ai_nonmen}, ___, at, {location, object}>
10. <action, event, after, concrete>
11. <at_ment, {abstract, event}, after, {event, no, time}>
12. <{at_nonmen, ai_nonmen}, {event, object}, between, {abstract, concrete, location}>
13. <{at_nonmen, ai_nonmen}, {event, object}, between, time>
14. <motion, ___, by, {location, object}>
15. <___, ___, by, manner>
16. <~motion, ___, by, {location, object}>
17. <___, ___, by, {abstract, event, object, vehicle}>
18. <___, ___, by, animate> passive voice
19. <___, ___, for, time>
20. <motion, ___, for, location>
21. <~linking, ___, for, {abstract, concrete, event}>
22. <___, ___, for, {abstract, event, object}>
23. <___, ___, for, animate}>
24. <{motion, speech_act}, ___, from, entity>

25. <motion, ___, in, {location, object}>
26. <___, ___, in, time>
27. <___, ___, in, vehicle>
28. <ai_nonmen, ___, in_front_of, {concrete, location}>
29. <ai_nonmen, ___, inside, {concrete, location}>
30. <___, ___, into, {abstract, concrete, location}>
31. <act, ___, like, ___>
32. <___, ~{location, object}, near, {location, object}>
33. <___, ~{location, object}, next_to, {location, object}>
34. <{at_nonmen, ai_nonmen}, {event, object}, on, {concrete, location, object}>
35. <___, ___, on, event>
36. <___, ___, on_to, {location, object}>
37. <motion, ___, out_of, {concrete, location}>
38. <___, ___, over, {abstract, event}>
39. <motion, ___, over, {concrete, location}>
40. <ai_nonmen, ___, through, {location, object}>
41. <{ai_nonmen, at_nonmen}, ___, under, {abstract, object}>
42. <___, ___, until, time}
43. <at_nonmen, ___, with, object>
44. <at_nonmen, ___, with, animate>
43. <at_nonmen, ___, without, object>
44. <at_nonmen, ___, without, animate>