

A Rule-Based and MT-Oriented Approach to Prepositional Phrase Attachment

Kuang-hua Chen and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University

1, SEC. 4, Roosevelt RD., 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 structural ambiguity. Recently, researches in corpora provide the lexical cue of prepositions with other words and the information could be used to partly resolve ambiguity resulted from prepositional phrases. Two possible attachments are considered in the literature: either noun attachment or verb attachment. In this paper, we consider the problem from viewpoint of machine translation. Four different attachments are told out according to their functionality: noun attachment, verb attachment, sentence-level attachment, and predicate-level attachment. Both lexical knowledge and semantic knowledge are involved resolving attachment in the proposed mechanism. Experimental results show that considering more types of prepositional phrases is useful in machine translation.

1 Introduction

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

Kevin watched the girl with a telescope.

Whether the prepositional phrase with a telescope modifies the head noun girl or the verb watch are not resolved by using only one knowledge source. Many researchers 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 on world knowledge sometimes.

Some approaches to determination of PPs are reported in literature (Kimball, 1973; Frazier, 1978; Ford et al., 1982; Shieber, 1983; Wilks et al., 1985; Liu et al., 1990; Chen and Chen, 1992; Hindle and Rooth, 1993; Brill and Resnik, 1994). The possible attachment they consider are NOUN attachment and VERB attachment. 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 Association (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 a parsing tree.

2. Semantics-based

- Lexical Preference (Ford et al., 1982)
The real attaching point must satisfy some constraints, e.g., verb features. Different verbs accompanying with the same PPs may have the different attaching points.
- Preference Semantics (Wilks et al., 1985)
Wilks and his colleagues argue the real attaching point must be determined by the preference of verbs and prepositions.
- Propagated Semantics (Chen and Chen, 1992)
The attachment of prepositional phrase is co-determined by the semantic usage of noun, verb, and preposition.

3. Corpus-based

- Statistical Score (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 attachment 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 erroneous attachments.

These approaches manage to resolve the PP attachment via only one language consideration. In contrast, we investigate this problem from viewpoint of machine translation and do not restrict ourselves in two possible attachment choices.

In the sections what follows, we will first present our viewpoint from machine translation to this problem. Section 3 will discuss the detail resolution to PPs attachment, which considers more different attachments. Section 4 will conduct experiments to investigate our approach. Section 5 will provide some concluding remarks.

2 Our Viewpoint from MT

From the viewpoint of machine translation, in particular, English-Chinese machine translation (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 PPs (PPP): PPs that serve as predicates.
He is at home.
Ta1 zai4 jia1.
He found a lion in the net.
Ta1 fa1xian4 shi1zi5 zai4 wang3zi5 li3.

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

There is no parking along the street.
Zhe4 tiao2 jie1 shang4 jin4zhi3 ting2 che1.
We had a good time in Paris.
Zai4 ba1li2 wo3men5 you3 yi1 duan4 mei3hao3 de5 shi2guang1.

- PPs Modifying Verbs (VPP)

I went to a movie with Mary.
Wo3 han4 ma3li4 qu4 kan4 dian4ying3.
I bought a book for Mary.
Wo3 wei4 ma3li4 mai3 le5 yi4 ben3 shu1.

- PPs Modifying Nouns (NPP)

The man with a hat is my brother.
Dai4 mao4zi5 de5 ren2 shi4 wo3 ge1ge5.
Give me the book on the desk.
Ba3 zhuo1 shang4 de5 shu1 gei3 wo3.

It is obvious that these four different prepositional phrases have their own appropriate positions in Chinese. That is after we determine the type of a prepositional phrase, the constituent to which PP is attached is known and its corresponding position in Chinese is also determined.

3 Resolving PP-Attachment

In the previous section, four different types of PPs are defined according to their functionality. 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) defines 32 different verb patterns to describe the usage of verbs. These verb frames are like skeleton of a sentence and the constituents are the flesh of sentence. Chen and Chen (1994) have proposed a method to determine the predicate argument structure of a sentence. The OALD-defined verb frames are regarded as the primary language knowledge source and an NP parser and a finite-state mechanism are cooperatively used to determine the plausible predicate-argument structure. Once the predicate-argument structure of a sentence contains prepositional phrase, the underlying prepositional phrase is PPP.

As for SPP, VPP, and NPP, the rules are dependent on the lexical knowledge and semantic usage. That is to say, the semantic tag should be assigned to each word. Figure 1 and Figure 2 describe the semantic hierarchy for noun and verb. 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 1. 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 definitions 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 $\langle V, N1, P, N2 \rangle$ to denote the relationship of a possible PP attachment, where V denotes semantic tag of verbs, N1 denotes the semantic tag of accusative noun, P denotes the preposition and N2 denotes the semantic tag of oblique noun. For example, the following sentence has the 4-tuple $\langle non_speech_act, human, with, instrument \rangle$.

Kevin watched the girl with a telescope.

Having the 4-tuple in advance, we could apply 67 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 46 rule-templates for VPP.
- (4) Otherwise, it is an NPP.

4 Experiments

The Penn Treebank (Marcus et al., 1993) is used as the testing corpus. 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 PPs contained in the aforementioned sentence are extracted as follows.

<from the Hampton County jail, VPP>

<in Springfield, NPP>

<with a special 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 these PPs. Finally, the attached PPs are compared to the standard set for performance evaluation. The results are shown in Table 2.

	Total	Correct
SPP	750	750
VPP	6392	4923
NPP	7230	7230
PPP	387	387
Total	14759	13290

Table 2: Experimental Results

First, NPP and VPP dominate the distribution of PPs (92%). The former occupies 49% population and the latter 43%. If we carefully process NPP and VPP, the result would be good. In fact, the proposed algorithm is based on the philosophy of model refinement. That is, we assume each PP is NPP except it is a PPP or it matches the 67 rule-templates. Table 2 shows that each NPP is

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	MATTER	Motion	264-315
	Number	84-105		In General	316-320
	Time	106-139		Inorganic	321-356
	Change	140-152		Organic	357-449
	INTELLECT	Causation	153-179	AFFECTIONS	In General
Formation of Ideas		450-515	Personal		827-887
VOLITION	Communication of Ideas	516-599	Sympathetic		888-921
	Individual	600-736	Moral		922-975
	Intersocial	737-819	Religious		975-1000

Table 1: Classification of Roget's Thesaurus

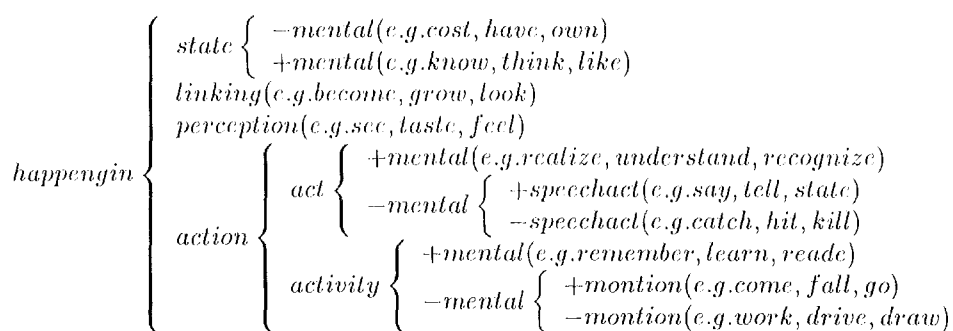


Figure 1: Semantics Tags for Verbs.

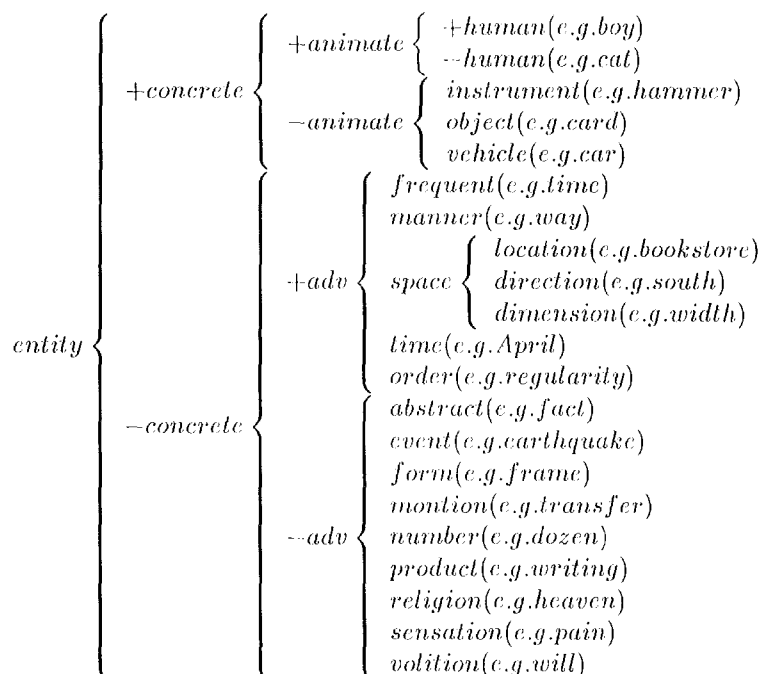


Figure 2: Semantics Tags for Nouns.

not misdetermined and this corresponds to the behavior to model refinement. However, many VPPs are not correctly resolved due to the rigidity of rule-templates. Therefore, relaxing these rules will result in more correct VPP, but less correct NPP. Another difficulty comes from the assignment of semantic tags. As everyone knows the sense ambiguity is a serious problem, to assign unique semantic tag is hard. We plan to resolve this problem in the near future and to use the semantic-tagged corpus to train the rule-templates.

5 Concluding Remarks

Prepositional phrases usually result in structural ambiguities and cost systems many resources to resolve the attachment. We develop a rule-based and MT-oriented model refinement algorithm to tackle this problem. We find PPP, SPP, VPP, and NPP are more realistic than only two attachment choices in machine translation. After large-scale experiments, the results show that rule-based system is also useful for difficult problem like PP attachment. However, the determination of VPP is relatively difficult under our algorithm. Another difficulty is how to assign unique semantic tag to word. The resolution for these two problems will greatly improve the performance of this work.

References

- E. Brill and P. Resnik. 1994. A Rule-Based Approach to Automated Prepositional Phrase Attachment Disambiguation. *Proceedings of COLING-94*, page 1198-1204.
- K. H. Chen and H. H. Chen. 1992. Attachment and Transfer of Prepositional Phrases with Constraint Propagation. *Computer Processing of Chinese and Oriental Languages: An International Journal of the Chinese Language Computer Society*, 6(2), page 123-142.
- K. H. Chen and H. H. Chen. 1994. Acquired Verb Subcategorization Frames. *Proceedings of the Second Conference for Natural Language Processing (KONVENS-94)*, page 407-410. Vienna, Austria.
- K. H. Chen and H. H. Chen. 1995. Machine Translation: An Integrated Approach. *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, page 287-294.
- M. Ford, J. Bresnan and R. Kaplan. 1982. A Competence-Based Theory of Syntactic Closure. *The Mental Representation of Grammatical Relations*, J. Bresnan, Eds., MIT Press, page 727-796.
- L. Frazier. 1978. *On Comprehending Sentences: Syntactic Parsing Strategies*, Doctoral Dissertation, University of Connecticut.
- D. Hindle and M. Rooth. 1993. Structural Ambiguity and Lexical Relations. *Computational Linguistics*, 19(1), page 103-120.
- A. S. Hornby. 1989. *Oxford Advanced Learner's Dictionary*, Oxford University Press.
- J. Kimball. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, 2, page. 15-47.
- C. L. Liu, J. S. Chang and K. Y. Su. 1990. The Semantic Score Approach to the Disambiguation of PP Attachment Problem. *Proceedings of ROCLING-90*, page 253-270. Taiwan, R.O.C.
- M. Marcus, B. Santorini, M. A. Marcinkiewicz. 1993. Building a Large Annotated Corpus of English: the Penn Treebank. *Computational Linguistics*, 19(2), page 313-330.
- S. Shieber. 1983. Sentence Disambiguation by a Shift-Reduced Parsing Technique. *Proceedings of IJCAI-83*, page 699-703. Karlsruhe, Germany.
- Y. Wilks, X. H. Huang and D. Fass. 1985. Syntax. Preference and Right Attachment. *Proceedings of IJCAI-85*, page 779-784. Los Angeles, CA.

Appendix

The following lists rule-templates for PP-attachment. Every template consists of four elements $\langle V, N1, P, N2 \rangle$. The curl bracket pair denotes OR, the underline denotes DON'T CARE and \neg denotes NOT.

1. Rule-template for SPP

1. $\langle \underline{\quad}, \underline{\quad}, about, time \rangle$
2. $\langle \underline{\quad}, \underline{\quad}, across, location \rangle$
3. $\langle \underline{\quad}, \underline{\quad}, after, time \rangle$
4. $\langle \underline{\quad}, \underline{\quad}, along, location \rangle$
5. $\langle \underline{\quad}, \underline{\quad}, among, location \rangle$
6. $\langle \underline{\quad}, \underline{\quad}, at, \{location, time\} \rangle$
7. $\langle \underline{\quad}, \underline{\quad}, before, time \rangle$
8. $\langle \underline{\quad}, \underline{\quad}, between, \{location, time\} \rangle$
9. $\langle \underline{\quad}, \underline{\quad}, by, time \rangle$
10. $\langle \underline{\quad}, \underline{\quad}, during, time \rangle$
11. $\langle \underline{\quad}, \underline{\quad}, in, \{location, time\} \rangle$
12. $\langle \underline{\quad}, \underline{\quad}, in_front_of, location \rangle$
13. $\langle \underline{\quad}, \underline{\quad}, near, location \rangle$

14. $\langle \text{---}, \text{---}, \text{next_to}, \text{location} \rangle$
15. $\langle \text{---}, \text{---}, \text{on}, \text{time} \rangle$
16. $\langle \text{---}, \text{---}, \text{out_of}, \{ \text{abstract}, \text{location} \} \rangle$
17. $\langle \text{---}, \text{---}, \text{over}, \{ \text{location}, \text{time} \} \rangle$
18. $\langle \text{---}, \text{---}, \text{through}, \{ \text{abstract}, \text{event}, \text{time} \} \rangle$
19. $\langle \text{---}, \text{---}, \text{under}, \text{time} \rangle$
20. $\langle \text{---}, \text{---}, \text{with}, \text{abstract} \rangle$
21. $\langle \text{---}, \text{---}, \text{without}, \text{abstract} \rangle$
11. Rule-template for VPP
1. $\langle \text{motion}, \text{---}, \text{about}, \{ \text{object}, \text{location} \} \rangle$
2. $\langle \text{at_ment}, \text{---}, \text{about}, \text{object} \rangle$
3. $\langle \text{action}, \text{event}, \text{after}, \text{concrete} \rangle$
4. $\langle \text{at_ment}, \{ \text{abstract}, \text{event} \}, \text{after}, \{ \text{event}, \text{no}, \text{time} \} \rangle$
5. $\langle \text{motion}, \text{---}, \text{across}, \{ \text{location}, \text{object} \} \rangle$
6. $\langle \{ \text{at_nonmen}, \text{ai_nonmen} \}, \text{---}, \text{along}, \{ \text{location}, \text{object} \} \rangle$
7. $\langle \neg \text{motion}, \neg \{ \text{concrete}, \text{location} \}, \text{among}, \{ \text{concrete}, \text{location} \} \rangle$
8. $\langle \neg \{ \text{at_nonmen}, \text{ai_nonmen} \}, \text{---}, \text{at}, \{ \text{animate}, \text{object} \} \rangle$
9. $\langle \{ \text{at_nonmen}, \text{ai_nonmen} \}, \text{---}, \text{at}, \{ \text{location}, \text{object} \} \rangle$
10. $\langle \text{action}, \text{event}, \text{after}, \text{concrete} \rangle$
11. $\langle \text{at_ment}, \{ \text{abstract}, \text{event} \}, \text{after}, \{ \text{event}, \text{no}, \text{time} \} \rangle$
12. $\langle \{ \text{at_nonmen}, \text{ai_nonmen} \}, \{ \text{event}, \text{object} \}, \text{between}, \{ \text{abstract}, \text{concrete}, \text{location} \} \rangle$
13. $\langle \{ \text{at_nonmen}, \text{ai_nonmen} \}, \{ \text{event}, \text{object} \}, \text{between}, \text{time} \rangle$
14. $\langle \text{motion}, \text{---}, \text{by}, \{ \text{location}, \text{instrument} \} \rangle$
15. $\langle \text{---}, \text{---}, \text{by}, \text{manner} \rangle$
16. $\langle \neg \text{motion}, \text{---}, \text{by}, \{ \text{location}, \text{object} \} \rangle$
17. $\langle \text{---}, \text{---}, \text{by}, \{ \text{abstract}, \text{event}, \text{object}, \text{vehicle} \} \rangle$
18. $\langle \text{---}, \text{---}, \text{by}, \text{animate} \rangle$ passive voice
19. $\langle \text{---}, \text{---}, \text{for}, \text{time} \rangle$
20. $\langle \text{motion}, \text{---}, \text{for}, \text{location} \rangle$
21. $\langle \neg \text{linking}, \text{---}, \text{for}, \{ \text{abstract}, \text{concrete}, \text{event} \} \rangle$
22. $\langle \text{---}, \text{---}, \text{for}, \{ \text{abstract}, \text{event}, \text{object} \} \rangle$
23. $\langle \text{---}, \text{---}, \text{for}, \text{animate} \rangle$
24. $\langle \{ \text{motion}, \text{speech_act} \}, \text{---}, \text{from}, \text{entity} \rangle$
25. $\langle \text{motion}, \text{---}, \text{in}, \{ \text{location}, \text{instrument} \} \rangle$
26. $\langle \text{---}, \text{---}, \text{in}, \text{time} \rangle$
27. $\langle \text{---}, \text{---}, \text{in}, \text{vehicle} \rangle$
28. $\langle \text{ai_nonmen}, \text{---}, \text{in_front_of}, \{ \text{concrete}, \text{location} \} \rangle$
29. $\langle \text{ai_nonmen}, \text{---}, \text{inside}, \{ \text{concrete}, \text{location} \} \rangle$
30. $\langle \text{---}, \text{---}, \text{into}, \{ \text{abstract}, \text{concrete}, \text{location} \} \rangle$
31. $\langle \text{act}, \text{---}, \text{like}, \text{---} \rangle$
32. $\langle \text{---}, \neg \{ \text{location}, \text{object} \}, \text{near}, \{ \text{location}, \text{object} \} \rangle$
33. $\langle \text{---}, \neg \{ \text{location}, \text{object} \}, \text{next_to}, \{ \text{location}, \text{object} \} \rangle$
34. $\langle \{ \text{at_nonmen}, \text{ai_nonmen} \}, \{ \text{event}, \text{object} \}, \text{on}, \{ \text{concrete}, \text{location}, \text{object} \} \rangle$
35. $\langle \text{---}, \text{---}, \text{on}, \text{event} \rangle$
36. $\langle \text{---}, \text{---}, \text{on_to}, \{ \text{location}, \text{object} \} \rangle$
37. $\langle \text{motion}, \text{---}, \text{out_of}, \{ \text{concrete}, \text{location} \} \rangle$
38. $\langle \text{---}, \text{---}, \text{over}, \{ \text{abstract}, \text{event} \} \rangle$
39. $\langle \text{motion}, \text{---}, \text{over}, \{ \text{concrete}, \text{location} \} \rangle$
40. $\langle \text{ai_nonmen}, \text{---}, \text{through}, \{ \text{location}, \text{object} \} \rangle$
41. $\langle \{ \text{ai_nonmen}, \text{at_nonmen} \}, \text{---}, \text{under}, \{ \text{abstract}, \text{object} \} \rangle$
42. $\langle \text{---}, \text{---}, \text{until}, \text{time} \rangle$
43. $\langle \text{at_nonmen}, \text{---}, \text{with}, \text{instrument} \rangle$
44. $\langle \text{at_nonmen}, \text{---}, \text{with}, \text{animate} \rangle$
45. $\langle \text{at_nonmen}, \text{---}, \text{without}, \text{instrument} \rangle$
46. $\langle \text{at_nonmen}, \text{---}, \text{without}, \text{animate} \rangle$