The Semantic Score Approach to the Disambiguation of PP Attachment Problem

*Chao-Lin Liu, *,**Jing-Shin Chang and *Keh-Yih Su

*Department of Electrical Engineering
National Tsing Hua University
Hsinchu, TAIWAN, R.O.C.

**BTC R&D Center
2F, 28, R&D Road II, Science-Based Industrial Park
Hsinchu, TAIWAN, R.O.C.

ABSTRACT

In a Natural Language Processing System which takes English as the source input language, the syntactic roles of the *prepositional phrases* in a sentence are difficult to identify. A large number of ambiguities may result from these phrases. Traditional rule-based approaches to this problem rely heavily on general linguistic knowledge, complicated knowledge bases and sophisticated control mechanism. When *uncertainty* about the attachment patterns is encountered, some *heuristics* and *ad hoc* procedures are adopted to assign attachment *preference* for disambiguation. Hence, although the literatures about this topic are abundant, there is no guarantee of the *objectiveness* and *optimality* of these approaches.

In this paper, a probabilistic semantic model is proposed to resolve the PP attachment problem without using complicated knowledge bases and control mechanism. This approach elegantly integrates the *linguistic* model for semantics interpretation and the *objective* characteristics of the probabilistic *Semantic Score* model. Hence, it will assign a much more *objective* preference measure to each ambiguous attachment pattern. It is found that approximately 90% of the PP attachment problem in computer manuals can be solved with this approach without resorting to any heuristics-based rules and complicated control mechanism. The mapping between the abstract Score Function paradigm and the real PP attachment problem will be addressed in this paper. Future expansion of the semantic score function for resolving general ambiguity problems is also suggested.

1. PP Attachment Problem

In a natural language processing system, there are many sources which may cause a given sentence to be multiple analyzed. One of the problems is the *uncertainty* on the placement of the modifiers. Such problem is known as the attachment problem. The most well-known attachment problem in English is the *PP attachment problem* (hereafter, PPAP), where a given prepositional phrase (PP) may modify either the main verb or the preceding noun phrase of the sentence. (Modification to other constituents, for example, the whole sentence, is also possible.) This uncertain characteristics on the placement of the prepositional phrase may lead to a large number of ambiguities.

The importance of resolving the PP attachment problem is twofold. First, the sentences with prepositional phrases are common in English. Secondly, the number of ambiguities resulted from PPs increases with the number of PPs. It is estimated that the number of ambiguities approximately follows a combinatoric series called the Catalan numbers (1, 1, 2, 5, 14, 42, 132, 469, 1430, 4862... and so on) [CHUR 82]. Therefore the importance of PP attachment disambiguation can not be overlooked.

In the past few years, there is an abundance of literature concerning the resolution of the PP attachment problem. For example, Frazier and Fodor [FRAZ 78] uses the well-known heuristic principles *Right Association* (RA) and *Minimal Attachment* (MA) to cope with the attachment problem in their *Sausage Machine* (SM). Marcus [MARC 80] uses a case-frame interpreter to decide the proper attachment pattern in his PARSIFAL parser, which relies heavily on selectional restriction. Ford, Bresnan and Kaplan [FORD 82] proposed their *Theory of Closure* to tackle this problem. Other solutions to the PPAP is numerous. Most of them are not explicitly aimed at the PPAP but aimed at more general semantic analyses. For example, Wilks applied *Preference Semantics* [WILK 75a, 75b, 83] in his intelligent analyzer and understander of English. More detailed information and comments on these approaches can be found in [HIRS 87].

2. Problems with Conventional Approaches

The rule-based approaches in the previous section do resolve certain attachment problem in some specific domain. However, these approaches share some common characteristics which make them difficult to adopt in a large practical system such as a commercialized machine translation system. The following problems are frequently encountered with such mechanisms :

[a] The rule-based systems are inappropriate for handling *uncertain* knowledge. When dealing with a large system with wide coverage, this problem becomes even worse.

[b] The heuristic measures used to assign *preference* to various ambiguities by rule-based systems are usually *ad hoc* or *heuristics-based.* There is no objective measure for evaluating the effectiveness of such rule systems, nor is there any formal way to predict whether the evolution of the rule system is toward the direction of the optimal solution.

[c] A large number of *rules* or *templates* are required which impose a heavy load on the linguists. Any variation in the rules may have unpredictable effect on the whole system. Hence, the large rule system also makes the maintenance of the system a hard task. In addition, complicated control mechanisms are usually required to handle such a rule system.

[d] In sum, these approaches are *non-systematic* in that there is no simple systematic approach for *extracting* the required linguistic knowledge, *verifying* the validity of these rules, maintaining the *consistency* of the rule system and *manging* the rules or templates in an effective way.

In the following sections, we will propose a probabilistic semantic model to resolve the PP attachment problem. Due to the inherent properties of *objectiveness*, *trainability* and *consistency* of a probabilistic model, such model will be more appropriate than rule–based systems when the above problems are taken into account.

The probabilistic semantic model is based on the *Score Function* paradigm suggested in [SU 88, 89, 90b, 90c], which combines both *semantics* and *statistics* to deal with general disambiguation problems. By taking *semantics* into account, we can avoid *blind* preference assignment as suggested in some heuristic approaches like RA and MA principles. By introducing *statistics*, the probability will provide a more *objective* preference measure than heuristically assigned scores in some other systems. Moreover, because a semantic score is given to each analysis, requirement for imposing rigid priority order on conflicting rules are eliminated in *uncertain* situations. Furthermore, the rules or templates will be revealed in the form of probability distribution if they do have some linguistic reality. Hence the linguists can be relieved of writing *exact* rules such as those used in selectional restriction or templates for lexical preference.

3. The Score Function and Semantic Score Approach

Due to the problems mentioned above on conventional disambiguation methods, we seek to find a systematic way to overcome these problems during the development of our English-Chinese Machine Translation System, **ArchTran** [SU 85, 87]. Since the ArchTran is meant to be an *operational* MTS rather than just a laboratory system, a *systematic* approach to semantic interpretation is very important. To achieve this goal, we have proposed a probability based Score Function as the overall evaluation function for *preference* measurement of a sentence [SU 88, 89, 90b, 90c].

To state formally, for a given parse tree (or more generally, a subtree) T which is annotated with semantic feature values on its nodes, the score associated with this particular interpretation of the sentence is given by the following Score Function :

$$Score(T) \equiv P(SEM, SYN, LEX|WRD) \qquad ...(1)$$

$$= P(SEM|SYN, LEX, WRD) \times P(SYN|LEX, WRD) \times P(LEX|WRD)$$

where SEM, SYN and LEX are the specific set of *semantic annotation*, *syntactic structure* and *lexical features* attached to the nodes of the parse tree, and WRD is the set of terminal words of the sentence. For example, in the analysis :

$$\left[[saw]_{v\{+stat\}}\ [the\ girl]_{np\{+anim\}}\ \left[with_{\{+means\}}\ [a\ telescope]_{np\{+tool\}}\right]_{pp}\right]_{vp}$$

the verb phrase is given a specific set of semantic annotation SEM = { saw/+stat(ive), [the girl]/+anim(ate), ... }; the syntactic structure SYN is identified by the subtree of the parse, namely VP[v NP[det n] PP[p NP[art n]]]; and the lexical feature is represented by the lexical categories (and other lexical information) of the lexical items as LEX = { v(erb), det(erminer), n(oun), p(reposition), art(icle), n(oun) }. The score for this subtree is then defined as the conditional probability of SEM, SYN, LEX, given the input words WRD = { saw, the, girl, with, a, telescope}. In this sense, we can measure the degree of preference of a semantically annotated parse tree with the conditional probability of any specific set of SEM, SYN and LEX, given the known WRD.

As shown in equation (1), the *score function* can be further divided into three product terms, which are called *semantic score*, *syntactic score* and *lexical score*, respectively. Intuitively, the *semantic score* P(SEM | SYN, LEX, WRD) corresponds to the control mechanism of the semantic analysis phase in traditional stratified analyses. By dividing the score function into these components, it is much easier to apply them to different phases of the analyses or to incorporate them into the system incrementally.

The score function can be shown to be *optimal* as a decision rule *in Bayesian sense*; and the function (or its component functions) has been adopted for several applications [SU 88, 89, 90a, 90b]. For example, a simulation has been conducted to select the preferred parse among a set of ambiguous constructions [SU 88] based solely on the *syntactic score*. The syntactic score paradigm successfully models the parsing process in which arbitrary degree of context sensitivity can be handled. The result is quite promising. It shows that the correct syntactic structures of more than 85% of the test sentences are successfully ranked at the *first* place when a total of *three* local left and right context symbols are consulted. In addition, over 93% of the correct syntax trees are ranked at the *first* or *second* place based on the syntactic score and *two* context symbols.

With this promising result, we were encouraged to develop the semantic score model for ambiguity resolution. In particular, in this paper, we show how the semantic score (more exactly, the *partial* semantic score for a *verb phrase* in a sentence) can be used to solve the PP attachment problem. To state briefly, the semantic score approach to PP attachment problem adopts a simplified version of the *semantic score* as the *preference* measure for possible attachment patterns. The attachment pattern with the highest semantic score is regarded as the most probable attachment. In the next section, we will give a more detailed introduction to the mapping between the PP attachment problem and the semantic score function.

4. Semantic Score for PP Attachment

To simplify the discussion of the semantic score approach, we shall only consider a special case of the PPAP which is characterized by the four major components, [V N_1 P N_2], of a verb phrase . A typical verb phrase of this type is : "*saw the girl with a telescope.*" The symbols V, N_1, P and N_2 refer to the main verb, the head noun of the object, the preposition, and the head noun in the PP, respectively. These four components are selected to characterize the attachment problem because the resolution of the attachment problem depends heavily on their semantic features. Some examples will be shown later. If V-PP is used to mean that PP is attached to the main verb and N_1-PP to mean that PP is attached to the preceding noun, then the (partial) semantic score associated with these attachment patterns can be formulated as :

$$
\begin{aligned}
& SC_{SEM}(X|V,N_1,P,N_2) \qquad \ldots(2)\\
& \equiv P(X|V,N_1,P,N_2)\\
& = \sum_{v,n_1,p,n_2} [P(X|v,n_1,p,n_2,V,N_1,P,N_2) \times P(v,n_1,p,n_2|V,N_1,P,N_2)]
\end{aligned}
$$

where X can either be V-PP or N_1-PP, and the summation is taken over all *semantic features* v, n_1, p, n_2 of V, N_1, P, N_2, respectively. In other words, we try to assign the attachment preference by evaluating the probability of a particular attachment pattern conditioned on the [V, N_1, P, N_2] 4–tuple. If $SC_{SEM}(V-PP|V,N_1,P,N_2)$ is greater than $SC_{SEM}(N_1-PP|V,N_1,P,N_2)$, then the PP will be attached to the main verb, otherwise the attachment to N_1 is preferred.

An alternative formulation is to compute the joint conditional probability of the attachment pattern and the possible combination of the semantic features [v n_1 p n_2] based on the input strings. The score can then be formulated as :

$$\begin{aligned}
&SC_{SEM}(X,v,n_1,p,n_2|V,N_1,P,N_2) \qquad \dots (2') \\
&\equiv P(X,v,n_1,p,n_2|V,N_1,P,N_2) \\
&= P(X|v,n_1,p,n_2,V,N_1,P,N_2) \times P(v,n_1,p,n_2|V,N_1,P,N_2)
\end{aligned}$$

which is exactly the individual terms in Eqn. (2).

If $SC_{MAX}(V-PP) \equiv \underset{v,n_1,p,n_2}{\mathrm{MAX}} [SC_{SEM}(V-PP,v,n_1,p,n_2|V,N_1,P,N_2)]$ is greater than $SC_{MAX}(N_1-PP) \equiv \underset{v,n_1,p,n_2}{\mathrm{MAX}} [SC_{SEM}(N_1-PP,v,n_1,p,n_2|V,N_1,P,N_2)]$, then the attachment pattern V-PP is preferred over N_1–PP, otherwise N_1–PP is the preferred attachment pattern. In other words, the preferred attachment pattern is determined by the most probable attachment pattern and the sequence of semantic classes given the [V N_1 P N_2] 4–tuple. Furthermore, the semantic feature [v n_1 p n_2] which corresponds to the maximal score is assigned to the input [V N_1 P N_2]. Hence, with this formulation, the *lexical ambiguity* on multiple word senses, can also be resolved at the same time when the most preferred attachment pattern is decided.

It is obvious that the second alternative requires less computation than the first one. However, due to the constraints on the amount of tagged corpus and the consideration of producing significant statistics, we use the first formulation as the basis in our simulation.

We can simplify Equation (2) further if we make the following assumptions [LIU 89] :

1. Once the semantic feature of a word is known, the word itself does not affect the score significantly. For example, if the semantic feature {v} of the verb is known, then the input {V} can be ignored from the conditional probability. If this is the case, we can assume that $P(X|v,n_1,p,n_2,V,N_1,P,N_2) \approx P(X|v,n_1,p,n_2)$ by ignoring {V, N_1, P, N_2}. In other words, the attachment pattern is more closely related to the semantic features of the words.

2. The semantic feature of a given word is not strongly influenced by other words or the semantic features of these words so that we can assume that

$$
\begin{aligned}
& P(v, n_1, p, n_2 | V, N_1, P, N_2) \\
& = P(v | n_1, p, n_2, V, N_1, P, N_2) \times P(n_1 | p, n_2, V, N_1, P, N_2) \\
& \times P(p | n_2, V, N_1, P, N_2) \times P(n_2 | V, N_1, P, N_2) \\
& \approx P(v|V) \times P(n_1|N_1) \times P(p|P) \times P(n_2|N_2)
\end{aligned}
$$

In other words, we assume that the dependency between the semantic feature of a given word and its context symbols is nearly context-free.

Under these assumptions, equation (2) can be simplified as :

$$
\begin{aligned}
& SC_{SEM}(X|V, N_1, P, N_2) \qquad \ldots(3) \\
& \approx \sum_{v,n_1,p,n_2} P(X|v, n_1, p, n_2) \times P(v|V) \times P(n_1|N_1) \times P(p|P) \times P(n_2|N_2)
\end{aligned}
$$

(See [LIU 89] for more details on the derivation of the simplified formula.)

Although it is not known whether the second assumption is true, we make the assumption so as to simplify the problem. The tests show that this assumption still leads to satisfactory results. To take contextual information into account, we can simply retain the items that are significant to the resolution of the PPAP, and extend the above formulation to an arbitrary degree of context sensitivity. In this paper, we will not discuss such topics.

Eqn. (3) is used in our tests to show the effects of the semantic score approach to PPAP when all four components in {V, N_1, P, N_2} are considered. We shall refer to such test scheme as [VNPN] in the following sections. To reduce the computational complexity and to show the individual effect of each component in {V, N_1, P, N_2}, we also conduct a series of tests with some of the terms in equation (3) ignored. The testing schemes and their simplified score functions are listed as follows :

[VxPN] (Ignore the contribution of the object)

$$
\begin{aligned}
& SC_{SEM}(X|V, N_1, P, N_2) \\
& \approx \sum_{v,p,n_2} [P(X|v, p, n_2) \times P(v|V) \times P(p|P) \times P(n_2|N_2)] \quad \ldots(4)
\end{aligned}
$$

[VxPx] (Ignore the contribution of the nouns)

$$
\begin{aligned}
& SC_{SEM}(X|V, N_1, P, N_2) \\
& \approx \sum_{v,p} [P(X|v, p) \times P(v|V) \times P(p|P)] \qquad \ldots(5)
\end{aligned}
$$

[xxPx] (Consider the contribution of P only)

$$SC_{SEM}(X|V, N_1, P, N_2) \approx \sum_{p} [P(X|p) \times P(p|P)] \quad \ldots (6)$$

(An "x" in the test scheme means to ignore the contribution of the corresponding component in [VNPN]; that is, "Don't Care".)

The [xxPx] scheme will cover the simplest cases in which the *preposition* strongly implies the attachment preference. For example, the preposition *of* usually leads to the N_1-PP attachment preference such as in :

- "change the format *of* the disk" (N_1-PP).

The [VxPx] scheme further includes the cases in which the *subcategorization* feature of the main *verb* or its *feature co-occurrence* characteristics with the *prepositional phrase* provides extra information for assigning attachment preference. This scheme will assign different preferences to the sentences such as :

- "*sent* the ticket *to* Taipei" (V-PP), and
- "*lost* the ticket *to* Taipei" (N_1-PP).

When the head *noun* (N_2) of the noun phrase in the prepositional phrase reveals strong evidence on the *case role* of the prepositional phrase, including N_2 will definitely be helpful for assigning attachment preference. The [VxPN] scheme formally encodes such preference. It can be useful in resolving such ambiguities as :

- "*eat* the apple *in* the *box*" (N_1-PP), and

  "*finish* the job *in* two *minutes*" (V-PP).

In the latter case, the noun *minutes* strongly implies a [+TIME] feature. Hence, the prepositional phrase "*in* two *minutes*" has the preference of being filled into the case slot of the verb with [+TIME] constraint. Hence, V-PP attachment is preferred.

Finally, when the case is so complicated that we must jointly take into account the *subcategorization features* of the main *verb*, the *predicate-argument structure* among the main *verb*, the object *noun* N_1 and the possible case filler N_2, and the *feature co-occurrence* constraints between the [VNPN] 4–tuple, then we might need a more complicated model such as the one suggested by the [VNPN] scheme.

Note that we have encoded the attachment preference with the *semantic attributes* of the [VNPN] 4–tuple only. Hence, we can easily determine the attachment preference without resorting to *complex knowledge bases* and *control mechanism* as traditional rule-based systems do. We can also benefit from such approach in that an *objective, trainable* and *consistent* system for assigning attachment preference can be easily acquired.

5. The Classification of Semantic Attributes

Before the computation of the required scores, the semantic features must be assigned to each of the four components V, N_1, P, N_2. Among the four components, the semantic features of the *verbs* are considered to be of most importance. To see how the semantic features of the verbs affect the resolution of PPAP, we have tried three different semantic feature sets/hierarchies which are suggested by Givón [GIVÓ 84], Tang [TANG 88] and Chodorow [CHOD 85], respectively.

According to Givón's classification, each *sense* of a lexical item is unique to the language. Hence, each word sense can be regarded as one class. Therefore, the verbs "contain", "have" and "hold", though all have the sense of "inclusion", will be regarded as three *different* verb classes which differ only slightly. With such criterion, the selected verbs in the test sentences are classified into 16 classes [LIU 89], corresponding to 16 word senses of the 14 most frequently used verbs in our test set and training set. The semantic classes thus defined is show in Figure 1 for the verbs used in our preliminary experiments.

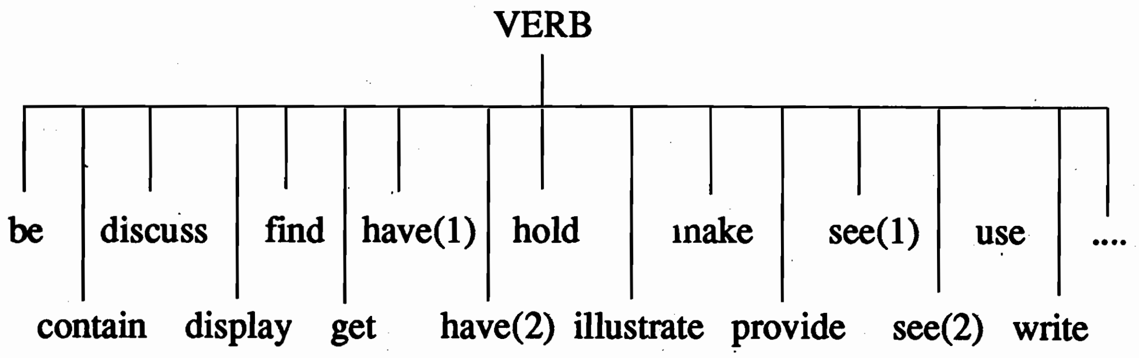

Figure 1 Verb Classification according to [GIVÓ 84].

In Tang's analyses, the verbs can be classified into *stative* and *non-stative* according to their semantic features. Non-stative verbs can be divided further into *non-dynamic* and *dynamic* verbs, which in turn consists of *accomplishment* verbs and *activity* verbs. Such classification forms a semantic hierarchy of the verbs which is characterized by the *syntactic functions* and *aspect features* of the verbs [TANG 88, LIU 89]. Figure 2 shows the hierarchy for the verbs used in our tests.

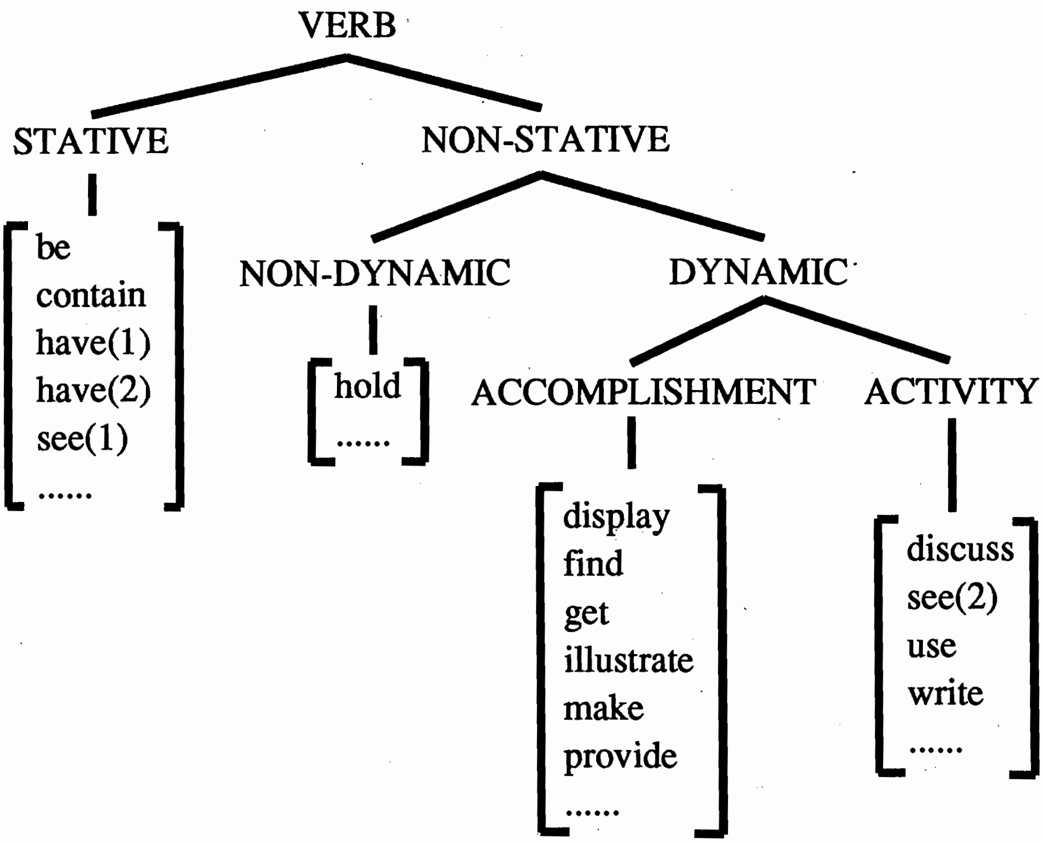

Figure 2 Verb Classification according to [TANG 88].

In the third classification system, Chodorow classifies the verbs which have "similar sense" into the same class. Hence, for example, the verbs "contain", "have" and "hold" will be classified into the *same* class as opposed to Givón's classification. The classification is meant for extracting *semantic hierarchies* from on-line dictionaries. In our testing, we adopted the definitions in Webster's Dictionary (1988 edition) to determine whether two verbs have "similar sense". According to such criterion, a list of 10 verb classes are selected for the test sentences. They are shown in Figure 3.

For prepositions and nouns, only one classification system for each category is adopted. The semantic features for prepositions came from Quirk [QUIR 85], where semantic features such as "duration", "manner" and "means" are used [LIU 89]. The semantic features for nouns, on the other hand, came from [CKIP 88]. They are classified into a hierarchical structure of "physical entity", "non-physical entity", "animate", "non-animate" ([LIU 89]), and so on.

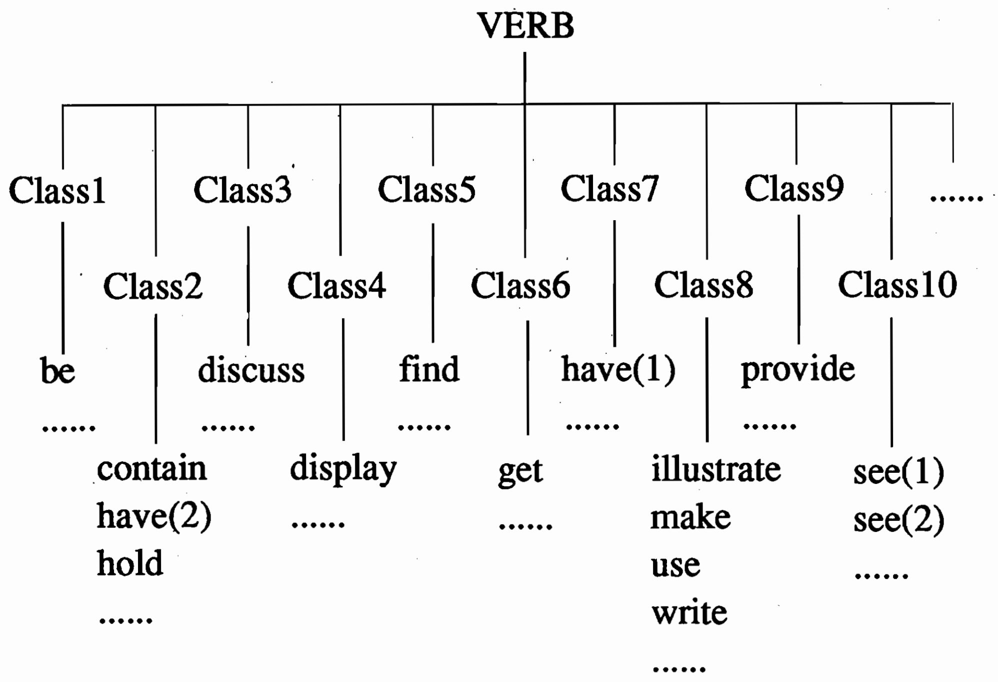

Figure 3 Verb Classification according to [CHOD 85].

With these semantic features, the test sentences in the corpus are either tagged by the ArchTran MTS or manually tagged for testing.

6. Simulation Results

To test the validity of the semantic score approach, we selected the most frequently used verbs, nouns and prepositions from *ten* books in the *computer* field and 1607 sentences parsed by the ArchTran MTS. The result is a list of 14 verbs, 10 prepositions, 6 nouns for N_1 and 18 nouns for N_2. Because not all the sentences have the required verbs, nouns and prepositions in the selected list, we divided the training data and test sentences into 6 databases for different tests. The four *training databases* are called 1607PC(1), 1607PC(2), 1607PC(3) and 1607PC(4), each containing 595, 370, 109 and 31 sentences, respectively. They are chosen from the 1607 sentences and *one* of the ten books. The 1607PC(1) database is a superset of 1607PC(2), which, in turn, is a superset of 1607PC(3). The 1607PC(4) database is the subset of all the other three. The 1607PC(1) database contains [VNPN] 4–tuples which include one of the prepositions in the list of common prepositions; it is used to train the probability for the [xxPx] scheme. Similarly, 1607PC(2), 1607PC(3) and 1607PC(4) contain the required patterns for training [VxPx], [VxPN] and [VNPN] schemes, respectively. There are two test sets, AL and B9, each containing 235 and 115 sentences, respectively. The AL database is selected from *two* of the ten books, while B9 is selected from *nine* out of the ten books.

The probability entries are estimated with relative frequency counts. If a null entry is found, it is replaced with the reciprocal of the number of semantic features of the corresponding lexical category. In other words, if we do not have any information about the semantic feature of a given word, we assume that it can be assigned with any of the possible semantic features with equal probability. Moreover, the probability of the form P(X | v n_1 p n_2) is assigned 1/2 if the [v n_1 p n_2] combination is not found. In other words, we assume that only attachment to V and N_1 is possible.

With the semantic features properly tagged, the following results are observed for the test schemes. Eqn. (3), (4), (5), (6) are used in the test schemes [VNPN], [VxPN], [VxPx] and [xxPx] respectively. Through the result of these schemes, we can estimate which of the four components in {V, N_1, P, N_2} is more significant, and evaluate the number of components required to solve the PPAP. In the following tables, we will show the effects of the score function on PPAP. The semantic features will be included gradually in the order of P, V, N_2 and N_1 in the tests.

TEST I

In this test, only the contribution of the *preposition* is considered in resolving the PPAP. The training data is the 1607PC(1) database, and it also serves as the testing data for *close* data set test. The *open* data set test, which takes sentences *not* in the training data, uses the AL database for testing. The success rates are shown unparenthesized in the table. To compare the semantic model with the RA-MA heuristics in Frazier and Fordor's work [FRAZ 78], the success rate for the RA-MA heuristics is shown *parenthesized* in the table. (For the following tests, these two types of success rates will be shown in the same manner.) It is evident that the semantic score approach outperforms the RA and MA heuristics drastically.

Test Data	[xxPx]
1607PC(1)	78.82 (33.45)
AL	81.28 (18.72)

Number of Sentences : 1607PC(1) = 595 & AL = 235

TEST II

In this test, the effects of both *verb* and *preposition* are considered. The effects of different semantic feature sets for verbs as proposed in [GIVÓ 84], [TANG 88] and [CHOD 85] are also shown in the table. It shows a success rate of about 90% in both open and close data set tests. The semantic feature set proposed by [TANG 88] is slightly better in the open data set test. However, the difference is not distinct.

Test Data	[VxPx]		
	GIVO	TANG	CHOD
1607PC(2)	88.92 (28.92)	86.22 (28.92)	87.84 (28.92)
AL	88.94 (18.72)	92.77 (18.72)	89.79 (18.72)

Number of Sentences : 1607PC(2) = 370 & AL = 235

TEST III

Another test scheme [VxPN] is conducted by taking account of the *noun* in a PP. Comparing the test schemes [xxPx], [VxPx] and [VxPN], the success rates increase as the head nouns of the PPs are taken into consideration. Over 90% of the attachment can be correctly decided in these cases.

Test Data	[xxPx]	[VxPx]			[VxPN]		
		GIVO	TANG	CHOD	GIVO	TANG	CHOD
1607PC(3)	79.82 (30.28)	95.41 (30.28)	86.24 (30.28)	94.50 (30.28)	96.33 (30.28)	90.83 (30.28)	96.33 (30.28)
B9	74.78 (6.96)	90.43 (6.96)	93.04 (6.96)	90.43 (6.96)	92.17 (6.96)	90.43 (6.96)	92.17 (6.96)

Number of Sentences : 1607PC(3) = 109 & B9 = 115

TEST IV

For the last test, the database 1607PC(4) is used. Because there are only 31 sentences in the training set, we do not conduct any open data set test. The results for different test schemes are shown in the table. The table shows a high success rate for all test schemes. Because the database size is small, this table is meant for reference only. However, it shows the trend of increasing success rate when more and more semantic features are involved.

	GIVO	TANG	CHOD
[xxPx]	90.32 (51.61)		
[VxPx]	100.00 (51.61)	90.32 (51.61)	96.77 (51.61)
[VxPN]	100.00 (51.61)	96.77 (51.61)	100.00 (51.61)

[VNPN]	100.00 (51.61)	100.00 (51.61)	100.00 (51.61)

DataBase = 1607PC(4) & Number of Sentences = 31

7. Perspectives and Conclusion

In this paper, we propose a *semantic score* approach to solve the Prepositional Phrase Attachment Problem (PPAP) without resorting to *complex knowledge bases* and *complicated control mechanism.* The probabilistic semantic model elegantly bridges the gap between linguistic knowledge and probability theory, and hence provides both *systematic* and *non-heuristic* approach to resolving the ambiguity problem. In the various simulations, about 90% of the attachment patterns can be correctly determined with this approach. In essence, the only semantic information used to acquire this performance is the *feature co-occurrence* distribution of the semantic classes of the [VNPN] 4-tuple. Hence, no complicated lexical entries and control mechanism are required in such paradigm. This is a very attractive property over conventional rule-based approaches.

We have also attempted to explore the disambiguation effects of different semantic *feature sets* (for verbs). Although the differences are not distinct in the preliminary tests, they do present another important issue in constructing a systematic mechanism for solving general ambiguity problem. Hence the selection of the most significant semantic features will be studied in greater detail and be incorporated into the ARCHTRAN MTS in the future.

In this paper, the semantic score approach is applied to the PP attachment problem only. For more general problems of disambiguation in which various sources of ambiguities can occur simultaneously, a *generalized* probabilistic semantic model, such as in [CHAN 90], will be required to deal with the semantic part of the ambiguity problems. In addition, the *integration* of *lexical* preference, *syntactic* preference and *semantic* preference will be very important for resolving more complicated ambiguity problems in various context. Some of the approaches of integration, such as [SU 90c], will be studied more extensively.

ACKNOWLEDGEMENTS

This paper is a revised version of the master thesis [LIU 89] of the first author. We would like to express our deepest gratitude to BTC R&D Center for providing the required text corpus and many helpfui supports from the ArchTran MTS group.

REFERENCES

[CHAN 90] Chang, J.-S., *GPSM : A Generalized Probabilistic Semantic Model for Ambiguity Resolution*, Master thesis, Institute of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, TAIWAN, R.O.C., 1990.

[CHOD 85] Chodorow, M.S., R.J. Byrd and G.E. Heidorn, "Extracting Semantic Hierarchies from a Large On-Line Dictionary," *ACL Proc. of 23rd Annual Meeting*, pp. 299-304, 8–12 July 1985.

[CHUR 82] Church, K. and R. Patil, "Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table," *AJCL*, vol. 8, no. 3-4, pp. 139-149, 1982.

[CKIP 88] CKIP group, *Syntactic Categories for Chinese (revised edition)*, Technical Report #2, Computing Center, Academia Sinica, Taipei, R.O.C., 1988.

[FORD 82] Ford, M., J.W. Bresnan and R.M. Kaplan, "A Competence-based Theory of Syntactic Closure," in Bresnan, J.W. (ed.), *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA, U.S.A., 1982.

[FRAZ 78] Frazier, L. and J.D. Fodor, "The Sausage Machine : A New Two-Stage Parsing Model," *Cognition* 6, pp. 291-325, 1978.

[GIVÓ 84] Givón, T., *SYNTAX : A Functional Typological Introduction*, John Benjamins Publishing Company, Amsterdan, Holland, 1984.

[HIRS 87] Hirst, G., *Semantic Interpretation and the Resolution of Ambiguity*, Cambridge University Press, Cambridge, U.K., 1987.

[LIU 89] Liu, C.-L., *On the Resolution of English PP Attachment Problem with a Probabilistic Semantic Model*, Master thesis, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, TAIWAN, R.O.C., 1989.

[MARC 80] Marcus, M.P., *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, MA, U.S.A., 1980.

[QUIR 85] Quirk, R., et al., *A Comprehensive Grammar of English Language*, Longman Group Limited, London, U.K., 1985.

[SU 85] K.-Y. Su, H.-H. Hsu and M.-L. Hsiao, "Development of an English-To-Chinese Machine Translation System," *Proc. of Natl. Computer Symposium (NCS) 1985*, vol. 2, pp. 997-1001, Kaohsiung, R.O.C., Dec 20-22, 1985.

[SU 87] Su, K.-Y, J.-S. Chang and H.-H. Hsu, "A Powerful Language Processing System for

English-Chinese Machine Translation," *Proc. of 1987 Int. Conf. on Chinese and Oriental Language Computing*, pp. 260-264, Chicago, ILL, U.S.A., 15–17 June 1987.

[SU 88] Su, K.-Y and J.-S. Chang, "Semantic and Syntactic Aspects of Score Function," *Proc. of COLING-88*, vol. 2, pp. 642-644, 12th Int. Conf. on Computational Linguistics, Budapest, Hungary, 22-27 August 1988.

[SU 89] Su, K.-Y, J.-N. Wang, M.-H. Su and J.-S. Chang, "A Sequential Truncation Parsing Algorithm Based on the Score Function," *Proc. of Int. Workshop on Parsing Technologies*, pp. 95-104, CMU, Pittsburgh, PA, U.S.A., 28-31 August 1989.

[SU 90a] Su, K.-Y. and T.-H. Chiang, "A Framework for a Chinese Phonetic Typewriter System," *Proc. of 1990 Int. Conf. on Computer Processing of Chinese and Oriental Languages*, Changsha, Hunan, China, 25–28 April 1990.

[SU 90b] Su, K.-Y., J.-N. Wang, M.-H. Su and J.-S. Chang, "GLR Parsing with Scoring," to appear in Tomita, Masaru (ed.) : *Generalized LR Parsing*, Kluwer Academic Publisher, 1990.

[SU 90c] Su, K.-Y., J.-S. Chang and Y.-C. Lin, "A Unified Approach to Disambiguation Using A Uniform Formulation of Probabilistic Score Functions," in preparation, 1990.

[TANG 88] Tang, T.-C., "The Semantic Attributes and Syntactic Functions of English Verbs," *A Cognitive Grammar of English : Form, Meaning and Function*, vol 1, pp. 173–245, Student Book Co., Ltd., Taiwan, R.O.C., 1988.

[WILK 75a] Wilks, Y., "Preference Semantics," In Keenan, Edward Louis, III (ed.), *Formal Semantics of Natural Language*, Cambridge University Press, 1975.

[WILK 75b] Wilks, Y., "An Intelligent Analyzer and Understander of English," *Communications of the ACM*, vol. 18, no. 5, pp 264–274, May 1975.

[WILK 83] Wilks, Y., "Preference Semantics, Ill-formedness, and Metaphor," *AJCL*, vol. 9, no. 3–4, pp 178–187, July-Dec. 1983.