Information-based Language Analysis for Thai

Virach Sornlertlamvanich
Researcher
Center of International Cooperation for Computerization
12-35, Shibaura 4-Chome, Minato-ku, Tokyo, Japan 108

Wantanee Phantachat
Researcher
National Electronics and Computer Technology Center
Rama 6 Rd., Phayatai, Bangkok, Thailand 10400

ABSTRACT

Thai language is recognized as an isolative language having neither lexicon inflection due to word agreement and tense (as in English) nor obvious syntactic case marker (as in Japanese). The position of word in a sentence is the important superficial syntactic information for recognizing the meaning and the syntactic role. In this paper, we are going to describe the methodology and algorithm to cope with the mentioned Thai specific language phenomena. The rather fixed relation of word position and its syntactic role gives a well-formed pattern of a phrase. Therefore, localization of pattern analysis helps much in phrasal recognition and works well in lexicon disambiguition. In a sentence, the relatively less ambiguous in concept of words (variety of concepts) are consecutively determinated to make up a bunch of concept. Then, according to the information retrieved from dictionary, subcategorization employed by the verb of the sentence will finally create the relation between those groups of concepts to build up a dependency structure to represent the meaning of the sentence. Besides the lexicon information from the dictionary, the grammatical rules are employed to identify the appropriate semantic relation between concepts with lexicon functional reasoning in the pair of provides and requires attribute.

Keywords: subcategorization, dependency structure, functional reasoning, interlingua

1 INTRODUCTION

This paper presents a methodology and algorithm of the parser in Thai sentence analysis in multilingual machine translation system. The language analysis methodology is partially based on dependency grammar,

representing the meaning of a sentence in an interlingual representation--in other word is called a conceptual dependency structure. As Thai language is an isolative language with richness of lexical ambiguities, the methodology is constructed to extract those ambiguities by interpreting both syntactic and semantic roles of the language. The presenting methodology is mainly considered in two approaches. Firstly, it concerns the subcategorization of verb and its arguments. The verb pattern table is created as the information based knowledge. Secondly, the provides and requires attributes are considered to define the semantic relation of two concepts by using lexicon functional reasoning which is implemented in the rule base.

The system is authorized in the name of the Machine Translation Project for Asian Languages, supported by the Ministry of Trade and Industry (MITI) of Japan, conducted by the Center of Cooperation for Computerization (CICC) cooperating with other four governments of the People's Republic of China, the Republic of Indonesia, Malaysia and Thailand. The parser, mentioned hereinafter, is an out-come of the co-research between CICC and NECTEC (National Electronics and Computer Technology Center, Ministry of Science, Technology and Energy of Thailand).

2 DIFFICULTIES IN THAI SENTENCE ANALYSIS

Isolative and mono-syllable characteristics in Thai sentence leave us so many levels of problems to solve in the computer system. One surface word usually has more than one meaning and/or more than one syntactic category. In the information preparation step, we have tried to identify the grammatical role of words in each sentential form. As the result, we realized that besides the meaning of the word itself, the word position is properly notified to be the grammatical role for itself. After testing the words with any arbitrary position in a sentence form, we grouped up a set of word category with the consideration of the implementation of grammatical rule when applying to the organization. The inventory of word category employed in this analysis system was presented in Computer Processing of Asian Languages '89 at AIT.

The difficulties in Thai sentence analysis, from language computing standpoint, may be raised in this prototyping analysis system to a summary of such:-

(1) *Polysemy phenomenon* which occurs in most of Thai single word. The more frequently the word appears, the more meaning derivation it has. This is the nature of the easy-to-use words. So that, formulating the

constrains for their usage distinction is needed. The constrains which is taken into considered can be its grammatical role (Word category; CAT, SUBCAT) or syntactic usage pattern (Verb pattern; VP) or the information of neighboring words in the sentence (in pragmatic rules). For instance, the word "/caak/" has at least three meanings as follows:

L1; /caak/ : #CAT.{V}, #CP.{LEAVE}
#CAT.{PREP}, #CP.{FROM}
#CAT.{N}, #CP.{NIPA^PALM}

(2) *Appropriate word, as well as sentence, boundary assignment.* Thai language has a nature of being written in a string of characters with no any remarkable word boundary or sentential marker. This really causes the difficulties in analysis as it must have been segmented into sentences or words. In addition, Thai language has neither punctuation marker to mark the clause boundary. To separate the clause, space between string of characters is proposed to be the marker determining the boundary of the clause or the sentence. But the word segmentation is still the problem in analyzing as how precise the word segmentation is. As the word formation in Thai language is formed by attaching each words together to form the new word, so the problem is how to keep the word in dictionary, single word or compound word. For instance, "/kaanplxxphaasaaduaikhoomphiuter/" is composed of 5 single words as "/kaan/", "/plxx/", "/phaasaa/", "/duai/","/khoomphiuter/". This word can be interpreted as follows:

L2; /kaanplxxphaasaaduaikhoomphiuter/
can be segmented in 4 different ways:-
5 words as /kaan/, /plxx/, /phaasaa/, /duai/, /khoomphiuter/
4 words as /kaanplxx/, /phaasaa/, /duai/, /khoomphiuter/
3 words as /kaanplxxphaasaa/, /duai/,/khoomphiuter/
1 words as /kaanplxxphaasaaduaikhoomphiuter/

(3) *No inflection, no verb agreement.* Thai language is an isolative and monosyllable language. There is no inflection to mark morphology of the language like English or Japanese. On the other hand, those morphology is designated by the lexical item. For example, the passive voice is indicated by a lexical item in the position of pre-verb.

S1; /nakrian/ /thuuk/ /khruu/ /longthoot/
student passive marker teacher punish
"The student is punished by the teacher."

Like the passive voice, Thai language expresses tense, aspect, modality in lexical items modifying verb in pre- or post-position.

(4) *Tense point of view.* In (3), we have mentioned that tense in Thai is expressed overtly by a lexical item as auxiliary category. Only one lexical item of "/ca/=will", is a marker expressing that the event is not yet occurred. So it can be summarized that Thai recognizes only two tenses:-

(a) Irrealis tense expresses that the event is not yet occurred, corresponding to future tense.

(b) Realis tense expresses that the event has already been occurred, corresponding to present and past tense which is not distinctive.

The difficulty appears in how to assign the universal tenses of present, past or future to the interlingual representation of Thai sentence.

3 ANALYSIS ARCHITECTURE

The target of this analysis system is to produce an interlingual representation (dependency tree structure) from a linear sequence of Thai character string. The output interlingual representation will then be transferred to sentence generation system to generate the any other specified languages. Therefore, the interlingua must be exhaustive to represent all the meaning units of the source language. The research on interlingua is carried on in other framework of the project. Here the detail of interlingual representation and the generation part will not be discussed.

Designing this analysis system scoped to process a syntactic sophisticated structure of Thai language needs a lot of tactics in the rule implementing and rather flexible parser with ample functions for data manipulating. The parser itself will be discussed in the next section of this paper. The followings are the postulates for system construction. These are realized in both of the parser capability and methodology implemented in the rules.

(1) *Sentence analysis*

This is a restriction narrowing the possible information which can be taken into account in the parse time. But, this restriction protects us from unpredictable calculation time and misinterpretation. Especially for the Thai language, there is no an sentential marker preferable whether it is a comma between phrases or a full stop at the end of sentence. Nevertheless, discourse analysis is believed to be another precise method to accurate the translation. The idea of discourse analysis is also in our extension plan.

(2) *Lookahead in parsing*

This thought positively supports the idea of using all the available information in parsing. The

full support of information from either of it own lexicon specific features or surrounding constituents means a lot in drawing the most appropriate result in any step of parsing.

S2; /khon/ /khian/ /nangsuuniyai/ /khon/ /nan/
person write -ing novel person that
/kamlang/ /dean/ /maa/
walk come
"The person who writes a novel is coming."
or "The author is coming."

Considering an example of noun phrase in S2 above, it is difficult to determine the end of the noun phrase if the system has no lookahead capability. The embedded sentence of "/khon/ /khian/ /nangsuuniyai/=A person writes a novel." will actually be parsed as a sentence followed by an another sentence of "/khon/ /nan/ /kamlang/ /dean/ /maa/=That person is coming.". It makes sense but it is better to be parsed as one sentence with a noun phrase of "/khon/ /khian/ /nangsuuniyai/ /kon/ /nan/=The person who writes a novel (or The author)." being the subject of the sentence. This lookahead function is very useful in selecting the suitable alternatives. Thus the informations of all the constituents in the sentence must be referable at any points in the parse.

(3) *Node instantiation*

The system has to be able to recognize each node identically.

S3; /khao/ /maa/ /caak/ /haatyai/ /doi/ /rotfai/
he come from Haadyai by train
"He comes from Haadyai by train."

The train "/rotfai/" as well as the others is instantiated as an object representing a train with the specific syntactic features while it appears in the sentence. In case of multiple concept of a node such as "/khao/", the concept of "he", "animal horn" and "mountain" are to be instantiated separately to maintain the 3 concepts attaching to the same node.

(4) *Bottom-up parsing in top-down parsing*

The significant ambiguity of word category in the language such as Thai language multiplies the search path of grammatical rules. This ambiguity cannot be easily reduced locally only with its lexicon features before taking it into the top-down parsing mechanism rule set. The top-down parsing is introduced to produce all the possible interpretations. As a result, this will allow a large number of candidate to be taken into account because of the significant feature of word ambiguity of Thai. Thus, the system needs some heuristic rules to disambiguate the word category and some kinds of sentence constituent

reducing rules in the bottom-up parsing mechanism according to the locality in analysis allowing in the language.

3.1 Lexicon Information Representation

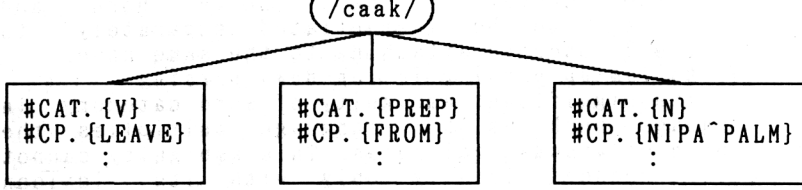
The static information of a lexicon is assigned in the lexicon dictionary having a general surface form (word spelling) as a key to retrieve. The information assigned in the lexicon dictionary is in the form of feature and its value, #feature.{feature_value}. Such as,

```
L3; /plxx/
    #CAT.{V}
    #SUBCAT.{VACT}
    #VP.{11}
    #MAPPING.{SUB=AGT,DOB=OBJ}
    #CP.{TRANSLATION}
    #AKO.{2111}
```

The feature can be word category (CAT), subcategory (SUBCAT), mapping constraints between syntax and semantic relation (MAPPING) or verb pattern (VP) for the information concerning the syntactic feature. And, there also includes word concept (CP) and conceptual hierarchy (AKO) for the information defining the meaning of word. The detailed classification of feature_value and dictionary construction is discussed in "Thai Dictionary for Multi-lingual Machine Translation System" presented at AIT 1989.

All of the information attached to each lexicon is instantiated when it is retrieved from the dictionary. The parser will treat every syntactic or semantic ambiguity to a surface word individually. For example, the word "/caak/" in L1, three instances of "/caak/" will be generated attaching to one same surface.

T1;



This nature of instantiation is very useful while a word appearing in the sentence has more than one meaning or one usage. Especially in the disambiguition process, the parser need to know the syntactic feature of "/caak/" which means "to leave" as a verb rather than the possibility of "/caak/" to be either verb or prep or noun and to mean "to leave" or "from" or "Nipa

palm" without any discrimination standard. Because, "Nipa palm" cannot be treated as a verb as well as "from" cannot be treated as a noun.

During the parse time, rule can assign additional information to the instance when its concept is augmented by other concept or its role in the sentence becomes clearer. For instance,

S4; /nangsuu/ /bon/ /to/ /nii/ /mii/ /raakhaa/ /phaeng/
book on table this have cost expensive
CAT: ... PREP N ... ...
"The book on the table is very expensive."

The prepositional phrase "/bon/ /to/=on the table" is reduced to be a "/to/=table" with the augmented value of #PSPL.{/bon/} to indicate that the current instance of "/to/" has something to be "/bon/=on". This point of view has come from the analyzing idea in case grammar theory which determines the prepositional phrase as the noun phrase.

3.2 Methodology

The parser refers to grammatical rules for examining the acceptable solution of a sentence. But, sometimes it cannot tell that which solution is favored over the others or near to the human preferences in parsing. To initiate the parse for selecting the best alternatives, we have implemented some of the results from psycholinguistic research concerning the human preferences in parsing as the parse principles. So far as this paper concerns we are not going to discuss the human preferences in detail. Followings are the parse principles implemented to supplement the grammatical rules.

(1) *Right Association*

S5; /nakrian/ /khit/ /waa/ /aacaan/ /ca/ /mai/ /maa/
student think that teacher will not come
/nai/ /wannii/
in today
"The student thinks that the teacher will not come today."

New constituents tend to be interpreted as being part of the current constituent under construction rather than part of some constituent higher. In S5, it is preferable to interpret that "the teacher will not come today" rather than "The student thinks ... today".

(2) *Lexicon Preferences*

S6; /chan/ /suu/ /nangsuu/ /nai/ /raan/ /nii/
I buy book in shop this
"I buy a book in this shop."

S7; /chan/ /yaakdai/ /nangsuu/ /nai/ /raan/ /nii/
I want book in shop this

"I want a book in this shop."

In S6, the prepositional phrase "/nai/ /raan/ /nii/=in this shop" is most likely to modify the verb phrase "/suu/ /nangsuu/=buy a book", which is interpreted as "buy (a book) in this shop", rather than the noun "/nangsuu/=a book", which is interpreted as "buy the book which is in this shop". But, there is no alternative at all for the prepositional phrase "/nai/ /raan/ /nii/=in this shop" in S7 to modify the verb phrase "/yaakdai/ /nansuu/=want a book", which is interpreted as "want (a book) in this shop".

This kind of information is lexicon dependence so we assign it into the lexicon feature in the dictionary.

(3) *No Dependency Crossing*

S8; /khao/ /rotnam/ /tonmai/ /thukwan/ /nai/ /suan/
he water plant everyday in garden

S9; /khao/ /rotnam/ /tonmai/ /nai/ /suan/ /thukwan/
he water plant in garden everyday

S8 is interpreted as "water in the garden" while S9 has to be interpreted as "plant in the garden". It is impossible to interpret S8 to have a noun phrase of "plant in the garden" because of the relation between "/rotnam/=water" and "/thukwan/=everyday". In other words, the adverb "/thukwan/=everyday" cannot modify any other constituent accept for the verb phrase "/rotnam/ /tonmai/=water a plant".

3.2.1 Bottom-up parsing in top-down parsing

This is the base mechanism of parsing in the system. All the analysis paradigms discussing in the rest subsections are implemented consistently on this base parsing principle. This subsection describes the interrelation of the rules for making hypotheses in top-down parsing and the rules for pattern determinating in bottom-up parsing. The top-down parsing has some distinctions from the conventional one. That is, all the possible interpretations are generated immediately as the hypotheses for following rules implementing. They are not generated one by one after the faulty detection and being caused backtracking as in the conventional parsing. This top-down parsing generates the hypotheses under the restriction on those lexicons. When all the possible hypotheses have been generated, the other processes (discussed in the next subsections) will then conduct the elimination of flawed hypotheses or selection of the effective hypotheses, and finally select the only one plausible solution.

Followings are small grammatical rule sets

simplified in phrase structure form.

(1) Rules in top-down parsing :-

(1.1) S <- NP VP
(1.2) VP <- V NP PP
(1.3) VP <- V NP
(1.4) VP <- V

(2) Rules in bottom-up parsing :-

(2.1) V <- LAUX V RAUX
(2.2) V <- V RAUX
(2.3) NP <- N NUM CLAS DET
(2.4) NP <- N VATT CLAS DET
(2.5) NP <- N CLAS DET
(2.6) NP <- N DET
(2.7) NP <- N VATT
(2.8) PP <- PREP NP

```
S10;  /chaang/ /yai/  /tua/   /nan/  /aasai/  /yuu/
CAT:      N       V    N,CLAS  DET      V      V,AUX
SUBCAT: NCMN     VATT  NCMN,   DDAC    VACT    VSTA,
                       CLAS                    XVAE
CP:   ELEPHANT   BIG   BODY    THAT    DWELL,  STAY,
                                       RESORT^TO STATE

          /nai/  /paa/  /luk/
CAT:      PREP     N      V
SUBCAT:   RPRE   NCMN   VATT
CP:       IN    FOREST  DEEP
```

"That big elephant lives in a deep forest."

After inspecting of all words through the whole sentence, the parser generates all of the possibilities for the verbs "/yai/", "/aasai/", "/yuu/" and "/luk/" beyond the information and constraints retrieved from the lexicon dictionary. As the results, "/yai/", "/yuu/" and "/luk/" activate the rule (1.4), and "/aasai/" activates the rules (1.3) and (1.4). The rule (1.3) is also consulted for "/aasai/" because it also has the meaning of "to resort to".

The lookahead feature of the parser allows the prediction in the bottom-up parsing process. This can be simulated as below :

"*" determines the position of the parse.
"f" determines the lookahead position.

Parse state 1:

```
     Parse position                Parse rule
   *
/chaang/ /yai/ /tua/ /nan/..  (1.1)S <-* NP VP
                              (2.3)NP<-* N NUM CLAS DET
                              (2.4)NP<-* N VATT CLAS DET
                              (2.5)NP<-* N CLAS DET
                              (2.6)NP<-* N DET
                              (2.7)NP<-* N VATT
```

Parse state 2:

```
     Parse position                Parse rule
```

```
   *       f
/chaang/ /yai/ /tua/ /nan/..  (1.1)S <-* NP f VP
                              (1.4)VP<-f V
                              (2.4)
                                 NP<-* N f VATT CLAS DET
                              (2.7)NP<-* N f VATT
Parse state 3:
     Parse position                    Parse rule
   *               f
/chaang/ /yai/ /tua/ /nan/..  (1.1)S <-* NP VP f
                              (2.4)
                                 NP<-* N VATT f CLAS DET
                              (2.7)NP<-* N VATT f
```

Rule (2.4) finally supports the decision to parse "/chaang/ /yai/ /tua/ /nan/" as an NP rather than a sentence because the existing of the verb "/aasai/" in the later part will break this sentence into two sentences of "/chaang/ /yai/" and "/tua/ /nan/ /aasai/ sentence at a time". And the longest parse preference will give the priority to the rule (2.4) rather then (2.7).

After a parse through the end of the sentence, the ambiguity still remains in what concept of the "/aasai/" is used whereas the "/aasai/" is firmly marked to be parse as the main verb of the sentence.

```
S10'; /chaang/ /yai/  /tua/  /nan/  /aasai/  /yuu/
CAT:     N       V    ~~N~~,CLAS  DET     V     ~~¥~~,AUX
SUBCAT: NCMN    VATT  ~~NCMN~~,    DDAC   VACT   ~~¥STA~~,
                      CLAS                      XVAE
CP:   ELEPHANT  BIG   BODY   THAT   DWELL,  ~~STAY~~,
                                RESORT TO  STATE

         /nai/  /paa/  /luk/
CAT:     PREP    N      V
SUBCAT:  RPRE   NCMN   VATT
CP:      IN    FOREST  DEEP
```

The bottom-up rules will then reduce "/chaang/ /yai/ /tua/ /nan/" to be an NP by (2.4), "/aasai/ /yuu/" to be a V by (2.2), and "/nai/ /paa/ /luk/" to be a PP by (2.7) and (2.8).

Up to this stage, the top-down procedure still maintains two planes of parsing possibility of the main verb "/aasai/". The parse is resumed in the next subsection to extract the plausible interpretation.

3.2.2 Subcategorization

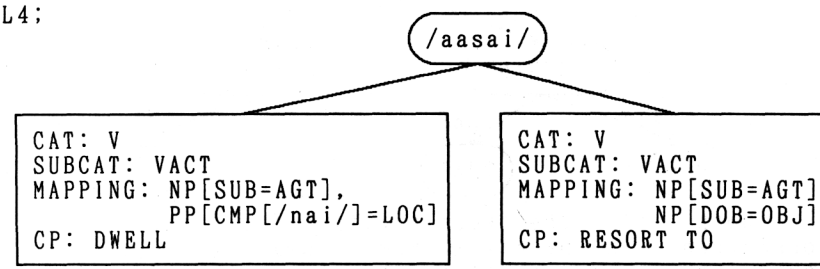
The disambiguition of a word among its different word categories is explicitly realized by extracting the most plausible usage pattern as described in the previous subsection. Though the word usage according to its category can reduce the category ambiguity to some extent, the ambiguity still remains especially to the

word which possibly acts as a verb. As a conclusion, the lexicon ambiguity is recognized in two different levels. First is the one in which a word occupies more than one category such as "/tua/" can be a common noun (NCMN) and a classifier (CLAS) or "/yuu/" can be a stative verb (VSTA) and a right auxiliary (XVAE) in S10. The system reduces this kind of ambiguity according to its plausible usage pattern. Second is the ambiguity occurring within the same word category but different meaning such as "/aasai/" which is an active verb (VACT) having the meaning of both "to dwell" and "to resort to" in S10. In this case, with just only the word category the system cannot distinguishes its meanings at all. Therefore, the category of verb is specially defined correspondingly to its distinctive natures.

A verb apparently needs some other constituents to fulfill its meaning for detailing an event. For instance, the "/chaang/=elephant" in S10 referred to a kind of animal has a complete meaning within itself whereas "/aasai/=to dwell" in S10 needs a significant agent of the action and a place to where the action is attached to complete the meaning describing an event. On the other hand, the verb "/aasai/=to resort to" needs an agent of the action and an object to which the action is directed. Therefore, a verb having more than one meaning such as "/aasai/" can be described as to the number and the syntactic and semantic nature of the elements it combines with. The dependencies that hold between the verb and its dependent elements are referred to as subcategorization restrictions.

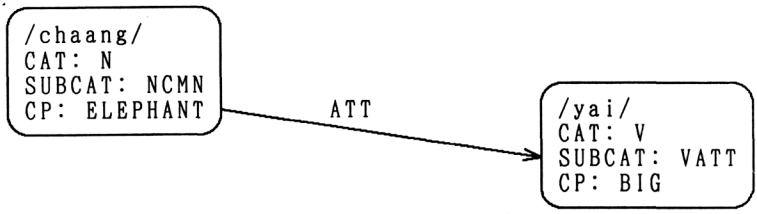
The top-down rule of (1.3) and (1.4) generate two planes of meaning for the verb "/aasai/" in S10. Both are different in the subcategorization restriction as simply depicted in the value of MAPPING below.

L4;

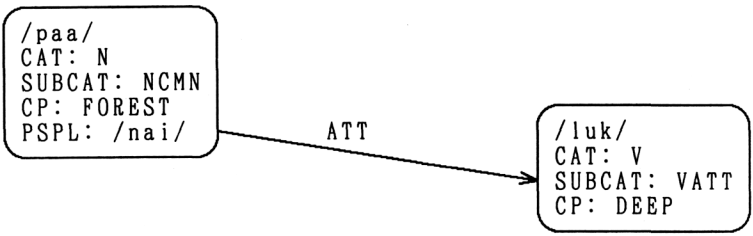


The partial structure and lexicon information of the other phrases are simplified as follows:

T2; /chaang/ /tua/ /yai/:

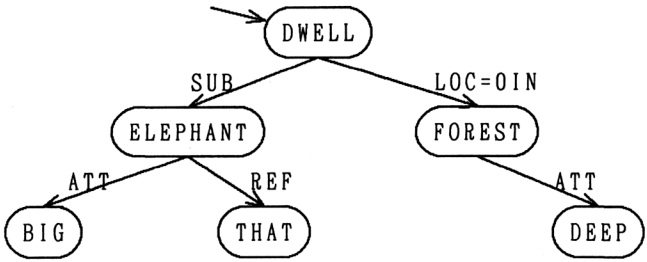


T3; /nai/ /paa/ /luk/:



The phrase "/chaang/ /tua/ /yai/" satisfies both MAPPINGs of the "/aasai/" because it provides the attribute of a noun phrase as required in the MAPPINGs. The phrase is subcategorized for as the subject of the verb while the phrase "/nai/ /paa/ /luk/" is a prepositional phrase (PP) which has the feature value of PSPL (preposition information) satisfies only the MAPPING of "/aasai/=to dwell" according to the constraint of CMP value (verb complement) in the MAPPING.

T4; /chaang/ /yai/ /tua/ /nan/ /aasai/ /yuu/ /nai/ /paa/ /luk/



Both of the possibility of "/aasai/" are considered in parallel when they are generated. Comparing the degree of satisfactory, the "/aasai/=to dwell" has full number of elements which satisfy all the requirement while "/aasai/=to resort to" has only one element which satisfies the need. Thus "/aasai/=to dwell" is selected to build a dependency structure.

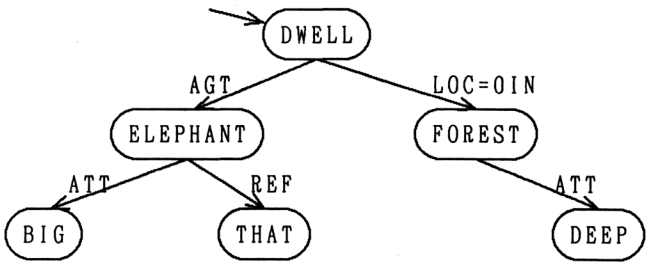
The syntactic dependency structure of S10 is therefore constructed as in the T4 above.

However, in the rule implementation we need more information than describing above to justify the subcategorization of verb. For example, "/tuu/=cabinet" cannot "/aasai/ /yuu/ /nai/ /baan/=live in the house" like a living thing but the sentence must be "/tuu/ /yuu/ /nai/ /baan/=cabinet is in the house". In this case, we need some kind of conceptual hierarchy (AKO) to mark the property of the object. Therefore, the verb "/aasai/" will subcategorize for a subject which is a kind of living thing to give the meaning of "to dwell or to live".

3.2.3 Case Mapping

T4 shows the syntactic dependency structure where the upper node is the head node and the lower node is the dependent node. To build a deep structure we have to compile the relation between nodes into a deep case representation (defined in the interlingua for multilingual machine translation). The syntactic relation obtained from MAPPING feature of verb determines the case between the nodes. Therefore, "/chaang/=elephant" in S10 is assigned to be the agent (AGT) of the action "/aasai/=dwell" and "/paa/=forest" is assigned to be the location (LOC) where the action takes place. The "/paa/=forest" is definitely not the object (OBJ) of the action "/aasai/=resort to" which gains the possibility according to the ambiguity of the word "/aasai/".

DS1; /chaang/ /yai/ /tua/ /nan/ /aasai/ /yuu/ /nai/ /paa/ /luk/



Therefore, the MAPPING feature of the selected meaning of "/aasai/=dwell" confirms the semantic relation in the tree of T4 to be as in the above DS1.

The analysis process usually ends here after generating the deep structure as DS1. In some cases, the tree structure generated according to MAPPING feature of verb is not logically acceptable as a deep structure (this also depends on the definition of case

set). The linguistic phenomena such as in S11 (is to be interpreted as S11') or contraction in S12 ("/yuu/ /nai/=be in" is to be interpreted to be a case of LOC) are considered to be in the case.

S11; /khruu/ /sang/ /nakrian/ /hai/ /tham/ /kaanbaan/
teacher order student to do homework
S11';[/khruu/ /sang/ /hai/[/nakrian/ /tam/ /kaanbaan/]]
teacher order that student do homework
S12; /khruu/ /yuu/ /nai/ /hong/
teacher be in room

3.2.4 Lexicon Functional Reasoning

We introduce the functional reasoning to be the fundamental guide to infer the appropriate semantic case which is to be set in between the nodes connection. Every node is treated individually as an existing object. And, each object node has its own syntactic/semantic functions which can be deleted or modified during the process of reasoning. The functions of the object node are realized according to its currently occupying syntactic/semantic features. This means that the initial functions change continually during whole analysis process. For instance, a node of noun (provides: N) will be the head node of prepositional phrase (provides: PP, PSPL) after the connection with a preposition.

This reasoning process has the propositions to achieve the goal as follows:

Goal: Construct a semantic tree having all node sconnected together with the appropriate semantic cas eand a node being the head node of the tree.

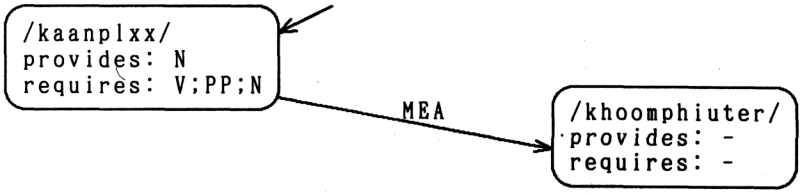
Propositions:

(1) Each object node provides a set of functions.
(2) For each function it provides, an object node requires a set functions.
(3) A functional connection can occur between two nodes if one provides a function required by the other.
(4) A constructed structure consists of a set of nodes having a node to be the head node of the structure and a set of functions installed in the head node.
(5) Semantic case indicates type of connection.

S13; /kaanplxx/ .... /duai/ /khoomphiuter/
translate with computer

```
/kaanplxx/
provides: N
requires: V;PP;N
```

```
/khoomphiuter/
provides: PP, PSPL.{/duai/}
          AKO.{Finished Product}
requires: V;N
```



4 ACKNOWLEDGEMENT

The system is the first prototype designed to process Thai language. The language model is hoped to be able to be extended to other languages in the isolative language family. Many syntactic restrictions are raised to make the system feasible at this stage. For the further development, we have realized the necessity of semantic analysis in deeper level performing interactively with the syntactic analysis. At present, we are preparing knowledge bases to accurate the semantic disambiguition process. And we are also interesting in the concept of reader's model and reader's background knowledge generation and retrieval. The further studies will be conducted to support the development of multilingual machine translation in the project and the attempt to form the natural language interfacing module.

REFERENCES

Allen, J. (1987) **Natural Language Understanding.** The Benjamin/Cumming Publishing Company.

Freeman, P., Newell, A. (1972) 'A model for functional reasoning in design' Proc 2nd IJCAI. London UK, 621-640.

Muraki, K., Sornlertlamvanich, V., et al. (1989) 'Thai Dictionary for Multi-lingual Machine Translation System' **Computer Processing of Asian Languages (CPAL).** AIT, 211-220.

Muraki, K. (1991) 'Concept Representation and Machine Translation' In Nomura H. (ed), **Language Processing and Machine Translation.** Kodansha (in Japanese)

Panupong, V. (1984) **The Structure of Thai Grammatical System.** Ramkhamhaeng Univ. Press.

Pollard, C., Sag I. A. (1987) **Information-based Syntax and Semantics.** Vol.1 Fundamentals, Center for the Study of Language and Information (CSLI).